Making the Little Guy Pay:

Payments System Networks, Cross-Subsidization,

and the Collapse of the Suffolk System

Howard Bodenhorn

Lafayette College

Easton, PA 18042

bodenhoh@lafayette.edu

Several recent studies have drawn lessons for modern payment systems from historical experience. One important example is the Suffolk System, which operated in New England for about 40 years prior to the American Civil War.[1] The Suffolk established a region-wide net clearing system for banknotes that resulted in par circulation for all system members. Historians offer nearly universal praise for the system, arguing that region-wide par redemption reduced the costs and risks to consumers of using a privately issued currency.[2] Modern economists generally agree with the historians, and develop several explanations for why and how the Suffolk operated as long and as well as it did. Mullineaux (1987) argues that the Suffolk sold an efficient franchise contract in which each party faced incentives to maintain the quality of the franchised good, namely banknotes that traded at par. Calomiris and Kahn (1996) argue that the Suffolk provided effective monitoring services, reduced the likelihood of bank runs, and encouraged currency usage that, in turn, enhanced

payment system efficiency.

Others provide less praiseworthy interpretations of the Suffolk. Bodenhorn (1998) argues that the Suffolk established its system through extralegal coercion and exploited its monopoly position by extracting rents from its members. Rolnick, Smith and Weber (1998) argue that large-scale payments clearing systems tend to be natural monopolies and that economic theory has not yet reached a consensus about the efficiency effects of unregulated natural monopolies. Smith and Weber (1999) develop a model in which a Suffolk-like agent enhances consumer welfare by reducing the use of unproductive specie in exchange and as bank reserves, but their model also suggests that a Suffolk-like system may produce endogenous bank liability volatility and will not be effective in smoothing either interest rates or inflation. They then produce some data from the Suffolk era consistent with their model's hypotheses.

Thus the debate about the Suffolk's effectiveness as a agent of payment system efficiencies remains unresolved. This paper does not address these issues directly. Instead, it steps outside the existing debate and considers two different but related questions: Why was the Suffolk system detested and undermined by its own members? And, why did the system come to an abrupt end in 1858?[3] From its inception, country banks were reluctant to join and many did so only under duress having suffered the indignity of what can only be characterized as a "bank run" organized and led by the Suffolk Bank and its agents. Even after they joined, members mutinied on occasion and bent the system's rules as much as the Suffolk allowed. Finally, in 1855 several Suffolk members spearheaded the establishment of a competitive clearing system, abandoned the Suffolk's system, and enticed about half of New England's banks to do the same. That its very members undermined it suggests that there may well have been a "dark side" to the Suffolk that has yet to be exposed

(Gorton 1999).

There was no sinister dark side to the Suffolk, but its methods were coercive. The Suffolk's fundamental error was that it either failed to recognize the network externalities inherent in its system and price its services accordingly, or it refused to price its services in this way because doing so would have undermined its support among Boston's banks. Because the Suffolk network involved external effects, efficient prices differed for seemingly similar members, depending on the value each member placed on participation. Banks that placed low individual values on membership but whose membership generated large (external) benefits for other subscribers should have had their membership fees subsidized by other subscribers. Generally, this would have meant that large, urban banks would have subsidized the membership of small, remote banks by paying most of the common costs of the network's operations. Instead, the Suffolk system allocated costs in the opposite fashion. City banks were unwilling to bear the costs of the system's operations, and unable to force them to do so, the Suffolk pushed the costs onto New England's country banks. Early complaints of the costs of note clearing and banknote discounts, then, were probably not complaints about the total costs of the clearing and redeeming country banknotes, but about their allocation. In the pre-Suffolk era, city banks paid the lion's share of the costs of bank liability clearing and settlement. In the Suffolk era, the costs were pushed onto country banks which ultimately led to the Suffolk's demise.

A Brief History of the Suffolk System

Historical accounts typically describe Boston circa 1810 as a city swamped with country banknotes. Because the regional balance of trade favored Boston, country banknotes flowed into the city where they competed with and, to some extent, displaced Boston banknotes in everyday

transactions.[4] Between 1796 and 1818 several plans were hatched by Boston's banks to return country notes to their place of issue, but country banks raised such obstacles -- the most common being an infuriatingly slow and methodical counting and recounting of specie -- that none proved successful (Redlich 1947, p. 67). Despite the city bank's best efforts, the influx of country notes continued.

Following several failed attempts to limit the circulation of country notes within the city, Boston's banks washed their hands of the business. Some Boston banks refused to accept country notes on deposit, others accepted them only conditionally. Dealings in country banknotes devolved to the market and several private note brokerages opened.[5] Note brokers bought banknotes at a discount from par and recirculated them or returned them to the issuing bank for specie.[6] Brokers profited when the discount exceeded redemption costs and reimbursed the broker for the risks he assumed.

With note brokers buying country banknotes at discounts ranging between one and six percent, the potential profits were sure to attract competitors and in 1813 the New England Bank of Boston re-entered country banknote brokerage (Whitney 1878). The New England Bank agreed to accept the notes of all regional banks at a discount equal to redemption costs and within a relatively short time reduced discounts to a uniform one percent. Believing that the New England Bank was profiting even at this low and uniform rate, the Tremont Bank of Boston entered into note brokerage a few years later. Competition between the two drove the average discount to about one-quarter of one percent. The two banks apparently entered into a cartel agreement and for a few years thereafter the average discount remained stable at about one percent (Redlich 1947, p. 69).

The Suffolk Bank was chartered in 1818 and began buying country banknotes at a discount

in 1819. Though initially profitable for the Suffolk, competition between it and the New England/Tremont alliance reduced the average discount to about three-eighths of one percent, making the business "hardly profitable" for any of the banks (Redlich 1947, p. 72). The Suffolk's actions, however, had not accomplished its main objective, which was the reduction of the volume of country banknotes circulating in the city. In fact, the Boston banks' actions probably had the opposite effect. By eliminating the discount on country banknotes, competition between the Suffolk and the New England/Tremont alliance made city and country banknotes even closer substitutes and made Boston consumers more willing to accept country banknotes (Mullineaux 1987).

In 1824 the Suffolk and six other Boston banks formed a coalition known as the Associated Banks who believed that concerted action could limit country banknote circulation and expand their own (Knox 1903, p. 366). Under the terms of the Association agreement, each bank sent all country banknotes received on deposit to the Suffolk which acted as a common redemption agent. Whenever the Suffolk's holdings of a particular bank's notes reached a predetermined level an agent was sent to redeem them at the issuing bank.

