THE THEORY OF MULTIPLE EXPANSION OF DEPOSITS: WHAT IT IS AND WHENCE IT CAME

Thomas M. Humphrey

Beginning students of banking must grapple with a curious paradox: the banking system can multiply deposits on a given base of reserves yet none of its member banks can do so. Let the reserve-to-deposit ratio be, say, 20 percent and the system can, by making loans, create $5 of deposit money per dollar of reserves received. By contrast, the individual bank receiving that same dollar on deposit can lend out no more than 80 cents of it. How does one reconcile the banking system's ability to multiply loans and deposits with the individual bank's inability to do so? Fully answering this question required the intellectual efforts of at least six economists writing in the period 1826-1921. The story of their contributions is the story of the evolution of the theory of the multiple expansion of deposits.

At the heart of banking theory is the notion of the multiple expansion of bank deposits. This idea consists of two interrelated parts. The first explains how the banking system as a whole creates deposits by making loans equal to a multiple of its cash reserves, the multiplier being the inverse of the reserve-to-deposit ratio. The second shows how the individual bank contributes to this expansion, not by multiplying its own deposits, but rather by making loans and losing reserves through the clearinghouse to other banks so that they too can expand. Taken together, these components reconcile the banking system's ability to multiply loans and deposits with the individual bank's inability to do so. For the individual bank, far from expanding its loans by several times any new cash deposits received, lends out only the fraction of those deposits remaining after required reserves have been set aside.

The preceding ideas are fairly well known. Many economics textbooks explain why a banking system having a required reserve ratio of, say, twenty percent can create five dollars of deposit money per dollar of cash reserves while at the same time no individual bank can lend more than eighty cents per dollar of deposits received. What the texts do not explain, however, is the origin and development of the theory. The result is to convey the impression that the theory has always existed in its present form, having been fully and correctly articulated from the start. Nothing, however, could be further from the truth. On the contrary, as Lloyd Mints notes in his authoritative *A History of Banking Theory* (1945), "The problem of the manner in which the banking system increases the total volume of the circulating medium, while at the same time the lending power of the individual banks is severely limited, has proved to be one of the most baffling for writers on banking theory" [10, p. 39]. Far from understanding how loans generate deposits, bankers throughout the nineteenth and early twentieth centuries insisted that banks lend only the funds entrusted to their care and therefore could not possibly multiply deposits. Economists, on the other hand, often went to the opposite extreme, arguing that individual banks were simply small-scale versions of the banking system at large and thus could multiply deposits per dollar of reserves just as the system does. Both views were wrong. Not until the 1820s did a more plausible view start to emerge. And not until the 1920s was it finally stated in a way that fully convinced the economics profession and thus enabled the theory to gain widespread acceptance. In an attempt to provide historical perspective and to show how earlier writers resolved the paradox of a banking system doing what none of its members could do, this article traces the evolution of the theory between those two dates. Before doing so, however, it reviews the essentials of the theory as a prerequisite to identifying what earlier writers had to say about them.

The Theory of Deposit Expansion

Suppose for simplicity that the banking system consists of a single monopoly bank constrained by a required reserve-to-deposit ratio r and desiring to be fully loaned up. Suppose further that the public never wishes to convert deposits into currency so that no cash withdrawals occur when deposits expand. Because the bank cannot lose reserves through the clearinghouse to other banks (of which there are none) or to cashholders via withdrawal, it faces no restriction on its ability to expand loans and deposits other than the requirement that it hold r percent of its deposits in reserves. Thus upon the receipt of C dollars of new reserves it can instantly expand loans and deposits D up to the full limit allowed by the reserve ratio—that is, up to the amount $D = (1/r)C$, where $(1/r)$, the inverse of the reserve ratio, is the deposit expansion multiplier. In this way the system as a whole multiplies deposits per dollar of reserves.

