

# The Fed's Entry into Check Clearing Reconsidered

---

Jeffrey M. Lacker, Jeffrey D. Walker, John A. Weinberg

The Federal Reserve's check clearing business has been a significant part of its operations since its founding and has set the precedent for other payment services offered by the Federal Reserve (Fed) to depository institutions. While there is considerable debate about the proper role for the Fed in the modern payment system, there seems to be much less disagreement concerning the Fed's entry into check collection.<sup>1</sup> Many scholars believe that when the Fed took on the check payment function during the first decade of the institution's existence, its entry served to enhance the efficiency of the check payment system. Indeed, Fed documents on its role in the payment system speak of the "breakdown of the check collection system" around the turn of the century.<sup>2</sup> According to the conventional view, check collection prior to the founding of the Fed was decentralized and defective in a number of ways. By centralizing the system, the Fed was able to eliminate many of the defects.

Our purpose in this article is to reexamine the facts concerning the Fed's entry into check clearing and to evaluate the conventional view in light of those facts. We find that the evidence of inefficiency in the pre-Fed check collection system is inconclusive. Further, inefficiency would imply that there was some form of market failure, yet most discussions of check clearing in the early part of this century are vague or silent on possible sources of market failure. Absent a clearly articulated explanation of why participants in the check collection system failed to achieve efficient results, we find the conventional view to be unconvincing.

---

■ The authors wish to thank Ned Prescott, John Walter, Alex Wolman, and Tom Humphrey for helpful comments. The views expressed do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. The authors remain solely responsible for the contents of this article.

<sup>1</sup> See, for example, Benston and Humphrey (1997).

<sup>2</sup> Board of Governors of the Federal Reserve System (1990).

We propose a different interpretation of the Fed's entry, one based on the network characteristics of check collection. Under this alternative view, pre-Fed check collection arrangements were relatively efficient, and the complaints many observers voiced about the excessive costs of the pre-Fed system should be understood as complaints about the distribution of the system's costs rather than about its aggregate costs. Moreover, this interpretation explains why the founders felt compelled to give the Reserve Banks check clearing powers, given the reserve requirements in the Federal Reserve Act. New light is also shed on the par collection controversy, and the tortuous process, spanning several years, by which the Reserve Banks established their check collection service. Hence, the economics of network organization gives a coherent account of the facts concerning check collection prior to the founding of the Fed and the process by which the Fed entered the check clearing business.

Before presenting our view of the Fed's entry into check clearing, we present some of the key facts and review the conventional view. The relevant facts concern the system before the Fed and the means by which the Fed became a significant provider of check clearing services. Section 1 discusses nonpar remittance and the importance of correspondent relationships for clearing out-of-town checks in the pre-Fed system. Nonpar remittance occurs when a bank on which a check is drawn pays a collecting bank less than the par value of the check. Correspondents are banks, usually larger city banks, that perform a variety of services, including check collection, for other banks.

The Fed's ability to successfully penetrate the check clearing market was dependent on Congress giving it the proper authority. While the Federal Reserve Act authorized the Reserve Banks to clear checks, the Fed's initial attempts at inducing fully voluntary participation by member banks were unsuccessful, as we discuss in Section 2. It was not until Congress granted the Reserve Banks a competitive advantage—the sole right to present by mail at par—that the Fed was able to become a significant participant in the market.

Sections 3 and 4 deal with the conventional view, according to which many of the features of the correspondent banking system represent inefficiencies that existed because of nonpar payment. Chief among these features was the observation that occasionally a check would pass through the hands of many widely dispersed intermediaries on its way to the paying bank. Such "circuitous routing" is often cited as unambiguous evidence of inefficiency. However, circuitous routing is quite consistent with our alternative view of the pre-Fed system as efficient, as we discuss in Section 5. In Section 6 we reconsider the Fed's entry into check clearing in light of our alternative view. The Fed was able to gain market share because it enjoyed a legal privilege in presentment that was unavailable to private collecting banks. Exercising that privilege had the effect of shifting the allocation of the common costs of check clearing away from collecting banks and toward small country banks and taxpayers.

## 1. CHECK CLEARING BEFORE THE FOUNDING OF THE FED

By the mid-nineteenth century, the use of checks had become a prominent means of payment in American banking and commerce. As early as 1855, the value of checkable deposits exceeded the value of bank notes in circulation (Spahr 1926, p. 84). In this earlier period, however, there were distinct geographic differences in payment practices. Checks were used primarily for payments within cities, and notes were used predominantly in the countryside.<sup>3</sup> Payments between the country and the city and across geographic regions were made using bank drafts. A bank draft is like a check, except it is drawn on an account held by one bank with another. Hence, a Midwesterner wishing to make a purchase from an East Coast city would go to his local bank and purchase a draft drawn on that bank's balances held with a bank in an eastern city.

Toward the end of the nineteenth century, checks began to replace drafts as means of payment in interregional transactions, a change that was nearly complete by the turn of the century (Preston 1920, p. 566). During this same period, the check also replaced the bank note as a means of payment among people in the countryside. By 1900, the value of demand deposits was more than quadruple the value of currency in circulation (Friedman and Schwartz 1963, p. 705). In the period just prior to the founding of the Fed, the check had become a dominant payment instrument for making both long distance and local payments.

Check clearing involves the delivery of items to the banks on which they are drawn. Many observers have pointed out that, in the United States prior to the existence of the Fed, clearing was affected by the different legal treatments accorded to different forms of delivery.<sup>4</sup> While a paying bank was obligated to make payment (remit) at par for checks presented in person (over the counter), there was no such obligation for checks presented through the mail.<sup>5</sup> Banks were free to extract a presentment fee (exchange charge) from their payments on such indirect presentments. This practice of *nonpar banking* is the focus of many discussions of the Fed's entry into the check collection business. The prospects of receiving less than par for a check gives a collecting bank an incentive to find a way of getting the check to the bank on which it is drawn without mailing it directly. Alternative means of transport and presentment will be used, presumably, if the total cost does not exceed the presentment fee.

---

<sup>3</sup> Spahr (1926, p. 60); Federal Reserve Bank of Richmond (1922, p. 1).

<sup>4</sup> Spahr (1926, pp. 103–05); Baxter (1983, p. 559); Duprey and Nelson (1986, p. 20); Summers and Gilbert (1996, p. 4); Weinberg (1997, p. 38); Gilbert (1998, pp. 123, 129); James (1998, p. 143).

<sup>5</sup> See the discussion in Spahr (1926, pp. 103–04). The requirement that checks presented over the counter be paid at par had its origins in English common law.

For a check drawn on a nearby bank, in-person presentment was relatively cheap. Hence the clearing and settlement of local checks presented few challenges for the banking system of the United States.<sup>6</sup> In the larger cities a number of banks typically would have frequent business with one another, making it worthwhile for them to form cooperative clearing organizations such as clearinghouses.<sup>7</sup> In such an arrangement, the banks' representatives would meet daily at a designated location to exchange items drawn on each other, avoiding the duplicative cost of bilateral contacts. By mutual agreement among participants, presentment at the clearinghouse was taken to be equivalent to presentment in person at the paying bank's premises.<sup>8</sup>

In less densely populated areas, a typical bank had a relatively small number of banks nearby from which it regularly received checks. Maintaining a clearinghouse arrangement among the smaller number of banks did not tend to be economical. Instead, rural banks usually made direct exchanges with other banks in the same general area (Spahr 1926, p. 98). Such direct presentations were made once or twice a week, or daily, depending on the distance between the banks and on the volume of checks flowing between them. Settlement of local checks, either in the city clearinghouses or among banks in rural communities, could be made through the exchange of currency or by debiting interbank balances (Cannon 1900, pp. 36–46).

The clearing of out-of-town, or interregional, checks was accomplished via a network of bilateral agreements on clearing terms that generally took the form of *correspondent banking* relationships.<sup>9</sup> This relationship involved a bank in a larger city serving as a *correspondent* bank for a bank in a smaller city or town. The latter bank, the *respondent*, would hold balances with its correspondent. When the respondent received a check drawn on a bank in the city or area where the correspondent conducted business, the respondent would send the check to the correspondent, who would, in turn, present the check directly to the paying bank. The correspondent would receive payment from the paying bank and credit the amount to the respondent's account. Often, the correspondent would credit the respondent for the par value of the check even if the paying bank did not remit at par (Spahr 1926, pp. 101, 111). Sometimes the respondent agreed, in return, to remit at par on checks sent to it by its correspondent. The correspondent's main form of compensation was typically the interest margin it could earn on the funds held as balances by the respondent bank (Spahr 1926, pp. 101, 111–12). It was not uncommon for a correspondent

---

<sup>6</sup> Spahr (1926, p. 98); Duprey and Nelson (1986, p. 19).

<sup>7</sup> Cannon (1900, pp. 148–54); Spahr (1926, pp. 79–82).

<sup>8</sup> Hallock (1903, p. 59); Spahr (1926, pp. 104–05).

<sup>9</sup> For descriptions of correspondent banking at the Fed's founding, see Watkins (1929, Ch. 6), Spahr (1926, pp. 99–101).

to pay presentment fees to its respondents while paying them par for all the checks it collected for them.

Often a bank would act as correspondent for banks outside the region, offering to collect checks drawn on any bank in the neighboring territory, with the proceeds credited to the account of the distant bank. There appears to have been active competition for collection business between correspondent banks, as evidenced by the many advertisements in bank directories from the late nineteenth century.<sup>10</sup>

The importance of correspondent relationships is reflected in the magnitude of interbank balances. In the years preceding the Fed's founding, the amount of deposits at national banks that were held for other banks was roughly 40 percent as large as national bank deposits held for individuals (Watkins 1929, pp. 10–18). These interbank balances were held predominantly at banks in the larger cities and especially in the major financial centers.

Correspondent banks, then, were linked together into a network of banks through which checks were collected. When a bank received a check drawn on an out-of-town correspondent bank—what might be termed a *regular* inter-regional check—it would be presented through the established clearing arrangement.<sup>11</sup> When a bank received a check drawn on a distant bank with which it did not have an established relationship—what might be termed an *irregular* interregional check—the check would most likely be sent on to a correspondent. If a correspondent received from a respondent a check drawn on a paying bank with which the correspondent did not have a relationship, then the correspondent would typically send the check to one of its correspondent banks located near the paying bank.<sup>12</sup> In fact, to aid routing, bank directories listed each bank's correspondents; the bank holding the check could look for correspondents it shared with the paying bank.<sup>13</sup> The next bank receiving the check might present it directly to the paying bank. Such indirect routing had two advantages for the correspondent. First, it avoided having to pay a presentment fee to a paying bank from which it received no compensating benefit. Second, it saved the cost of sending a single item to the paying bank instead of bundling the item with others being sent on a normal shipment (Cannon 1900, p. 76).

---

<sup>10</sup> Williams (1901). Typical notices: “Prompt and careful attention given to collections throughout Mississippi and Alabama” (First National Bank of Meridian, Mississippi); “Unsurpassed facilities for handling collections, especially Oklahoma and Kansas” (Kansas National Bank, Wichita, Kansas); “Send us your Southwest Collections” (Home Savings Bank & Trust Co., Phoenix, Arizona).

<sup>11</sup> The term “correspondent” is sometimes used narrowly to refer to a small number of distant banks that a bank formally designates to receive items drawn on it. A bank typically had no more than two or three correspondents in this sense, and they were listed in banking directories. The term is often used more broadly, however, to refer to any bank with which a bank regularly exchanges items by mail. We will use the term “correspondent” in this broader sense.

<sup>12</sup> Cannon (1900, p. 76); Spahr (1926, pp. 111–12).

<sup>13</sup> Williams (1901).

Irregular interregional items, then, were often handled by sending them to banks with whom the correspondent conducted regular interregional business. Concerns about the operation of the check clearing system prior to the founding of the Fed focused largely on these irregular interregional checks. In such a system, there were inevitably cases in which the next bank to receive a check was not a correspondent of the paying bank. Hence, the process of sending the check to a correspondent might be repeated more than once. Furthermore, some mistakes in judgment and in handling of items were inevitable. Accordingly, there are documented examples of checks traveling circuitous routes, through the hands of many banks, in making their way from the bank of first deposit to the paying bank.<sup>14</sup>

Settlement of an irregular interregional check began when the paying bank remitted to the bank that finally presented the check. If the presenting bank was a correspondent for the paying bank, then settlement could be made by debiting the paying bank's account balances. If there was no account relationship between the paying and presenting banks, then payment would typically be made in the form of a draft on the paying bank's account with one of its correspondents, often a New York bank (most banks maintained a relationship with at least one New York bank).<sup>15</sup> The presenting bank could send the draft to its New York correspondent who would credit the amount to the presenting bank's account. The presenting bank's New York correspondent would, in turn, present the draft to the paying bank's correspondent through the New York clearinghouse. In this way, New York reserve balances served increasingly as a universal settlement medium. This use of bank drafts in settlement stands in contrast to the greater use of specie earlier in the nineteenth century.<sup>16</sup>

Collecting and paying banks incurred a variety of expenses in the process of clearing and settling interregional checks. The resource costs incurred by collecting banks, including the costs of recordkeeping and postage, were estimated to have amounted on average to about 3/4 of 1 percent to 1 percent of the value of collected items (Spahr 1926, p. 113). Other costs borne by collecting banks resulted from the negotiated arrangements for clearing and collection. For instance, many collecting banks gave credit to their account holders at the time of an item's deposit. By doing so, the collecting bank would incur the float costs that accrued until settlement was received from the paying bank. In addition, national banks faced a tax of 1/2 of 1 percent on deposits. These costs, together with the presentment fee charged by some paying banks (1/8 of 1 percent on average), added roughly another 3/4 of 1 percent to the costs incurred by collecting banks (Spahr 1926, p. 114). On the other side of the

---

<sup>14</sup> We discuss the examples of circuitous routing below.

<sup>15</sup> Cannon (1900, p. 46); Spahr (1926, p. 100); Watkins (1929, p. 104); Williams (1901, *passim*). Alternatively, payment could be made in specie or banknotes.

<sup>16</sup> Preston (1920, p. 565); Spahr (1926, pp. 45–51).

transaction, paying banks incurred clerical costs in receiving and remitting for checks presented on them (Langston 1921, pp. 13–39).

It seems apparent that a substantial share of the cost of clearing inter-regional checks was borne by collecting banks (Spahr 1926, p. 113). Customers of collecting banks typically received the full par value for their deposits of out-of-town checks. Even if the collecting bank passed along some charge to the depositor of the check, there appears to be little evidence of systematic price discrimination by businesses between customers paying for goods and services with out-of-town checks and those paying by other means. Most observers conclude that competition for both local deposits and correspondent business drove city banks to absorb much of the cost of collecting country checks.<sup>17</sup>

## 2. THE FED'S ENTRY INTO CHECK CLEARING

A flurry of banking and monetary reform proposals around the turn of the century ultimately led to the passage of the Federal Reserve Act on December 23, 1913. The central motive of the Act, and many other reform proposals as well, was to prevent recurrent financial panics of the kind typical of the late-nineteenth and early-twentieth centuries in the United States. The Federal Reserve System was designed to prevent such panics by providing “an elastic currency” that allowed a relatively rapid expansion of the supply of notes when needed.<sup>18</sup>

Nationwide Federal Reserve check clearing was not envisioned in the earliest versions of the Act introduced in Congress.<sup>19</sup> The bill introduced by Senator Carter Glass in February 1913, for example, required Reserve Banks to accept from their members at par any checks drawn on any other member's deposit at a Reserve Bank. This would have allowed member banks to settle obligations to other member banks with checks drawn on their own Reserve Bank deposits, much the way interbank obligations were settled using drafts drawn on bank deposits at New York banks. Ultimately, however, the final version of the Federal Reserve Act allowed Reserve Bank check clearing. Each Reserve Bank was required to accept checks written by *depositors* at member banks, and was permitted, though not required, to accept checks written by depositors at member banks of other Reserve Banks.<sup>20</sup>

---

<sup>17</sup> Watkins (1929, p. 106); Fellows (1940, p. 18); Miller (1949, p. 11); Duprey and Nelson (1986, p. 20); Summers and Gilbert (1996, p. 4).

<sup>18</sup> See Willis (1923) or Timberlake (1978), for example.

<sup>19</sup> See Stevens (1996) for an account of the legislative history of the check clearing provisions of the Federal Reserve Act.

<sup>20</sup> Board of Governors (1915, pp. 23–44). Section 16 of the Federal Reserve Act as finally passed on December 23, 1913, stated, in part, “Every Federal reserve bank shall receive on deposit at par from member banks or from Federal reserve banks checks and drafts drawn upon any of its depositors, and when remitted by a Federal reserve bank, checks and drafts drawn by

It was a number of years before the Reserve Banks successfully penetrated the check collection business. On March 4, 1915, the Federal Reserve Board announced that it had directed the Reserve Banks to establish a “voluntary reciprocal plan” for intradistrict check clearing.<sup>21</sup> Member banks would be able to collect at par from all other member banks in their district that joined the plan. A member bank joining the plan had to agree to accept their own checks at par. Details of the scheme were left up to the individual Reserve Banks. The hope was that many member banks would join, making the plan attractive to those that at first held out. The voluntary reciprocal plan proved disappointing, however. By late July the Reserve Banks were reporting that most banks said they would not join unless the plan were made mandatory. Reported participation in the plan peaked in October 1915 at only 2,456 banks, out of about 7,600 member banks, and withdrawals exceeded additions every month thereafter.<sup>22</sup>

A new plan was introduced early the next year. On May 1, 1916, the Federal Reserve Board released a circular to member banks—Circular No. 1, Series of 1916—detailing a “compulsory plan” to be put into effect in June (later postponed to July 15).<sup>23</sup> Member banks would now be *required* to pay at par on checks presented to them by their Reserve Bank. Presentation by a Reserve Bank through the mail would be construed as presentation at their counters. The Reserve Bank would defray the paying bank’s cost of sending payment, either in notes or acceptable checks on other banks, if the bank’s reserve balance was insufficient. Each Reserve Bank would accept checks from

---

any depositor in any other Federal reserve bank or member bank upon funds to the credit of said depositor in said reserve bank or member bank. Thus each Reserve Bank was *required* to accept on deposit *at par* from member banks or other Reserve Banks checks and drafts drawn upon any of its depositors. Section 16 went on to state, “The Federal Reserve Board . . . may at its discretion exercise the functions of a clearing house for such Federal reserve banks, or may designate a Federal reserve bank to exercise such functions, and may also require each bank to exercise the functions of a clearing house for its member banks.” Section 13 adds that “Any Federal Reserve Bank may receive from any of its member banks . . . checks and drafts upon solvent member banks, payable upon presentation; or solely for exchange purposes, may receive from other Federal reserve banks . . . checks and drafts upon solvent member or Federal reserve banks, payable upon presentation.” Thus each Reserve Bank was *permitted* to accept from its member banks checks or drafts drawn upon solvent member banks, or, “solely for exchange purposes,” checks or drafts from other Federal Reserve Banks drawn upon solvent member banks or other Federal Reserve Banks.

<sup>21</sup> Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin* (1915, pp. 6–9) [hereafter “*Federal Reserve Bulletin*”].

<sup>22</sup> *Federal Reserve Bulletin* (1915, pp. 192–95); Spahr (1926, pp. 174).

<sup>23</sup> *Federal Reserve Bulletin* (1916, pp. 259–60, 262–64). Each Reserve Bank issued its own circular detailing the operation of the plan in its district; Chicago’s is reprinted in the *Federal Reserve Bulletin* (1916, pp. 312–14). Regulation J was released later in 1916, incorporating the September 7, 1916, amendment (see below) and superceding Circular No. 1. To this day Regulation J codifies the Board’s requirement that member banks accept checks at par (see Board of Governors [1998], 12 CFR 210.9).

its member banks at par if they were drawn on other member banks or on nonmember banks that paid at par, although a small per-item service charge was allowed.<sup>24</sup> The plan was nationwide in scope; each Reserve Bank would accept at par checks drawn on banks in other districts if they were Federal Reserve members or they agreed to pay at par. Member banks were still free to collect checks outside of the Federal Reserve System as they saw fit, and they were free to carry balances with other banks for purposes of clearing. They were still free to charge presentment fees as well, but the Board's Circular prevented them from charging presentment fees to the Reserve Banks.

The original Federal Reserve Act did not permit the Reserve Banks to accept checks drawn upon nonmembers. Early in 1916 the Board recommended to Congress various changes in the Act, including alteration of Section 13 of the Act to allow Reserve Banks to accept "checks and drafts payable upon presentation within its district."<sup>25</sup> The amendment passed without change on September 7, 1916. The amendment also removed the qualification that Reserve Banks could only accept checks drawn on "solvent" banks, further expanding the field of acceptable checks. The Fed's strategy was to make the service as attractive as possible by increasing the number of banks on which they could collect.<sup>26</sup>

Although nonmember banks still could not deposit checks directly with the Reserve Banks, they were entitled to send checks to the Fed through their correspondents that were member banks, and many apparently did (Spahr 1926, p. 197). The Board viewed the inability to accept checks directly from nonmember banks as an impediment to the success of the compulsory plan: "Any clearing and collection plan to be effective must be so comprehensive as to include all checks."<sup>27</sup> The Reserve Bank check collection service was at first intended as a benefit of membership, but the Board decided that it would be better to offer clearing services to nonmembers to entice them to remit at par on checks sent to them by the Reserve Banks through the mail. As part of a package of suggested amendments sent to the Congress in December 1916, the Board included an amendment to Section 16 that would permit nonmember banks to use the Fed's clearing service, provided they agreed to pay their own checks at par and kept a "compensating balance" with the Reserve Bank. The amount of the compensating balance was to be determined by the Federal Reserve Board.

---

<sup>24</sup> The Board's circular stipulated that Reserve Banks keep an accurate account of the cost of the clearing service and that the Board would fix the charge by rule. Charges ranged from 0.9 cents to 2.0 cents per item. Charges were ultimately lowered and then abolished on July 1, 1918 (Spahr 1926, pp. 192–93, 211).

<sup>25</sup> *Federal Reserve Bulletin* (1916, pp. 323–24).

<sup>26</sup> In discussing the compulsory plan in June 1916, the Board said "it is thought that in the near future checks upon practically all banks throughout the United States can be handled at par by Federal Reserve Banks" (*Federal Reserve Bulletin* 1916, p. 263).

<sup>27</sup> *Federal Reserve Bulletin* (1917, p. 100).

Congress ultimately enacted the suggested check clearing provisions, but not without a fight that revealed the depth of opposition to the Fed's check collection plans.<sup>28</sup> After being called back into session in April, Congress again took up the Board's suggested amendments. Representatives of the American Bankers Association took the opportunity to lobby for a bank's right to charge presentment fees, even against the Reserve Banks. As a result, Senator Hardwick of Georgia introduced an amendment that would have added to Section 16 the proviso that nothing in the Act shall be construed as prohibiting a member or nonmember bank from "making reasonable charges, but in no case to exceed 10 cents per \$100 . . . for collection or payment of checks and drafts. . . ."<sup>29</sup> As drafted, the Hardwick Amendment would have effectively negated the provision of Circular No. 1 requiring banks participating in the compulsory plan to pay at par on checks sent by the Reserve Banks through the mail.

The Hardwick Amendment was ruled out of order in the House, where the Board's suggested amendments were passed on May 5. The Senate then approved the bill but included the Hardwick Amendment, despite the opposition of the Federal Reserve Board. In conference, Glass prevailed upon the conferees to modify the Amendment in two ways. The Federal Reserve Board was given the authority to determine and regulate the charges, and a clause was added stating that "no such charges shall be made against the Federal reserve banks." The House passed the resulting bill, as did the Senate, after the reading of a letter from President Wilson that described the original Hardwick Amendment "as most unfortunate and as almost destructive of the function of the Federal reserve banks as a clearing house for member banks."<sup>30</sup> The Board's ban on charging presentment fees against Reserve Banks was now law.<sup>31</sup>

As part of the compulsory plan, the Board directed Reserve Banks to maintain so-called "par lists" consisting of the nonmember banks in their districts that accepted checks at par. The par list and the Fed's campaign for universal par presentment became the center of the celebrated "par collection controversy" (see Spahr [1926], Ch. 7). At first, only banks that explicitly agreed to remit at par were added to the System's list, but in early 1919 a concerted effort was begun to expand the list. Reserve Banks took aggressive measures to

---

<sup>28</sup> See Spahr (1926, p. 200) and Wyatt (1944).

<sup>29</sup> A contemporary account of the legislative action appears in "The Hardwick Amendment" (1917, pp. 40–41).

<sup>30</sup> Miller (1949, p. 20); *Congressional Record* (1917, p. 3761).