Most country banks interpreted the Suffolk's plan as provocative, even predatory, and devised various schemes to thwart it. When the Lincoln Bank of Wiscasset, Maine was called on to redeem $3,000 of its notes, it first tendered a Boston draft that the Suffolk's agent promptly refused, asserting the Lincoln Bank's legal duty to redeem its notes in specie. The Lincoln Bank's cashier then delayed redemption by methodically counting out small coins. Only $500 in notes were redeemed by the hour of closing. The South Royalton Bank of Vermont adopted a similar strategy. When the Suffolk's agent appeared demanding specie, he was informed that the bank would first redeem its $1 notes. By the time this was done the hour of closing was reached. Overnight, the

South Royalton obtained specie from two neighboring banks sympathetic to its plight and the following day, when the Suffolk's agent reappeared, the bank's cashier announced that he stood ready to redeem its $2 notes. Even while these events unfolded, the South Royalton's attorney filed suit against the Suffolk charging it with a "malicious intent to break the Bank without cause" (Harper 1969, pp. 33-34). The South Royalton eventually met the Suffolk's demands, but pressed its suit. Ultimately, the Suffolk prevailed when Vermont's supreme court found in its favor, deciding the malicious intent unproved.

As a result of these and sundry other stalling tactics among country banks the Suffolk soon abandoned its random, unannounced redemption calls with an organized system. Initially, the Suffolk agreed to accept the notes of any of the region's banks at the current market discount if the bank agreed to maintain a $5,000 interest-free, permanent deposit in addition to an interest-free balance equal to its average clearings (Knox 1903, p. 365). Sometime later the permanent deposit fell to 2 percent of bank capital, with a minimum of $2,000 for banks with paid-in capital of $100,000 or less, and banknotes were redeemed at par (Whitney 1878, p. 19). Most of the region's banks eventually joined the system and the Suffolk established a region-wide clearing system in which nearly all banknotes circulated at par.

Although Calomiris and Kahn (1996, p. 771) characterize the Suffolk's clearing system as coordination without coercion, it is clear that the Suffolk imposed its will through persuasion, intimidation, and embarrassment of nonmember banks. When the Worcester Bank of Massachusetts refused to join, the Suffolk purposely stockpiled about $38,000 of the Worcester's notes (about one-half it total circulation) and demanded immediate redemption. The Worcester Bank, which then had about $39,000 on deposit with the New England Bank of Boston offered a Boston draft in exchange

for its notes. The Suffolk's agent refused. He remained intransigent in his lawful right to specie. The Worcester paid the Suffolk's agent $28,000 in specie and offered to redeem the remaining $10,000 the following day at the Suffolk's office. Refusing even this offer, the Suffolk's agent sent for the sheriff and had the Worcester Bank's property attached (Dewey 1910, p.85).[7]

Given such practices it is not surprising that the Suffolk found few friends among country banks. More surprising, however, was the mixed response the Suffolk got from the region's legislators. Vermont passed a law requiring that state's banks to join or pay an annual tax equal to 1 percent of paid-in capital. Several banks held out and paid the tax until 1850. Maine's legislature responded differently. In 1831 it overhauled that state's banking regulations, including one statute that explicitly protected that state's banks from the Suffolk's unannounced redemption calls. The law required that banks meet "usual" redemption calls in specie on demand, but it gave the banks the right to delay by 15 days any "unusual" redemption call (Root 1895, p. 279). By 1837, Maine's bank commissioners still attacked the Suffolk's actions and asked whether further legislation was needed to protect the banks.[8] Though some of Maine's bank commissioners had grown more tolerant of the Suffolk by the 1840s, others still characterized it as "tryanical [sic] and oppressive" (Chadbourne 1936, pp. 44-45).

Several of Maine's banks entered into a compact of mutual support and refused to join. They remained opposed to membership because, in their opinion, the Suffolk's plan worked too well. In the absence of the Suffolk's clearings, Maine banknotes would have sold at a discount outside the immediate vicinity of the issuing bank and would have tended to circulate in a smaller circle. With Suffolk clearings and region-wide par redemption, however, the notes tended to travel toward Boston from whence they were quickly returned (Simonton 1971, p. 25). Calomiris and Kahn (1996, p. 777)

argue that nonmember Maine banks found a limited market for their notes. But this might have been a purposeful, perhaps even a profitable, strategy on their part. Suffolk membership, after all, meant more frequent redemptions and shorter circulation periods.

If other banks shared Maine's concerns, their fears were soon realized. Within a year of having the system in place, the Suffolk redeemed and returned more than $17.4 million in banknotes. Between 1825 and 1831 the Suffolk's business grew rapidly, but remained manageable. The following half-decade, however, witnessed the opening of more than 90 new banks. The Suffolk was quickly swamped with redeemed notes, country banks were often slow in making their remittances, and many were seriously overdrawn on their redemption accounts (Whitney 1878, pp. 23-25). This forced the Suffolk to revise its overdraft policy, limiting overdrafts to the amount of a bank's permanent deposit. Once its overdraft exceeded this amount, notes were bundled and sent to the issuing bank for specie redemption along with a demand to replenish its account. This change reduced the Suffolk's risks, but had no noticeable effect on the growth of bank money (Fenstermaker and Filer 1986). In 1837 the Suffolk redeemed about $105 million in country banknotes; by 1847 its clearings increased to $165 million; and in 1857 to $376 million. The increased volume of redemptions followed from two factors: an increased number of issuing banks; and, more distressing to country banks, a notable reduction in the circulation period.

Table 1 reports statistics on total and average New England bank circulation, and the total and average Suffolk redemptions for selected years. Straight-forward calculations suggest that in 1825 an average New England banknote was cleared about 1.7 times each year. By 1835 a note cleared 4.9 times, a figure consistent with the Connecticut bank commissioners' estimate of 5 times per year (Hasse 1957, pp. 22-23). In 1850 each note cleared about 6.9 times, an estimate again

largely consistent with the Connecticut bank commissioners' estimate of 8.6 times.

Even as the velocity of note turnover accelerated, the Suffolk grew increasingly intransigent, refusing to offer even the smallest short-term accommodation to its members. It also alienated the few friends its had when it began dictating policy without consulting its Boston and Providence agents. In 1855 country banks banded together and obtained a charter for the Bank of Mutual Redemption from the Massachusetts legislature. The Bank of Mutual Redemption's charter was unique in several respects. First, the bank was strictly a banker's bank. Its capital was wholly owned by New England's other banks and it acted as their clearing agent. Second, the Bank of Mutual Redemption was given sharply limited note issue privileges. It could issue banknotes in value only up to one-half its paid-in capital, and all notes were large denomination, which were useful for interbank settlements but not for hand-to-hand transactions. Although chartered in 1855, it did not commence operations until 1858 (Lake 1947, p. 196).

The rules laid down by the Bank of Mutual Redemption in 1858 provide insights into the country banks' principal grievances. Nonshareholding, nonmember banks were required to maintain a permanent deposit in the same amount required by the Suffolk. For share holding members, the interest-free permanent deposit was eliminated. In addition, both shareholders and nonshareholders were required to maintain an additional redemption account sufficient on average to meet redemption calls. Unlike the Suffolk, however, the Bank of Mutual Redemption promised to pay 3 percent on these balances (Lake 1947, pp. 203-4).