Next suppose that the system consists of many small banks, each of which loses through the clearinghouse reserves equal to the full amount of loans made. Because of these adverse clearing balances, no bank can safely lend out more than $(1-r)$ of each dollar of deposits received, this sum being the amount remaining after r percent has been put in required reserve. Thus the first individual bank receiving C dollars of new cash deposits lends $(1-r)C$ of that amount after putting rC dollars in reserve. When borrowers write checks on the proceeds of the loans in favor of recipients who deposit the checks in a second group of banks, the latter banks gain $(1-r)C$ dollars in new deposits. They in turn keep r percent of the new deposits in reserve and lend out the remaining $(1-r)$ percent so that their loans equal $(1-r)(1-r)C$. This amount they lose through the clearinghouse to a third group of banks whose deposits accordingly rise by $(1-r)^2C$, and who, after setting aside a fraction r for reserve, lend out the remaining $(1-r)^3C$. And so it goes from bank to bank in ever-diminishing amounts until excess reserves are zero and all the new cash reserves C are absorbed in backing deposits in the ratio of 1 to r. Summing over the successive groups of banks in the dwindling, never-ending chain gives total new deposits D for the system of $D = [1 + (1-r) + (1-r)^2 + (1-r)^3 + \ldots + (1-r)^n]C$ which, when the number of banks n gets large, converges to the limit $D = (1/r)C$, the same expression that holds for the single monopoly bank.

In short, multiple expansion occurs in the multibank case because the excess reserves that form the basis for loans, though lost to the individual bank, are not lost to the system as a whole. They are simply transferred to other banks that use them for further expansion. As the expansion proceeds from bank to bank, each institution retains the reserves required to back the new deposits that brought it the extra reserves in the first place and lends out the remainder. The result is multiple expansion, the same as that achieved in the monopoly case. The only difference is that in the multibank case each individual bank does not multiply its own deposits. Rather it creates them for other banks by making loans and allowing its reserves to shrink to a fraction of the initial deposit. In a word, the banking system collectively multiplies deposits per dollar of new reserves while the small individual bank fractionalizes reserves per dollar of new deposits.

Historical Evolution

Having outlined the theory itself, we are now prepared to trace its origin and development. Retrospectively, one can discern a certain logical progression. First came the perception that deposits are a multiple of reserves, followed by a rudimentary exposition of the lending, redeposit, and multiplier aspects of the expansion mechanism. Next appeared a specification of the limits to deposit expansion and a definition of the limit value of the multiplier. There followed an analysis of how expansion spreads from bank to bank in a multibank system. Then came the first algebraic statement of the theory followed by the first clear distinction between the expansion power of a monopoly bank and a competitive bank. Finally came the persuasive restatement of the theory that, by consolidating, refining, and elaborating its key ideas, established it in mainstream banking analysis. Each stage saw a different innovator—Pennington, Torrens, Joplin, Marshall, Davenport, and Phillips are the key names here—advance the theory.

Multiple Deposits Recognized

The initial step in the theory's evolution came in the eighteenth century when writers such as John Law (1671-1729), Bishop Berkeley (1685-1753), and Alexander Hamilton (1755-1804) observed that bank deposits were several times larger than the underlying cash base and inferred from this that banks create deposits (see O'Brien [11, p. 15]). These writers, however, did not explain the mechanism that works to multiply deposits. They simply assumed that multiple deposit expansion would somehow occur for both the individual bank and the banking system as a whole. They failed to state that deposit multiplication occurs through the successive lending and redeposit of excess reserves. Not until 1826 was this point made clear.

James Pennington (1777-1862)

It was James Pennington, a British currency expert and confidential monetary advisor to the government, who advanced the theory into its second stage. He did so with his rudimentary exposition of the lending, redeposit, and multiplier mechanics of deposit expansion. His contribution appears in his 1826 memorandum to the English statesman and financier William Huskisson. There he shows (1) that with fractional reserve banking cash deposits produce excess reserves, (2) that such excess reserves lead to loans, and (3) that the proceeds of the loans when redeposited in the system augment the volume of deposits per dollar of cash base. To illustrate these points he argued that if banks receive a cash deposit of which half must be held in reserve the rest will go to purchase earning assets (loans and investments). The sellers of these assets will, upon receiving the cash, redeposit it in their banks thus increasing the volume of deposits. At the end of this first round of the expansion process, the cash reserves of the banks will be the same as before, but the sum total of deposits—including the initial cash deposit plus the additional deposits created by loan—will already be increased by fifty percent. In his words:

> of the money entrusted to their [bankers'] care. . . . if a reserve of one half were sufficient. . . the other half would be employed in discounting bills [i.e., making loans]. . . . But the Persons to whom these advances. . . were made, would, for their own convenience, deposit the money. . . in the hands of their respective bankers, and the aggregate amount of the outstanding [deposit] balances. . . would. . . be encreased 50 per cent. . . . The money due to all the depositors would be 50 per cent more than it was previously to the commencement of these operations. . . [12, pp. xlv-xlvi].