<sup>31</sup> At the request of the Federal Reserve Board, the Attorney General issued an opinion on the scope of the new language in Section 13. He said that the Federal Reserve had no power to regulate the exchange charges of nonmember banks who were not depositors under the clearing system, and if nonmembers insisted on making charges, the Reserve Banks could not handle checks drawn on them, since Section 13 now prohibited such charges (*Federal Reserve Bulletin* 1918, pp. 367–70).

attempt to collect at par on all banks in their districts.<sup>32</sup> Some Reserve Banks put recalcitrant banks on the par list without their explicit permission; they accumulated their checks and had them presented directly over the counter, where banks were generally required to pay par.<sup>33</sup> The nonmember par list grew from about 10,000 in December 1918 to over 19,000 at the high-water mark in November 1920, leaving only about 1,700 nonpar banks.<sup>34</sup>

The opposition to the Federal Reserve's methods was fierce in some quarters, however. Some banks refused to cooperate, and the resulting litigation—including cases that reached the Supreme Court—established limits on the measures the Reserve Banks could employ to obtain par remittance.<sup>35</sup> Checks could no longer be accumulated for presentation at the counter to “coerce” banks into paying par. Banks could pay checks at their counter by draft rather than lawful money. Moreover, the Supreme Court ruled that the Federal Reserve was under no congressional mandate to bring about universal par clearance. In response, the Board ordered the Reserve Banks to cease using agents other than banks in making collections and to stop accepting checks drawn on nonpar banks.<sup>36</sup> Banks withdrew from the par list until nonpar banks numbered nearly 4,000.<sup>37</sup> Nonpar banking persisted thereafter, chiefly in small one-bank towns or in small towns with only nonpar banks.<sup>38</sup> The number of nonpar banks declined sharply in the early 1970s and finally sank to zero in 1980.<sup>39</sup>

### 3. THE CONVENTIONAL VIEW

Many payments system researchers have described the pre-Federal Reserve check collection system as inefficient in the sense that real resource costs were higher than they would have been under a centrally run system. Likewise, it is conventionally argued that the Fed-imposed clearing arrangements resulted in lower real resource costs than private arrangements. Walter Spahr (1926)

---

<sup>32</sup> See Harding (1921); Tippetts (1924, pp. 635–36); Tippetts (1929, pp. 277–80); Preston (1920, pp. 571–78).

<sup>33</sup> Under some state laws, however, the bank was not required to pay in legal tender but could pay instead with a draft. See Spahr (1926, pp. 284–86).

<sup>34</sup> Spahr (1926, p. 248) displays data compiled from various issues of the *Federal Reserve Bulletin*.

<sup>35</sup> The major cases are described in Spahr (1926, pp. 249–82, 284–86) and Tippetts (1924, 1929). The key decision came in the “Richmond case,” *Farmers and Merchants National Bank of Monroe, North Carolina, et al. v. Federal Reserve Bank of Richmond, Virginia*, in which the U.S. Supreme Court upheld the constitutionality of a North Carolina law that authorized state banks to charge a presentment fee of no more than 1/8 of 1 percent and specifically allowed payment by draft for checks presented over the counter by a Federal Reserve Bank, post office, or express company.

<sup>36</sup> *Federal Reserve Bulletin* (1923, pp. 903–04, 1194).

<sup>37</sup> *Federal Reserve Bulletin* (1928, p. 535).

<sup>38</sup> Stevens (1998, p. 19); Jessup (1967, p. 26).

<sup>39</sup> The reasons for the decline of nonpar banking have not, to our knowledge, been studied.

compiled the most comprehensive and widely cited statement of this view 12 years after the founding of the Federal Reserve. Most recent proponents of the conventional view echo Spahr.

Spahr claimed that one of the most serious problems with the pre-Fed system was the excessive cost of collecting country checks and that nonpar checking was to blame. The practice of nonpar checking had its origins in earlier times when banks incurred significant shipping costs in remitting specie to distant banks for settlement. The amount of the charges had been steadily declining during the 50 years immediately prior to the Fed as transportation costs fell; nevertheless, nonpar remittance practices remained a significant source of revenue to some country banks and a significant source of irritation to some city banks in 1914. The most widespread criticism of exchange charges was that technological advances had removed most of the country banks' settlement costs, eliminating the once-valid justification for the fees. Spahr describes the fees as "excessive" and out of proportion to the costs incurred by the levying banks (Spahr 1926, p. 241).

Besides imposing excessive costs, presentment fees were said to cause costly and inefficient collection practices. Chief among these was circuitous routing: some checks were sent to banks in roundabout ways through a number of different banks in order to avoid presentment fees, resulting in excessive postage and clerical costs and extended check float. Although there are no data on the extent of circuitous routing to avoid exchange charges, the literature contains a number of examples. Cannon (1900) describes an example of a check drawn on The Peconic Bank of Sag Harbor, Long Island, and deposited in a bank in Hoboken, New Jersey. The now-famous check traveled from the Hoboken bank to a New York City bank, and then to banks in Boston, Tonawanda, Albany, Port Jefferson, Far Rockaway, New York City (again, but a different bank), Riverhead, and Brooklyn, before finally arriving at the Sag Harbor bank. James Hallock (1903) cites the Sag Harbor check and three others as well. Spahr (1926) cites the Sag Harbor check, Hallock's Stonington, Connecticut, check, plus two additional examples. Other writers typically cite Cannon's Sag Harbor check.<sup>40</sup>

Most of the pre-Fed writers on check clearing were city bankers; there is almost no early academic literature on the subject. Nonetheless, virtually all authors around that time supported the thrust of Spahr's argument. In his 1890 annual report, the Comptroller of the Currency conveys an early official opinion on exchange charges by saying a conservative estimate of their total amount "would constitute a heavy burden upon the commercial interests of the country." Hallock (1903, p. 17) asserts that the "avoidance of collection

---

<sup>40</sup> See, for example, Conway and Patterson (1914, p. 324), Miller (1949, p. 10), or Baxter (1983, p. 560).

charges is the motive for shunting a check up and down the country” and that “the practice is not unusual.” Oliver Sprague (1910, p. 42) observes that “collections and payments are subject to delay and involve heavy expense.” Similarly, turn-of-the-century *Banker’s Magazine* and *A.B.A. Journal* articles, some of which Spahr references in his book, confirm the common perception that bankers saw a need for reform. An array of historical descriptions of the U.S. payments system written since 1914 have either referenced or endorsed Spahr’s evaluation of the pre-Fed clearing system.<sup>41</sup>

Many of the founders of the Federal Reserve System shared Spahr’s assessment. In a debate prior to the Federal Reserve Act, Carter Glass revealed that he thought centralized clearing by the government would reallocate costs in a welfare-improving way:

Precisely how much difficulty and cost will be incurred by the Federal Reserve Banks in carrying out the provisions of this section cannot be precisely calculated. It can, however, be positively stated that such expenditures will be very much less than those incurred by banks at the present day in carrying through their exchanges. The proposed provision will eliminate the numerous and well-founded complaints of unjust charges for exchange; and, while it will prevent certain banks from profiting as they do by exchange transactions it will correspondingly benefit the community.<sup>42</sup>

Glass expresses the idea here that although exchange-charging country banks will be made worse off, a Federal Reserve clearing system will make others better off by lowering their clearing costs. He appears to have in mind the proposition that aggregate costs will be lower overall, and thus the community will benefit. H. Parker Willis, Glass’s advisor and the first Secretary to the Federal Reserve Board, made similar criticisms of the old system in his post-1914 works. For example, Willis wrote the lead article in the March 1914 *American Economic Review* on the new Federal Reserve legislation and referred to exchange charges as “extortion.”<sup>43</sup>

---

<sup>41</sup> For example, Conway and Patterson (1914) explain the “disadvantages” of the old methods, highlighting circuitous routing, excessive exchange charges, and unnecessary check float. Gidney (1916, p. 607) states, “Important economies are expected to be effected in the total cost of check collection, through having checks reach the paying bank by a reasonably direct route and after having passed through relatively few banks. . . .” Kemmerer (1928) describes a “defective exchange and transfer system,” adding “large shipments of currency” as another inefficiency. Tippetts (1929) gives a similar account, calling the system “the source of a number of evils.” Watkins (1929) emphasizes that exchange charges “operated under the old system to lessen its efficiency” and goes on to argue that such charges caused unduly large bankers’ balances. See also Miller (1949), Jessup (1967), Duprey and Nelson (1986), and Moore (1990).

<sup>42</sup> U.S. Congress, House (1913) pp. 55–56.

<sup>43</sup> For Willis’s opinion on Spahr (1926) and Cannon (1900), see Willis, et al. (1933), p. 238. W. P. G. Harding, then Governor of the Federal Reserve Board, stated that “the establishment of a universal country wide par-collection system” would result in the “elimination of the burdensome delays and expenses incident to the old indirect routing system” (Harding 1921, p. 338).

In the conventional view, the Fed's entry was a struggle by the progressive forces of banking reform against the vested interests of nonpar bankers. Duprey and Nelson (1986, p. 18) write:

At the turn of the century . . . the private banking sector was widely acknowledged to have produced an inefficient and counterproductive arrangement for collecting checks beyond the local level. The invisible hand wasn't working. This failure to produce an adequate solution for collecting out-of-town checks efficiently was one reason that the Congress, as part of its banking reform measures developed between 1908 and 1913, gave the Federal Reserve System both a regulatory role and an operating role in check clearing and collection.

They go on to argue that the Fed's quest for a universal par clearance system was frustrated by "stiff opposition and competition" from the correspondent banking system.<sup>44</sup> Nonetheless, the Fed did enjoy "some success in improving the efficiency of intercommunity check collection"—the Fed reduced the number of nonpar banks "and probably helped limit abuses" in their practices. Summers and Gilbert (1996), drawing on Spahr, note "widespread dissatisfaction" with the settlement of interregional transactions pre-Fed and cite enhancing payment system efficiency as an important purpose for creating the Federal Reserve. In a similar vein, Gilbert (1998) concludes that, based on the fall in reserve holdings at banks joining the system, "evidence from the period when the Fed was founded suggests that the Fed's services improved payments system efficiency."<sup>45</sup>

#### 4. A PROBLEM WITH THE CONVENTIONAL VIEW

One reason to question the conventional view is that (with only a few exceptions) most writers do not explain why the participants in the check collection system were unable to implement efficiency improvements, if they were available. In other words, it is not clear why there would be a market failure in check

---

<sup>44</sup> Duprey and Nelson (1986, p. 19).

<sup>45</sup> Gilbert (1998, p. 137). One view which we do not discuss at length here is that Congress wanted the Fed to collect checks in order to prevent disruptions in check collection that accompanied financial panics, like the one in 1907 that occasioned widespread suspension of cash payments at banks (Corrigan 1983, pp. 345–48). While preventing financial panics was clearly the central motive behind the provisions moving reserve accounts over to the Reserve Banks, we know of no evidence that any of the founders perceived the functioning of the check collection system during panics, per se, as a motive for granting check clearing powers to the Reserve Banks. Moreover, it is hard to see why the rediscounting and open market powers of the Reserve Banks should not be sufficient to prevent financial disruption. Walter (1988, p. 57) reports that in congressional debate on the Federal Reserve Act there is no mention of the Fed providing check collection services to produce a safer payment system. "Senator Bristow and O. M. W. Sprague agreed, in an exchange during Senate hearings, that the problems with inter-city check collection during panics were caused by a lack of a lender of last resort" (U.S. Congress, Senate, Senate Committee on Banking and Currency 1913, pp. 512–13), cited in Walter (1988).

collection. One might argue that it was beyond the capability of participants to create the Federal Reserve's clearing system, since the Fed is a collective nonprofit institution.<sup>46</sup> And yet the Reserve Bank's check collection activities employed precisely the same technology and organizational techniques employed by the private sector. The Reserve Banks were essentially correspondent banks for the members of their clearing system, and their relationship with their respondents was organized in essentially the same way as private correspondent relationships. The Reserve Banks did set up a wire transfer system for moving funds rapidly between Reserve Banks, but private banks had been moving funds via wire transfer prior to the founding of the Fed (Langston 1921, pp. 168–72).

Private clearinghouses were collective nonprofit institutions set up by banks, often endowed with quasi-regulatory power over their members.<sup>47</sup> In fact, in the decades prior to the founding of the Fed, clearinghouses were making moves to expand their clearing activities to encompass checks drawn on country banks.<sup>48</sup> The Boston clearinghouse had been clearing New England country checks for years.<sup>49</sup> In short, banks could have set up the equivalent of the Federal Reserve Bank clearing system on their own. Presumably they would have done so if it would have made some participants better off without making others worse off—for example, if it would have appreciably lowered the costs of collecting checks. The fact that they did not do so seems to suggest that it would not have lowered costs. If the Reserve Bank clearing system lowered check clearing costs, why couldn't the private sector do the same?

Some believe one possible answer is that nonpar country banks enjoyed monopoly power. Presentment fees were set inefficiently high in order to extract rents from collecting banks. Such fees, it is said, “can lead to costly and complicated countermeasures” to avoid nonpar transactions.<sup>50</sup> This view envisions a bank facing a choice between mailing directly to a nonpar bank with which it does not have a correspondent relationship and sending the check on to a correspondent who can present it at par. But it is not clear that this was always the case. In fact, presentment fees were often paid to nonpar banks by their correspondents. These collecting banks appeared to pay fees willingly in exchange for the respondent banks' reserve balances (Spahr 1926, p. 111). Hence, presentment fees were often voluntarily agreed to as a component of a broader correspondent-respondent relationship. It is not clear how to

---

<sup>46</sup> Duprey and Nelson (1986) seem to advocate this position when they argue that “well-established rivalries” between city and country banks somehow prevented banks from voluntarily agreeing to a mutually beneficial national clearing system.

<sup>47</sup> Gorton (1985); Gorton and Mullineaux (1987).

<sup>48</sup> Cannon (1900); Hallock (1903); Spahr (1926, pp. 119–30); Duprey and Nelson (1986, pp. 22–23).

<sup>49</sup> *Federal Reserve Bulletin* (1916, p. 317). The Federal Reserve Bank of Boston took over the operations of the Boston clearinghouse by unanimous consent in 1916.

<sup>50</sup> McAndrews (1995, p. 56). See also Gilbert (1998).

reconcile this market power view with the documented features of correspondent relationships.

Check collection routes were determined by the pattern of correspondent relationships. Thus, it is the market for correspondent relationships—as opposed to the market for the clearance of a particular check—that is relevant to the question of market power. What a country bank had to offer as a respondent—reserve balances—was available from other country banks as well. It appears unlikely that small country banks exercised any monopoly power in the market for correspondent relationships.

## 5. AN ALTERNATIVE VIEW

The conventional view sees the pre-Fed check collection system as inefficient and disjointed, with much of the inefficiency being driven by the practice of nonpar remittance. An alternative interpretation is possible if we view the check collection system as a whole and focus on its *network* characteristics.<sup>51</sup> While such terms as “network effect” or “network externality” are used widely by economists in reference to a variety of market settings, a check collection system (or any other payment clearing system) literally is a communication network. Two key characteristics are central to understanding the organization and performance of network communications industries: joint benefits and common costs.

For most goods, a unit of consumption provides benefits to a single user. Some other goods or services provide simultaneous benefits to many people. For most such goods, a musical performance for example, one person’s utility does not depend on whether anyone else is partaking. One unit of a communication service, however, necessarily involves two “consumers”: a sender and a receiver. Neither party derives a benefit from communication unless the other one does. Similarly, the clearing of a check provides benefits jointly to both the payor and the payee. Note that the presence of joint benefits affects the criterion for judging whether provision of a unit of a good is economically efficient. For an ordinary good, we would say that a unit’s provision is efficient if the buyer’s willingness-to-pay exceeds the incremental resource costs of the good. For a network communication service, we would say that the provision of a unit of service is efficient if the sum of the willingness-to-pay of the sender and the receiver (or payor and payee in the case of a payment instrument) exceeds the incremental cost.<sup>52</sup>

---

<sup>51</sup> For a discussion of the characteristics of network services as applied to payment systems, see Weinberg (1997) and Lacker and Weinberg (1998). James McAndrews (1995) first suggested that check collection at the founding of the Fed should be viewed as a network communications industry, analogous to credit card and ATM clearing networks. See also Summers and Gilbert (1996, pp. 6–7).

<sup>52</sup> This distinction is emphasized by Baxter (1983).

The fact that check clearing services provide joint benefits to pairs of users implies that there are common costs even at the level of the individual unit of service. Common costs are costs that cannot be uniquely attributed to the provision of service to particular users or groups of users. The incremental cost of a unit of a payment clearing service is common to the payor and payee in that it cannot be uniquely attributed to either. In network services markets, common costs tend to exist at a variety of levels. The technology for such services often includes substantial fixed infrastructure costs. The physical transport of items such as checks involves common costs since the cost of a trip cannot be attributed to particular items or particular pairs of senders and receivers. Many common costs in such markets are fixed relative to the quantity of a service provided. For example, the cost of a transportation node facility, such as a terminal, or (in the case of checks) a bank branch, often cannot be uniquely attributed to any particular item passing through it.

The presence of substantial common costs implies that it is impossible to specify precisely an individual user's share of total costs. Consequently, there is some ambiguity in determining the "right" price for a particular user to face. Efficiency requires that no individual or group pay less than its incremental cost, defined as the cost of extending service to the group in question given the level of services provided to all other users. Otherwise they might inefficiently overuse the service.<sup>53</sup> If all users pay incremental cost, however, the service will not recover all of the costs that are common across groups of users. In order to cover common costs, the service must charge some users more than their incremental cost. There are often many ways to allocate common costs, all of which are consistent with efficient provision of the service.<sup>54</sup>

The presence of joint benefits and common costs gives rise to what are often called "network effects." One person's participation in a network brings benefits to that person as well as to all others who wish to communicate with him. It is important to note, however, that while these benefits may be external to the individual's action, they are internal to the network in which he participates. A network's participants, as a group, may have an incentive to shift common costs away from some individual participants: those who place

---

<sup>53</sup> This principle must be modified in the presence of joint benefits: see discussion on the next page.

<sup>54</sup> We use the term "efficient" in the sense of Pareto efficiency. An allocation is Pareto efficient if no party can be made better off without making some other party worse off. Some readers may be familiar with the Ramsey cost allocation principle, which states that each price charged by a multi-product provider should be set at a markup over costs that is inversely proportional to demand elasticity. This would appear to prescribe a uniquely efficient allocation of common costs. The Ramsey allocation is optimal, however, only under a particular assumption about the way in which social benefits are calculated, specifically, by adding up the utilities of individual agents. The Pareto criterion is less restrictive, and there will tend to be multiple Pareto efficient allocations.

a relatively low private value on participation but who bring large external benefits to other participants. Efficient participation may even require that an individual pay less than their incremental cost.

To be more precise, consider an existing network whose incremental cost of adding a particular new member is  $c_n$ , while the mutual benefit to existing members of adding this new member is  $v_n$ . For the potential new member, the corresponding incremental costs and benefits are  $c_i$  and  $v_i$ . It is efficient to add this member if  $v_n + v_i > c_n + c_i$ . Suppose that the network charges the new member the price  $p$ . The new member is willing to participate if  $v_i > c_i + p$ . The network is willing to add the new member if  $v_n + p > c_n$ . Thus, any price satisfying  $v_i - c_i > p > c_n - v_n$  induces efficient participation. Note that if  $v_i - c_i < c_n$ , then the network must charge less than the network's incremental cost of adding the new member, because the new member's private net benefit from joining is low. Participation is efficient nonetheless because of the value,  $v_n$ , the member brings to the network. Since the new member's participation in the network brings joint benefits to all network participants, both  $c_n$  and  $c_i$  are common costs. The price determines who bears the common costs.

## 6. PRE-FED CHECK CLEARING

When seen in terms of the allocation of the common costs of a network service, a very different view of the pre-Fed check collection system emerges. The centerpiece of the conventional view is the practice of nonpar collection and the resulting circuitous routing of checks. The level of presentment fees, however, helps determine the allocation of the common costs of check collection. Moreover, circuitous routing is not obviously wasteful, given the common costs of shipping check bundles. And complaints about excessive costs appear to be motivated by dissatisfaction with the allocation of costs among participants rather than the overall level of aggregate costs.

Correspondent banking relationships were central to the clearing of checks before the founding of the Fed and can be understood quite clearly in terms of common costs. The correspondent relationship bundled together a number of distinct functions: the respondent used the correspondent to clear checks drawn on banks in the vicinity of or that had relationships with the correspondent; the correspondent presented checks drawn on the respondent; and the respondent held balances with the correspondent, which were used to settle clearings in either direction (Spahr 1926, p. 111). Settlement via interbank balances has clear advantages over settlement by remittance of specie or exchange draft, since the common cost of correspondent balances serves both investment and settlement functions (Watkins 1929, pp. 3–5). Holding balances with a bank in a financial center, where they could earn interest, was preferable to holding sterile reserves in the vault. Combining several items into a single

shipment saved shipping costs. Selecting a limited number of correspondents was advantageous because it economized on the fixed costs associated with any given relationship; bilateral clearing and settlement arrangements with the universe of depository institutions would obviously be far too costly. The overall advantage of a correspondent relationship is that certain common costs are spread among a number of distinct payment services rather than duplicated across multiple service providers.

Presentment fees can be easily understood as a means of allocating common costs. The presentment fee was the price paid to the paying bank for accepting presentment by mail rather than over the counter, where the paying bank was obligated to pay at par. Presentments, whether over the counter or by mail, were generally paid by debits to correspondent balances or by drafts. Mailing drafts to presenting banks involved postage costs, and so for some checks there was a positive incremental cost to the paying bank of accepting mail presentment. Otherwise the paying bank was largely indifferent about the means by which checks arrived for payment (Spahr 1926, pp. 99–101). In terms of our earlier notation, interpreted here as the incremental benefits and costs of accepting presentment by mail rather than over the counter,  $v_i$  was approximately zero and  $c_i$  was slightly positive. A collecting bank, on the other hand, was likely to place considerable value on having a means to avoid the real resource costs of making over-the-counter presentments at long distances. Thus we would expect a large value for  $-c_n$ , the positive cost savings associated with mail presentment to country banks. Apart from costs, the collecting bank should be relatively indifferent about means of presentment, so  $v_n$  should be approximately zero as well.

If  $q$  is the presentment fee, then using our earlier notation,  $q = -p$ . With this change of variables, the condition for efficient participation is now  $c_i - v_i < q < v_n - c_n$ . The presentment fee must exceed the paying banks' cost of participating, net of benefits,  $c_i - v_i > 0$ . Similarly, the presentment fee must not exceed  $v_n - c_n \approx -c_n > 0$ , the net incremental benefit to the collecting bank of adding the paying bank to the mail presentment network. The net benefit of switching to mail presentment is  $v_n + v_i - c_n - c_i \approx -c_n - c_i$ , which is positive when the cost savings to the presenting bank,  $-c_n$ , exceeds the incremental cost to the paying bank,  $c_i$ . It seems likely that for many country checks, direct presentment was more costly than postage for remittance, and therefore mail presentment was economically efficient.

Under the property rights inherent in the pre-Fed check clearing, the paying bank was free to choose the presentment fee  $q$ . In this setting one would predict that paying banks would set  $q$  as high as possible. We therefore should have observed  $q = -c_n$ , presentment fees equal to the collecting bank's net willingness

to pay to avoid over-the-counter presentment.<sup>55</sup> For checks drawn on country banks the cost savings associated with mail presentment were substantial, and thus  $-c_n$  was large. When the paying bank was nearby, the resource cost of over-the-counter presentment was likely to be quite low, and the collecting bank's willingness to pay for mail presentment would be correspondingly small. This is consistent with country bank presentment fees that are close to the cost to collecting banks of making a direct presentment through an agent or an express company, and the general absence of presentment fees in the city. The presentment fee effectively passes the collecting bank's cost savings on to the country paying bank as an inducement to participate via mail presentment. All of the joint benefits of mail presentment,  $-c_n - c_i = q - c_i$ , accrue to the paying bank.

Although some banks complained that presentment fees exceeded the direct outlays of the *paying bank* for remittance (postage, the cost of exchange, and so on),<sup>56</sup> there is no economic efficiency reason why they should not. It is clearly possible for the presentment fee to exceed  $c_i$ , the paying bank's direct outlays for remittance, without violating the condition for efficient participation. Thus presentment fees that are "excessive" in this sense are not necessarily evidence of monopoly power as some have claimed.<sup>57</sup>

Did the system in fact economize on the costs of moving checks? Much of the commentary on this topic has argued the contrary, based on well-documented instances of circuitous routing. But it is not at all obvious that the examples of circuitous routing constitute strong evidence of excessive costs. Cannon (1900, p. 76), commenting on the famous Sag Harbor check, writes that

The reason why banks forward checks in this apparently unreasonable way, often getting the items far out of their regular course, is easy to explain. It sometimes appears cheaper to the one who has the check in hand to enclose it with other items to some regular correspondent, who, assumedly, is nearer the bank on which the check is drawn, than to hunt up a special correspondent for it alone.