Except for the elimination of permanent deposits for shareholders and the payment of interest on redemption accounts, the Bank of Mutual Redemption's plan did not depart significantly from the Suffolk's. Such seemingly small differences apparently mattered though. Within a relatively

short time, nearly one-half of the Suffolk's members had defected to the Bank of Mutual Redemption. The Suffolk staged a brief redemption war, followed by an equally brief refusal to engage in any transactions whatsoever with the interloper, followed by limited and grudging interactions with the Bank of Mutual Redemption, and, finally, by retrenchment and withdrawal from the redemption and clearing business in 1860.

Rolnick, Smith and Weber (1998) argue that the Suffolk's quick exit from the industry following the entry of a rival suggests that the redemption and clearing of banknotes was a natural monopoly. Unable to capture economies of scale and scope as a duopolist, the Suffolk retired from the business and focused on its retail banking business. This may be true, though the New York experience where eight large city banks redeemed that state's country banks suggests otherwise.[9] Accepting for the moment that clearing was a natural monopoly, even monopolies must charge appropriate efficient prices to survive. When viewed in the light of modern network theory, the Suffolk system emerges as a monopolist that may have charged inefficient prices given the inherent network externalities. In effect, the Suffolk charged country members too high a connection fee and city members too low a fee. Before the details of the Suffolk's pricing structure is discussed, the next section provides a summary of contemporary network theory and its application to payments systems.

Payments System Networks

Envision a number of economic agents who wish to communicate or interact with one another. In the absence of a network, communication is necessarily bilateral in nature, that is, each agent interacts with others directly. Moreover, if she interacts with k other agents, she will establish

[k(k-1)]/2 distinct connections. Suppose instead that a network forms so that members communicate through a central node rather than bilaterally. Some agents may join the network, some may not. The establishment of the network does not alter the optimal aggregate volume of interactions. Even after the network is in place, an individual will still interact with the same agents as before (ceteris paribus) regardless of whether they subscribe to the network, but the benefit of network membership is the resource savings of routing at least some communications through the jointly operated and maintained network. In addition, network membership often generates aggregate benefits greater than the sum of the private benefits.

Networks are composed of links and nodes and the analysis of a network is typically carried out under the assumption that these links and nodes are combined in such a way as to produce a marketable good or service. Economides (1996, p. 676) calls this network compatibility and argues that a necessary precondition for networks is the ability to combine links and nodes in economically meaningful ways. Everyday examples of networks are telephone systems, railroads, automated teller machine (ATM) systems, and airlines.

Suppose we have a basic "star" network with a central switch S and spokes radiating out to nodes a, b, c, d, etc. If connections operate in both directions, that is from b to S to c (call this connection bSc) as well as from c to S to b (cSb), we have what is commonly referred to as a two-way network.[10] There are several important features of two-way networks. First, all components are complementary so that any two of them can be combined or connected to form a composite good (such as bSc). Second, the components are complementary even though they are very similar goods. Third, seemingly similar goods such as bSc and cSb are distinct and are not interchangeable. Fourth, composite goods that share components (such as bSc and bSd) are generally not close substitutes.

Fifth, proper network operation requires compatible components, which may require the adoption of certain technical standards.

The most important feature of any network, however, is the presence of network effects or externalities, which have two essential elements. The first is that the benefit to an individual user increases with the number of distinct nodes or locations connected to the network. Saloner and Shepard (1995, p. 480) label this the "accessability" effect. Classic examples are telephone networks with geographically dispersed connection points (hence, phones in hotel rooms and coin-operated pay phones in bus terminals) or retail distribution where the value of a commonly consumed good increases with the number of retail outlets carrying the good. The second network effect is the "production-scale" effect and is the source of network externalities for subscribers to a particular network. In effect, the value of subscribing to a network increases with the number of potential communicants (recall, MCIs "Friends and Family" discount plan).

The production-scale effect generates externalities for existing members by increasing the size of an exiting network. In a simple star network, the externality is direct. If there are n existing network nodes, there are n(n-1) potential composite goods. An additional location (i.e., node n+1) provides direct external benefits to all existing customers by adding 2n potential new goods by providing a complementary link to all existing nodes. For a telephone system, the more subscribers the more people a given individual can communicate with. For an airline operating a simple hub and spoke (a basic star) network, the addition of a single new destination doubles the number of possible connections for passengers.

One potential problem arising from the presence of this externality, however, is the possibility of underproduction and mispricing. So long as the network is strictly internal, its

organizer faces an incentive to include an optimal number of nodes because the network externality is internalized.[11] If, on the other hand, the network is composed of several autonomous firms or individuals the network may be suboptimally small. Adding another member to an existing network benefits all existing members, but if the existing members do not offer a subsidy to some potential subscribers they will not participate (McAndrews 1997, p. 18).

Joint provision of the network goods and shared benefits implies common costs for the subscribers, and sharing the fixed costs among users becomes a nontrivial problem (Henriet and Moulin 1996, p. 332). Common costs are those costs that cannot be explicitly attributed to a single user.[12] For example, the cost of providing a telephone connection is common to both the initiator and receiver of the call, but cannot be attributed to either separately. Common costs take many forms, such as the laying of infrastructure in a communications network, the cost of physically transporting goods from one place to another, or record keeping costs. In addition, many common costs are fixed relative to the volume of traffic on the network (a switching station) and cannot be allocated to a single item passing through the network.

Common costs, then, make it impossible to identify an individual subscriber's share of the network's total costs. It follows that determining the efficient price is difficult. For most goods, efficiency requires that consumers pay marginal cost; set price less than marginal cost and they will tend to over utilize the good, set price above marginal cost and they will underutilize it. For network goods, however, the price equals marginal cost condition may not lead to the efficient output. Some subscribers must pay more than marginal cost or the network provider will not recover all common fixed costs. The fundamental pricing questions are: Which subscribers pay more than marginal cost? And, how are those costs distributed?

The solution requires a recognition of the network effects -- that one subscriber's participation in the network provides benefits to all other subscribers who might communicate with her. Recognizing that these benefits are external to the individual subscriber, but internal to the network leads to the optimality of shifting some of the costs from some subscribers to others. Generally, this cost shifting takes a form so that those who place a low individual value on network participation but who bring large external benefits to other members pay a smaller share of the common costs. If the external benefits of a particular individual's membership are large enough, it might even be efficient for that subscriber to pay less than marginal cost.

Lacker, Walker and Weinberg (1999, pp. 18) develop the following example.[13] Consider an existing network for which the marginal cost of adding a new member is C_n. Suppose that this individual's subscription to the network would provide benefits to existing subscribers of B_n and that the marginal costs and benefits to the new subscriber are C_i and B_i. It is efficient for this individual to subscribe if:

$$B_n + B_i > C_n + C_i$$

The network operator must also charge the new subscriber a price p such that it is willing to join the network. She will subscribe so long as $B_i > C_i + p$, and the network operator will offer a subscription so long as $B_n + p > C_n$. Thus any price that satisfies:

$$B_i - C_i > p > C_n - B_n$$

will result in efficient subscriptions. That is, the net private benefits to an additional subscriber must exceed the cost of subscribing even while the price exceeds the aggregate net costs to current subscribers of including the additional subscriber in the network.