Pennington did not trace the expansion process beyond the first round. But he did indicate how the individual bank contributes to expansion in a multibank system. He pointed out that as one bank expands its loans it either recovers the proceeds in the form of redeposits or else it loses reserves to other banks so that they too can expand. Either way, deposits increase. As he put it in a letter published in Volume 2 of Thomas Tooke's *History of Prices* (1838), if, after a bank receives an initial cash deposit and makes a loan,

> a cheque be drawn upon the. . . banker for the amount of the advance. . . . [and] be paid into his hands by some other depositor, and placed at the credit of that other depositor. . . the whole amount of the book credits [i.e., deposits] of that banker will be increased to the extent of this new advance. And even if the cheque be paid into the hands of some other banker, the [initial] amount of the book credits of the banker who has paid the cheque will not be diminished, while the book credits, as well as the reserved fund of the banker, to whom it is paid, will be increased by its amount [13, p. lvi].

In other words, reserves lost by one bank show up as new deposits in another. In this way deposits gradually multiply on the given increase in the reserve base as it shifts from bank to bank. To illustrate, he showed that if the first bank in a system of two identical banks lends and loses through the clearinghouse half its initial cash reserve to the second that subsequently does the same, deposits of both banks expand although the reserve base remains unchanged [12, pp. xlvii-xlviii].

Pennington's failure to trace the expansion process to its completion accounts for his failure to specify the limit value of the multiplier. Far from defining it as the reciprocal of the reserve ratio, he was content merely to demonstrate that its value was greater than one. He also denied that he viewed the multiplier as a rigid mechanical relationship. This view was attributed to him by Robert Torrens, who cited Pennington as the source of the notion that London banks always hold in the form of notes of the Bank of England a one-fifth cash reserve against deposits, resulting in a multiplier of five. In correcting Torren's misapprehension, Pennington said:

> It never occurred to me, as appears to have been supposed by Colonel Torrens, that every million of notes issued by the Bank of England forms the basis of five millions of deposits; and that every million withdrawn from circulation, by the Bank, occasions a five-fold diminution of those deposits. On the contrary, it is perfectly consistent with my view of the subject, to suppose that the deposit accounts of the London bankers may be materially diminished, while the circulation of the Bank of England is greatly enlarged, or vice versa [13, p. lii].

Pennington contended that bankers' desired reserve ratios (and thus the multiple relationship between deposits and reserves) vary with the state of business confidence. In so doing, he originated the notion of a flexible multiplier.

Pennington's contemporaries quickly grasped the significance of his pioneering work. Torrens referred to it as "a subject of the greatest practical importance" [19, p. 12]. The Banking School likewise shared this opinion. While not accepting his definition of deposits as money, they used his notion of a flexible multiplier to argue that the credit superstructure (of which deposits were the chief component) could expand

and contract independently of the narrow monetary base such that control of the base did not imply control of the superstructure.

Robert Torrens (1780-1864)

Pennington was the first to outline the lending, redeposit, and multiplier aspects of bank credit creation. But Robert Torrens was the first to specify the limits to deposit expansion and to define the limiting value of the multiplier. Torrens, a professional soldier, newspaper proprietor, member of Parliament, promoter of schemes for the colonization of Australia, co-discoverer of the theory of comparative advantage, and one of the ablest monetary theorists of his generation, presented his analysis in his 1837 *Letter to Lord Melbourne*. There, in a section bearing the caption "A given amount of circulating Cash becomes the basis of a much greater amount of Bank Deposits," he wrote that deposits expand until they reach that particular ratio to reserves that bankers deem "safe and legitimate" [19, p. 16]. In other words, the desired deposit/reserve ratio together with the available quantity of reserves fixes the upper limit to expansion. He also explained how deposits grow up to this limit. Stressing the successive lending and redeposit of excess reserves, he wrote that given

> a reserve. . . in coin. . . more than sufficient to meet. . . occasional demands. . . . a part of this coin would be again advanced upon securities, and would be again returned upon the banks, in the form of new deposits, restoring their reserve. . . to the original sum. . . [19, p. 15].