This reasoning suggests that the cost comparison implicit in the conventional interpretation of the circuitous routing examples is not the relevant one. The appropriate comparison for the bank holding the check is between the cost of sending the item directly to the paying bank (or a "special correspondent" for this check alone) and the cost of including the item with a batch of checks being sent to an established correspondent in the hope that the correspondent

---

<sup>55</sup> An alternative property rights regime in which collecting banks were entitled to mail presentment at par would result in  $q = c_i - v_i$ . In this case the cost savings from mail presentment would accrue to the collecting banks rather than to the paying banks.

<sup>56</sup> Spahr (1926, p. 240-43).

<sup>57</sup> Stevens (1998), for example.

would be better able to get the check to its ultimate destination. For an irregular check—one drawn on a bank with whom one does not have a correspondent relationship—the incremental cost of sending the check directly to the paying bank would include the postage on the letter, along with the cost of preparing a separate shipment. The incremental cost of adding a check to a regular shipment to a nearby correspondent was probably negligible. Note also that settlement was probably less costly through established correspondent relations. Consequently, the latter was almost certainly less costly than direct presentment for a bank holding a check drawn on a country bank. Critics of pre-Fed check clearing implicitly attribute to the wayward check *all* of the common cost of the shipment to the correspondent.<sup>58</sup>

The bank deciding where to send an item would not consider the costs incurred by the next bank to hold the check. The next bank could send it on to another correspondent, which could send it on again to one of its correspondents, and so on. Could this lead to excessively costly check routes? From a social point of view, the correct cost comparison is between the *expected* cost of sending the check along to the next correspondent bank (including the cost of sending it to subsequent correspondents) and the cost of a more direct route. Again, there is no reason to believe that the expected cost of indirect routing was not almost always less than that of direct routing of irregular interregional checks. The fact that some items ended up following routes that look excessively costly *ex post* does not mean that routing choices were inefficient *ex ante*.

Some mistaken routing choices were inevitable in a decentralized system with thousands of banks. The average costs of such mistakes constitute a valid part of the social cost of the pre-Fed system. The mere existence of such costs, however, does not imply inefficiency for the system as a whole. Only if the costs of the decentralized system exceeded the costs of creating a centralized system would inefficiency be implied. None of the critiques of the pre-Fed system has presented evidence on this dimension.<sup>59</sup> Evidence that it is possible to reduce some collection costs is not, by itself, conclusive evidence that an alternative arrangement would be superior. Yet as implied in the comments by Carter

---

<sup>58</sup> A possibility we do not pursue here is that the cost of postage does not represent the social cost of mailing checks. Since postage rates, then as now, are uniform across destinations, it is quite likely that postage on the irregular country checks in question was lower than incremental social cost, in the sense that the total postal revenue on shipments to the country bank's location failed to cover the incremental cost of service to that location. This would provide yet another reason to question the cost comparison implicit in the circuitous routing evidence. It would also cast doubt on the social value of the movement to shift to direct mail presentment as opposed to direct presentment by an agent such as an express company.

<sup>59</sup> There appear to be no available estimates of the frequency of such circuitous routing. Cannon (1900), Hallock (1903), and Spahr (1926) cite only eight examples between them. Referring to the Sag Harbor check, James (1998, p. 144) notes that "Given the paucity of other examples, one might be suspicious of this example's general applicability."

Glass above, to build an alternative system that reduces the costs associated with collecting certain checks would itself be costly. Similarly, some pre-Fed writers, such as Cannon (1900) and Hallock (1903), used evidence on the costs of collecting country checks to argue for the value of creating and operating a country clearinghouse. As noted above, the frustratingly slow pace of such efforts suggests that the costs of creating a country clearinghouse exceeded the available cost savings.

If the status quo was efficient, why was there so much dissatisfaction with it? Here, it is important to note that complaints about the system are almost entirely voiced from the point of view of *city banks*. It was argued that the cost of collecting country checks was too high, and it was widely acknowledged that a par collection system would reduce the earnings of the nonpar country banks.<sup>60</sup> In other words, par collection would reduce net costs for city banks and raise net costs for nonpar country banks. The complaints of the city banks seem to have been driven by dissatisfaction with the *allocation* of costs implied by the status quo arrangements, rather than by dissatisfaction with *aggregate cost*.<sup>61</sup>

What city banks wanted was for presentment through the mail to have parity with direct presentment, in which case the paying bank would be obligated to pay at par, without deducting a presentment fee. Such a regime would have inevitably shifted costs toward country banks and away from city banks. Although city banks were generally unable to obtain par presentment legislation, the Federal Reserve Board ultimately granted that right to Reserve Banks by regulatory fiat.

## 7. THE FED'S ENTRY INTO CHECK CLEARING RECONSIDERED

According to the conventional view, self-evident inefficiencies motivated and rationalized the Fed's entry into the check collection industry. The process involved a struggle between a progressive reform effort and the vested interests of nonpar banks. Our alternative view suggests a very different interpretation

---

<sup>60</sup> Most of Spahr's (1926, pp. 240–43) "arguments for par collection" amount to claims that costs borne by collecting city banks could and should be reduced. For example, he argues that par collection under the Federal Reserve would "relieve trade . . . of the burden" of clearing costs because "the costs would fall to the Federal Reserve Banks and reduce to that extent the earnings that go to the government" (Spahr 1926, p. 240). That is, the government should defray collecting bank costs.

<sup>61</sup> City banks' complaints about the costs of collecting country checks should also be viewed in the context of secular trends in the structure of the banking industry. The total number of banks in the country grew from 12,424 in 1902 to 26,765 in 1914, and many of these new banks were small country banks (U.S. Treasury, Office of the Comptroller of the Currency 1903, 1915). Thus there were a growing number of country banks to contend with.

of the process by which the Fed entered the check collection industry. In our view, the motive was to attract membership, and the process involved a struggle over the allocation of the common costs of check collection.

All national banks were required to join the Federal Reserve System, but for state-chartered banks, membership was optional. From the beginning, attracting members, and their reserve balances, was viewed as critical to the success of the institution (White 1983, p. 130). A key perceived defect of the previous system was the “pyramiding” of reserves in financial centers, which left the latter vulnerable to sudden widespread withdrawals. Through rediscounting the Reserve Banks would provide an elastic supply of balances in response to rapid demand shifts, preventing financial panics (White 1983, pp. 63–125).

In this context, the Reserve Bank check clearing service authorized by the Act could help attract members, as the early leaders of the Federal Reserve clearly understood. A mid-1915 report to the Reserve Bank Organizing Committee spelled out the link between Reserve Bank check clearing services and the membership question:

It must be borne in mind that the banking power of the United States will divide more sharply than it has ever done before into two groups—members and non-members. It is the intent of the Act itself to bring non-members into the system. But so long as there is any considerable body of non-member banks, the two groups will of necessity be in competition with one another, producing two parallel clearing systems. . . . (T)he domestic exchange business of the Federal reserve system must be so arranged as to offer constant inducements to non-members to enter the system. At the same time, members must find it more profitable to use the Federal reserve system than to make collections as at present. The situation is more complex when it is taken into consideration that member banks are in a position to deal on favorable terms either with the Federal reserve banks and their members or with non-members.<sup>62</sup>

The Fed’s check clearing service should aim to reduce costs to members and attract nonmembers to join the system. To do so, they would need to attract the check clearing business of their members, who were under no obligation to clear through the Fed.

The reserve requirements in the Federal Reserve Act, while essential to the monetary goals of the Act, made it more difficult to attract members. Prior to the founding of the Fed, banks kept reserves with correspondents in addition to specie and notes in their vaults. Correspondent balances could be used to satisfy legal reserve requirements, up to a limit, under state laws and the National Bank Act (Watkins 1929, pp. 67, 96). Moreover, correspondents would often grant

---

<sup>62</sup> Preliminary Committee on Organization (1914, pp. 58–59). The organizing committee consisted of the Comptroller of the Currency and the Secretaries of the Treasury and Agriculture. The Preliminary Committee on Organization reported to them and was chaired by H. Parker Willis.

immediate credit for deposited checks, and these counted toward the respondent's required reserves. Under the Federal Reserve Act, reserve requirements had to be met with balances held at the Reserve Banks; correspondent balances would no longer count. If the Reserve Banks did not offer check clearing services, member banks would have to hold separate correspondent balances in order to clear checks, and these balances would have a higher opportunity cost, since they would no longer do double duty. Willis (1923, p. vi) described the implications in dramatic terms:

It was recognized that, without these powers, [referring to Reserve Bank check clearing authority] the reserve banks would become merely the holders of dead balances carried for the member banks without any service to them; and, since the business public abhors an idle or unnecessary institution, just as nature is traditionally said to abhor a vacuum, it would not submit long to the needless burden created by such emergency institutions designed to put out financial fire.

Failure to offer attractive check clearing services to justify member bank reserve balances would threaten support for the System.<sup>63</sup>

The reserve requirement provisions of the Federal Reserve Act were to be phased in over three years, so the System had time to develop a strategy. According to the original Act, member bank required reserves would be transferred over from correspondents in annual steps from November 16, 1914, to November 16, 1917.<sup>64</sup> These provisions were revised by the amendments of June 21, 1917, in connection with measures to aid financing of the government's war effort, lowering the requirements but making them effective immediately.<sup>65</sup>

The Fed's struggle to establish its check clearing service is readily understandable from our alternative perspective. The first national check clearing venture, the "voluntary reciprocal plan" initiated by the Board in March 1915, was unsuccessful, never attracting more than a third of the member banks. In exchange for agreeing to accept mail presentment at par, member banks were able to clear at par on members that joined. This was essentially a voluntary clearinghouse, modeled after the city clearinghouses. That it failed should be no surprise, given the terms that were offered. Reciprocal par presentment allocates common costs according to each bank's outlays. This allocation successfully attracts members where the cost of over-the-counter presentment is low, as among city banks. In this case, a bank joining the scheme gives up little in presentment fee income. But where over-the-counter presentment costs are

---

<sup>63</sup> Stevens (1998) also discusses the role of reserve requirements in the evolution of the Fed's check clearing activities.

<sup>64</sup> Starting on November 16, 1914, correspondent balances could count towards a maximum of 6/15 of required reserves in the first year, 5/15 the second year, 4/15 the third year, 3/15 the third year, and not at all after November 16, 1918.

<sup>65</sup> The Board gave banks until July 15, 1917, to comply (*Federal Reserve Bulletin* 1917, pp. 508–09).

high and banks have the right to charge for remittance outside the scheme, this allocation of common costs may fail to induce participation; some prospective members would have no incentive to join, even if their participation would be worthwhile. Two-thirds of member banks apparently did not want to join under the reciprocal par presentment cost allocation. Some might have brought large benefits to the other participants in the scheme by lowering presentment costs, even though they themselves placed a relatively low value on participation. As we noted above, these are just the types of banks that charge presentment fees. The Fed needed to find a way to induce their participation.

With its second venture, the “compulsory” plan, the Fed found a solution to the problem: under Circular No. 1 (May 1, 1916), members were required to accept mail presentment at par. Having failed to induce more than a third of their members to voluntarily give up charging presentment fees against the Reserve Banks, the Fed prohibited such fees outright. The compulsory plan was more attractive to members than the voluntary-reciprocal plan on two counts. First, the Reserve Banks were offering to clear checks on any member bank, regardless of how many member banks joined the scheme. The Reserve Banks immediately had over 7,000 par endpoints. Second, the cost of joining the second plan was much lower for many banks. Under the previous plan they would have had to give up presentment fees. Under the new plan, the Board had already taken away their right to charge presentment fees against the Reserve Banks. Joining was less of a sacrifice now.

Success was not yet assured. In about a year, reserves would be transferred over from correspondents to the Reserve Banks under the new reserve requirements. As we noted above, correspondent balances were a key component of the bundle of mutual clearing services that made up the typical correspondent banking relationship. Settling cleared checks by crediting or debiting correspondent balances was less costly than remitting specie or exchange drafts. Members would need to retain some correspondent balances to clear checks on nonmembers, and such balances would no longer do double duty as required reserves. The new reserve requirements would break apart some of the shared common costs built into pre-Fed correspondent banking arrangements, raising member bank costs.<sup>66</sup> The Fed’s strategy now was to rebuild that cost sharing around member bank balances at Reserve Banks. The objective was to offer to collect at par checks drawn on every bank in the country.<sup>67</sup> In theory, member banks would no longer need external correspondent balances, the Reserve Banks having taken over all the essential clearing functions connected with them.

To this end, the Board sought the amendments passed on September 7, 1916, allowing Reserve Banks to clear checks drawn on nonmembers. But the

---

<sup>66</sup> Martin, et al. (1915, pp. 369–70).

<sup>67</sup> “Any clearing and collection plan to be effective must be so comprehensive as to include all checks” (*Federal Reserve Bulletin* 1917, p. 100).

Reserve Banks had trouble getting nonmember banks to accept presentment at par, as one would expect; the Reserve Banks had nothing of value to offer in exchange. This led the Board in early 1917 to seek amendments allowing the Reserve Banks to accept checks deposited by nonmembers. Perhaps reciprocal check clearing privileges would entice nonmembers to give up their presentment fee income. By early 1917, however, the Board's ban on presentment fees against the Reserve Banks made clear that check collection costs would be reallocated as a by-product of the Fed's strategy to eliminate presentment fees. Bankers who would be disadvantaged by such a reallocation mobilized to push the Hardwick Amendment. The resulting legislative battle exposed the divergent interests in the allocation of check collection costs. The Hardwick Amendment was effectively defeated following a "nationwide campaign . . . by the Credit Men's Association, the mail-order houses, manufacturers, jobbers, wholesalers and merchants in the large centers."<sup>68</sup> Payees, in other words, lobbied in favor of shifting costs towards payor banks.<sup>69</sup>

An ensuing opinion by the Attorney General spelled out the new distribution of property rights.<sup>70</sup> Reserve Banks could not pay presentment fees. Nonmember banks could decide for themselves whether to charge fees, but assessing fees against a Reserve Bank for mail presentment was the equivalent of not accepting mail presentment from them. In this environment, the Reserve Banks attempted to exercise as much leverage as possible to persuade nonmembers to pay at par. During the period from early 1919 through 1923, the Reserve Banks resorted to a number of costly collection techniques, such as sending Reserve Bank employees to present over the counter or hiring local agents to make direct presentation. These were characterized in litigation as outside the bounds of customary banking practice; in many cases the expenditures on such techniques exceeded the presentment fees that were avoided and thus would not have been undertaken by private sector collecting banks.<sup>71</sup> The list of banks accepting par presentment was naturally largest in the presence of such measures. When court decisions struck them down, Reserve Bank leverage was commensurately reduced and the par list shrank.

---

<sup>68</sup> "The Hardwick Amendment" (1917, p. 40). See also Tippetts (1929, pp. 272–74).

<sup>69</sup> Note that costs were shifted to taxpayers as well, since the Fed stopped recovering its costs when collection fees were eliminated in June 1918 (Spahr 1926, pp. 192–93).

<sup>70</sup> *Federal Reserve Bulletin* (1918, pp. 367–70).

<sup>71</sup> Spahr (1926); Tippetts (1929).

## 8. CONCLUSION

The Reserve Banks were able to achieve what reformers had been unable to bring about—a more centralized clearing of interregional checks. Although early reform-minded writers, like Cannon (1900) and Hallock (1903), had argued that the cost savings from eliminating what they saw as inefficiencies in clearing irregular interregional checks would exceed the cost of setting up a more centralized arrangement, no such schemes had emerged. The Fed succeeded where earlier efforts failed, under our alternative view, because the Board arrogated the right to present at par on member banks. No such unilateral ability to reallocate property rights was available to private sector collecting banks. Reallocating property rights had the effect of shifting the common costs of check collection away from collecting banks that used the Reserve Bank system, toward member banks and, after fees were eliminated in 1918, toward federal taxpayers.

The par presentment right granted to the Reserve Banks by the Board in Circular No. 1, Series of 1916 (now Regulation J), is essentially a barrier to competition in the sense that it allows the Reserve Banks to offer check collection services at lower costs than competitors. Other banks could obtain the right to present at par, but they would have to offer paying banks material compensation in order to do so. The Reserve Banks did not need to offer any compensation. Note that this barrier to competition persists today in the form of differential presentment times (Lacker and Weinberg 1998). Private sector collecting banks must present by 8:00 a.m. in order to obtain same-day funds (at par), while Reserve Banks can present until 2:00 p.m. for same-day funds; the Reserve Banks have a “six-hour monopoly.” Interestingly, the Board of Governors recently revisited the presentment time differential.<sup>72</sup> The statement announcing the Board’s decision to retain the competitive advantage notes that any equalization would reallocate costs in a way that would be disadvantageous to some segment of the industry. Moving the private presentment time later, for example, would make collecting banks better off and paying banks worse off.<sup>73</sup> Ironically, the Fed’s original entry into check collection appears to have been accomplished by reallocating the common costs of check collection in just this fashion.

---

<sup>72</sup> Board of Governors of the Federal Reserve System (1998).

<sup>73</sup> Board of Governors of the Federal Reserve System (1998, p. 12).

---

---

**REFERENCES**

- Baxter, William F. "Bank Interchange of Transactional Paper," *Journal of Law and Economics*, vol. 26 (October 1983), pp. 541–88.
- Benston, George J., and David B. Humphrey. "The Case for Downsizing the Fed," *Banking Strategies* (January/February 1997), pp. 30–37.
- Board of Governors of the Federal Reserve System. "Collection of Checks and Other Items by Federal Reserve Banks and Availability of Funds and Collection of Checks" (Docket No. R-1009). Available: <http://www.bog.frb.fed.us/boarddocs/press/BoardActs/1998/19981209-/R-1009.pdf> [March 3, 1999].
- \_\_\_\_\_. *Regulations*. "Regulation J, Collection of Checks and Other Items by the Federal Reserve Banks and Funds Transfers through Fedwire," as amended 1998.
- \_\_\_\_\_. "The Federal Reserve in the Payments System," *Federal Reserve Bulletin*, vol. 76 (May 1990), pp. 293–98.
- \_\_\_\_\_. *Seventh Annual Report, 1920*. Washington: Government Printing Office, 1921.
- \_\_\_\_\_. *Annual Report of the Federal Reserve Board, 1914*. Washington: Government Printing Office, 1915.
- Cannon, James G. *Clearing-Houses: Their History, Methods and Administration*. London: Smith, Elder, & Co., 1900.
- Congressional Record*. Washington, D.C., 1917.
- Conway, Thomas, Jr., and Ernest M. Patterson. *The Operation of the New Bank Act*. Philadelphia: J. B. Lippencott Co., 1914.
- Corrigan, E. Gerald. "Background Materials to the Statement of E. Gerald Corrigan," in *The Role of the Federal Reserve in Check Clearing and the Nation's Payments System*. Finance and Urban Affairs, U.S. Congress, House of Representatives. Subcommittee on Domestic Monetary Policy of the Committee on Banking, and the Subcommittee on Commerce, Consumer, and Monetary Affairs of the Committee on Government Operations. Hearing, 98 Cong. 1 Sess. Washington: Government Printing Office, 1983.
- Duprey, James N., and Clarence W. Nelson. "A Visible Hand: The Fed's Involvement in the Check Payments System," *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 10 (Spring 1986), pp. 18–29.

- Federal Reserve Bank of Richmond. "The Development of Par Collections by Federal Reserve Banks," *Letters to College Classes in Economics and Banking Discussing The Practical Operations of the Federal Reserve System*, no. 6 (May 1922). Fellows, E. G. The Exchange Charge on Checks. Dissertation, Rutgers University, 1940.
- Friedman, Milton, and Anna Jacobson Schwartz. *A Monetary History of the United States: 1867–1960*. Princeton: Princeton University Press, 1963.
- Gidney, Ray. "Notes: Federal Reserve Clearings and Collections," *Journal of Political Economy*, vol. 24 (June 1916), pp. 606–08.
- Gilbert, R. Alton. "Did the Fed's Founding Improve the Efficiency of the U.S. Payments System?" *Federal Reserve Bank of St. Louis Review*, vol. 80 (May/June 1998), pp. 121–42.
- Gorton, Gary. "Clearinghouses and the Origin of Central Banking in the United States," *Journal of Economic History*, vol. 45 (June 1985), pp. 277–83.
- \_\_\_\_\_, and Donald J. Mullineaux. "The Joint Production of Confidence: Endogenous Regulation and Nineteenth Century Commercial-Bank Clearinghouses," *Journal of Money, Credit, and Banking*, vol. 19 (November 1987), pp. 457–68.
- Hallock, James C. *Clearing Out-of-Town Checks*. St. Louis: Author, 1903.
- Harding, W. P. G. "Reply to Senate Resolution Relative to Check Clearing," in Board of Governors of the Federal Reserve System, *Seventh Annual Report, 1920*. Washington: Government Printing Office, 1921, pp. 335–57.
- "The Hardwick Amendment," *Journal of the American Bankers Association* (July 1917), pp. 40–41.
- James, John A. "Commentary," *Federal Reserve Bank of St. Louis Review*, vol. 80 (May/June 1998), pp. 143–50.
- Jessup, Paul F. *The Theory and Practice of Nonpar Banking*. Evanston, Ill.: Northwestern University Press, 1967.
- Kemmerer, Edwin W. *The ABCs of the Federal Reserve System*. Princeton: Princeton University Press, 1928.
- Lacker, Jeffrey M., and John A. Weinberg. "Can the Fed be a Payment System Innovator?" *Federal Reserve Bank of Richmond Economic Quarterly*, vol. 84 (Spring 1998), pp. 1–25.
- Langston, L. H. *Practical Bank Operations*. New York: Ronald Press Co., 1921.
- Martin, Wm. McC., Frederic H. Curtiss, and Pierre Jay. "Report of the Federal Reserve Agents' Committee on Clearings," *Federal Reserve Bulletin*, vol. 1 (November 1915), pp. 369–72.
- McAndrews, James J. "Commentary," *Federal Reserve Bank of St. Louis Review*, vol. 77 (November/December 1995), pp. 55–59.

- Miller, Melvin C. *The Par Check Collection and Absorption of Exchange Controversies*. Cambridge: The Bankers Publishing Co., 1949.
- Moore, Carl H. *The Federal Reserve System*. London: McFarland & Co., 1990.
- Preliminary Committee on Organization. *Report to the Reserve Bank Organizing Committee*. New York: Reserve Bank Organizing Committee, June 1, 1914.
- Preston, Howard H. "The Federal Reserve Banks' System of Par Collection," *Journal of Political Economy*, (July 1920), pp. 565–90.
- Spahr, Walter Earl. *The Clearing and Collection of Checks*. New York: Bankers Publishing Co., 1926.
- Sprague, Oliver M. *History of Crises under the National Banking System, U.S. National Monetary Commission, Senate Document No. 538, 61 Cong. 2 Sess.* Washington: Government Printing Office, 1910.
- Stevens, Ed. "Non-Par Banking: Competition and Monopoly in Markets for Payments Services," Working Paper 9817. Cleveland: Federal Reserve Bank of Cleveland, November 1998.
- \_\_\_\_\_. "The Founders' Intentions: Sources of the Payments Services Franchise of the Federal Reserve Banks," Working Paper 03–96. Cleveland: Federal Reserve Bank of Cleveland, Financial Services Working Paper Series, December 1996.
- Summers, Bruce J., and R. Alton Gilbert. "Clearing and Settlement of U.S. Dollar Payments: Back to the Future?" *Federal Reserve Bank of St. Louis Review*, vol. 78 (September/October 1996), pp. 3–27.
- Timberlake, Richard H. *The Origins of Central Banking in the United States*. Cambridge, Mass.: Harvard University Press, 1978.
- Tippetts, Charles S. *State Banks and the Federal Reserve System*. New York: D. Van Nostrand Co., 1929.
- \_\_\_\_\_. "The Par Remittance Controversy," *American Economic Review*, vol. 14 (December 1924), pp. 629–48.
- U.S. Congress, House of Representatives, *House Report 69*, 63 Cong. 1 Sess. Washington: Government Printing Office, 1913.
- U.S. Congress, Senate, Senate Committee on Banking and Currency, *Hearings on H. R. 7837 (S. 2639)*, 63 Cong. 1 Sess., November 6, 1913. Washington: Government Printing Office, 1913, pp. 512–13.
- U.S. Treasury, Office of the Comptroller of the Currency. *Annual Report*, 1914. Washington: Government Printing Office, 1915.
- \_\_\_\_\_. *Annual Report*, 1902. Washington: Government Printing Office, 1903.