It follows from the above as well that if $B_i - C_i < C_n$ the network should charge the new

member a price lower than the network's marginal costs of including her because the value of membership to the individual is low, but her inclusion in the network creates substantial external benefits that accrue to other members. In effect, existing members will be asked to subsidize this individual's entry into the network if the external benefits she generates exceed the costs of connecting her. Similarly, the network operator should charge a potential subscriber a price greater than marginal cost if $B_i - C_i > C_n$ because the individual's private net benefits of membership are high and the network operator should charge a price reflecting his willingness to pay. It is efficient for both the high low-value and high-value user to join the system, because of the external benefits their membership entails, but each should pay a different price even if they are similar in other respects.

Henriet and Moulin's (1996, pp. 334-35) network model provides an analogous pricing rule. They show that networks can use a "private cost" pricing formula (charging each subscriber the cost of connecting her to the network), an "external cost" pricing formula (charging each subscriber a fraction of all other subscribers' connection costs), or any convex combination of these two. All such pricing schemes involve some cross-subsidization, with the exact pricing mechanism determining the direction, extent, and efficiency of the cross-subsidization. Suppose there are two types of subscribers, rural and urban, and the costs of connecting urban subscribers is strictly less than the cost of connecting rural subscribers. If most traffic is rural-to-urban, Henriet and Moulin (1996, pp. 343-45) show that optimality requires that urban subscribers pay everyone's connection costs and rural subscribers connect for free. A more realistic situation is one in which urban subscribers pay membership fees greater than their connection costs and rural subscribers pay less. Thus, subscribers who place a high individual value on connecting to the network subsidize the

connections of those who place a low individual value on connecting. Cross-subsidization is inevitable if the network operators desire universal participation.

The basic model of a communications network is easily extended to a payments system network. McAndrews (1997) and Weinberg (1997) argue that clearing and settlement functions involving financial institutions separated by time and distance are inherently network services. Thus, instead of viewing the Suffolk as a proto-central bank intent on controlling the region's money supply, as some historians such as Hammond (1957) and Redlich (1949) have, it is more useful to view it as a private supplier of a clearing and settlement network established and operated to generate profits.[14] Furthermore, organizations involved in the clearing and settlement of payments, whether a strictly private organization (the Suffolk) or a quasi-public organization (the Federal Reserve), generate network externalities.

One of the principal functions of banks is their provision of the means of payment. Banknotes and demand deposits will become important means of payment only if a reliable method for clearing, settlement, and redemption arises (Weinberg 1997, p. 25). As agricultural economies develop, exchange increasingly involved transactions separated by time and space, implying that the location at which trade occurred differed from the location at which the means of payment was issued. In such a world, settlement required communication between places. The Second Bank of the United States engaged in some limited clearing and settlement operations but never emerged as a national or regional clearinghouse for bank liabilities. The Suffolk Bank, on the other hand, became New England's principal banknote clearinghouse in the first half of the nineteenth century.

With little or no branch banking, correspondent relationships were used to clear banknotes when the issuing bank was distant from the receiving bank.[15] If both banks had a common

correspondent, the collection and clearing of banknotes proceeded at a significantly lower cost than if redemption required personal presentment or even redemption by mail.[16] Thus any two banks sharing a common correspondent formed a network, however rudimentary. We can easily generalize this to several country banks sharing the same city correspondent if they regularly accepted each other's notes or their notes tended to travel in a particular direction in the course of trade.

Which banks would voluntarily join such a network? Clearly, the value of belonging to a clearing and settlement network depended on the frequency with which the bank received notes issued by distant banks. Consider the problem faced by a bank receiving on deposit the notes of a distant bank with which it had few dealings. The receiving bank could send an agent to present the banknotes at the issuing bank or it could send them by post, but this could be risky if specie were shipped in return. If both banks were members of the same network, the receiving bank forwarded the issuing bank's notes to the common correspondent for clearing and collection. The receiving bank had its correspondent account credited; the issuing bank's account was debited and its notes returned.

Clearing and settlement did not require every country bank have the same city correspondent. New York and Philadelphia witnessed the development of competitive clearing operations where a handful of city banks performed overlapping clearing and settlement functions (Myers 1931). If one country bank remitted notes of another country bank that did not have the same city correspondent, clearings required a second level of interbank cooperation. The city bank receiving notes of a nonrespondent bank simply cleared those notes with its city correspondent. Such roundabout methods may have sacrificed some of the benefits of economies of scale in network

clearings, but the costs were still lower than simple bilateral exchange.[17]

Correspondent networks generated two sources of resource savings. First, direct redemption or presentment costs (either travel or post and insurance) were avoided. Second, there were savings arising from economies of scale in the transport of banknotes. The marginal cost of clearing and placing an additional banknote in a bundle of banknotes already addressed to the issuing bank was practically nil and certainly smaller than the cost of sending it separately. At the same time, the correspondent may have been able to exploit economies of scale in clearing and settlement functions. In the presence of significant fixed network costs, the average costs of providing clearing services fell rapidly as the network expanded.[18] Moreover, establishment of networks may have fostered long-term interbank relationships and lowered interbank monitoring costs (Calomiris and Kahn 1992).

In a payments system network, the value of communicating between two points depended on the volume of transactions between those points. Bankers in Bangor, Maine valued a network that included members in Providence, Rhode Island or Williamstown, Massachusetts only if a significant volume of transactions took place between Bangor and Providence or Bangor and Williamstown. Each endpoint or network node was differently valued by each member. An efficient network included only those endpoints for which the total value of participation (private and external benefits) exceeded the resource costs of participation (Weinberg 1997, pp. 26-27). That is, if banknotes issued by the Agricultural Bank of Pittsfield, Massachusetts rarely ended up in the hands of the Marine Bank (or any of the three other New Bedford banks), and vice versa, the private and external benefits were potentially small relative to the costs of connecting the Agricultural Bank. The Agricultural Bank then may not have voluntarily participated. If, on the other hand, there was

a substantial coasting trade between New Bedford and Plymouth, implying that notes originating in each place traveled to the other, the private and external benefits of including all four New Bedford and both Plymouth banks almost surely outweighed the connection costs.

It follows that the value to a given bank of subscribing to a payments system network depended on the network's existing subscribers. Generally, agents in one location placed a high value on network membership if the network connected it to a large number of agents with which it regularly interacted. Efficient pricing forced banks to recognize and react appropriately to the network externalities involved. In a network is made up of members with different valuations, seemingly similar subscribers should have paid different prices for similar services (Weinberg 1997, p. 34). The questions addressed below are: What prices did the Suffolk charge for connecting to its network? And, did these prices generally accord with efficient network pricing? That many of the Suffolk system's participants were always unhappy and near revolt suggests that it may not have priced its services properly.