It follows that

> Whatever sums they may advance upon securities in the morning, the same sums will be returned to them in the evening, in the form of new deposits; and in this way the amount of their deposits must continue to increase, until they bear that proportion to the fixed amount of the returning cash, which the experience of the bankers may suggest as safe and legitimate [19, p. 16].

That is, expansion proceeds via the successive lending and redeposit of excess cash reserves until the desired deposit/reserve ratio is attained.

As for the deposit multiplier itself, Torrens expressed it as the inverse of the reserve ratio. He saw, for example, that a reserve ratio of one-tenth would produce a multiplier of ten. Observing that

> in ordinary times, one-tenth, or even one-twentieth, of the money deposited with a banker, is a sufficient rest [reserve] for meeting occasional demands; and that nine-tenths, or even nineteenth-twentieths, of the sums deposited with a bank may be lent out on securities [19, p. 18],

he concluded:

> I should not be arguing on an extreme case, were I to assume that the cash originally deposited. . . with bankers, will be successively re-issued upon securities, by the banks, and successively returned to them, in the form of new deposits, until the proportion between the amount of the deposits, and the amount of the cash, is as ten to one [19, pp. 18-19].

Here is the first clear statement of the multiplier as the reciprocal of the reserve ratio.

In his theoretical analysis, Torrens treated the multiplier as a potentially variable magnitude, fluctuating in value from a high of twenty to a low of five depending on the state of business confidence and its impact on bankers' desired deposit/reserve ratios. As he put it, these ratios

> will necessarily vary with the variations of commercial confidence. When trade is prosperous, when few failures are occurring, and when commercial bills are promptly paid as they fall due, bankers might consider it safe to continue to re-issue, upon securities, the cash returning upon them as deposits, until the proportion between their deposits and their cash, became as fifteen to one, or even as twenty to one. In periods of commercial pressure, on the other hand, bankers would be disposed to contract their liabilities, until the deposits. . . bore to their cash a proportion, not exceeding seven to one, or even five to one [19, pp. 17-18].

Owing to these potential multiplier fluctuations, "a fixed amount of circulating money may be the basis of a fluctuating amount of credit money" [19, p. 17]. Yet in his practical policy analysis he treated the multiplier (or deposit/reserve ratio) as a more-or-less-fixed constant, arguing that control of the reserve base constituted automatic control of the deposit superstructure.

This last idea proved especially influential. The Currency School used it to argue that bank reserves controlled an inverted credit pyramid (with deposits the chief component) resting on a gold and banknote base. Through the writings of the Currency School, Torrens's doctrines of deposit multiplication on a reserve base and deposit control via that base became sufficiently well established by the mid-nineteenth century to be bequeathed to future generations of monetarists (see O'Brien [11, p. 16]). In short, the modern monetarist notion of base control derives straight from Torrens by way of the Currency School.

Thomas Joplin (1790-1847)

The next step in the theory's evolution was taken by Thomas Joplin, a British banker and co-originator of the principle of "metallic fluctuation" around which much of nineteenth century monetary controversy raged. He advanced a view markedly different from Torrens's of the way deposits expand to the limit set by bankers' desired reserve ratios. As documented above, Torrens focused on the lending-redeposit mechanism of the banking system as a whole; he did not trace the expansion process from bank to bank. He merely stated that banks as a group expand loans, then recoup the proceeds in the form of redeposits, and then expand again and again until the limit is reached. He did not identify individual banks nor did he mention the distribution of reserves among them.

By contrast, Joplin explained how expansion proceeds from one bank to the next, each lending out its excess reserve and losing it to another bank which also expands and so on until excess reserves are eliminated and all cash is absorbed in backing deposits at the ratio desired by bankers. Joplin developed his analysis in his 1841 book *The Cause and Cure of Our Commercial Embarrassments*. He starts out by establishing the limits to expansion and defining the deposit multiplier as the inverse of the reserve ratio.

> Every banker. . . has therefore the power of creating bank money, and. . . there is no other limit to the exercise of this power than his own prudence. . . . I apprehend that bank money is always created by the bankers to the full extent that prudence will permit. If one-fifth of their deposits in cash be sufficient to meet any demand for payment by their depositors, for every thousand pounds of cash deposited with them, they discount to the extent of £5,000, and create £5,000 of bank money [7, pp. 33, as quoted in Mints 10, p. 105].