- Walter, John R. *An Examination of the Factors Motivating the Federal Reserve System's Entrance into the Provision of Check Collection Services*. Dissertation, University of Delaware, 1988.
- Watkins, Leonard L. *Banker's Balances*. Chicago: A. W. Shaw, 1929.
- Weinberg, John A. "The Organization of Private Payment Networks," *Federal Reserve Bank of Richmond Economic Quarterly*, vol. 83 (Spring 1997), pp. 25–43.
- White, Eugene Nelson. *The Regulation and Reform of the American Banking System, 1900–1929*. Princeton: Princeton University Press, 1983.
- Williams, Charles R. *The Bankers Encyclopedia*. Chicago: The Bankers Encyclopedia Co., 1901.
- Willis, Henry P. "The Federal Reserve Act," *American Economic Review*, vol. 4 (March 1914), pp. 1–24.
- Willis, Henry Parker. *The Federal Reserve System: Legislation, Organization and Operation*. New York: The Ronald Press Co., 1923.
- Willis, H. Parker, John M. Chapman, and Ralph West Robey. *Contemporary Banking*. New York: Harper & Brothers Publishers, 1933.
- Wyatt, Walter. "The Par Clearance Controversy," *Virginia Law Review*, vol. 30 (June 1944), pp. 361–97.



# A Forward-Looking Monetary Policy Reaction Function

---

---

Yash P. Mehra

**T**he Federal Reserve's reaction function, which summarizes how the Federal Reserve (Fed) alters monetary policy in response to economic developments, plays an important role in macroeconomic and policy analyses. It can be helpful in predicting actual policy actions, thereby serving as a benchmark for assessing the current stance and the future direction of monetary policy. Also, in macro models, the reaction function is central in evaluating Fed policy and determining effects of other macro policies or economic shocks, implying macroeconomic performance may itself depend upon the conduct of monetary policy. Consequently, there is considerable interest in identifying the nature of actual policy pursued by the Fed and determining whether the estimated reaction function fostered or hindered macroeconomic stability.<sup>1</sup>

Although numerous monetary policy reaction functions have been estimated, in this article I estimate one that sheds new evidence on the nature of Fed policy since 1979. In particular, I present and estimate a forward-looking

---

■ The views expressed are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

<sup>1</sup> See, for example, Clarida, Gali, and Gertler (1998b) and Taylor (1998). Clarida, Gali, and Gertler estimate a forward-looking version of the Taylor rule for the post-World War II U.S. economy and find that the interest rate policy in the Volcker-Greenspan period was more responsive to changes in expected inflation than it had been in the pre-Volcker period. They then embody these estimated policy rules within a standard business cycle model and analyze the statistical properties of inflation and output. They show that the estimated pre-Volcker rule permits larger fluctuations in inflation and real output (and therefore greater macroeconomic instability) than does the Volcker-Greenspan rule. Similarly, Taylor (1998) argues that U.S. interest rate policy during the period 1986 to 1997 is well predicted by the Taylor rule and that this period in U.S. monetary policy had the greatest degree of economic stability. Both inflation and real output had smaller fluctuations during this period. He then identifies two other periods, 1879 to 1914 and 1960 to 1979, when policy deviated significantly from the Taylor rule in that the nominal interest rate was not very responsive to both inflation and real output fluctuations. During those periods macroeconomic performance was not as good.

monetary policy reaction function that predicts the actual path of the funds rate during most of the period from 1979 to 1997. The distinguishing characteristic of this policy reaction function is that policy responds to movements in long-term inflationary expectations as evidenced by the behavior of the bond rate, an issue discussed first in Goodfriend (1993) but ignored in the recent empirical work on estimated monetary policy rules.<sup>2</sup> I also examine whether the policy reaction function has changed significantly during the 1990s, especially during Alan Greenspan's tenure as Fed Chairman. Finally, since this reaction function predicts actual policy actions fairly well, I discuss whether policy during the most recent period 1997Q1 to 1998Q2 is consistent with prior Fed behavior. This period is of interest because during this period the Fed did not adjust the funds rate in response to above-trend real growth.

The policy reaction function that I consider here has both backward- and forward-looking components. It assumes that the funds rate responds to actual inflation, increases in expected future inflation, expected output gap, and the bond rate. The funds rate response to the bond rate captures the influence of long-term inflationary expectations on policy. The empirical work here, which focuses on the behavior of the funds rate over two sample periods, 1960Q2 to 1979Q2 and 1979Q3 to 1998Q2, broadly supports this specification. However, policy responses differ across these sample periods. The most significant difference is that the funds rate has responded to movements in the bond rate after 1979 but not before. This indicates that since 1979 the Fed has been very sensitive to long-term expected inflation; so much so that for most of this period the nominal funds rate has moved more than one-for-one with actual inflation. Hence the real as well as the nominal funds rate increased in response to inflation. That is not the case in the pre-1979 period, when the nominal funds rate did not adjust one-for-one with actual inflation. In that period the real funds rate declined in response to actual inflation.

The policy reaction function given here tracks the actual behavior of the funds rate more closely since 1979 than it does in the period before. The sample period 1979Q3 to 1998Q2 spans the tenures of Paul Volcker and Alan Greenspan as Fed Chairmen. The results, however, indicate that the policy reaction function has not changed much between the Volcker and Greenspan periods. Finally, policy during the most recent subperiod 1997Q1 to 1998Q2 is consistent with prior Fed behavior. While the U.S. economy has grown at a very strong rate during this period, actual inflation has fallen steadily and

---

<sup>2</sup> Clarida, Gali, and Gertler (1998a) and Clarida, Gali, and Gertler (1998b) estimate forward-looking versions of the Taylor rule for G7 countries including the United States, but they do not examine the role of the bond rate. Nor do they examine issues relating to the stability of the reaction function and the ability to predict actual policy. Mehra (1997), on the other hand, does consider the response of policy to the bond rate and finds that the bond rate is significant in the reaction function.

long-term inflationary expectations as measured by the behavior of the bond rate have remained well behaved. Furthermore, there is also some evidence that the economy's underlying trend growth rate may have increased somewhat during the '90s. Once we consider together the influences of all these economic factors on the funds rate target, then the actual funds rate, which hovered around 5.5 percent during this subperiod, is not too different from the value predicted by the policy reaction function. On a more intuitive level, the surprisingly good performance of the economy on the inflation front combined with well-behaved long-term inflationary expectations and higher estimates of trend growth worked to neutralize the tighter policy response indicated by above-trend growth. Accordingly, the absence of any policy move during this subperiod is not out of line with prior Fed behavior.

## 1. THE MODEL AND THE METHOD

### A Forward-Looking Specification

The policy reaction function considered here builds upon the work in Taylor (1993), Mehra (1997), and Clarida, Gali, and Gertler (1998b). The particular specification estimated here can be derived using the following two equations:

$$FR_t^* = \bar{r} + a_{11}INFL_{t-1} + a_{12}(INFL_{t-1} - INFL^*) + a_{21}(EINFL_{t+1} - INFL_{t-1}) \\ + a_{31}(BR_t - EINFL_{t+k}) + a_{41}EGAP_{t+k}, \text{ and} \quad (1)$$

$$FR_t = (1 - \rho)FR_t^* + \rho FR_{t-1} + v_t; \quad 0 \leq \rho \leq 1, \quad (2)$$

where  $FR_t$  is the actual funds rate;  $FR_t^*$  is the Fed's nominal funds rate target for period  $t$ ;  $INFL$  is the inflation rate;  $INFL^*$  is the Fed's inflation target;  $GAP$  is the output gap;  $BR$  is the bond rate;  $\bar{r}$  is the economy's underlying equilibrium real interest rate;  $v_t$  is a stochastic disturbance term; and  $E$  is the expectations operator. Equation (1) specifies the economic determinants of the funds rate target. It assumes that the Fed has a target for inflation and a target for the level of output. It also assumes that the Fed pays attention to actual inflation as well as to the expected change in its future direction. Equation (1) thus hypothesizes that the funds rate target each period is determined as a function of the real rate of interest, actual inflation, and gaps between actual inflation and expected output and their respective target levels. Since the Fed is concerned with the expected future direction of inflation, equation (1) also posits that the funds rate target depends upon the change in expected future inflation and the bond rate. The other important assumption implicit in (1) is that the economy's underlying real rate of interest and the Fed's short-term target for inflation are constant in the short run.

Equation (2) specifies the actual funds rate as a weighted average of the last-period funds rate and the current-period funds rate target. It assumes the Fed smooths short-run changes in interest rates and hence the actual funds rate adjusts gradually to the target implied by economic fundamentals specified in (1) (Goodfriend 1991). The magnitude of the parameter  $\rho$  measures the degree of interest rate smoothing in Fed behavior. In equation (2) there is also a shock term  $v_t$ , indicating that the Fed may deviate transitorily from its systematic rule in (1). I, however, assume these transitory deviations from the policy rule are not serially correlated, as in true “policy shocks.” If we substitute (1) into (2), we get (3), which is the policy reaction function investigated here.

$$\begin{aligned} FR_t = & (1 - \rho)(\bar{r} - a_{12}INFL^*) + \rho FR_{t-1} + (a_{11} + a_{12})(1 - \rho)INFL_{t-1} \\ & + a_{21}(1 - \rho)(EINFL_{t+1} - INFL_{t-1}) + a_{41}(1 - \rho)(EGAP_{t+1}) \\ & + a_{31}(1 - \rho)(BR_t - EINFL_{t+1}) + v_t, \end{aligned} \quad (3)$$

where all variables are defined as before. Equation (3) is the short-run policy reaction function where the funds rate in period  $t$  is determined in part by its actual value in the previous period and in part by previous, current, and expected future values of economic factors. The reaction function indicates that the funds rate rises if actual inflation increases, if future inflation is expected to increase, if the bond rate moves relative to its value consistent with the Fed’s current forecast of near-term expected future inflation, or if the expected future output gap is positive. The parameters  $a_{ij}(1 - \rho)$ ,  $i = 2, 3, 4, j = 1, 2$ , measure these short-run responses, their magnitudes being determined in part by the degree of interest rate smoothing in Fed behavior. If the Fed does not smooth interest rates, then  $\rho$  is zero and the funds rate adjusts each period in response to changes in its economic determinants. The coefficients  $a_{ij}$ ,  $i = 2, 3, 4, j = 1, 2$ , then measure the responses of the funds rate target to economic fundamentals within each period. Consequently, the period  $t$  responses are the long-term responses.

The key feature of the short-run reaction function (3) is that the funds rate is assumed to respond to long-term inflationary expectations imbedded in the bond rate. Goodfriend (1993) has convincingly argued that in order to establish and maintain credibility to low inflation, the Fed has reacted to the long bond rate. However, since an expected future inflation variable is already included in the reaction function, the bond rate should influence the funds rate only if it contains information beyond that which is already imbedded in the Fed’s current forecasts of the future inflation rate. As a result, the funds rate target is assumed to respond to deviations of the bond rate as measured from the expected future inflation rate.

We define the steady state as the one in which the Fed has achieved its short-term objectives for inflation and real output and in which the public’s expectations of inflation are stabilized whereby long-term expected inflation

equals the constant inflation target. Clearly the bond rate has no role in the steady state because the long-term expected inflation also equals the constant target inflation. Under this definition of the steady state (with  $EGAP_{t+1} = 0$ ,  $EINFL_{t+1} = INFL_{t-1} = INF^*$ ,  $a_{31} = 0$ ,  $a_{11} = 1$ ), the policy reaction function (1) has the property that the nominal funds rate target equals the inflation rate plus the economy's underlying real rate of interest, i.e.,  $\bar{r} + INFL_t$ .<sup>3</sup> The policy reaction function (3) has thus embedded in it the Fisher relation as applied to the nominal funds rate target, indicating that economic fundamentals such as the inflation rate and the real rate of interest are the steady-state determinants of the funds rate target.

The component  $\bar{r} + (a_{11} + a_{12})INFL$  in (1) is of interest, however, for one more reason. It provides information about the long-run response of the nominal funds rate to inflation.<sup>4</sup> In particular, the estimated coefficient  $(a_{11} + a_{12})$  that appears on the level of actual inflation in (1) measures the net response of the funds rate to three inflation variables.<sup>5</sup> If its estimated value is above unity during a given sample period, it indicates that, as a result of the Fed's short-term reactions to inflation indicators, both the nominal and real funds rate target rose in response to inflation. On the other hand, if the estimated value is less than unity, then it indicates the real funds rate target declined in response to inflation. This information can be useful in assessing whether or not monetary policy is neutral during a given sample period.

It is also worth pointing out that the short-term policy reaction function (3) studied here is similar in some respects to the Taylor rule recently estimated in Taylor (1998). The policy rule estimated there is given in (4):

$$FR_t^* = \bar{r} + INF_t + d_1(INF_t - INF^*) + d_2GAP_t; FR_t = FR_t^*, \quad (4)$$

where  $INF$  is measured by the average inflation rate over the past four quarters and all other variables are defined as before. First note that the reaction function (3) studied here collapses to the policy rule (4) if we substitute actual for expected output gap and set  $\rho = 0$ ,  $a_{11} = 1.0$ ,  $a_{12} = d_1$ ,  $a_{21} = a_{31} = 0$ ,  $a_{41} = d_2$ . In other words, according to the policy rule (4), the Fed does not smooth interest rates, responds only to actual inflation and the output gap, and ignores altogether the behavior of forward-looking inflation indicators in setting the funds rate target. Therefore one can interpret the policy rule estimated here as one that relaxes the restrictions implicit in the estimated Taylor rule.

---

<sup>3</sup> I am implicitly assuming that the bond rate has been stripped of its assumed constant real rate component. In the empirical work, the bond rate variable ( $BR - EINFL_{t+1}$ ) used has been demeaned.

<sup>4</sup> This is the long-run funds rate equation estimated in Mehra (1997) and is also the steady-state component of the Taylor rule (Taylor 1993).

<sup>5</sup> To see this result, rewrite equation (1) so that all three inflation variables are in levels. Then the coefficients that appear on levels of these three variables ( $INFL, EINFL, BR$ ) sum to  $(a_{11} + a_{12})$ .

### Estimating the Forward-Looking Reaction Function

The reaction function (3) contains unobserved expected values of inflation and output gap. In estimating (3), I replace unobserved expected values with actual values and assume that the Fed's expectations of these variables are rational and hence uncorrelated with time  $t - 1$  information known to the central bank as in (5):

$$\begin{aligned} INFL_{t+1} &= E(INFL_{t+1}/I_{t-1}) + vp_{t+1} \\ GAP_{t+1} &= E(GAP_{t+1}/I_{t-1}) + vy_{t+1}, \end{aligned} \quad (5)$$

where  $vp_{t+1}$  and  $vy_{t+1}$  are forecast errors that are uncorrelated with  $t - 1$  dated information used by the central bank to forecast inflation and the output gap. If we eliminate the unobserved expected values from (3), we can rewrite (3) as (6):

$$\begin{aligned} FR_t &= (1 - \rho)(\bar{r} - a_{12}INF^*) + \rho FR_{t-1} + (a_{11} + a_{12})(1 - \rho)INFL_{t-1} \\ &\quad + a_{21}(1 - \rho)(INFL_{t+1} - INFL_{t-1}) + a_{41}(1 - \rho)GAP_{t+1} \\ &\quad + a_{31}(1 - \rho)(BR_t - INFL_{t+1}) + vv_t, \end{aligned} \quad (6)$$

where  $vv_{t+1} = v_t - (a_{21} - a_{31})(1 - \rho)vp_{t+1} - a_{41}(1 - \rho)vy_{t+1}$ . The composite error term,  $vv_t$ , is serially uncorrelated as both  $vp_{t+1}$  and  $vy_{t+1}$  are serially uncorrelated. But it is correlated with period  $t + 1$  values of actual inflation and output gap. That is, the disturbance term in (5) is correlated with the right-hand side explanatory variables. However, it can be verified that  $vv$  satisfies orthogonality conditions expressed in (7):

$$E(vv_t/I_{t-1}) = 0. \quad (7)$$

That is, the composite error term  $vv_t$  is uncorrelated with  $t - 1$  dated information used by the central bank to forecast one-period-ahead inflation and the output gap. That suggests equation (5) can be estimated by instrumental variables, using variables in the information set  $I_{t-1}$  as instruments. In particular, I follow Clarida, Gali, and Gertler (1998b) and estimate (5) using Hansen's (1982) generalized method of moments estimator. Under the identifying assumptions (7), this procedure produces efficient instrumental variables estimates. Furthermore, the procedure generates a test of identifying restrictions (7) used to estimate the model parameters.<sup>6</sup>

<sup>6</sup> I performed the generalized method of moments (GMM) estimation using the statistical package Regression Analysis for Time Series (version 4). The GMM is an efficient instrumental variables estimator. If we specify the list of instruments that are assumed to be uncorrelated with the disturbance term ( $vv_t$  in [7]) and if we know exactly the covariance matrix of this disturbance term, then the GMM estimator is in fact the generalized instrumental variables estimator of the form given below:

$$\beta = [(X'Z)(Z'\Omega Z)^{-1}(Z'X)]^{-1}(X'Z)(Z'\Omega Z)^{-1}Z'y, \quad (a)$$

### Data, Definition of Economic Variables, and Empirical Specifications of the Funds Rate Equation

The empirical work here estimates the reaction function over two sample periods, 1960Q2 to 1979Q2 and 1979Q3 to 1998Q2. It is widely believed that in the second subperiod the Fed made serious efforts to reduce the trend rate of inflation and contain inflationary expectations and that this deflationary policy was set in motion when Paul Volcker became Fed Chairman in late 1979. It is also believed that such policy has continued through the current regime of Alan Greenspan. Hence the estimated monetary policy reaction function is likely to differ between pre- and post-1979 periods.<sup>7</sup>

With regard to data and definitions, the funds rate variable (*FR*) is the average quarterly value of the effective funds rate. Inflation (*INFL*) is measured by the behavior of the (chain-weighted) GDP deflator. The output gap variable (*GAP*) is measured as the excess of actual over potential GDP. I consider two alternatives about the Fed's estimate of potential GDP. In one I follow the evidence in Hodrick and Prescott (1997) that potential GDP has a smoothly varying trend and that this trend is well approximated by passing real GDP through the Hodrick-Prescott (HP) filter with the smoothness parameter  $\lambda$  set at 1600.<sup>8</sup> Taylor (1998) also estimates policy rules with series on the potential output generated with the HP filter. Alternatively, following Clarida, Gali, and Gertler (1998b), I also consider results using potential GDP estimated from a fitted quadratic function of time. This specification assumes that trend GDP is

---

where  $X$  is the matrix of observations on the explanatory variables,  $Z$  is the matrix of observations on instruments,  $\Omega$  is the covariance matrix of the disturbance term, and  $y$  is the vector of observations on the dependent variable. In the special case where  $\Omega = \sigma^2 I$ , so that the disturbance term  $v_t$  is both homoscedastic and serially uncorrelated, the estimator (a) reduces to the simple instrumental variables estimator. However, in practice we do not know the form of  $\Omega$ . But as Hansen (1982) shows, it is possible to compute consistent estimators using a procedure that imposes little structure on the matrix  $\Omega$ . In particular, the estimates here are generated using a two-step procedure. In step one, the policy reaction function is estimated using the instrumental variables with  $\Omega = \sigma^2 I$ , and the residuals are computed. In step two, the matrix  $(Z'\Omega Z)$  is estimated using residuals as suggested in Hansen (1982) and the GMM estimator is computed replacing the component  $(Z'\Omega Z)^{-1}$  in (a) by its estimated value. In addition to specifying the list of instruments, one has to specify the length of lags on the instruments. Moreover, if the model suggests the presence of serial correlation, then one can take that into account in specifying the matrix  $\Omega$ . The empirical work here is performed using four lags of instruments. Since the policy reaction function focuses on one-step-ahead expected values, the disturbance term under that assumption is serially uncorrelated.

<sup>7</sup>The empirical work in previous research also indicates that policy responses differ significantly across these two sample periods (Mehra 1997; Clarida, Gali, and Gertler 1998b; and Taylor 1998).

<sup>8</sup>The magnitude of the smoothness parameter  $\lambda$  determines the variability of the trend component. The larger the value of  $\lambda$ , the smaller the variability of its trend component. If  $\lambda$  is chosen to be infinity, then the filtered series approaches the least squares fit of a linear trend model. Hodrick and Prescott (1997), however, show that small changes in the value of the smoothness parameter chosen ( $\lambda = 1600$ ) do not much alter the business cycle properties of real GDP.

deterministic as opposed to being stochastic. The long-term bond rate ( $BR$ ) is measured here by the nominal yield on ten-year U.S. Treasury bonds.

In some previous studies, including Mehra (1997), lagged money growth is significant when included in the reaction function. In order to investigate this issue, I also include money growth in the policy reaction function. As in previous studies, money is defined by M1 until 1982Q3 and by M2 thereafter. Moreover, as in previous studies, money growth for the period 1979Q3 to 1982Q3 is included interacting with a slope dummy variable that is defined to be unity over this subperiod and zero otherwise. This formulation is consistent with the popular view that the Fed's "new operating procedures" paid considerable attention to M1, and consequently such procedures may have been a source of movements in the funds rate target during this period. McCallum and Nelson (1998) also report that the New Operating Procedure dummy is generally significant when included in policy rules estimated there.<sup>9</sup> Given the above-noted considerations, the policy reaction function estimated here is expressed in the following form:

$$FR_t = (1 - \rho)a_0 + \rho FR_{t-1} + a_1(1 - \rho)INFL_{t-1} + a_2(INFL_{t+1} - INFL_{t-1}) \\ + a_3(BR_t - INFL_{t+1}) + a_4GAP_{t+1} + a_5M1_{t-1} + vv_t, \quad (8)$$

where  $a_0 = \bar{r} - a_{12}INFL^*$ ,  $a_1 = (a_{11} + a_{12})$ ,  $a_2 = a_{21}(1 - \rho)$ ,  $a_3 = a_{31}(1 - \rho)$ , and  $a_4 = a_{41}(1 - \rho)$ .  $M1$  in (8) is money growth measured by the behavior of M1 and all other variables are defined as before. The instrument list consists of a constant and four lagged values of the funds rate ( $FR$ ), the inflation rate ( $INFL$ ), the output gap ( $GAP$ ), the bond rate ( $BR$ ), and the growth rate of real GDP. For the subperiod 1960Q2 to 1979Q2, the instrument list also included four lagged values of M1 growth. For the other sample period the instrument list includes money growth interacting with a slope dummy that is defined to be 1 over 1979Q3 to 1982Q3 and zero otherwise.<sup>10</sup>

## 2. EMPIRICAL RESULTS

### Estimates of the Forward-Looking Reaction Function

Table 1 presents GMM estimates of the short-run monetary policy reaction function (8) for two sample periods, 1960Q2 to 1979Q2 and 1979Q3 to 1998Q2.

<sup>9</sup> In their empirical work, the New Operating Procedure dummy is simply an intercept dummy, defined to be unity over 1979Q3 to 1982Q3 and zero otherwise. Here, the New Operating Procedure dummy is a slope dummy on M1 growth.

<sup>10</sup> The choice of instruments is motivated by the view that the Fed's forecast of expected inflation and the output gap depends upon the past history of inflation, the output gap, real growth, monetary growth, and the bond rate. In addition, the history of policy actions measured by the past behavior of the funds rate is also relevant. As is clear, many of the economic variables included in the Fed's information set are consistent with the Phillips curve and monetarist views of the inflation process.

**Table 1 GMM Estimates of the Forward-Looking Reaction Function**

<b>Panel A: HP Trend</b>									
<b>Sample Period</b>	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$\rho$	$a_0$	$x_1^2$	$J$
1960Q2–1979Q2	0.71 (3.8)	0.34 (3.3)	−0.11 (1.0)	0.20 (3.6)	−0.04 (1.0)	0.76 (9.6)	2.6 (2.9)	2.3 (0.13)	24.0 (0.15)
1979Q3–1998Q1	0.64 (4.2)	0.26 (2.2)	0.27 (7.4)	0.41 (4.8)	0.28 (7.5)	0.69 (13.5)	4.6 (11.7)	5.3 (0.02)	17.9 (0.27)
<b>Panel B: Quadratic Time Trend</b>									
1960Q2–1979Q2	0.70 (3.2)	0.32 (2.6)	−0.06 (0.5)	0.12 (4.1)	−0.05 (1.0)	0.77 (8.5)	2.5 (2.5)	1.8 (0.17)	22.4 (0.21)
1979Q3–1998Q1	1.2 (8.7)	0.40 (4.2)	0.46 (7.1)	0.26 (7.2)	0.26 (6.9)	0.59 (12.1)	2.9 (7.4)	2.8 (0.09)	13.1 (0.60)

Notes: The coefficients (t-values in parentheses below) reported above are from the funds rate equation (8) of the text:

$$FR_t = (1 - \rho)a_0 + \rho FR_{t-1} + (1 - \rho)a_1 INFL_{t-1} + a_2(INFL_{t+1} - INFL_{t-1}) + a_3(BR_t - INFL_{t+1}) + a_4 GAP_{t+1} + a_5 M1_{t-1},$$

where  $FR$  is the federal funds rate;  $INFL$  is the inflation rate;  $M1$  is M1 growth;  $GAP$  is the output gap; and  $BR$  is the bond rate. For the sample period 1979Q3–1998Q1, money growth is included interacting with a slope dummy variable  $D^*M1$ , where  $D$  is a dummy that is 1 over 1979Q3–1982Q3 and 0 otherwise. The instrument set consists of a constant, four lagged values of the funds rate, inflation, the bond rate, money growth, output gap, and real  $GDP$  growth.  $J$  is the test of overidentifying restrictions and is distributed Chi-squared.  $x_1^2$  is the Chi-squared statistic that tests the null hypothesis  $a_1 = 1$ . Significance levels of these statistics are reported in parentheses below. The constant term  $a_0$  is  $(rr - a_{12}INF^*)$ , where  $rr$  is the real rate of interest and  $INF^*$  is the Fed's inflation target.