Network Service, Connection Fees, and the Collapse of the Suffolk System

The principal network benefit accruing to New England's consumers from the Suffolk was the accessability effect. Banknote holders were better off the larger and more geographically inclusive the network because it increased the likelihood that any banknote received in the course of trade would be "current" or traded at par. This is similar to ATM networks described by Saloner and Shepard (1995, p. 480) where ATM users are better off the larger the ATM network because a more geographically expansive network lowers a depositor's cost of obtaining currency. An analog in the Suffolk era was interbank correspondent relationships where banks agreed to accept others'

notes at par, which effectively extended the area across which its notes retained their currency. The larger and wider the network, holding all else constant, the more convenient it was for consumers to use a bank's notes. The more convenient the use of its notes, the greater the demand for them and, under the appropriate cost and demand conditions, the greater the bank's profits.

Because of differences in location, customer spending habits, and the bank's own lending patterns, each bank's optimal correspondent network would be of a different size. A remote bank with few customers who traded outside the immediate neighborhood of the bank would be unlikely to enter into a correspondent agreement because its customers' demands (hence, its own) for such services would have been small. On the other hand, more centrally located banks had a greater derived demand for correspondent redemption agreements. In fact, clearinghouses formed in the 1850s in New York, Philadelphia, and Baltimore (Gorton 1985). Under clearinghouse agreements, each bank sent a representative to an agreed upon location where each presented notes to the other. Doing so avoided the costly duplication of exchanges that would have occurred under bilateral exchange. Moreover, banks generally considered clearinghouse presentations as equivalent to presentations at the bank's own counter so notes were cleared at par.

Establishment of clearinghouses in sparsely populated rural areas was uneconomical. Instead, each bank engaged in irregular clearings with banks in their own vicinity. Notes drawn on distant banks were probably rarely seen and were either redeemed by post or forwarded to a note broker. As country banks grew, as their customers' spending habits changed, as their customers' trading circles widened, and as the bank's notes circulated over an ever-expanding area, the banks' demands for inclusion in a correspondent-redemption network increased. As the payments technology diffused throughout New England and more people were brought into the monetary

economy, the more valuable it became for even remote banks to join a redemption network. The network would have naturally expanded as the cost of subscribing to the network fell or the benefits rose. A reasonable expectation, then, is that banks expecting to use the network more often or to a larger extent would join first. In antebellum New England, these were banks whose customers were engaged in more extensive dealings with the center than with the periphery. A naturally evolving correspondent network would begin small and expand out from the center.

Pre-Suffolk clearing arrangements in New England invariably imposed some of the expenses of clearing and collection on urban correspondents. Such costs included postage, record keeping and, of course, redemption costs. If clearing and redemption services were supplied competitively, the discount on banknotes provides an upper bound to the resource cost (including a risk component) of banknote collections. It was not unusual for interregional discounts to exceed 10, sometimes even 20, percent. Intraregional banknote discounts, however, were usually small, often as little as one-eighth of one percent though they could reach one to two percent on occasion.[19]

City banks bore some of the clearing and redemption costs. The New England/Tremont Bank coalition accepted country banknotes at small discounts and then returned them to the issuing bank, bearing the actual redemption and presentment costs. Early on, the Suffolk adopted a similar policy, and the discount was generally sufficient to cover the bank's costs. Occasionally, however, the Suffolk or one of the other Boston bank suffered losses if a bank failed before its notes were redeemed. Moreover, many country banks refused to enter into redemption arrangements. In other words, an endogenous clearing and collection network was expanding more slowly than city banks preferred. The Suffolk's 1824 plan accelerated the process.

Contemporary complaints about nonpar banknote exchange were not necessarily complaints

about the inefficiency of the system or the total costs associated with banknote redemptions. Rather they were complaints by city banks about the distribution of those costs. Under certain circumstances, nonpar banknote exchange was efficient. A banknote discount was the price paid by a collecting bank to a broker to avoid the cost of presenting the note in person. Using the notation developed in the previous section, we can analyze non-network clearings in the following manner.[20] Let the current banknote discount be d = -p. There were costs (C_i) to the issuing bank of direct presentation (interest costs of redemption balances), and costs (C_n) to city banks of redeeming the notes (record keeping and presentment costs). In the absence of an organized collection network, the internal and external benefits of participating in a nonpar, broker-based redemption scheme were practically nil so that the terms B_i and B_n were effectively zero. Nonpar banknote circulation was efficient if:

$$C_i < d < -C_n.$$

That is, so long as the banknote discount was less than the cost to city banks of direct presentment and redemption, nonpar circulation of country banknotes was efficient. Note, as well, that an efficient discount might exceed the cost of redemption to the issuing bank. The latter condition surely infuriated collecting banks. Thus, nonpar circulation imposed costs on Boston's banks which felt aggrieved at having to bear those costs. The early nineteenth century history of Boston banking was a continual search for a method of shifting more of the costs on to country banks, something the Suffolk finally accomplished.

The Suffolk charged neither explicit network subscription fees nor network usage charges. Instead, it required it members to maintain a deposit that can be separated into two distinct components. Each member was required to maintain a permanent deposit equal to 2 percent of the

bank's paid-in capital; banks with less than $100,000 in capital deposited the $2,000 minimum. In addition, each member was required to maintain a redemption account sufficient to redeem its clearings. The Suffolk loaned the deposited funds at market rates to offset the network's operating costs.

Although balances in the redemption accounts were regularly drawn down, the Suffolk demanded that they were large enough, on average, to cover clearings. When the redemption account was exhausted, the Suffolk returned the member's notes and asked it to replenish its account. Doing so imposed costs on member banks because remittors of funds typically paid commission fees of one-half of one percent for Boston drafts. The Suffolk offered some members overdraft privileges, which Smith and Weber (1999, p. 650) argue constituted a subsidy, but by the late 1830s the Suffolk offered overdrafts only grudgingly and rebuffed banks who abused the privilege.[21]

Not all the network's costs were indirectly paid by country banks. Boston banks were also expected to maintain a permanent deposit. In 1824, this deposit was set at $30,000 regardless of the bank's size (Whitney 1878, pp. 23-25). In 1833 it was lowered to $15,000; it was further reduced to $10,000 in 1834, and finally reduced to $5,000 in 1835. Suffolk members, therefore, were required to keep substantial deposits with the Suffolk, though the Boston banks' share fell markedly in absolute terms and more markedly compared to their aggregate capital.