He then proceeds to trace the expansion process across a succession of banks until the limit is reached. Assuming a reserve ratio of 20 percent, he states that a bank receiving a new cash deposit of £1,000 will immediately put £200 in reserves and lend out the remaining £800. The borrowers, upon receiving this sum,

> pay the amount, we shall assume, to the credit of their account with some other banker, who. . . finds his cash increased £800, and his deposits £800, and he has in consequence £640 to spare, which he lends accordingly. This again being paid into another bank, the same operation, again occurs, and so it goes on from bank to bank until the thousand pounds has created for itself deposits to the extent of £5,000 [7, pp. 33-34, as quoted in Mints 10, p. 105].

Here are all the elements found in modern textbook treatments of the multiple expansion process: (1) the initial cash deposit that generates excess reserves, (2) the lending out and subsequent loss of those reserves to other banks who repeat the process, (3) the resulting diminution of excess reserves at each successive bank as they are absorbed in backing the extra deposits created by their arrival, and (4) the cumulative rise in deposits until they reach their limit ratio to cash reserves, at which point excess reserves vanish. All that was missing was a mathematical statement of the process.

Alfred Marshall (1842-1924)

The mathematical statement referred to above constituted the next stage of the theory. The key name here is that of the great English neoclassical economist Alfred Marshall, who provided the algebraic basis for the theory and who used the standard mathematical technique to derive the deposit expansion multiplier as the summation of a geometrical series. Marshall used the symbol n to denote the multiplier, defined by him as the ratio of deposits to reserves (i.e., the inverse of the reserve ratio). In a note scribbled in the margin of his personal copy of Robert Giffen's *Stock Exchange Securities* (1877), he wrote:

> Let it [bankers' desired reserve/deposit ratio] be 1/n th: Let A be the original amount of deposits without credit: then credit can be spread to $A\left[1 + \frac{n-1}{n} + \left(\frac{n-1}{n}\right)^2 + \ldots\right] = A \cdot n$. This is true if there is only one bank, as well as if many, except that if there are many banks n cannot be very large in any one bank, while on the other hand if the banks pool their reserves (theoretically or practically) they count as cash what they have in the pool and the pool lends much of that again [quoted in Eshag 4, pp. 9-10].

He elaborates the substance of this brief note in his evidence before the Gold and Silver Commission of 1887. He says:

> I should consider what part of its deposits a bank could lend and then I should consider what part of its loans would be redeposited with it and with other banks and, vice versa, what part of the loans made by other banks would be received by it as deposits. Thus I should get a geometrical progression; the effect being that if each bank could lend two-thirds of its deposits, the total amount of loaning power got by the banks would amount to three times what it otherwise would be. If it could lend four-fifths, it will then be five times; and so on. The question how large a part of its deposits a bank can lend depends in a great measure on the extent to which the different banks directly or indirectly pool their resources [8, p. 37, as quoted in Eshag 4, p. 10].

In these passages Marshall makes three main points. First, to find the multiplier, one simply adds to each dollar of initial cash deposit the proportion of that dollar that successive banks can lend as it goes in dwindling amounts from bank to bank. In this connection it should be noted that the terms $\frac{n-1}{n}$, $\left(\frac{n-1}{n}\right)^2$, etc., of Marshall's equation are the same as the terms $(1-r)$, $(1-r)^2$, etc., which show the proportion of each dollar of initial deposit that successive banks can lend out after required reserves have been set aside. The resulting multiplier, Marshall notes, is the same whether the system is composed of a single monopoly bank or many small competing banks. Second, the proportion of its deposits a bank can lend is determined by its reserve ratio. If that ratio is, say, one-fifth, the bank can lend out the remaining four-fifths of its deposits. Third, reserve ratios and the resulting power to lend vary by type of bank. Small isolated banks, because of their potentially greater exposure to cash drains and adverse clearings, will operate with larger reserve ratios than big banks or those having ready access to a central reserve pool.