Panel A in Table 1 contains results that occur when potential output is measured with the HP trend, and Panel B contains results that occur when instead the quadratic trend is used. For all variables the coefficients reported are the short-run coefficients, with the exception of the one for actual inflation. For that variable the coefficient reported is the long-run coefficient  $a_1$ ; the short-run coefficient can be recovered by multiplying the reported coefficient by  $(1 - \rho)$ , i.e.,  $a_1(1 - \rho)$  in (8).<sup>11</sup> The J-statistic reported there tests the null hypothesis that

<sup>11</sup> As noted before, the policy reaction function (8) is nonlinear in parameters, with  $\rho$  appearing in front of many variables including  $FR_{t-1}$ . One could estimate (8) with or without imposing these nonlinear restrictions. If restrictions are imposed, then one gets estimates of the long-term coefficients and the short-term smoothness parameter  $\rho$ . Given long-term estimates, the short-term coefficients are recovered by multiplying the estimated long-term coefficients by  $(1 - \rho)$ . Alternatively, one may ignore these nonlinear restrictions and estimate directly the following version of equation (8):

$$FR_t = d_0 + d_1 FR_{t-1} + d_2 INFL_{t-1} + d_3(INFL_{t+1} - INFL_{t-1}) + d_4(BR_t - INFL_{t+1}) + d_5 GAP_{t+1} + d_6 M1_{t-1} + v_t,$$

orthogonality restrictions imposed under GMM estimation are consistent with data. If this statistic is small, then it indicates the restrictions are not rejected by the data; therefore GMM estimates are consistent.

If one focuses on post-1979 estimates, one can see that all estimated coefficients have expected signs and are generally statistically significant (see t-values in parentheses below estimates in Panels A and B, Table 1). Those estimates indicate that the funds rate rises if actual inflation rises, if future inflation is expected to increase, if output is expected to be above trend, or if the current-period bond rate moves relative to the expected future inflation rate. The estimated short-run coefficients ( $a_2, a_3$ ) that appear on two future inflation indicators are positive, indicating the funds rate target responded to expected changes in the future direction of actual inflation. In particular, the estimated short-run coefficient ( $a_3$ ) on the bond rate is 0.3 to 0.4, which indicates the funds rate increases 30 to 40 basis points in response to 1 percentage point increase in the bond rate.<sup>12</sup> The M1 growth dummy is significant, which shows the influence of “new operating procedures” on the funds rate. The point-estimate of the coefficient that appears on the lagged funds rate ( $\rho$ ) is between 0.6 to 0.8, which indicates the presence of considerable interest rate smoothing in Fed behavior. Finally, the reported J-statistic is small, which indicates the restrictions imposed in deriving GMM estimates are consistent with data.<sup>13</sup>

The results do not change qualitatively between two alternative measures of potential output used here. However, magnitudes of individual coefficients that measure responses of the funds rate target to different economic fundamentals are sensitive to the measure of trend used. The one difference to highlight is the long-run, estimated coefficient ( $a_1$ ) that appears on the level of actual inflation. This coefficient is greater than unity if output gap is estimated with the quadratic trend, suggesting that in the post-1979 period, the real funds rate target rose in response to inflation. However, the point estimate of  $a_1$  falls below unity if output gap is estimated instead with the HP trend (compare estimates of  $a_1$  in Panels A and B, Table 1). Nonetheless, as discussed later, the point

---

where all variables are defined as before. The estimated coefficients  $d_i, i = 0, 1, 2, 3, 4, 5, 6$ , are then estimates of short-term coefficients. The long-term coefficients can then be recovered by multiplying the short-term estimates by  $[1/(1 - d_1)]$ . Both these procedures yield qualitatively similar results. The empirical work reported here is based mostly on estimates generated using nonlinear restrictions.

<sup>12</sup> Since the Fed smoothes interest rates in the short run, the long-term estimated response measured by  $a_3/(1 - \rho)$  is stronger. In the HP trend case, the estimated long-run response is  $0.26/(1 - 0.69)$ , i.e., 0.87.

<sup>13</sup> The overall fit of the regression as measured by the standard error of estimate is better when four lagged values of the instruments are used in estimation than it is when two or three lagged values are used. The results, however, do not change qualitatively if the policy reaction function is instead estimated using two to three lagged values of the instruments. In particular, the bond rate remains significant in the post-1979 period. The J-statistic continues to confirm the condition in (7) that the error term is not correlated with instruments.

estimate of  $a_1$  has remained above unity in most other subperiods ending in the '90s (see estimates of  $a_1$  in Tables 3 and 4 for various subperiods). Together these estimates indicate that during most of the Volcker-Greenspan period the real funds rate target increased in response to actual inflation.<sup>14</sup>

If one focuses on pre-1979 estimates, one can see that, like the post-1979 estimates, these estimates indicate that the funds rate target rises in response to actual inflation, to increases in expected future inflation, and to the positive expected output gap. But pre-1979 estimates differ significantly from post-1979 estimates in several ways. First, the bond rate is not significant, indicating that in the pre-1979 period the Fed did not adjust the target in response to movements in the bond rate. This result on the absence of long-term inflationary expectations on policy is robust to changes in the measure of trend used. Second, money growth is also not significant.<sup>15</sup> Third, the long-run, estimated coefficient ( $a_1$ ) that measures the response of the funds rate to inflation is economically less than unity, indicating that the funds rate did not adjust one-for-one with inflation. That is, the real funds rate declined in response to actual inflation prior to 1979. This decline occurred despite the evidence here that in the pre-1979 period the funds rate is responsive to movements in near-term expected future inflation (see estimates of  $a_2$  in Table 1). The absence of response to the bond rate may explain why the real funds rate target declined in response to actual inflation during the pre-Volcker period.

#### **Assessing the Predictive Content of the Policy Reaction Function: Pre- and Post-1979**

The key hypotheses posited about Fed behavior here are that the fund rate responds to economic fundamentals specified in (3). The estimates discussed above lend support to these hypotheses. In order to assess further the empirical plausibility of these hypotheses, I examine how well the policy reaction function predicts the actual behavior of the funds rate during the two sample periods. The predictive content is evaluated with the following regression:

$$FR_t = c + dPFR_t + ee_t, \quad (9)$$

<sup>14</sup> Table 1 also reports estimates of the constant term  $a_0 = \bar{r} - a_{12}INFL^*$  in (8). But, as is obvious, it is not possible to recover estimates of the assumed constant real rate  $\bar{r}$  without information about the Fed's constant inflation target.

<sup>15</sup> In the forward-looking reaction function estimated here, money growth, when included, is generally not significant, with the exception of the brief "new operating procedures" period 1979Q3 to 1982Q3. This result is in contrast with the one reported in many previous studies, where money growth is significant. One explanation of these different results is that the reaction functions in these studies are backward-looking. Hence money growth may be significant in these studies not because the funds rate responds to money growth but because past money is giving information about future economic factors that are omitted from the reaction function.

where  $PFR$  is the funds rate predicted by the policy reaction function and  $ee$  is the disturbance term. The predicted values used in (9) are the dynamic within-sample simulated values of the policy reaction function (8), generated using actual values of explanatory variables. However, in order to highlight the importance of the effect of interest rate smoothing on the funds rate target, I also generate predictions of the funds rate with the smoothing parameter  $\rho$  in (8) set to zero. The predicted funds rate is an unbiased predictor of the actual funds rate if  $c = 0$  and  $d = 1$ .<sup>16</sup>

Table 2 reports estimates of the regression (9) for two sample periods, 1961Q1 to 1979Q2 and 1981Q1 to 1998Q1. The results are reported for both measures of the output gap and with and without accounting for the effect of interest rate smoothing (see Panels A and B). (Figures 1 through 4 chart predicted and actual values of the funds rate for the HP filtered output gap.) If one focuses on post-1979 sample results with smoothing, one can see that the coefficient that appears on the predicted fund rate variable  $PFR$  in (9) is close to unity.  $\chi^2_1$  is the Chi-squared statistic that tests the null hypothesis that  $(c, d) = (0, 1)$ . This statistic is not significant at the 5 percent level, suggesting the predicted funds rate is an unbiased predictor of the actual funds rate. The result is not sensitive to the measure of trend used or to the presence of interest rate smoothing in Fed behavior (compare results in Panels A and B, Table 2). Figures 2 and 4 tell the same story, which is that the funds rate moves closely with the level determined by economic fundamentals as specified in (1). The results, however, are different in the pre-1979 period. When the focus is on results without smoothing, the coefficient that appears on the predicted funds rate  $PFR$  is significantly below unity and this result is not sensitive to the measure of the output gap. If one allows for the effect of interest rate smoothing on the funds rate, the coefficient that appears on the predicted funds rate does however move closer to unity (compare Figure 1 with Figure 3). But the predicted funds rate is still a biased predictor of the actual funds rate in one specification of the output gap (compare estimates in Panels A and B, Table 2). Hence the hypothesis that the funds rate target is a function of economic variables as specified in the policy reaction function (1) is a better description of Fed policy in the post-1979 period than it is in the previous period.

---

<sup>16</sup> The result that the predicted funds rate may be a biased predictor of the actual funds rate in (9) does not imply that the stochastic disturbance term in the estimated policy reaction function (8) is biased and the estimation procedure used here is therefore invalid. The reason is that the result above—that the predicted funds rate is a biased predictor—may arise because the predicted values used in regression (9) are dynamic, not static. The policy reaction function (8) instead is estimated using actual values of explanatory variables including the lagged funds rate. Hence the disturbance term, though unbiased in (8), may appear biased in (9) if predicted values used are dynamic.

**Table 2 In-Sample Predictability of the Forward-Looking Reaction Function**

<b>Panel A: Actual and Predicted Funds Rate, Without Smoothing</b>						
<b>Sample Period</b>	<b>HP Trend</b>			<b>Quadratic Trend</b>		
	<i>c</i>	<i>d</i>	$x_1^2(2)$	<i>c</i>	<i>d</i>	$x_1^2(2)$
1961Q1–1979Q2	2.7 (3.5)	0.46 (4.9)	38.5*	2.2 (3.5)	0.53 (5.7)	25.9*
1981Q1–1998Q1	1.0 (1.3)	0.89 (10.1)	1.8	0.64 (1.6)	0.93 (20.9)	2.6
<b>Panel B: Actual and Predicted Funds Rate, With Smoothing</b>						
1961Q1–1979Q2	0.4 (1.2)	0.89 (18.9)	5.3**	0.2 (0.2)	0.88 (8.2)	4.3
1981Q1–1998Q1	-0.27 (0.5)	1.0 (16.7)	0.5	0.2 (0.6)	0.96 (22.7)	0.8

\*Significant at the 5 percent level.

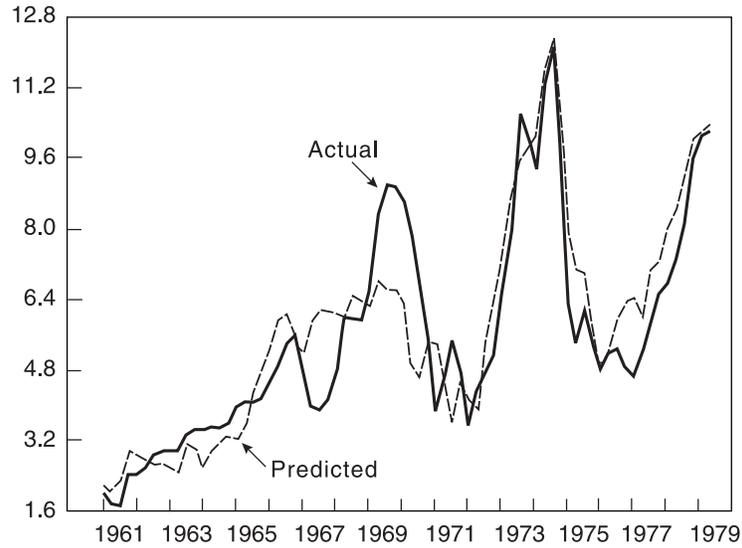
\*\*Significant at the 10 percent level.

Notes: The coefficients (t-values in parentheses below) reported above are from regressions of the form  $FR_t = c + dPFR_t$ , where  $FR$  is the actual funds rate, and  $PFR_t$  is the predicted funds rate. The predicted values are generated using the policy rule (8) of the text, with  $\rho$  set to zero (see Panel A) and with  $\rho$  set to its estimated value (see Panel B). The predicted values used in Panel B regressions are within-sample but dynamic.  $x_1^2$  is the Chi-squared statistic that tests the null hypothesis  $(c, d) = (0, 1)$ .

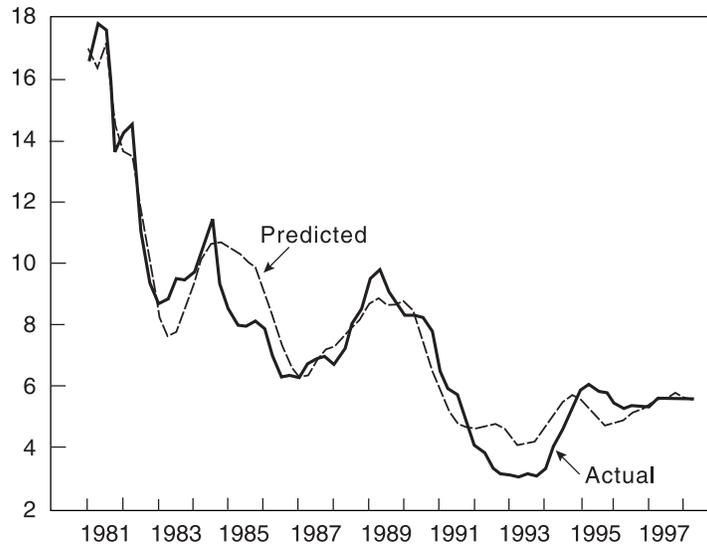
### Has the Reaction Function Changed during the Greenspan Period?

It is commonly believed that the Fed under Alan Greenspan has maintained an anti-inflationary stance set in motion when Paul Volcker became the Fed Chairman. In fact, as discussed above, the forward-looking reaction function (8) estimated here is consistent with the actual path of the funds rate over the Volcker-Greenspan period, indicating the policy reaction function may not have changed much over this period. Nevertheless, I investigate this issue further by comparing the reaction function between Volcker and Greenspan periods, 1979Q4 to 1987Q3 and 1987Q3 to 1998Q2. The main problem with estimating separate reaction functions over the Volcker and Greenspan periods is that those estimates may be subject to the small sample bias. In fact, during the Greenspan period both the bond rate and inflation have not varied much. In view of these considerations I estimate the reaction function using the technique of rolling regressions, which generates somewhat larger subsamples. In particular, I begin with estimating the reaction function over the Volcker period 1979Q4 to 1987Q3 and then continually reestimate it advancing the end date by four quarters, keeping the start date fixed. The resulting estimates of key coefficients

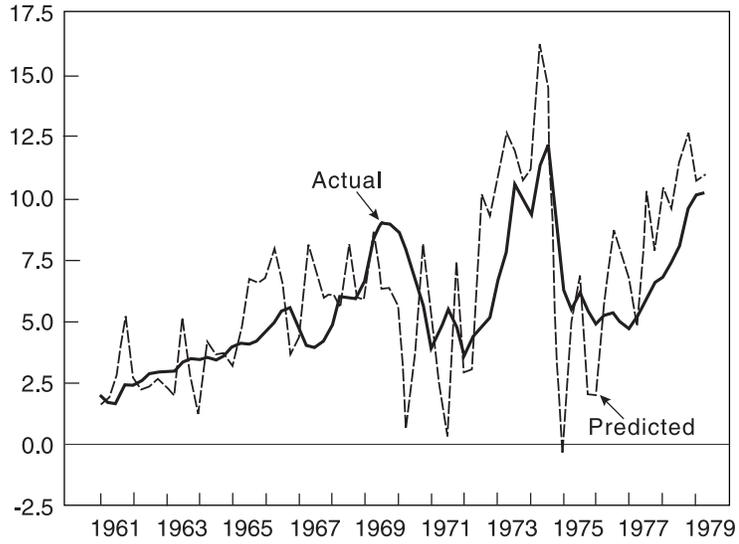
**Figure 1 Actual and Predicted Funds Rate, With Smoothing;  
HP Trend Pre 1979**



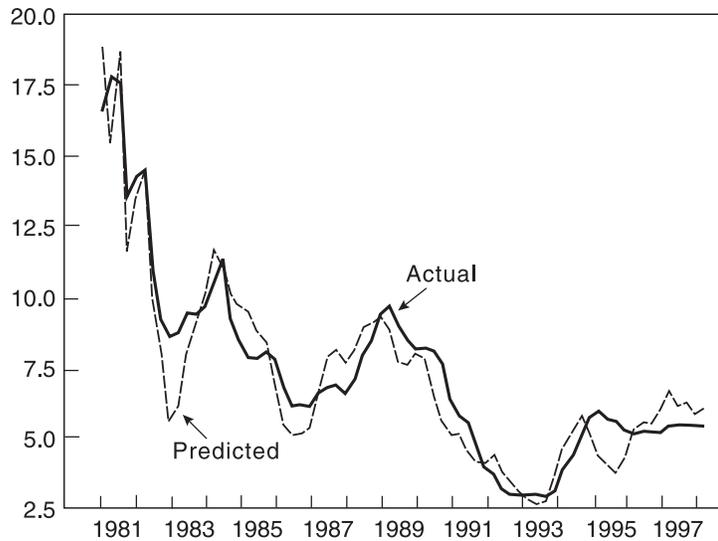
**Figure 2 Actual and Predicted Funds Rate, With Smoothing;  
HP Trend Post 1979**



**Figure 3 Actual and Predicted Funds Rate, Without Smoothing; HP Trend Pre 1979**



**Figure 4 Actual and Predicted Funds Rate, Without Smoothing; HP Trend Post 1979**



are reported in Tables 3 and 4. Table 3 reports estimates using the HP filtered output gap, and Table 4 gives estimates using the quadratic trend output gap.

If we focus on short-run estimated coefficients,  $a_i, i = 2, 3, 4$ , they all have theoretically expected signs in all subsamples considered here and are generally statistically significant (t-values not reported). The estimated sizes of the individual coefficients are relatively stable over different subsamples, with few exceptions. Of particular interest is the long-run, estimated coefficient  $a_1$  that appears on the level of actual inflation. It is mostly above or close to unity in all subsamples, with the exception of those ending in 1996 and 1997. In those two subsamples the magnitude of the estimated coefficient is sensitive to the measure of the output gap used: it falls below unity if the output gap is estimated with the HP trend (compare estimates of  $a_1$  in Tables 3 and 4). These estimates thus suggest that the Fed during most of the Greenspan period has responded aggressively enough to expected inflation indicators that the real funds rate has increased in response to actual inflation.

### **Predicting the Funds Rate during the Greenspan Period**

The results discussed in the previous section indicate that estimates of the individual coefficients that measure responses of the funds rate to economic fundamentals do display considerable subsample variability during the Volcker-Greenspan period. But despite such variability, a cursory look at Figure 2 indicates that the policy reaction function here tracks the actual behavior of the funds rate fairly well in the '80s and the '90s. In this section I provide additional evidence on this issue by examining the out-of-sample predictive performance of the reaction function over the period 1988Q1 to 1998Q2, which for the most part spans the current tenure of Alan Greenspan as the Fed Chairman.

Table 5 presents the predicted values of the funds rate. The predicted values are the dynamic one-year-ahead forecasts of the funds rate that are conditional on actual values of the economic fundamentals and are generated using rolling regressions over the forecast period. Panel A in Table 5 presents the predicted values generated with the HP trend and Panel B presents the predicted values with the quadratic trend. (Figures 5 and 6 chart the quarterly values for this period.) Actual values of the funds rate, prediction errors, and summary error statistics are also presented. As shown, the reaction function tracks the actual funds rate fairly well over the forecast period. The mean value of the prediction error is very small and the root mean squared error is about 0.4 of a percentage point. The average annual prediction errors are not statistically significant, with the exception of the year 1995. In 1995 the prediction error is positive, and it is more than twice the root mean squared error. During that year real growth decelerated from its rapid pace of the previous year, but the Fed did not lower the funds rate in response to such a slowdown. Nevertheless, since 1995 the magnitude of the prediction error has steadily declined.

**Table 3 Rolling Regression Estimates of the Reaction Function during Volcker-Greenspan Periods, GDP Deflator and HP Trend**

Sample Period Ends in	$a_1$	$a_2$	$a_3$	$a_4$	$\rho$	$a_0$	$J(sl)$
1987Q4	1.0	0.6	0.48	0.25	0.42	3.6	17.4 (0.29)
1988Q4	1.0	0.6	0.45	0.28	0.43	3.4	21.6 (0.12)
1989Q4	1.1	0.5	0.37	0.40	0.50	3.5	19.3 (0.20)
1990Q4	1.2	0.6	0.35	0.32	0.50	3.1	16.0 (0.38)
1991Q4	1.2	0.6	0.35	0.36	0.51	3.1	14.1 (0.52)
1992Q4	1.2	0.6	0.35	0.41	0.56	2.7	15.0 (0.45)
1993Q4	1.4	0.7	0.39	0.39	0.57	1.9	16.1 (0.37)
1994Q4	1.5	0.7	0.39	0.38	0.59	1.6	15.5 (0.41)
1995Q4	1.1	0.5	0.33	0.38	0.61	3.2	14.7 (0.47)
1996Q4	0.8	0.4	0.29	0.41	0.64	4.1	16.1 (0.38)
1997Q4	0.7	0.4	0.28	0.32	0.66	4.5	17.1 (0.31)
1987Q3–1998Q1	0.7	0.1	0.28	0.35	0.81	4.7	17.8 (0.27)

Notes: The coefficients above are GMM estimates of the forward-looking reaction function given in Table 1. Unless stated otherwise the estimation period for all regressions begins in 1979Q3 and ends in the year as shown in the first column above. All reported coefficients have significant t-values (not reported), with one exception:  $a_2$  is not significant over 1987Q3–1998Q2.

**Table 4 Rolling Regression Estimates of the Reaction Function during Volcker-Greenspan Periods, GDP Deflator and Quadratic Trend**

Sample Period Ends in	$a_1$	$a_2$	$a_3$	$a_4$	$\rho$	$a_0$	$J(sl)$
1987Q4	1.2	0.7	0.54	0.17	0.39	3.3	14.6 (0.49)
1988Q4	1.2	0.7	0.54	0.17	0.39	3.1	15.4 (0.42)
1989Q4	1.1	0.6	0.47	0.25	0.50	3.6	14.3 (0.50)
1990Q4	1.3	0.6	0.47	0.23	0.48	3.2	14.5 (0.48)
1991Q4	1.3	0.7	0.49	0.26	0.50	3.0	14.1 (0.52)
1992Q4	1.4	0.7	0.48	0.29	0.54	2.6	15.3 (0.43)
1993Q4	1.5	0.7	0.49	0.30	0.54	2.3	14.3 (0.50)
1994Q4	1.5	0.7	0.49	0.29	0.54	2.2	14.1 (0.52)
1995Q4	1.3	0.5	0.43	0.30	0.59	3.1	13.7 (0.55)
1996Q4	1.2	0.5	0.44	0.31	0.59	3.3	12.6 (0.63)
1997Q4	1.3	0.5	0.45	0.28	0.57	2.9	12.1 (0.60)
1987Q3–1998Q1	1.4	0.2	0.28	0.20	0.73	2.5	17.5 (0.15)

Notes: See notes for Table 3.