Given the cross-subsidization inherent in a network, the issue at hand is how the Suffolk distributed the network's costs across its members. The two principal costs of Suffolk membership were the interest foregone from keeping a specie deposit with the Suffolk that could have been used to finance additional lending. The interest cost of the permanent deposit was a clear loss to the subscribing bank because it lost the use of these funds regardless of the Suffolk's effect on banknote

circulation and redemption patterns. Calculating the interest and other costs associated with the redemption account is less straight-forward. Member banks lost some interest income if the redemption account exceeded the specie reserves it would have held in its own vault in the absence of the Suffolk's operations. It is generally accepted that the Suffolk increased redemptions or note turnover, so that in its absence country banks may have expected fewer or slower redemptions. For simplicity, it is assumed that the Suffolk doubled note turnover. Thus, the interest cost of the redemption account is calculated as one-half the average weekly redemption balance times the market interest rate. Finally, country banks, unlike Boston banks, had to pay commissions on drafts to replenish their redemption accounts. It is assumed that they tendered weekly drafts.[22]

Table 2 provides two estimates of the cost of membership as a percent of paid-in capital for four classes of country banks and two classes of Boston banks. One estimate (Columns 1) reports the interest cost of the permanent deposit, which is a lower-bound estimate of Suffolk membership. The second estimate (Columns 2) includes estimates of the interest cost of the permanent deposit, the redemption account, and the commission charges associated with keeping the latter current. The second estimates provide approximate upper bounds of the cost of Suffolk membership.

Two features are immediately evident. First, the lower-bound estimates for 1827 imply that city banks paid proportionately more (as a percent of capital) of the Suffolk's network costs than country banks. Half-million dollar Boston banks, in fact, paid the highest proportionate connection fees, equal to about one-half of one percent of paid-in capital. The upper-bound estimates, on the other hand, show that country banks paid a higher proportionate share, generally between one-third and three-quarters of one percent of paid-in capital, though the differences are not marked. After 1835, and the substantial reduction in the Boston banks' permanent deposit, country banks paid an

even greater, and ever increasing, share of the system's costs. By 1855, the lower bound cost estimates show that a country bank with $100,000 capital paid twice as much as a percentage of capital as a $500,000 Boston bank. The smallest country banks paid nearly four times as much. The upper-bound estimates show that the country banks costs were proportionately even larger. With membership costs to country banks approaching 2.5 to 3 percent of paid-in capital it was not surprising that open revolt broke out in 1855 with the chartering of the Bank of Mutual Redemption. It is also apparent why the Boston banks fought so hard to keep the Suffolk system intact. While a country bank might pay an annual membership price equal to 2 percent of its capital, Boston's banks paid less than one-third of one percent. The Suffolk effectively pushed the costs of clearing and collection on to country banks, and more so (as a percentage of capital) the smaller the bank.

Recall that efficient network pricing involves cross-subsidization where high net benefit subscribers subsidize the participation of low net benefit users. Boston's banks were high individual but low external benefit subscribers, along with those country banks whose customers engaged in extensive trade with the center. The external benefit generated by the participation of a Boston bank was relatively low because it was not significantly more costly to redeem a Boston banknote at the Suffolk or to walk across State Street and tender it at the issuing bank's own counter. The low individual but high external benefit subscribers, on the other hand, were remote country banks with relatively few interactions with the center, and with relatively high direct presentation costs. Efficient pricing in the Suffolk network, therefore, would have enticed remote members to subscribe by waiving a connection fee (permanent deposit) and possibly even its usage fee (redemption deposit). But connecting remote country banks imposed costs on the network provider, costs it should have recovered from high individual benefit users (Boston banks) by charging them

connection and usage fees in excess of the incremental costs they imposed on the network provider. Instead, the Suffolk effectively reversed the pricing mechanism. Large city banks paid a proportionately small share of the network's costs, even while small country banks paid a proportionately large share. Moreover, by reducing the permanent deposit required of Boston banks beginning in 1833, the Suffolk pushed even more of the costs on to low individual and high external benefit users.

Viewed from a network perspective, the Suffolk's method of allocating the common costs takes on a perverse aspect. With little to gain from participating in the network, small rural banks were reluctant to join, which reduced the usefulness of the Suffolk as a payments system network. Because the Suffolk mispriced its services it could not elicit voluntary membership, but it recognized that the system would be effective only if it attained nearly universal participation. If country banks failed to join of their own volition, how did it attract members? Instead of prices, the Suffolk used threats, intimidation, and coercion. Almost all commentators on the Suffolk have portrayed it from the outset as unpopular with most of its members. Its activities were viewed as war against country banks waged by an "Holy Alliance" of Boston banks. The Suffolk employed a very effective device against country banks reluctant to join -- an organized bank run. If a bank, like the previously discussed Worcester Bank, refused to join, the Suffolk purposively collected a large volume of its outstanding circulation that it presented without warning. Few banks could meet such large calls on such short notice, and the Suffolk typically asked the sheriff to attach the bank's property, notified the state's bank commissioners that it was unable to redeem its notes (an offense that could be punished by charter revocation), and sued the bank as a bankrupt. Most relented and joined the system rather than risk further embarrassment, charter revocation, or another Suffolk

attack.

There were clearly instances where the Suffolk employed these tactics when others may have worked. The Suffolk continually received applications for reduced permanent deposits or for interest-bearing accounts. The bank curtly rebuffed every such application, saying that no preferences would be given any member. In the summer of 1852, for example, the South Royalton Bank of Vermont enquired into making an arrangement with the Suffolk on better than standard terms. The Suffolk refused and notified the bank that unless it made arrangements to redeem its notes immediately, the Suffolk would send back $10,000 of its notes for redemption. An agent was dispatched to South Royalton, but the bank had him arrested and filed a civil case against the Suffolk. Apparently, the Suffolk responded by punishing some of the South Royalton's allies in Vermont, because at "the request of the other banks in Vermont the difficulties were finally adjusted, [with] the South Royalton Bank acceding to the usual terms" (Whitney 1878, p. 45), but not before the South Royalton had pushed its case all the way to Vermont's Supreme Court.

In hiring agents to travel to country banks and redeem their notes at par at their counters, the Suffolk was acting outside normal banking practice. It was likely that the costs of these redemption raids exceeded the costs of nonpar acceptance of country banknotes. If it were otherwise, private clearing operations elsewhere would have pursued the same strategy as the Suffolk. To the best of our knowledge none did, which explains why no Suffolk-like system arose elsewhere. A pale shadow of the system was recreated in New York in the 1840s, but it was imposed by legislative fiat and allowed competitively supplied clearing and collection arrangements as well as nonpar circulation of country banknotes. A system of voluntary clearing arose in Pennsylvania at about the same time the Suffolk appeared, but it had neither universal participation nor universal par

circulation of banknotes.