Herbert Joseph Davenport (1861-1931)

The theory progressed to its sixth stage with University of Missouri economist H. J. Davenport's distinction between the expansion power of a single monopoly bank versus that of a small competitive bank in a multibank system. "Modern developments," writes F. A. Hayek, "follow the exposition of H. J. Davenport" [6, p. 153]. On page 261 of his *Economics of Enterprise* (1913) Davenport shows that a monopoly bank in a closed community can do what a whole banking system can do but what a competitive bank cannot do, namely multiply loans and deposits per dollar of cash reserves received. The monopoly bank, he says, loses no reserves to other banks; all checks written on it return in the form of redeposits. Consequently the only restriction on its ability to expand is that it keep r percent of cash reserves against deposits. Thus upon the receipt of C dollars of new reserves it can expand deposits D up to the limit $D=(1/r)C$.

To illustrate, he shows that a new monopoly bank, being the only bank in an isolated town and facing a reserve requirement of 15 percent, will, upon opening for business, engineer a 6⅔-fold expansion of loans and deposits per dollar of initial cash reserves contributed by the stockholders. He then applies this same multiplier to a cash deposit of $100,000, showing how the bank puts $15,000 in reserve, lends out an amount equal to six and two-thirds of the remaining $85,000, and realizes a deposit expansion (primary plus loan-derived) of $666,666. The monopoly bank, he explains, expands up to the limit allowed by the reserve ratio for one reason: it loses no reserves through the clearinghouse or through cash drain.

> For the. . . customers of the bank make payments through checks upon the bank, and these credits are deposited in turn to the credit of other customers. . . . And if some customers draw out cash, other customers will probably receive it and return it to the bank [3, p. 261].

Having described the multiplicative power of a monopoly bank, he turns his attention to the competitive bank. He notes that a competitive bank cannot expand to the extent of a monopoly bank since its attempts to do so will result in reserve losses through the clearinghouse. The competitive bank, he says, cannot expect the proceeds of its loans to be redeposited with it. On the contrary,

> When the check drawn by the borrowing depositor may be deposited in other banks and collected by them against the lending bank, its granting of credits rapidly draws down its reserves to swell the reserves of its competitors [3, p. 263].

These reserves, he notes, go to other banks, which also try to expand; in this way the system as a whole ultimately expands in the same ratio as the monopoly bank. He also suggests that when all banks simultaneously expand their loans approximately in balance, their reserve losses will tend to cancel each other.

> Each bank, as it, in turn lends to its customers, is losing reserves to other banks, but is, in turn, gaining reserves at the expense of the other banks—if at the same time the banking activity of these other banks is maintained [3, p. 287].

To the extent this happens, the group of banks together can (like a monopoly bank) quickly expand to the limit allowed by the reserve ratio.

Chester Arthur Phillips (1882-1976)

The theory of deposit expansion reached its zenith with the publication of C.A. Phillips's *Bank Credit* in 1921. There in the famous Chapter III entitled "The Philosophy of Bank Credit" he stated the theory with a power, precision, and completeness unmatched by his predecessors. In particular, it was Phillips more than anyone else who brought home to the economics profession the crucial distinction between the reserve loss of a competitive bank that expands its loans versus multiple expansion by the banking system as a whole. In so doing, he advanced the theory in at least three ways.

First, he refuted the view, held by Horace White, H. D. McLeod, and other banking writers of the time, that an individual bank multiplies its deposits on a given reserve base just as the banking system does. Not so, said Phillips. An individual bank can not multiply deposits. For its attempts to do so by making loans of several times the amount of new reserves received will simply result in reserve losses to other banks equal to the amount of the loans made (or slightly less if a small fraction of the loans returns to the bank as deposits). No bank, he said, could tolerate such losses that imperiled its legal reserve position.

> Let us suppose that the Hanover National Bank of New York acquires a deposit of \$1,000,000 in gold imported and lends \$10,000,000 to its customers, an amount suggested by the approximate ratio of 1 to 10 between reserves and deposits. . . . Perhaps not more than \$100,000 out of all the checks drawn against the \$10,000,000 borrowed would be deposited at the Hanover National Bank. The remainder of the manifold loans supposedly extended on the basis of the imported gold. . . would represent cash that the bank would lose through unfavorable clearing house balances, an amount that would be scattered widely among the banks of the system. It is clear that an individual bank attempting to lend greatly in excess of the amount of an addition to its reserves would do so at its peril [14, pp. 37-38].