**Table 5 Actual and Predicted Funds Rate: 1988Q1–1998Q2**

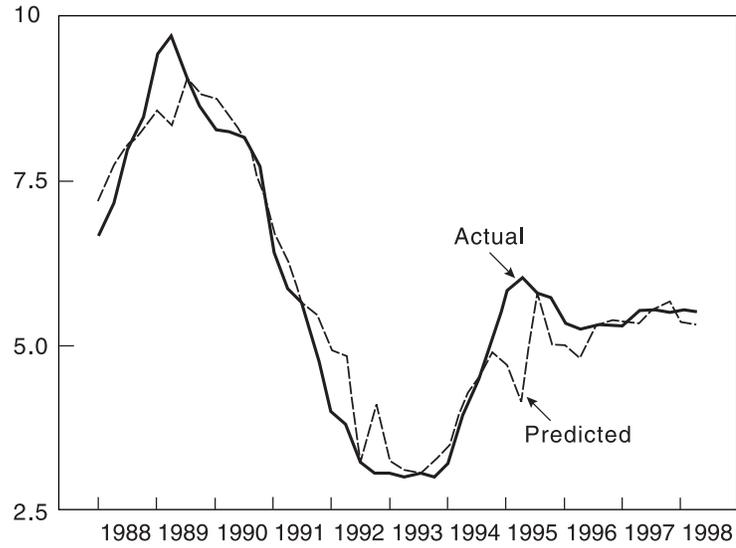
Year	Actual Funds Rate	Panel A: HP Trend		Panel B: Quadratic Trend	
		Predicted	Error	Predicted	Error
1988	7.6	7.8	−0.20	7.9	−0.30
1989	9.2	8.7	0.50	8.6	0.60
1990	8.1	8.2	−0.01	8.1	0.00
1991	5.7	6.0	−0.30	6.0	−0.30
1992	3.5	4.3	−0.80	4.1	−0.60
1993	3.0	3.1	−0.01	2.9	0.10
1994	4.2	4.2	0.00	4.2	0.00
1995	5.8	4.9	0.90*	5.1	0.70
1996	5.3	5.1	0.20	5.4	−0.10
1997	5.5	5.4	0.10	5.9	−0.40
1997Q1	5.3	5.3	0.0	5.9	−0.60
1997Q2	5.5	5.3	0.2	6.5	−1.0
1997Q3	5.5	5.5	0.0	6.7	−1.2
1997Q4	5.5	5.6	−0.1	6.2	−0.7
1998Q1	5.5	5.3	0.2	5.7	−0.2
1998Q2	5.5	5.2	0.3	5.9	−0.04
Mean Error (1988–1997)			0.00		−0.02
Root Mean Squared Error			0.44		0.41

\*The prediction error is twice the root mean squared error.

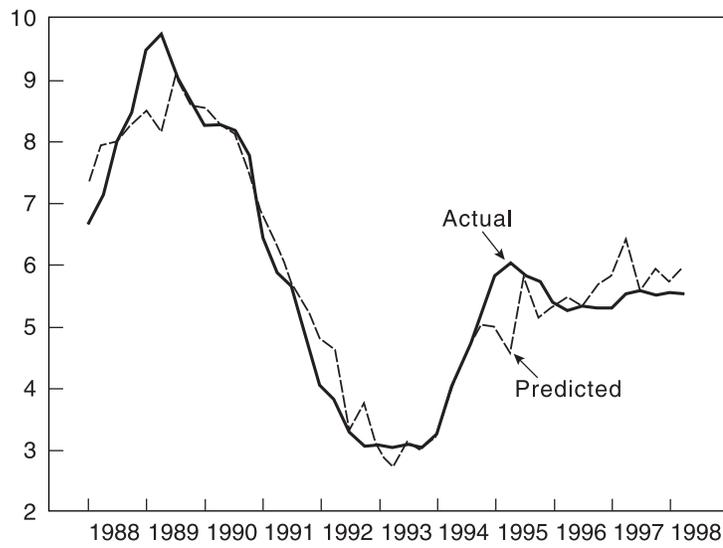
Notes: The predicted values above are generated using the forward-looking reaction functions reported in Tables 3 and 4. The predicted values are the dynamic, one-year-ahead forecasts, generated using rolling regressions over sample periods that all begin in 1979Q3 but end in the year before the forecast period.

Table 5 also presents the quarterly values of the predicted funds rate for the most recent period 1997Q1 to 1998Q2. During this period the U.S. economy has grown at an above-trend growth rate, while inflation has steadily declined. The Fed, however, did not adjust the funds rate target during this subperiod, which remained at 5.5 percent. The funds rate predicted by the reaction function is 5.4 percent if the Fed's estimate of trend GDP is generated with the HP filter, and it is 6.2 percent if instead the quadratic trend is used. There is a difference in prediction because in this subperiod the Fed's estimate of trend GDP is 3.0 percent if the HP trend is used and 1.9 percent if instead the quadratic trend is employed to estimate the output gap. Consequently, the magnitude of the positive output gap is smaller with the HP trend than with the quadratic, thereby suggesting a less tight response to observed above-trend growth. If one believes the popular view that the economy's underlying trend growth rate

**Figure 5 Actual and Predicted Funds Rate, Out-of-Sample; HP Trend**



**Figure 6 Actual and Predicted Funds Rate, Out-of-Sample; Quadratic Trend**



may have increased in the '90s, then one may have more confidence in the reaction function estimated with the HP filtered output gap than in the one with the quadratic. Therefore, the absence of policy response to recent above-trend growth is not inconsistent with prior Fed behavior if one takes seriously the proposition that the economy's trend growth in recent years may be higher than 2 percent estimated using the quadratic trend.

### 3. CONCLUDING OBSERVATIONS

In this article I estimate a forward-looking monetary policy reaction function that quite accurately predicts the actual behavior of the federal funds rate since 1979. The key property of the estimated reaction function in the Volcker-Greenspan period is that the funds rate target is responsive to movements in long-term inflationary expectations evidenced by the behavior of the bond rate. During this period the Fed has responded aggressively enough to future inflation indicators that the real funds rate target increased in response to actual inflation. Such is not the case in the pre-1979 period, when the real funds rate target declined in response to inflation.

It has been suggested that the U.S. economy experienced macroeconomic stability in the Volcker-Greenspan period because the interest rate policy pursued during this period was very responsive to expected inflation; so much so that the real funds rate increased in response to inflation (Clarida, Gali, and Gertler 1998b; Taylor 1998). The results here indicate that the Fed's willingness to move the funds rate preemptively to react to movements in the bond rate may explain why the funds rate was more responsive to inflation in the Volcker-Greenspan period than it had been in the previous period.

---

---

### REFERENCES

- Clarida, Richard, Jordi Gali, and Mark Gertler. "Monetary Policy Rules in Practice: Some International Evidence," *European Economic Review*, vol. 42 (June 1998a), pp. 1033–67.
- . "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," NBER Working Paper 6442, March 1998b.
- Goodfriend, Marvin. "Interest Rate Policy and the Inflation Scare Problem 1979–1992," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 79 (Winter 1993), pp. 1–24.
- . "Interest Rates and the Conduct of Monetary Policy," *Carnegie-Rochester Conference Series on Public Policy*, vol. 34 (Spring 1991), pp. 7–30.

- Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, vol. 50 (July 1982), pp. 1029–54.
- Hodrick, Robert J., and Edward C. Prescott. "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit, and Banking*, vol. 29 (February 1997), pp. 1–16.
- McCallum, Bennett T., and Edward Nelson. "Nominal Income Targeting in an Open-Economy Optimizing Model," Mimeo. A paper prepared for the Riksbank-IIES Conference on Monetary Policy Rules. June 1998.
- Mehra, Yash P. "A Federal Funds Rate Equation," *Economic Inquiry*, vol. XXXV (July 1997), pp. 621–30.
- Taylor, John B. "An Historical Analysis of Monetary Policy Rules," National Bureau of Economic Research conference on Monetary Policy Rules, January 1998.
- \_\_\_\_\_. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, vol. 39 (December 1993), pp. 195–214.



# Mercantilists and Classics: Insights from Doctrinal History

---

---

Thomas M. Humphrey

Economists typically view their discipline as a progressive science in which superior new ideas relentlessly supplant inferior old ones in a Darwinian struggle toward the truth. Thus it came as something of a shock when Milton Friedman challenged this belief in the May 1975 issue of the *American Economic Review*. In response to the question “What have we learned in the past 25 years?”, Friedman argued that what monetary economists have learned since 1950 are hardly new ideas but rather a rediscovery of old ideas inherited from David Hume and his contemporaries more than 200 years ago.

Three years later, the British economist Ivor F. Pearce shocked his readers even more. He denied that the Keynesian Revolution had contributed a single new or useful idea to monetary economics. Instead, he insisted that “human history is guided not by new ideas, for there are none,” but rather by “some ephemeral sub-group of . . . old ideas.” Such old ideas, “often believed to be new,” are “seized upon as the . . . solution to whatever difficulties immediate experience has made to seem important, and congealed into a crust of dogma by endless repetition and obeisance” (Pearce 1978, p. 93).

The above sentiments express what every doctrinal historian knows, namely that much of what passes for novelty and originality in monetary theory and policy is ancient teaching dressed up in modern guises. To be sure, the increasing application of mathematical modeling has given these concepts greater rigor and precision. Likewise, better data and more powerful empirical techniques have improved our statistical estimates of the relevant quantitative magnitudes. Still, the basic ideas themselves often remain much the same. Thus instead

---

■ Tom.Humphrey@rich.frb.org. An earlier version of this article appeared in this Bank’s *1998 Annual Report*. For valuable comments the author is indebted to his Richmond Fed colleagues Alice Felmlee, Marvin Goodfriend, Bob Hetzel, Rowena Johnson, Elaine Mandalers, and Ned Prescott. The views expressed do not necessarily reflect those of the Federal Reserve System.

of a steady progression of new paradigms, we see repeated cycles of existing ones whose periodic rise and fall perpetually casts them in and out of fashion.

By itself, this recycling of established ideas need be no cause for alarm. Theories may survive because experience indicates that they possess a high degree of validity and because no better theories have been found. The trouble is, however, that sound theories are not the only ones to survive. Unsound theories may coexist with the sound ones.

Unfortunately, policymakers and the general public are in no position to realize as much. Preoccupied by the pressing problems of the day, they have neither the time, inclination, or training, nor indeed the duty to trace the history of the ideas they employ or endorse. They have no reason to be aware of earlier policy debates in which sound theories were distinguished from fallacious ones. The result is that policymakers may subscribe to old theories under the mistaken impression that those theories are new. Worse, they may unwittingly deploy policies whose underlying theory has been challenged and found wanting in earlier policy debates.

Here is where the doctrinal historian can help. His comparative advantage lies in identifying the origin and tracing the evolution of rival monetary doctrines across a succession of writers, events, episodes, and policy controversies. Each such incident constitutes a test, or observation, of the relative strengths and weaknesses of the competing doctrines. While no single test can yield conclusive results, many such tests may do so. Taken together, they reveal which doctrine has emerged from past experience as the more robust analytically. By demonstrating as much, the historian specifies those ideas that seem to offer the most effective basis for public policy. Of course, there is no assurance that the policymaker will heed the doctrinal historian and employ the best ideas. On the contrary, he may reject them or temporarily accept and subsequently abandon them. Here again the historian has something to say. His study of the forces influencing the receptivity and implementation of ideas permits him to predict a doctrine's prospective success or failure. In this manner, the unique perspectives of doctrinal history may prove their worth.

This article puts those perspectives to work. It shows that from a broad standpoint much of the history of monetary theory reduces to a struggle between opposing mercantilist and classical camps. Mercantilists, with their fears of hoarding and scarcity of money together with their prescription of cheap (low interest rate) and plentiful cash as a stimulus to real activity, tend to gain the upper hand when unemployment is the dominant problem. Classicals, chanting their mantra that inflation is always and everywhere a monetary phenomenon, tend to prevail when price stability is the chief policy concern.

Currently, the classical view is in the driver's seat. By all rights it should remain there since it long ago exposed the mercantilist view as fundamentally flawed. It is by no means certain, however, that the classical view's reign is secure. For history reveals that, whenever one view holds center stage, the

other, fallacious or not, is waiting in the wings to take over when the time is ripe. In this manner, the mercantilism of John Law and Sir James Steuart gave way to the classicism of David Hume and David Ricardo, the Currency School's classicism bowed to John Maynard Keynes's mercantilism, the mercantilist doctrines of Keynes's disciples yielded to Milton Friedman's classical monetarism, and so forth. Even today, with central bankers in several nations expressing commitment to the classical goal of price stability and monetarists advocating systematic, zero-inflation rules for monetary policy, mercantilist undercurrents still run strong. Supply-siders who argue that monetary policy must be accommodative to allow tax cuts to work their magic echo mercantilist opinion. So too do those who contend that, with global competition and rapid technological progress holding inflation in check, monetary policy is free to pursue nonprice objectives such as boosting growth and achieving full employment. Finally, observers who believe that monetary policy is powerless to stimulate the currently depressed Japanese economy harbor mercantilist fears of unspent hoards of idle cash.

The following paragraphs attempt to spell out the core propositions of the original mercantilist and classical views and to establish the centrality of those propositions in the famous Currency School-Banking School and Keynesian-monetarist controversies—the two leading monetary policy debates of the nineteenth and twentieth centuries.<sup>1</sup> From this doctrinal historical exercise, three themes emerge. First, with some exceptions, classicals tend to be quantity theorists; mercantilists, anti-quantity theorists. Second, classicals prefer rules; mercantilists, discretion. Third, for all their cogency, classicals may be doomed to face a perpetual mercantilist challenge. As long as some observers continue to believe, rightly or wrongly, that inflation and deflation are nonmonetary, or real, phenomena and that unemployment is a monetary one capable of correction by the central bank, the debate will be unending.

## **1. MERCANTILIST AND CLASSICAL MONETARY DOCTRINES**

The roots of the debate trace back to the original mercantilist writers of the preclassical era 1550–1770. Those writers argued that a nation's stock of precious metals constituted the source of its plenty (wealth), power, prestige, and prosperity. For countries possessing no gold mines, augmentation of those conditions required the accumulation of specie through foreign trade. Accordingly, mercantilists advocated protectionist policies in the form of export

---

<sup>1</sup> Additional famous policy debates pitting mercantilists and classicals include (1) the Swedish Bullionist controversy (1755–1765), (2) the English Bullionist-Antibullionist, or Bank Restriction, dispute (1797–1821), (3) the Bimetallism debate (1880–1896), and (4) the German hyperinflation debate (1922–1923).

promotion and import restriction schemes to obtain a permanent trade balance surplus matched by corresponding persistent inflows of specie from abroad.

This policy prescription was of course the mercantilists' main claim to fame. But the hallmark that secures them a permanent niche in the history of monetary doctrines was their contra- or anti-quantity theory of money.<sup>2</sup> They used that theory to deny that money determines prices and to tout the employment benefits of money-stock expansion fueled either by specie inflows or by paper money creation should those inflows languish. Consisting of at least seven propositions, the mercantilists' contra-quantity theory held that (1) money stimulates trade, (2) real cost-push forces determine the price level and the inflation rate, (3) the interest rate is a purely monetary variable whose level, high or low, is proof of the scarcity or abundance of money, (4) idle hoards absorb any cash not employed in driving trade, (5) causality runs from prices and real activity to money such that the money stock passively adapts to the needs of trade, (6) overissue is impossible when the money stock is backed by the nominal value of real property, and (7) discretion outperforms rules in the conduct of monetary policy.

### **John Law (1671–1729)**

The clearest and most emphatic statements of the foregoing propositions came from John Law and Sir James Steuart, two economists writing near the close of the mercantilist era.<sup>3</sup> Of the two, Law's name is synonymous with the money-stimulates-trade doctrine that forms the central core and theme of his 1705 *Money and Trade Considered; with a Proposal for Supplying the Nation with Money*. Writing against the backdrop of a chronically depressed and under-employed Scottish economy (his home country), he argued that a shortage of metallic money was to blame, that a bank-issued paper currency must replace the deficient metallic one, and that the resulting expansion of the stock of paper notes would permanently increase the level of output and employment without raising prices.<sup>4</sup> His argument stemmed from his assumptions of (1) the

---

<sup>2</sup> Because anti-quantity theory elements also characterize the fixed-exchange-rate, small-open-economy case of the modern *monetary approach to the balance of payments*, some observers may be tempted to equate mercantilism with that approach. In fact, however, the two theories differ markedly. First, the monetary approach applies the quantity theory, rather than its opposite, to closed-economy and inconvertible-paper floating-exchange-rate regimes. By contrast, mercantilists, with few exceptions, tended to apply the anti-quantity theory indiscriminately to all regimes. Second, the monetary approach rejects the mercantilist money-stimulates-trade doctrine.

<sup>3</sup> On Law's monetary theory, see Murphy (1997, Chs. 6 and 8) and Hutchison (1988, pp. 134–40). On Steuart's theory, see Eltis (1986), Hutchison (1988, pp. 341–51), Meek (1967), and Skinner (1981).

<sup>4</sup> Law's fear of monetary shortage under a metallic standard is incompatible with the monetary approach to the balance of payments. The latter sees a small open economy, like Scotland, taking its price level as given from the closed world economy with money then flowing in through the balance of payments to support that price level such that no monetary shortage occurs. Of these

availability of idle resources at unchanged resource prices and (2) constant returns to scale in production. Given these conditions, it followed that the economy's long-run aggregate supply curve was perfectly horizontal up to the point of full employment. It likewise followed that money-induced increases in aggregate commodity demand would, via rightward shifts along the supply curve, generate matching increases in equilibrium real output without raising prices. Indeed, Law suggested that the price level might even fall if scale economies in production rendered the aggregate supply curve negatively sloped.<sup>5</sup> In no case, however, would expansion of the stock of paper money raise prices.

Having argued that causation runs from money to output, Law perceived that it could be made to run in the opposite direction too. With appropriate financial linkages put in place, output could induce the very monetary means of its own expansion. Indeed, Law thought this outcome was assured provided that banks issued money on productive loans secured by claims to future product or its equivalent. Coaxed forth by real output in this fashion, the paper money stock would grow in step with the real demand for it such that its purchasing power would be preserved unchanged. To ensure that the nominal money stock automatically expanded equally with the real demand for it, he advocated that paper notes be backed dollar-for-dollar with the nominal value of land. Collateralized by land, money would, he thought, enjoy stability of value. When economic development or cyclical recovery brought more land into cultivation, the money stock, secured by the extra land, could expand to meet the growing needs of trade at unchanged prices. Here was the prototype of the real bills doctrine later attacked so vigorously by classical writers.

As for the doctrine that low interest rates spell monetary ease and high rates monetary tightness, Law accepted it without reservation. Anticipating Keynes's liquidity preference theory of interest, Law saw interest rates as the price of money's use, a price that varied inversely with the quantity available to use. Being purely monetary phenomena, low rates unambiguously signified an abundance of money and high rates a scarcity of it. Law, an ardent advocate of low rates, argued that they reduced the businessman's cost of capital and so spurred investment and real activity. For him, money exerted its stimulus through indirect interest rate channels as well as through direct expenditure ones.

### **Sir James Steuart (1721–1780)**

To Law's doctrines, Steuart in his 1767 *An Enquiry into the Principles of Political Oeconomy* added four more. First was his explicit rejection of a monetary for a real cost-push theory of inflation. Tracing a causal chain from the degree

---

two propositions, Law recognized the first but denied the second. He also argued, contrary to the monetary approach, that expansion of the domestic stock of paper money would, by stimulating production of goods for export, improve a country's trade balance. See Murphy (1997, Ch. 8).

<sup>5</sup> See Blaug (1996, p. 16).

of competition in labor markets to wage rates to unit labor cost to product prices, he concluded that cost and competition determine the prices of all goods and thus the price level as a whole. Likewise, he held that the monopoly power of producers determines their profit margins as embodied in the profit mark-up component of individual and aggregate prices. In other words, he alleged that the same real forces—market power and cost—that govern relative prices account for absolute prices as well. He advanced a relative price theory of the absolute price level.<sup>6</sup>

Steuart's second contribution was his doctrine of the hoards which he used to bolster his denial that money determines prices. He argued that idle hoards of specie absorb excess cash from circulation just as they release into circulation additional coin to correct a monetary shortage. Consequently, there can be no monetary excess or deficiency to spill over into the commodity market to affect prices. The hoarding-dishoarding mechanism ensures as much.<sup>7</sup> For those occasional increases in the money stock that do manage to elude the hoarding mechanism and spill over into the commodity market, he argued, like Law, that they produce matching shifts in commodity demand along a horizontal supply schedule such that equilibrium real output alters at unchanged prices.

Third was his reverse causation doctrine according to which causality runs from prices to money and its circulation velocity rather than vice-versa as in the quantity theory. Positing a two-step process, he said that cost and competition first determine prices. Then, with prices settled, the circulation velocity of coin adjusts to render the existing stock sufficient to accommodate the prevailing level of real activity at the given prices.<sup>8</sup> If the money stock is excessive, wealth-holders remove the excess from active circulation and either hold it idle so that velocity falls or melt it down into plate and ornaments such that the money stock contracts. Conversely, if coin is deficient, the resulting recourse to paper substitutes and other expedients allows transactors to economize on coin whose velocity therefore rises. Via such devices, velocity adjusts to ensure that the stock of coin is just enough to purchase all the goods offered for sale at the predetermined level of prices. In this way, causation runs from prices to money and velocity. Here is the origin of the notion that changes in the stock of circulating media (coin and its paper substitutes) merely validate price changes that have already occurred and do nothing to produce such changes.

---

<sup>6</sup> On Steuart's cost-push theory, see Screpanti and Zamagni (1993, p. 53).

<sup>7</sup> Not all mercantilists were as sanguine as Steuart on hoards. Indeed they were somewhat ambivalent on the subject. Hoards to them could be either desirable or undesirable. On the one hand, hoards, by draining excess cash from circulation, would tailor the remaining stock precisely to the needs of trade. On the other hand, if output and so the needs of trade were expandable under the impact of a monetary stimulus, such hoards, by removing the source of that stimulus, could unduly constrain real activity. Even so, such hoards would see to it that no monetary excess ever developed to spill over into the commodity market to bid up prices.

<sup>8</sup> See Screpanti and Zamagni (1993, p. 53).

Finally, there was Stuart's uncompromising stance on the perennial issue of rules versus discretion in the conduct of policy. Like all mercantilists, Stuart sided with discretion. Monetary rules, whether of fixed or feedback variety, met with his skepticism as did all self-correcting adjustment mechanisms, natural or designed. To him, nothing but discretionary fine-tuning would do.<sup>9</sup> Such enlightened intervention was the hallmark of his omnipotent, ever-active, benevolent *statesman* whose job was to manipulate the volume of real activity in the national interest.<sup>10</sup> Stuart's statesman alone possessed the detailed knowledge necessary to conduct what today is known as a successful cheap-money, full-employment policy. The gap between actual and potential output, the monetary injection required to close the gap, and the interest rate necessary to draw the required metal from idle hoards: all revealed themselves to the statesman's astute and vigilant scrutiny. So too did the ever-changing circumstances to which he tailored his actions.

These propositions formed the core of mercantilist monetary theory which Law and Stuart deployed to analyze the underemployed economies of their time. Of the two writers, only Law, the paper money mercantilist, was able to translate his theory into action. His famous Mississippi scheme, which merged France's national bank of issue with a trading and land development firm (the Mississippi Company) while simultaneously promising to reduce the French public debt, involved paper money expansion on a mammoth scale.<sup>11</sup>

The resulting spectacular inflationary boom and collapse of Law's system had three consequences.<sup>12</sup> It revealed that the initial output stimulus of a monetary expansion eventually vanishes leaving only inflation in its wake. It served to discredit paper money and financial innovation schemes for many years to come. It, together with the similar debacle of the *assignats*, a nominally land-backed paper currency issued by the French revolutionary government to inflationary excess in the years 1794 to 1796, provoked classicals to reject mercantilist trade and monetary theory root and branch.

---

<sup>9</sup> Stuart of course never resorted to such modern terminology. Nevertheless, the concepts were his.

<sup>10</sup> On Stuart's statesman, see Eltis (1986) and Skinner (1981).

<sup>11</sup> Law denied that the monetary expansion was excessive on the grounds that much of it went to redeem outstanding government bonds and equity claims to his trading firm. Since to him bonds and stocks shared money's characteristic as a transactions medium, he saw all three instruments as exerting the same influence on spending. In his view, money swapped for bonds and equities leaves the total supply of financial purchasing power—money, bonds, and stocks—unchanged. Such monetary issue therefore is noninflationary. He erred. Bonds and stocks hardly qualify as transactions media and thus are far from perfect substitutes for money in spending. Monetizing them can be inflationary. See Niehans (1990, p. 51).