The Suffolk's actions against banks unwilling to join its network foreshadowed actions by the Federal Reserve in the late 1910s when it engaged in similar practices in an attempt to establish nearly universal par check collection (Lacker, Walker and Weinberg 1999). Fed members were required by law to accept checks at par, but nonmembers could choose to join its network or not. From the Fed's perspective, adoption of par check collection was infuriatingly slow. To expand participation, several Reserve banks collected checks on nonpar banks and presented them en masse at the bank's counters where they were required by law to pay par. As a result, the number of participating nonmember banks quickly increased from about 10,000 to about 19,000. By the end of 1920, only about 1,700 nonpar banks remained.

Opposition to the Fed's actions, like those of the Suffolk's a century earlier, was fierce. Some banks refused to join and sued the Fed, and several cases eventually reached the U.S. Supreme Court. The Court found in the nonpar banks' favor. Reserve Banks could no longer accumulate checks of nonpar banks and coerce them into participating in the Fed's presentation network. The Court also determined that country banks could pay in drafts on urban correspondents rather than lawful money. The Federal Reserve Board subsequently ordered Reserve Banks to abandon their practices. Not surprisingly, banks withdrew from the par collection network until by 1928 the number of nonparticipating banks reached 4,000. It was not until the 1980s that nonpar collection finally disappeared entirely (Lacker, Walker, and Weinberg 1999, p. 11).

In the Suffolk's heyday, only Maine allowed its banks to resist the Suffolk's organized bank runs. By allowing its banks a 15-day grace period to meet unusual (read Suffolk) redemption calls, Maine's legislators effectively by-passed the twin requirement of immediate redemption and

redemption in lawful money, just as the U.S. Supreme Court did a century later. A 15-day grace period allowed its banks an opportunity to draw on funds in Boston and meet the call. Other states, like Vermont, were less willing to make concessions to its banks. Vermont even encouraged participation by placing a special tax on nonmembers. That the Suffolk's network connection fees were too high is seen by the fact that several of Vermont's banks opted to pay the tax. Thus the Suffolk established a privilege not available to other collecting banks. "Exercising that privilege had the effect of shifting the allocation of common costs of [banknote redemption] away from collecting banks and toward small country banks (Lacker, Walker and Weinberg 1999, p. 2).

Concluding Comments

The Suffolk had two principal motivations in establishing its region-wide clearing and collection network. First, the circulation of banknotes at a discount imposed collection and redemption costs on Boston's banks, which the Suffolk aimed to reduce. Such costs, however, do not necessarily imply payments system inefficiency. Under certain conditions, notes circulating at a discount are consistent with efficiency. Second, contemporary complaints about the costs of nonpar circulation of banknotes were not necessarily complaints about the aggregate costs of nonpar circulation. It is important to note that most complaints originated from city banks. Switching from nonpar to par circulation of banknotes did not reduce the costs of note collection, it restructured them, shifting most of the costs of note clearing, collection, and redemption from city to country banks. With par redemption in Boston, city banks paid a relatively small share of the common network costs while country banks paid a relatively large share. This was a reversal of the situation when country banknotes exchanged at a discount, and it effectively inverted efficient network pricing

rules.

There were a number of methods that might have been used by the Suffolk to elicit membership of institutions providing large external benefits. It could have waived the permanent deposit requirement for some banks; it might have paid some below-market interest rate on the permanent deposit of others; and it may have paid an above-market rate on the deposit, effectively charging it a negative membership fee for a select few. Such a pricing strategy implies that remote country banks -- those banks with the largest banknote discounts and the highest presentation costs in a freely operating market -- would have paid little to join the network. The same pricing rule would have charged higher fees to Boston's banks. But because Boston's banks were searching for a method of reducing its clearing, collection, and redemption costs, the Suffolk could not impose this pricing scheme and maintain the support of Boston's banks. Faced with an intractable pricing problem, the Suffolk opted for the next-best solution. What it could not accomplish through an appropriate set of prices, it accomplished through coercion. And since it was unlikely to coerce its Boston neighbors through organized bank runs, it was forced to threaten and intimidate country banks.

The fundamental flaw in the Suffolk's plan, as Weinberg (1997, pp. 34-36) and Henriet and Moulin (1996) make clear, is that every network operator must respect the ability of its participants to form a competing network. Some members will almost inevitably generate large external benefits for others, and those members may need to receive a share of the net benefits that appears inordinately large relative to the price they pay. Unless, the network operator makes this concession, it faces the risk that its members will defect and form a competing network that recognizes those benefits and prices them accordingly. The Suffolk failed to set appropriate prices and its members,

particularly those whose participation generated the largest external benefits (namely, country banks), established a competing network whose prices better reflected the external effects inherent in a network.

References

Bodenhorn, Howard. "The End of the Suffolk System: A Reappraisal." Working paper, Lafayette College, June 1998.

Calomiris, Charles W. and Charles Kahn. "The Role of Demandable Debt in Structuring Optimal Banking Arrangements." American Economic Review 81:3 (June 1991), 497-513.

Calomiris, Charles W. And Charles Kahn. "The Efficiency of Self-Regulated Payment Systems: Learning from the Suffolk System." Journal of Money, Credit, and Banking 28:4 (November 1996), 766-97.

Chadbourne, Walter W. A History of Banking in Maine, 1799-1930. Orono, Maine: University of Maine Press, 1936.

Dewey, Davis R. State Banking Before the Civil War. Washington, D.C.: GPO, 1910.

Economides, Nicholas. "Network Externalities, Complementarities, and Invitations to Enter." European Journal of Political Economy 12:2 (September 1996), 211-33.

Economides, Nicholas. "The Economics of Networks." International Journal of Industrial Organization 14:6 (October 1996), 673-99.

Economides, Nichoas and Lawrence J. White. "Networks and Compatibility: Implications for Antitrust." European Economic Review 38:3 (April 1994), 651-62.

Fenstermaker, J. Van. The Development of American Commercial Banking: 1782-1837. Kent, Ohio: Kent State University Bureau of Economics and Business Research, 1965.

Fenstermaker, J. Van and John E. Filer. "Impact of the First and Second Banks of the United States and the Suffolk System on New England Bank Money, 1791-1837." Journal of Money, Credit, and Banking 18:1 (February 1986), 28-40.

Gorton, Gary. "Clearinghouses and the Origin of Central Banking in the United States." Journal of Economic History 45:2 (June 1985), 277-83.

Gorton, Gary. "Reputation Formation in Early Bank Note Markets." Journal of Political Economy 104:2 (April 1996), 346-97.

Gorton, Gary. "Comment on 'Private Money Creation and the Suffolk Banking System." Journal of Money, Credit, and Banking 31:3 (August 1999, part 2), 663-67.

Hasse, William F., Jr. A History of Banking in Connecticut. New Haven: privately printed, 1957.

Henriet, Dominique and Herve Moulin. "Traffic-Based Cost Allocation in a Network." RAND Journal of Economics 27:2 (Summer 1996), 332-45.

Knox, John Jay. A History of Banking in the United States. New York: Bradford Rhodes & Company, 1903.