Second, he explained with greater rigor and exactness than his predecessors how the individual bank contributes to systemwide multiple expansion even though it cannot itself multiply deposits. "How," he asked, "can a given amount of cash become the basis of manifold loans and deposits in a banking system if the acquisition of that amount by an individual bank has little or no multiplicative importance?" [14, p. 34]. His answer is that excess cash reserves obtained by one bank will, upon being lent out, provide another bank with excess cash with which it expands and so on until all cash is employed in supporting deposits at the ratio of one to r.

> The sudden acquisition of a substantial amount of reserve by a representative individual bank. . . tends to cause that bank to become out of tune with the banks in the system as a whole. As the individual bank increases its loans in order to re-establish its normal reserve-deposits ratio, reserve is lost to other banks and the new reserve, split into small fragments, becomes dispersed among the banks of the system. Through the process of dispersion it comes to constitute the basis of a manifold loan expansion [14, p. 40].

In short,

> Manifold loans are not extended by an individual bank on the basis of a given amount of reserve. Instead, as a consequence of lending, the reserve of the individual bank overflows, leaving only the equivalent of a fractional part of the additional volume of loans extended, the overflow cash finding its way to other and still other banks until it becomes the "residualized," yet shifting, foundation of manifold loans and deposits [14, p. 73].

To emphasize the point, he contrasted the way the banking system and the individual bank reach their desired reserve-deposit ratios—the system by expanding its deposit denominator; the bank by shrinking its reserve numerator.

Third, he was the first to publish algebraic formulas expressing the loan and deposit expansion potential of both the banking system and the individual bank. Then he used the standard mathematical technique of summation of a series to show that aggregation across the individual banks yields the systemwide formulas. His formulas for the banking system are straightforward and need only be summarized here. According to him, a system facing a required reserve ratio r can, upon the receipt of a new cash deposit C, immediately expand its loans L and deposits D by the amounts $L = \left(\frac{1}{r} - 1\right)C$, and $D = (1/r)C$, where the latter parenthesized multiplier is one larger than the former since it takes account of the initial primary deposit as well as deposits created by loan.

His expansion formulas for the individual bank, however, require some explanation. He noted that the expansion power of the individual bank depends not only on its reserve ratio r but also on the fraction k of its loans that remain with it as deposits. This fraction, he argued, depends upon such things as compensating balance requirements, the accumulation of balances in borrowers' accounts in anticipation of loan repayment, and the redeposit of checks in the same bank upon which drawn. Given these factors, it is an easy matter to trace Phillips's derivation of the bank's loan and deposit expansion formulas.

Thus for an individual bank having a reserve ratio r and an initial cash deposit C, let k be the fraction of loan-created deposits retained by the bank, and L the extra loans made. Once the loans are granted and $(1-k)$ of them withdrawn, final deposits (original plus the retained fraction of those created by loan) of $C+kL$ must, because deposited funds are either held in reserve or lent out, equal loans L plus required reserves $r(C+kL)$ obtained by applying the reserve ratio to deposits. In short, $C + kL = L + r(C + kL)$. Solving this equilibrium condition for loans yields Phillips's loan expansion formula $L = \left[\frac{1-r}{kr+1-k}\right]C$, which, when substituted in the preceding definition of final deposits, results in the deposit expansion expression $D = \left[\frac{1}{kr+1-k}\right]C$, where the bracketed terms are the loan and deposit multipliers.

Using the preceding formulas, Phillips showed that if cash deposits C equal \$1,000, and r and k equal 10 and 20 percent, respectively, then the individual bank can expand its loans L and deposits D by \$1,097.25 and \$1,219.51. These sums are somewhat larger that those of the hypothetical atomistic bank of the textbooks, whose k-factor of zero reduces its loan and deposit multipliers to $(1-r)$ and 1.0, respectively. On the other hand, the loan and deposit sums of Phillips's example are smaller than their counterparts in the case of a single monopoly bank, whose k-factor of 1.0 yields loan and deposit multipliers of $\left(\frac{1-r}{r}\right)$ and $(1/r)$, respectively. Thus Phillips's k-factor, varying as it does between one and zero, essentially indicates the extent to which any one bank can act as a monopoly bank, expanding loans and deposits as if it were the banking system as a whole (see Timberlake [18, pp. 10-12]).

Finally, in a demonstration similar to Marshall's, Phillips showed that the summation of the loan- and deposit-creation series across all individual banks yields the multiple expansion formulas for the system as a whole. Phillips's definitive exposition essentially established the theory once and for all in the form found in economics textbooks today.