<sup>12</sup> See Murphy (1997) for an exhaustive account of the rise and fall of Law's system.

### **Classical Counterpropositions**

Denouncing the mercantilist identification of wealth with precious metals, Adam Smith observed that national wealth consists not of specie or bullion but rather of stocks of productive resources—land, labor, and capital—and the efficiency with which they are used. With respect to the mercantilist prescription of protectionism as the path to opulence, both Smith and David Ricardo noted that wealth-enhancing, efficient resource allocation requires not protectionism but rather free trade in order to exploit comparative advantages stemming from specialization and division of labor.<sup>13</sup>

### ***Price-Specie-Flow and Quantity Theory Propositions***

Other classicals joined the attack. David Hume (1752) used his price-specie-flow mechanism to demonstrate the impossibility of the mercantilist goal of a permanently favorable trade balance and corresponding persistent specie inflow. Hume (pp. 62–63) noted that the additional specie, by raising domestic prices relative to foreign ones and so discouraging exports and spurring imports, would render the trade balance unfavorable and reverse the specie flow.<sup>14</sup> The resulting drain of monetary metal would continue until domestic prices fell to the level consistent with trade balance equilibrium. Similarly, Hume (pp. 33, 37, 48) showed that the mercantilist fear of scarcity of money was unwarranted since any quantity of money, via a proportionate adjustment in the price level, could drive the trade of a nation. To prove as much, Hume (pp. 62–63) advanced a rigid version of the quantity theory according to which an exogenously given one-time reduction in the stock of money has no lasting effect on real activity but leads ultimately to a proportionate change in the money price of goods.

### ***Distinction between Absolute and Relative Prices***

Hume's classical followers immediately seized upon his quantity theory and deployed it against the mercantilists. David Ricardo applied it to refute cost-push theories of the price level.<sup>15</sup> Accusing cost-pushers of confounding relative prices (market exchange ratios) with the absolute, nominal, or general level of prices, Ricardo flatly denied that a rise in costs—wage costs in particular—could raise general prices without an accompanying expansion of the money

---

<sup>13</sup> Thus a follower of Smith might attribute Scotland's penury not to monetary deficiency and the absence of banks, but rather to lack of specialization and division of labor resulting from a small population.

<sup>14</sup> Cesarano (1998) argues that Hume actually rejected the price-specie-flow mechanism and its attendant changes in relative national price levels for the monetary approach to the balance of payments. By contrast, the standard view emphasized here holds that neither Hume nor his classical followers subscribed to the approach's proposition of instantaneous purchasing power parity, or law of one price.

<sup>15</sup> See Ricardo (1951–1973, I, pp. 46, 61–63, 104–05, 126, 302–03, 307–08, 315).

stock. True, he did acknowledge that a wage hike might raise the prices of labor-intensive goods and so require consumers to spend more on those goods. But he also insisted that without accommodating increases in the money stock to foster spending, consumers would have less to spend on capital-intensive goods whose prices would therefore fall. The upshot was clear. Given a constant money stock, any wage-induced rise in some relative prices would be offset by compensating falls in others leaving the general average of all prices unchanged.

### *Short-Run Nonneutrality and Long-Run Neutrality Propositions*

Classicals reserved their severest criticism for John Law's money-stimulates-trade doctrine. Hume insisted that the doctrine holds in the short run but not the long.<sup>16</sup> At first, money-stock changes indeed affect output and employment. Eventually, however, the output stimulus vanishes and only higher prices remain. Law's doctrine holds in the short run because prices are temporarily sticky, or inflexible, in response to money stock changes. Such stickiness Hume attributed to the imperfect information price-setters possess on money-stock changes and their resulting failure to perceive and act upon the changes. Distribution effects constituted for him another source of temporary nonneutrality, or transitory influence on real activity, inasmuch as new money is initially concentrated in few hands and only gradually becomes dispersed throughout the economy.<sup>17</sup>

With prices sticky and money's circulation velocity given, it follows that changes in the money stock are absorbed by output which accordingly deviates temporarily from its natural equilibrium level. Prices only begin to adjust when price-setters discover that their inventories of goods and labor are abnormally high or low. Eventually, monetary and price-perception errors are corrected as are initial distribution effects. At that point, the price level fully adjusts to the new money stock and output returns to its natural equilibrium level. Here is the source of the classical doctrine of the short-run nonneutrality and long-run neutrality of money.<sup>18</sup>

---

<sup>16</sup> See Hume ([1752] 1955, pp. 37–38, 47–48).

<sup>17</sup> Classicals recognized still other sources of short-run nonneutrality including sticky nominal interest rates, fixed nominal charges such as rents and taxes, fixed nominal incomes of wage earners and rentiers, confusion of relative price for absolute price changes, market size encouragement to specialization and division of labor, and deliberate efforts on the part of organized groups to maintain real incomes. See Humphrey (1993, pp. 251–63).

<sup>18</sup> Hume ([1752] 1955 pp. 39–40) admitted that money might exhibit long-run superneutrality. Being partly unanticipated (perhaps because agents formulate their expectations adaptively in a backward-looking way), a *steady succession* of money stock changes might perpetually frustrate the attempt of prices to catch up and therefore permanently affect the level of real output.

### *Classical Case for Rules*

Four remaining mercantilist arguments clamored for demolition. Classical were glad to oblige. First was the mercantilist claim that discretion was superior to rules. Classical countered with the opposite claim that rules replaced destabilizing activist intervention with smoothly operating, or stabilizing, automatic adjustment mechanisms. Unlike Stuart, classical held a low opinion of the knowledge, capabilities, and motivation of the policy authorities. In particular, classical, especially Ricardo, John Wheatley, and other Bullionist critics of the Bank of England, feared that central bankers operating under the kind of floating exchange rate, inconvertible paper regime prevailing in England during the Napoleonic Wars, would, if left to their own discretion, pursue inflationary policies.

Since classical regarded stability of the value of money as the overriding policy objective, they advocated rules obligating policymakers to achieve that goal. One such rule was the gold standard. By requiring the maintenance of a fixed currency price of gold, this rule, provided that the gold price of goods also remained fairly steady, was tantamount to stabilizing the money price of goods. And with the price level stable, money could function reliably as a unit of account and medium of exchange. In so doing, it could make its maximum contribution to the efficient operation of the real economy and cease to be a source of financial crises and panics.

### *Say's Law of Markets*

Next in line for rejection was the mercantilist claim that deficient aggregate demand condemns cash-poor economies to perpetual unemployment. Not so, wrote the classicist Jean-Baptiste Say in his 1803 *Traité d'économie politique*. The value of goods produced equals the cost of the inputs absorbed in their fabrication. It follows that the very act of production creates, in the form of factor payments, incomes sufficient to buy the goods off the market. And those incomes indeed will be spent. The insatiability of wants together with the unlikelihood that rational people would hoard their savings indefinitely in the form of sterile money ensures as much.

Far from going unspent, saving automatically translates itself into investment. People deposit their savings with banks to earn interest. Those intermediaries, upon lending the saving to capitalist entrepreneurs to finance investment projects, guarantee that it enters the spending stream just as surely as if it were consumption spending. The upshot is that full-capacity supply creates its own demand such that mercantilist fears of general gluts and permanent stagnation are unfounded. Say's Law of Markets identifies the natural level of real activity with full employment.<sup>19</sup>

---

<sup>19</sup> Perhaps too cavalierly, classical dismissed or minimized the problem of unemployment. To them joblessness, while it certainly occurred from time to time, was necessarily short-lived and

*Real Interest Rate*

As for the mercantilist argument that the interest rate is purely a monetary phenomenon, Hume, Ricardo, and Henry Thornton all repudiated it.<sup>20</sup> They contended (1) that the natural equilibrium rate of interest is a real magnitude determined by productivity and thrift, and (2) that money, being neither of those variables, cannot affect the natural rate whose level is therefore resistant to monetary control. True, they conceded that a one-time monetary injection could temporarily depress the loan rate of interest below its equilibrium level. But they stressed the transience of this effect. They pointed out that the monetary injection puts upward pressure on prices. And since with higher prices more loans are needed to finance a given real quantity of investment projects, it follows that loan demands increase. The rise in loan demands reverses the initial fall in the loan rate and restores it to its natural level thereby frustrating attempts to keep it low. Supplementing the price-induced rise in loan demand is a fall in loan supply. For as prices rise, people need more cash, or coin, to mediate hand-to-hand transactions. The resulting conversion of notes and deposits into coin precipitates a cash drain from banks that diminishes bank reserves. To protect their reserves from depletion, banks raise their loan rates. Or what is the same thing, they contract their loan supply. The contraction of loan supply combines with the rise in loan demand to restore the interest rate to its natural equilibrium level determined by productivity and thrift.

*Criticism of Backing Theories of Money*

Last but not least was Law's idea of a land-collateralized paper money stock. Henry Thornton was merciless in his criticism. He excoriated the plan on the grounds that it would fail to limit the money supply and in so failing would render the price level indeterminate.<sup>21</sup> The plan's flaw, wrote Thornton, is that it ties money to the nominal or dollar value, rather than to the fixed physical acreage, of land. By anchoring each dollar to another dollar, it sets up a dynamically unstable price-money-price feedback loop whose elements are free to expand or contract without limit. The result is that any random shock which raises land's price would, by raising land's value, increase money's backing and so justify an expansion of its supply. The consequent expansion would further bid up land's price thereby justifying still further increases in the money

---

self-correcting through automatic wage, price, and interest-rate reductions. Only their inflationist, full-employment-at-any-cost counterparts of the Birmingham School, especially the Attwood brothers, Thomas and Matthias, were gravely concerned with it.

<sup>20</sup> See Hume (1752, pp. 47–59); Ricardo (1951–1973, I, pp. 363–64; III, pp. 88–89, 91, 92; IV, p. 233; V, p. 445); Thornton ([1802] 1939, pp. 253–56).

<sup>21</sup> Thornton ([1811] 1939, p. 342). He ([1802] 1939, pp. 244, 253–56) applies the same criticism to the real bills doctrine which ties the issue of bank money (notes and checking deposits) to the nominal volume of commercial paper that borrowers offer as collateral for bank loans.

stock which would raise prices again and so on ad infinitum. In short, backing money with the nominal value of land—or, for that matter, with commercial paper representing the nominal value of goods in the process of production and distribution—would destabilize prices rather than stabilize them. Price stability required another principle of monetary limitation.

Thornton's refutation of the nominal backing idea completed the list of the original classical rebuttals of mercantilist monetary doctrine. Having contested this doctrine once, however, classicals and their descendants were called upon to counter it repeatedly throughout the nineteenth and twentieth centuries. Mercantilist views, despite their devastating initial rejection, reemerged to form the Banking School position in the famous Currency School-Banking School controversy that took place in England in the mid-1800s. Most of the usual suspects—cost-push, hoarding, reverse causality, discretion, nominal backing—appeared in the Banking School's roundup. In opposing them, classicals, in their Currency School guise, found occasion to deploy the same quantity theoretic, price-specie-flow concepts they had earlier deployed against Law and Stuart.

## **2. CURRENCY SCHOOL-BANKING SCHOOL DEBATE (1830–1850)**

Ending a 24-year experiment with inconvertible paper, Britain had restored the gold convertibility of her currency in 1821. The ensuing Currency School-Banking School debate focused on whether the note component of such a convertible, gold-standard currency required statutory regulation to prevent overissue.<sup>22</sup> The Currency School's classical predecessors, notably David Ricardo, Henry Thornton, and others, had assumed that a convertible currency needed no such protection. If the currency were convertible, they reasoned, any excess note issue which raised British prices relative to foreign prices would be converted into gold to make cheaper purchases abroad.<sup>23</sup> The resulting loss of specie reserves would immediately force banks to contract their note issue thus quickly arresting the drain and restoring the money stock and prices to their pre-existing equilibrium level. Given smooth and rapid adjustment (monetary self-correction), convertibility alone was its own safeguard.

A series of monetary crises in the 1820s and 1830s, however, convinced the Currency School that adjustment was far from smooth and that convertibility per se was by no means a guaranteed safeguard to overissue. It was an inadequate safeguard because it allowed banks, commercial and central, too

---

<sup>22</sup> For classic accounts of the Currency School-Banking School debate, see Viner (1937, Ch. 5), Fetter (1965, Ch. 6), Robbins (1958, Ch. 5), and Mints (1945, Ch. 6). For recent interpretations, see O'Brien (1975, pp. 153–59) and Schwartz (1987).

<sup>23</sup> With the exception of John Wheatley, classicals held that national price levels could deviate temporarily from their purchasing power parity, or long-run equilibrium, levels.

much discretion in the management of their note issue. Banks, facing no minimum required reserve ratio and willing to sacrifice safety for profit, could and did continue to issue notes even as gold was flowing out, delaying contraction until the last possible moment, and then contracting with a violence that sent shock waves throughout the economy.

### **Currency School's Monetary Rule**

What was needed, the Currency School thought, was a rule removing the note issue from the discretion of bankers and placing it under strict regulation. To be effective, this rule should require the banking system to contract its note issue one-for-one with losses of gold reserves so as to put a gradual and early stop to specie drains. Such a rule would embody the Currency School's *principle of metallic fluctuation* according to which a mixed currency of paper and coin should be made to behave exactly as if it were wholly metallic, automatically expanding and contracting to match inflows and outflows of gold.<sup>24</sup>

Departure from this rule, the Currency School argued, would permit persistent overissue of paper. Such overissue, by forcing a protracted efflux of specie through the balance of payments, would in turn endanger the gold reserve, threaten gold convertibility, compel the need for sharp contraction, and thereby precipitate financial panics. Such panics would be exacerbated if internal gold drains coincided with external ones as domestic money holders, alarmed by the possibility of imminent suspension of cash payments, sought to convert paper currency into gold. No such consequences would ensue, the School felt, if the currency conformed to the metallic principle. Forced to behave like gold (regarded by the School as the stablest of monetary standards), the currency would be spared those sharp procyclical fluctuations in quantity that amplified disturbances arising from real shocks.

The Currency School scored a triumph when its monetary rule was enacted into law. The Bank Charter Act of 1844 embodied its prescription that, except for a small fixed amount of notes issued against government securities, bank notes were to be backed by an identical value of gold. In modern terminology, the Act established a marginal gold reserve requirement of 100 percent behind note issues. With notes rigidly tied to gold in this fashion, their volume would start to shrink as soon as specie drains signaled the earliest appearance of overissue. Monetary overexpansion would be corrected automatically, swiftly, and gently before it could do much damage. Here was a practical policy application of Hume's quantity theoretic, specie flow doctrines. Here was the notion of a channel of influence running from note overissue to rising prices to trade deficits to gold drains to corrective reductions in the note issue, reductions that

---

<sup>24</sup> O'Brien (1975, p. 153) credits Joplin, Drummond, Page, Pennington, and McCulloch with the simultaneous enunciation of the metallic principle.

restore general prices to their target equilibrium level. Here too was the classical preference for rules—in this case a 100 percent gold reserve requirement rule—rather than discretion in the conduct of banking policy.

### **Banking School**

The rival Banking School flatly rejected the Currency School's prescription of mandatory 100 percent gold cover for notes. Indeed, the Banking School denied the need for statutory note control of any kind. Instead, the School argued that a convertible note issue was automatically regulated by the needs of trade and required no further limitation. This conclusion stemmed directly from the *real bills doctrine* and the *law of reflux* which together posited guaranteed safeguards to overissue obviating the need for monetary control.

The School's real bills doctrine stated that the money stock could never be inflationary or deflationary if issued by way of collateralized loans advanced to finance transactions in the nominal volume of real goods and services. Similarly, the law of reflux asserted that overissue was impossible because any excess notes would be returned instantaneously to the banks for conversion into coin or for repayment of loans. Both doctrines embodied the notions of a passive, demand-determined money supply and of reverse causality running from prices and economic activity to money rather than vice versa as in the Currency School's view.<sup>25</sup> According to the reverse causality hypothesis, changes in the level of prices and production induce corresponding shifts in the demand for bank loans which banks accommodate via variations in their note issue. In this way, prices help determine the note component of the money stock, the expansion of which is the result, not the cause, of price inflation. As for the price level itself, the Banking School attributed its determination to factor incomes or costs (wages, interest, rents, etc.), thus positing a cost-push theory of price movements. The importance of cost-push theorizing to the Banking School cannot be overestimated. It even led Thomas Tooke, the School's leader, to argue that high-interest-rate tight-money policies were inflationary since they raised the interest component of business costs, costs that passed through into higher prices.<sup>26</sup>

### ***Mercantilist Ideas***

The concepts of cost inflation, reverse causality, and passive money are the hallmarks of an extreme anti-quantity theory of money to which the Banking School

---

<sup>25</sup> Because these doctrines are consistent with those of the monetary approach to the balance of payments, Skaggs (1999) interprets the Banking School as early anticipators of that approach. Even so, the School hardly derived its conclusions from the logic of the monetary approach. The conclusions may have been the same, but they were reached by a different route.

<sup>26</sup> On Tooke's interest cost-push theory and Knut Wicksell's definitive critique of it, see Humphrey (1998, pp. 60–64).

adhered. Additional mercantilist hallmarks included the School's propositions (1) that international gold movements are absorbed by idle hoards of excess specie reserves without affecting the volume of money in active circulation, (2) that gold drains stem from real shocks to the balance of payments rather than from domestic price inflation, (3) that changes in the stock of money are offset by compensating changes in the stock of money substitutes leaving the total circulation unchanged, and (4) that discretion is superior to rules in the conduct of monetary policy.

The Banking School put these propositions to work in its critique of the classical monetary doctrines of the Currency School. Those doctrines, of course, contended that note overissue is the root cause of domestic inflation and specie drains. In opposing them, the Banking School argued as follows: Overissue is impossible since the stock of notes is determined by the needs of trade and cannot exceed demand. Therefore, no excess supply of money exists to spill over into the goods market to bid up prices. In any case, causality runs from prices to money rather than vice versa. Finally, specie drains stem from real rather than monetary shocks to the balance of payments and are totally independent of domestic price-level movements.

These arguments severed all but one of the links in the Currency School's monetary transmission mechanism running from money to prices to the trade balance, thence to specie flows and their impact on the monetary base, and finally back again to the money stock. The final link was broken when the Banking School asserted that gold flows come from idle hoards—buffer stocks of excess specie reserves—and not from the volume of money in circulation. Falling solely on the hoards, gold drains would find their monetary effects neutralized (sterilized) by the implied fall in excess reserves. To ensure that these hoards would always be sufficient to accommodate gold drains, the Banking School recommended that the Bank of England hold larger metallic reserves.

With regard to the Currency School's prescription that discretionary policy be replaced by a fixed rule, the Banking School rejected it on the grounds that rigid rules would prevent the banking system from responding to the needs of trade and would hamper the central bank's power to deal with financial crises.

Finally, the Banking School asserted the impossibility of controlling the monetary circulation via control of the gold and bank note component alone since limitation of that component would simply induce the public to resort to money substitutes (deposits and bills of exchange) instead. In other words, the circulation is like a balloon; when squeezed at one end, it expands at the other. More generally, the Banking School questioned the efficacy of base control in a financial system that could generate an endless supply of money substitutes.

The Currency School, however, rejected this criticism on the grounds that the volume of deposits and bills was rigidly constrained by the volume of gold and notes and therefore could be controlled through the latter alone. In short, the total circulation was like an inverted pyramid resting on a gold and bank

note base, with variations in the base inducing equiproportional variations in the superstructure of money substitutes. In counting deposits as part of the superstructure, the Currency School excluded them from its concept of money. It did so on the grounds that deposits, unlike notes and coin, were not generally acceptable in final payments during financial crises.

### *Evaluation*

In retrospect, the Currency School erred in failing to define deposits as money to be regulated like notes. This failure enabled the Bank of England to exercise discretionary control over a large and growing part of the circulating medium, contrary to the School's intentions. The School also erred in failing to recognize the need for a lender of last resort to avert liquidity panics and domestic cash drains. By the end of the nineteenth century it was widely recognized that the surest way to arrest an internal drain was through a policy of liberal lending. Such drains were caused by panic-induced demands for high-powered money (gold coin and Bank of England notes) and could be terminated by the Bank's announced readiness to satiate those demands. The Currency School nevertheless remained opposed to such a policy, fearing it would place too much discretionary power in the hands of the central bank. These shortcomings in no way invalidated the School's monetary theory of inflation which was superior to any explanations its critics had to offer.

As for the Banking School, it rightly stressed the importance of checking deposits in the payments mechanism. But it was wrong in insisting that the real bills doctrine, which tied note issues to loans made for productive purposes, would prevent inflationary money growth. Like Henry Thornton, the Currency School triumphantly exposed this flaw by pointing out that rising prices would generate a growing demand for—and corresponding nominal collateral backing of—loans to finance the same level of real transactions. These loan demands, when accommodated in the form of deposit and note creation, would enlarge the money stock. In this way inflation would justify the monetary expansion necessary to sustain it and the real bills criterion would fail to limit the quantity of money in existence. Also, by 1900 Knut Wicksell and Irving Fisher had rigorously demonstrated the same point made by Thornton in 1802, namely that an insatiable demand for loans and a corresponding inexhaustible supply of eligible bills results when the loan rate of interest is below the expected rate of profit on capital. In such cases, the real bills criterion provides no bar to overissue.

### **3. THE KEYNESIAN REVOLUTION AND MONETARIST COUNTER-REVOLUTION (1936–1985)**

Classicals won the Currency-Banking dispute. Their victory lasted until ex-classical John Maynard Keynes, having defected to the opposite side, routed

them in 1936.<sup>27</sup> But they regained their crown when monetarists (with help from the new classical school) dislodged Keynesian macroeconomics in the 1970s and 1980s.

Keynes launched his attack in the midst of the Great Depression when the stark conditions of stagnation, poverty, and mass unemployment mocked the classical notion of a self-equilibrating, fully employed economy. Clearly the time was ripe for a mercantilist revival. That revival took the form of the Keynesian Revolution with the leader's *General Theory* as its bible. In that book, Keynes replaced the full capacity, quantity theoretic doctrines of the classicals with at least four propositions inherited from Law and Steuart.

### **Keynes's Mercantilist Propositions**

First, like Law, he argued that in times of mass unemployment the primary stimulative effects of expansionary monetary policy fall on real output and employment rather than on prices. That is, they do so unless negated by liquidity traps and interest-insensitive investment demand schedules, both of which cause velocity reductions to absorb the impact of monetary expansion. Absent such phenomena, however, Keynes's model implied that monetary stimuli affect real activity rather than prices. Like Law, he stressed that the stimulus works through an interest rate channel. More money means lower interest rates, a cheapened cost of capital, and thus a rise in investment spending. The increased investment induces additional rounds of consumption spending causing aggregate demand to rise by a multiple of the new investment spending. With idle resources available to draw upon, production expands to meet the increased aggregate demand. In expounding his interest rate transmission mechanism, Keynes praised his mercantilist forebears for anticipating it. Indeed, the "Notes on Mercantilism" section of his *General Theory* argues that the notion of a linkage running from money to interest rates to investment to output constituted the rationale for the mercantilists' advocacy of export surpluses financed by specie inflows.

Second, like Steuart, Keynes held that product prices, individual and aggregate, are determined by unit labor cost plus a markup to cover profits and nonlabor costs. Here is the mercantilist notion of the price level as a nonmonetary phenomenon.<sup>28</sup> True, Keynes admitted that monetary expansion through its stimulus to employment might, because of diminishing returns to labor, raise unit labor costs and so prices. But he tended to minimize or disregard money's price-raising effects. Instead, he treated the price level as an institutional datum

---

<sup>27</sup> Before he abandoned classicism, Keynes was one of its luminaries. Both his 1923 *A Tract on Monetary Reform* and his 1930 *A Treatise on Money* are squarely in the classical tradition. He returned to the classical fold shortly before his death in 1946.

<sup>28</sup> Keynes applied this notion to a closed economy. He was not referring to the case where, with foreign prices given and the exchange rate fixed, the real terms of trade drives the price level in a small open economy.

governed by nominal wage rates which autonomous forces—union wage-setting policy, worker money illusion, and the like—render downwardly inflexible at low levels of employment. By expressing prices in terms of exogenously given factor costs, he pointed the way to a cost-push theory of the price level. His immediate followers, Joan Robinson, Nicholas Kaldor, and Richard Kahn, certainly interpreted him this way and accordingly denied money a role in price determination.<sup>29</sup>

Third, Keynes restated Steuart's doctrine of hoarding in the form of his concept of the liquidity trap. The trap, he wrote, might come into operation in deep depressions when the interest rate falls to a level so low that everybody unanimously believes it cannot stay there but must return to its conventional normal height. At the floor rate, all are indifferent between holding cash or earning assets whose prices, which vary inversely with the interest rate, are expected to fall. Indeed, asset prices are expected to fall by an amount such that the resulting anticipated capital loss just equals (and so offsets) the interest return on the assets. As there is no advantage to holding such assets instead of zero-yield cash, the latter becomes a perfect substitute for the former in individuals' portfolios. At this point, the demand for money becomes insatiable and infinitely sensitive to the slightest change in interest rates. Keynes called this pathological condition *absolute liquidity preference*.