Lacker, Jeffrey M., Jeffrey D. Walker, and John A. Weinberg. "The Fed's Entry into Check Clearing Reconsidered." Federal Reserve Bank of Richmond Economic Quarterly 85:2 (Spring 1999), 1-32.

Lake, Wilfred S. "The End of the Suffolk System." Journal of Economic History 7:4 (December 1947), 183-207.

Massachusetts. General Court. "True Abstract of the Returns of Several Banks to the Secretary of

the Commonwealth." Boston: 1827, 1835, 1845, 1855.

McAndrews, James J. "Network Issues and Payment Systems." Federal Reserve Bank of Philadelphia Business Review (November/December 1997), 15-25.

Mullineaux, Donald J. "Competitive Monies and the Suffolk Bank System: A Contractual Perspective." Southern Economic Journal 53 (April 1987), 884-898.

Myers, Margaret G. The New York Money Market: Volume I, Origins and Development. New York: Columbia University Press, 1931.

Redlich, Fritz. The Molding of American Banking: Men and Ideas. New York: Hafner Publishing Company, 1947.

Rhode Island. General Assembly. "Abstract of the Returns from the Several Banks in the State of Rhode Island, made to the General Assembly." Acts and Resolves (1827, 1835, 1845, 1855).

Rolnick, Arthur J., Bruce D. Smith, and Warren E. Weber. "Lessons from a Laissez-Faire Payments System: The Suffolk Banking System (1825-58)." Federal Reserve Bank of Minneapolis Quarterly Review 22:3 (Summer 1998), 11-21.

Root, L. Carroll. "New England Bank Currency." Sound Currency 2:13 (June 1895), 253-83.

Saloner, Garth and Andrea Shepard. "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines." RAND Journal of Economics 26:3 (Autumn 1995), 479-501.

Simonton, William G. Maine and the Panic of 1837. Unpublished Master's thesis, University of Maine, 1971.

Smith, Bruce D. and Warren E. Weber. "Private Money Creation and the Suffolk Banking System." Journal of Money, Credit, and Banking 31:3 (August 1999, part 2), 624-59.

U.S. Comptroller of the Currency. Annual Report of the Comptroller of the Currency. Washington, D.C.: GPO, 1876.

Weinberg, John A. "The Organization of Private Payments Networks." Federal Reserve Bank of Richmond Economic Quarterly 83:2 (Spring 1997), 25-43.

Whitney, David R. The Suffolk Bank. Cambridge, Mass.: Riverside Press, 1878.

Endnotes

1. Mullineaux (1987), Calomiris and Kahn (1996), Bodenhorn (1998), Rolnick, Smith and Weber (1998), and Smith and Weber (1999).

2. Whitney (1878), Dewey (1910), Redlich (1947), and Hammond (1957).

3. The issues are related because, as Gorton (1999, p. 666) notes, if the Suffolk was as welfare enhancing as commonly portrayed, it is hard to explain why it was not longer lived or imitated elsewhere.

4. The extent to which country banknotes displaced city banknotes depended on the cross elasticity of demand between the two. The historical presumption is that the two were close substitutes which, prior to the establishment of the Suffolk system itself, seems questionable.

5. Gorton (1996) provides a detailed study of antebellum note brokerage.

6. Chadbourne (1936, pp. 41-42) reprints an agreement between the Augusta Bank of Maine and the Boston brokerage of Wyman and Stone, who agreed to recirculate the bank's notes after taking them in at a discount not to exceed 1 percent.

7. Calomiris and Kahn (1996, p. 788) argue that low specie reserve holdings by country banks counters the argument that country banks lived in constant fear of unexpectedly large unannounced redemption calls by the Suffolk. Actually, low specie reserve holdings do not show anything of the sort once most of the region's banks had joined the system. The Suffolk and its allies only employed the threat against banks that refused to join or threatened to withdraw. It was against the basic covenant of the Suffolk agreement to engage in unannounced redemption calls at a member bank's counter, except in cases where the member abused its overdraft privileges or was on the verge of collapse and the Suffolk was acting as the informed depositor discussed in Calomiris and Kahn (1991).

8. Statements made by bank commissioners in this era need to be read cautiously. Regulatory capture was very real in the antebellum era. It was not uncommon for the banks themselves to elect bank commissioners and many commissioners were retired bankers.

9. Myers (1931) provides a description of antebellum New York's money market and the clearing system.

10. Economides and White (1994, p. 652).

11. Mester (1987, p. 18) recognizes this for airlines operating hub and spoke networks.

12. Except where otherwise indicated, this discussion is taken from Lacker, Walker and Weinberg (1999), pp. 16-18.

13. Henriet and Moulin (1996, pp. 343-45) provide a derivation in the same spirit.

14. Rolnick, Smith and Weber (1998) show that the Suffolk generated profits far in excess of other Boston banks. Moreover, Fenstermaker and Filer (1986) show that the Suffolk did not place an effective brake on banknote expansion. This is consistent with Mullineaux's (1987) argument that the Suffolk may have increased the demand for country banknotes by making them near-perfect substitutes for Boston banknotes.

15. This and the next paragraph follow the discussion in Weinberg (1997, p. 38).

16. Redemption by mail was not unheard of, but was probably avoided in antebellum America. Given the risks of returning banknotes through the mail, the sending bank often tore the banknotes in half, sending the halves in different packages. Postal redemption thus proved expensive to all involved, including the issuing bank that effectively lost the use of the torn banknotes and had to have new ones printed.

17. Economides (1996, pp. 212-13, 221) argues that in the face of significant network externalities, a monopolist may have an incentive to invite competition. The intuition is that the monopolist cannot convince potential members that it will expand the network to an efficient size, because monopolists profit by restricting output. By inviting competition, potential members are more likely to subscribe to a network, which expands the network size and, under the appropriate conditions, implies larger profits for the monopolist. As will be seen below, the Suffolk found an alternative solution to the monopoly problem.

18. Bodenhorn (1998) reports evidence that the Suffolk exploited economies of scale in note clearing.

19. This is about the same proportionate cost as collecting nonpar checks in the 1920s (Lacker, Walker, and Weinberg 1999, p. 6).

20. Lacker, Walker, and Weinberg (1999, p. 19) analyze nonpar check collection in the early days of the Federal Reserve System similarly.

21. Smith and Weber (1999, p. 650) provide a back-of-the-envelope calculation suggesting that the subsidy would have taken the form of a below-market rate on overdrafts, but the rate could not have been less than 0.83 percentage points below the market loan rate. The subsidy was thus small indeed. If a country bank was never overdrawn by more than its permanent deposit (a Suffolk limit), the subsidy could not have exceeded $16.60 per annum for a bank with $100,000 paid-in capital. And we know that the subsidy was smaller than this because the Suffolk did not allow the overdraft to run for more than a few weeks. The effective subsidy, with only periodic overdrafts was more likely to have amounted to less than $5 per annum, perhaps considerably less.

22. Tables detailing the calculation of these figures are available from the author on request.