When this condition rules, no increase in the money stock, no matter how large, can reduce the interest rate. Suppose the central bank expands the money stock by purchasing bonds on the open market. Such bidding puts incipient upward pressure on bond prices. But the slightest rise of the latter induces bondholders to sell to the central bank and then to hoard the cash proceeds. Since at the floor rate of interest the demand for money is insatiable and the willingness to sell bonds absolute, no amount of open market operations can overcome absolute liquidity preference and reduce interest rates. And with rates at their irreducible minimum, they cannot fall any lower to stimulate real activity. Here is Keynes's expression of the mercantilist fear that monetary expansion cannot be counted upon to stimulate spending because the new money may disappear into idle hoards.

Fourth, Keynes found still another obstruction to block the interest rate channel. Even if monetary injections were successful in lowering interest rates, those injections still might fail to stimulate real activity if investment spending were unresponsive to the lower rates. If so, then two obstacles—an interest-insensitive investment schedule as well as a liquidity trap—could render monetary policy ineffective in a depression. In both cases, a rise in the money stock would be offset by a fall in velocity leaving total spending unchanged. With

---

<sup>29</sup> On the cost-push pricing theories of Keynes and his followers, see Tavlas (1981, pp. 324–30).

variable velocity absorbing the impact of money stock changes, none would be transmitted to nominal income. The rigid links connecting money to nominal income and prices as postulated by the classics would be severed or severely weakened. Steuart had said exactly the same thing in 1767.

### **Post-Keynesian Extensions**

To Keynes's own mercantilist doctrines, Keynes's followers writing in the inflationary post-World War II period added others. Some interpreted inflation as a cost-push phenomenon emanating from union bargaining strength, business monopoly power, oligopoly administered prices, commodity shortages, supply shocks, and other real and institutional forces putting upward pressure on factor costs and profit mark-ups. Then too, "cheap money" advocates held that expansionary monetary policy could be used to peg interest rates at low levels so as to minimize the interest burden of the public debt while simultaneously stimulating real activity. An alternative version of the same argument, associated with the Phillips curve trade-off approach to policy questions, held that monetary policy could peg the unemployment rate at permanently low levels at the cost of a stable (nonaccelerating) rate of inflation.

Underlying all these arguments were the presuppositions (1) that full employment is the dominant policy concern, (2) that the employment benefits of monetary stimuli exceed their inflationary costs, and (3) that disinflationary monetary policy, because entrenched inflation is so resistant to it, would produce intolerably large and protracted reductions in output and employment. John Law of course held similar presuppositions, as did other mercantilists.<sup>30</sup>

There remained the mercantilist ideas of reverse causation, passive money, and futility of base control of money and of inflation. Nicholas Kaldor supplied these ideas in his 1982 *The Scourge of Monetarism*. Representing the peak of post-Keynesian skepticism of the relevance of the quantity theory, Kaldor's *Scourge* denied the possibility of base control given the central bank's duty to guarantee bank liquidity and the financial sector's ability to engineer changes in the turnover velocity of money via the manufacture of money substitutes. Kaldor's transmission mechanism runs from trade unions to wages to prices to money and thence to bank reserves. Unions determine wages, wages determine prices, prices influence loan demands, and loan demands, via their accommodation in the form of bank-created checking deposits, determine the money stock, with central banks permissively supplying the necessary reserves. Far from exerting an activating influence, money appears at the end of the causal chain.

---

<sup>30</sup> On the mercantilists' policy goal of full employment, see Grampp (1952).

**Monetarists' Response to Keynes and the Keynesians:  
The Classical Comeback**

Even as Keynesianism was riding high, critics were sniping at it from the sidelines. Eventually these criticisms would culminate in a monetarist counter-revolution that would dethrone mercantilist doctrines and restore classical ones. At least eight mileposts mark the route of the classical comeback.

First came the theory of the real balance effect. Enunciated by Gottfried Haberler, A. C. Pigou, and Don Patinkin, it denied that Keynesian liquidity traps and interest-insensitive investment schedules could bar full employment.<sup>31</sup> That is, it denied they could do so provided (1) wealth in the form of real money balances influences consumers' spending decisions, and (2) prices possess some downward flexibility. The latter condition should hold in a slump since a depressed economy implies an excess supply of goods exerting downward pressure on prices. Lower prices in turn raise the real value, or purchasing power, of cash balances in consumers' wealth portfolios. The rise in real cash balances stimulates consumption spending until full employment is reached.

Indeed, it is unnecessary to wait for falling prices to activate the real balance effect. The central bank can achieve the same result directly by increasing the money supply. In principle, then, Say's Law holds and money is hardly powerless to affect aggregate demand even under extreme Keynesian conditions. Keynes might have realized as much had he incorporated real balances into his consumption function.

Second came the empirical work of Clark Warburton, Milton Friedman, and Anna Schwartz confirming money's power to affect spending. Contrary to Keynes's claim that idle hoards and offsetting velocity movements might negate money's impact on nominal expenditure, Warburton established that (1) an erratic money stock through its impact on spending had been the chief factor causing most U.S. recessions, (2) money's initial impact was on output, and (3) with a lag, prices eventually adjusted to fully absorb the money stock change.<sup>32</sup> Friedman and Schwartz (1963) then corroborated Warburton by showing that a one-third contraction of the money stock caused or intensified the Great Depression of the 1930s. These studies, together with Friedman's findings that persistent inflation is largely or solely the result of excessive monetary growth, effectively reestablished the classical doctrine of the short-run nonneutrality and long-run neutrality of money. They also showed that classical doctrine could account for the Great Depression.

Third came Karl Brunner's and Allan Meltzer's 1967 critique of the Law-Keynes theory of interest rates as a policy guide. That theory claimed that

---

<sup>31</sup> See Haberler (1941, pp. 242, 389, 403), Pigou (1943, 1947), and Patinkin (1948, 1965).

<sup>32</sup> See Warburton (1966) for a collection of his relevant papers, many published between 1944 and 1953.

the interest rate, a purely monetary variable, accurately measures the degree of monetary ease or tightness. Brunner and Meltzer disagreed. The rate, they said, is an unreliable indicator of monetary ease or tightness. It is unreliable because it registers the impact of nonmonetary determinants—notably business loan demands—as well as monetary ones. The rate might be low or high not because money was easy or tight but rather because loan demand was weak or strong. Neglect of this important consideration could lead to perverse, destabilizing policy. For example, in times of depression, when slack business loan demands rendered the rate low, the authorities, misinterpreting the low rate as signifying easy money, might contract the money stock and thereby intensify the depression.

Contrariwise, in times of inflation when booming credit demands rendered the interest rate high, the authorities, misinterpreting the high rate as signaling tight money, might expand the money supply and so escalate the inflation. By confounding the effects of loan demands with those of monetary ease or tightness, the central bank would engineer a perverse, procyclical monetary policy. This critique did much to discredit the Law-Keynes theory of the interest rate.<sup>33</sup>

Milton Friedman's case for monetary rules constituted the fourth monetarist milestone. Friedman (1960) argued that long and variable time lags render discretionary countercyclical monetary policy destabilizing. Because such lags make forecast errors inevitable, the central bank cannot predict the short-run impact of its moves. The result is that expansionary actions aimed at fighting recessions may take effect at precisely the wrong time when the economy is booming just as contractionary anti-inflation actions may hit the economy when it is already mired in a slump. Friedman's solution was to recommend a rigid rule fixing the money stock's growth rate equal to the trend growth rate of output. Such a rule would operate as an automatic stabilizer working to restore aggregate spending to its long-run noninflationary full-employment path. Inflationary spending that outruns the rule-determined money stock could not be sustained and must slacken. Conversely, spending that falls short of money stock growth, as in recessions, would eventually quicken under the impact of the monetary stimulus. In this way, such rule-induced corrections would ensure that money acts to smooth cyclical fluctuations in spending and that long-run aggregate demand grows at the same trend rate as real output such that prices remain stable.

The fifth milestone, and the one that more than any other turned the tide in favor of the classicals, was the stagflation experience of the 1970s. That episode saw the simultaneous appearance of rapid monetary growth, rising

---

<sup>33</sup> As did a related critique attributing high rates to the inflationary anticipations of the public. Embodied in the inflation-premium component of interest rates and fueled by premonitions of policy permissiveness, such anticipations would be realized if the central bank, in a misguided attempt to lower rates, subsequently engineered rapid monetary expansion.

unemployment, and accelerating inflation—an impossible combination according to the predictions of John Law and the Keynesian school. This experience did much to discredit mercantilist beliefs that money stimulates trade and that the price level is independent of the money supply.

### *Natural Rate Hypothesis*

The sixth milestone was the monetarists' natural rate hypothesis according to which unemployment returns to its natural equilibrium level regardless of the inflation rate. Milton Friedman (1968) and Edmund Phelps (1967) established this conclusion with the aid of an expectations-augmented Phillips curve. They showed that when inflationary expectations are incorporated into the Phillips curve, no permanent inflation-unemployment trade-offs remain to be exploited. True, like David Hume, they acknowledged that short-run trade-offs might still exist. Unanticipated rises in inflation, by lowering real wages, could stimulate employment and output temporarily. But once the increased inflation was fully perceived, anticipated, and therefore incorporated into nominal wage rates, the resulting rise in real wages would restore unemployment to its natural equilibrium level. In this way, the adjustment of expected to actual inflation transforms downward-sloping Phillips curves into a vertical line at the natural rate of unemployment. The classicals were right. Inflationary stimuli are temporary, never permanent. One cannot use a higher stable rate of inflation to peg the unemployment rate at arbitrarily low levels since there are no permanent employment gains to be had at any steady rate of inflation. Such gains can be had, if indeed they are available at all, only at the cost of ever-accelerating inflation.

Many Keynesians eventually came to accept the natural rate hypothesis. Even so, they still contended that disinflation was too costly to pursue. Their fear stemmed from early versions of the expectations-augmented Phillips curve.<sup>34</sup> Those versions embodied the assumption that agents revise their inflationary anticipations downward in mechanical, or adaptive, error-learning fashion only when actual, reported inflation turns out to be lower than expected. Accordingly, if the authorities sought to eradicate inflationary expectations—an absolute requirement of any successful disinflationary policy—they would have to force actual inflation below expected inflation thereby inducing the latter to adjust toward the former as it converged on the desired target rate. This sequence required the central bank to employ contractionary monetary policy to raise unemployment above its natural level. The resulting excess unemployment would put downward pressure on the actual rate of inflation to which the expected rate would adjust with a lag. Through this long and painful error-learning adjustment process, both actual and anticipated inflation eventually would be squeezed out of the economy, albeit at the cost of much lost output and employment.

---

<sup>34</sup> See Taylor (1997, pp. 278–79).

***Rational Expectations Lower the Cost of Disinflation***

The seventh monetarist/new classical milestone disposed of this Keynesian concern. Pairing John Muth's (1961) seminal work on rational expectations with Friedman's natural rate hypothesis, Robert Lucas (1972) and Thomas Sargent and Neil Wallace (1975) showed that if expectations are formed rationally rather than mechanically then disinflation need not be a painful drawn-out process. On the contrary, the unemployment cost of disinflation might be far less than Keynesians feared. For if people formed their anticipations rationally, they would take into account all systematic, and therefore predictable, future disinflationary policy actions and embody them in their price forecasts. Provided policymakers behaved in a nonhaphazard, credible fashion, actual and expected rates of inflation and disinflation would coincide such that no gap would develop between them. With no gap, there would be no need for excess unemployment to generate it. Consequently, inflation, actual and expected, would be brought to its zero target level with no cost in terms of excess unemployment. In actuality, of course, this conclusion proved to be a bit too facile and sanguine. In a world in which wages and prices are to some degree sticky or inflexible such that markets fail to clear instantaneously, even rationally expected disinflation would incur some unemployment cost. Nevertheless, the analysis showed that these costs could be much lower than Keynesians feared.

***Time Inconsistency Case For Rules***

The last milestone was the time inconsistency argument which strengthened the classical case for rules by showing how they reinforce policy credibility. Enunciated by Finn Kydland and Edward Prescott (1977) and by Robert Barro and David Gordon (1983a, b), the argument is simplicity itself. Suppose a discretionary, fine-tuning central bank wants to eradicate inflationary expectations so it can have a favorable temporary inflation-unemployment trade-off to exploit. The bank announces its intention to pursue a policy of price stability. It assumes people will believe the announcement and revise their inflation predictions accordingly. The announcement, however, lacks credibility. Private agents realize that once they formulate and act upon such new price predictions, the bank will be tempted to renege on its promise and create a surprise inflation in order to boost output and employment. Such knowledge induces the rational public to discount the announcement and to maintain inflationary expectations at levels high enough to remove the bank's temptation to cheat. The result is that equilibrium unemployment is no lower than it otherwise would be, and yet equilibrium inflation is too high. What prevents inflation from immediately dropping to zero at the natural rate of unemployment is the central bank's inability to promise credibly not to create surprise inflation. Needed is something to convince the public that the central bank will not succumb to the temptation to inflate. That something is a monetary rule replacing the bank's discretionary power with a

precommitment binding it irrevocably to price stability.<sup>35</sup> In demonstrating as much, the time inconsistency argument reinforced the classical case for rules.<sup>36</sup>

The cumulative effect of the foregoing developments was to shift mainstream monetary opinion away from the extremes of Keynesian mercantilism toward classical monetarism. Not all Keynesian doctrines were abandoned, of course. Nor were all monetarist ones embraced. On the contrary, mainstream opinion assimilated an eclectic amalgam of competing views. But a new consensus definitely had emerged. After four or five decades of mercantilist dominance, the classical view was at the wheel once again.

#### 4. CONCLUSION

Three centuries of monetary controversy and experience have established certain hard-won classical truths. Inflation and deflation are monetary rather than cost-push phenomena. There are no long-run inflation-output trade-offs to exploit; central banks cannot permanently peg real variables at disequilibrium levels. Attempts to do so produce explosive, ever-worsening inflation or deflation. Money-stock changes at best affect output and employment temporarily. The output effect vanishes when prices adjust; all that remains is a changed rate of inflation. Stability of the value of money is a prerequisite of an efficiently functioning real economy. All nonnegligible inflation rates violate this prerequisite and are therefore harmful. Monetary rules contribute to such stability.

Presently these truths are in the driver's seat. The proof is that many central bankers now view their primary mission as providing a stable price-level environment within which businesspeople can receive accurate market signals and allocate resources efficiently. Still the classical wisdom, though ruling, is hardly secure. For mercantilist views continue to abound. Even today, some economists still insist that it is better to live with inherited inflation than to fight it because disinflation is too costly to pursue. Others echo Steuart's cost-push theory, attributing the disinflation of the 1990s to such nonmonetary forces as increased global competition, rapid technological progress, falling computer and health-care costs, weakened power of labor unions, and the like. Still others evoke the Steuart-Keynes image of liquidity traps in holding that monetary policy is powerless to stimulate the currently depressed Japanese economy. Commentators even parrot Law's monetary theory of interest when they cite Japan's low interest rates as proof that the country is awash with money when the opposite is true. And always there are those who argue that, with prices

---

<sup>35</sup> Alternatively, an established reputation as a zealous inflation fighter would do.

<sup>36</sup> The time consistency case for rules differs a bit from Friedman's argument. He sees rules as overcoming the central bank's inability to predict the short-run impact of its actions. By contrast, the time inconsistency argument holds that rules are good for commitment reasons even when central bankers have full knowledge of the impact of their moves.

determined by real considerations, monetary policy should be free to pursue nonprice objectives such as achieving full employment and maximizing real growth.

The challenge then is to ensure that the classical truths will not be forgotten. But that is a tall order given that memories fade, that central bank leadership changes, that the current generation of economists familiar with the Keynesian-monetarist controversy is passing from the scene, that revisionist scholars can be counted upon to reinterpret the record radically, and that future generations may well be as reluctant as the present one to study the lessons of the past. The task of countering these influences and preserving the classical wisdom falls to the doctrinal historian. As curator of the stock of eclipsed and unfashionable ideas, he has his work cut out for him.

An even more important challenge is to embed, or lock, the classical truths into enduring institutional arrangements that allow no room for mercantilist policy alternatives. To this end, proponents of the classical view propose a variety of possible arrangements. These include (1) congressional mandates for price stability, (2) formal contracts between elected governments and central banks fixing quantitative targets for price-level behavior, (3) guaranteed independence for central bankers to insulate them from the political pressure to inflate, and (4) the appointment of conservative, inflation-averse central bankers committed to the goal of price stability. The trouble is, however, that none of these proposed arrangements can assure that classical policies will reign supreme for all time. Mandates can be changed, contracts terminated, guarantees revoked, and appointments altered. The upshot is that it is too early to declare a permanent victory for the classical view. Indeed, there may always be a market for the opposing view that central banks need not and must not be bound to the goal of price stability. For better or worse, that view will challenge the classical view whenever the public perceives unemployment or sluggish real growth rather than inflation to be the dominant economic problem.

Still, the inherent cyclicity of ideas suggests an inevitable classical response to that challenge. Classicism, in short, will return to prominence to be confronted anew. For history shows it to be nothing if not resilient. Over long spans of time, it has proved resistant to the kinds of economic shocks that occasionally propel mercantilists to prominence. That is one of the chief insights of doctrinal history.

---



---

**REFERENCES**

- Barro, Robert J., and David B. Gordon. "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, vol. 91 (August 1983a), pp. 589–610.
- \_\_\_\_\_. "Rules, Discretion, and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, vol. 12 (June 1983b), pp. 101–22.
- Blaug, Mark. *Economic Theory in Retrospect*, 5th ed. Cambridge: Cambridge University Press, 1996.
- Brunner, Karl, and Allan H. Meltzer. "The Meaning of Monetary Indicators," in G. Horwich, ed., *Monetary Process and Policy: A Symposium*. Homewood, Ill.: R. D. Irwin, 1967.
- Cesarano, Filippo. "Hume's Specie-Flow Mechanism and Classical Monetary Theory: An Alternative Interpretation," *Journal of International Economics*, vol. 45 (June 1998), pp. 173–86.
- Eltis, Walter. "Sir James Steuart's Corporate State," in R. D. Collison Black, ed., *Ideas in Economics*. London: Macmillan, 1986.
- Fetter, Frank W. *Development of British Monetary Orthodoxy, 1797–1875*. Cambridge, Mass.: Harvard University Press, 1965.
- Friedman, Milton. "25 Years after the Rediscovery of Money: What Have We Learned?" Discussion, *American Economic Review*, vol. 65 (May 1975), pp. 176–79.
- \_\_\_\_\_. "The Role of Monetary Policy," *American Economic Review*, vol. 58 (March 1968), pp. 1–17.
- \_\_\_\_\_. *A Program for Monetary Stability*. New York: Fordham University Press, 1960.
- \_\_\_\_\_, and Anna J. Schwartz. *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press for the National Bureau of Economic Research, 1963.
- Grampp, William D. "The Liberal Elements in English Mercantilism," *Quarterly Journal of Economics*, vol. 66 (November 1952), pp. 465–501.
- Haberler, Gottfried. *Prosperity and Depression: A Theoretical Analysis of Cyclical Movements*. Geneva: League of Nations, 1941.
- Hume, David. "Of Money," "Of the Balance of Trade," and "Of Interest," (1752), in Hume's *Writings on Economics*, E. Rotwein, ed. Madison, Wis.: University of Wisconsin Press, 1955.
- Humphrey, Thomas M. "Historical Origins of the Cost-Push Fallacy," *Federal Reserve Bank of Richmond Economic Quarterly*, vol. 84 (Summer 1998), pp. 53–74.

- . *Money, Banking and Inflation: Essays in the History of Monetary Thought*. Aldershot, England: Edward Elgar Publishing Ltd., 1993.
- Hutchison, Terence W. *Before Adam Smith: The Emergence of Political Economy, 1662–1776*. Oxford: Basil Blackwell, 1988.
- Kaldor, Nicholas. *The Scourge of Monetarism*. Oxford: Oxford University Press, 1982.
- Keynes, John Maynard. *The General Theory of Employment, Interest and Money*. London: Macmillan, 1936.
- Kydland, Finn, and Edward Prescott. “Rules Rather Than Discretion: The Inconsistency of Optimal Plans,” *Journal of Political Economy*, vol. 85 (June 1977), pp. 473–91.
- Law, John. *Money and Trade Considered, with a Proposal for Supplying the Nation with Money*. Edinburgh, 1705.
- Lucas, Robert E., Jr. “Expectations and the Neutrality of Money,” *Journal of Economic Theory*, vol. 4 (April 1972), pp. 103–24.
- Meek, R. L. “Rehabilitation of Sir James Steuart,” in his *Economics and Ideology and Other Essays*. London: Chapman and Hall, 1967.
- Mints, Lloyd W. *A History of Banking Theory*. Chicago: University of Chicago Press, 1945.
- Murphy, Antoin E. *John Law: Economic Theorist and Policy-Maker*. Oxford: Clarendon Press, 1997.
- Muth, John F. “Rational Expectations and the Theory of Price Movements,” *Econometrica*, vol. 29 (July 1961), pp. 315–35.
- Niehans, Jurg. *A History of Economic Theory: Classic Contributions, 1720–1980*. Baltimore: Johns Hopkins University Press, 1990.
- O’Brien, Denis P. *The Classical Economists*. New York: Oxford University Press, 1975.
- Patinkin, Don. *Money, Interest, and Prices*, 2d ed. New York: Harper and Row, 1965.
- . “Price Flexibility and Full Employment,” *American Economic Review*, vol. 38 (September 1948), pp. 543–64.
- Pearce, Ivor F. “Confrontation with Keynes,” in *The Coming Confrontation: Will the Open Society Survive to 1989?* London: The Institute of Economic Affairs, 1978.
- Phelps, Edmund S. “Phillips Curves, Expectation of Inflation and Optimal Unemployment Over Time,” *Economica*, vol. 34 (August 1967), pp. 254–81.
- Pigou, Arthur C. “Economic Progress in a Stable Environment,” *Economica*, vol. 14 (August 1947), pp. 180–90.

- \_\_\_\_\_. "The Classical Stationary State," *Economic Journal*, vol. 53 (December 1943), pp. 343–51.
- Ricardo, David. *The Works and Correspondence of David Ricardo*, Vols. I and II, edited by P. Sraffa with M. H. Dobb. Cambridge: Cambridge University Press, 1951–1973.
- Robbins, Lionel. *Robert Torrens and the Evolution of Classical Economics*. London: Macmillan, 1958.
- Sargent, Thomas J., and Neil Wallace. "Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *Journal of Political Economy*, vol. 83 (April 1975), pp. 241–54.
- Say, Jean-Baptiste. *Traité d'économie politique: Ou, simple exposition de la manière dont se forment, se distribuent, et se consomment les richesses*, Vol. 1. Paris: Deterville, 1803. Translated under the title *A Treatise on Political Economy*, 4th ed. Philadelphia: G. R. Elliot, 1834.
- Schwartz, Anna J. "Banking School, Currency School, Free Banking School," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave—A Dictionary of Economics*, Vol. 1. London: Macmillan, 1987.
- Screpanti, Ernesto, and Stefano Zamagni. *An Outline of the History of Economic Thought*, translated by David Field. New York: Oxford University Press, 1993.
- Skaggs, Neil. "Changing Views: Twentieth-Century Opinion on the Banking School-Currency School Controversy," *History of Political Economy*, vol. 31 (Summer 1999), pp. 361–91.
- Skinner, Andrew S. "Sir James Steuart: Author of a System," *Scottish Journal of Political Economy*, vol. 28 (February 1981), pp. 20–42.
- Tavlas, George S. "Keynesian and Monetarist Theories of the Monetary Transmission Process: Doctrinal Aspects," *Journal of Monetary Economics*, vol. 7 (May 1981), pp. 317–37.
- Taylor, John B. "Comment," in Christina and David H. Romer, eds., *Reducing Inflation: Motivation and Strategy*. Chicago: University of Chicago Press, 1997, pp. 276–80.
- Thornton, Henry. *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain (1802). Together with His Speeches on the Bullion Report, May 1811*. F. A. von Hayek, ed. London: George Allen & Unwin, 1939.
- Viner, Jacob. *Studies in the Theory of International Trade*. New York: Harper Brothers, 1937.
- Warburton, Clark. *Depression, Inflation and Monetary Policy, Selected Papers 1945–1953*. Baltimore: Johns Hopkins University Press, 1966.