The Theory Since Phillips

Since Phillips, at least three innovations have enhanced the theory of deposit expansion. First, economists James Harvey Rogers [15], Procter Thomson [17], and James Angell and Karel Ficek [1] incorporated into the deposit multiplier the public's currency-to-deposit ratio, c, to account for cash drains induced by deposit expansion itself. Using the resulting augmented multiplier expression $D = \left(\frac{1}{c+r}\right)C$, they showed that *both* the cash and reserve ratios c and r act to limit deposit expansion, which is therefore smaller than it otherwise would be if limited by the reserve ratio alone. Still other writers have incorporated time deposit and excess reserve ratios into the multiplier thus further diminishing its magnitude. Second, James Meade [9], Milton Friedman and Anna Schwartz [5, pp. 784-94] as well as Phillip Cagan [2, p. 12] have extended the idea of the deposit expansion multiplier into the broader concept of the *money multiplier*, m, relating the total money stock (currency plus demand deposits), M, to the so-called high-powered monetary base, B, consisting of bank reserves plus currency held by the public according to the expression $M=mB$. Third, Paul Samuelson [16, p. 283] has observed that the small bank "expands" in symmetry with the system, not by multiplying deposits on a given new reserve but by fractionalizing its reserve on a given new deposit.

But these extensions, important as they are, are merely recent refinements made to the fundamental core of ideas laid down by Pennington and his successors. The key ideas of that core—namely that a fractional reserve banking system multiplies deposits, that the mechanics of multiplication involve the successive lending and redeposit of excess reserves, that some crucial ratio or ratios exist to limit the expansion, and that the individual bank contributes to the expansion process not by multiplying its own deposits but by creating them for others when it makes loans and loses reserves through the clearinghouse—were already enunciated more than a century ago. Even today, one finds these ideas indispensable to a full understanding of how the supply of bank money expands and contracts.

References

1. Angell, James W. and Ficek, Karel F. "The Expansion of Bank Credit. I." *Journal of Political Economy* 41 (February 1933): 1-32.

2. Cagan, Phillip. *Determinants and Effects of Changes in the Stock of Money, 1875-1960.* New York: National Bureau of Economic Research, distributed by Columbia University Press, 1964.

3. Davenport, Herbert J. *The Economics of Enterprise.* New York: Macmillan, 1913.

4. Eshag, Eprime. *From Marshall to Keynes.* Oxford: Basil Blackwell, 1963.

5. Friedman, Milton and Schwartz, Anna J. *A Monetary History of the United States, 1870-1960.* Princeton: Princeton University Press, 1963.

6. Hayek, Friedrich A. *Monetary Theory and the Trade Cycle* (1933). New York: Kelley, 1966.

7. Joplin, Thomas. *The Cause and Cure of Our Commercial Embarrassments.* London, 1841.

8. Marshall, Alfred. *Official Papers.* London: Macmillan, 1926.

9. Meade, J. E. "The Amount of Money and the Banking System." *Economic Journal* 44 (1934): 77-83. Reprinted in American Economic Association. *Readings in Monetary Theory.* Homewood: Irwin, 1951, pp. 54-62.

10. Mints, Lloyd. *A History of Banking Theory.* Chicago: University of Chicago Press, 1945.

11. O'Brien, Dennis P. "Monetary Economics." In *Economic Analysis in Historical Perspective,* pp. 3-45. Edited by J. Creedy and D. P. O'Brien. London: Butterworths, 1985.

12. Pennington, James. "Observations on the Private Banking Establishments of the Metropolis: First Memorandum to Huskisson" (1826). In *Economic Writings of James Pennington,* pp. xlv-li. Edited by R. S. Sayers. London: The London School of Economics and Political Science, 1963.

13. ————. "Letter Addressed to the Author by James Pennington, Esq.," Appendix C of Volume 2 of Tooke's *History of Prices* (1838). In *Economic Writings of James Pennington,* pp. lii-lxii. Edited by R. S. Sayers. London: The London School of Economics and Political Science, 1963.

14. Phillips, Chester A. *Bank Credit.* New York: Macmillan, 1921.

15. Rogers, James H. "The Absorption of Bank Credit." *Econometrica* 1 (1933).

16. Samuelson, Paul A. *Economics.* 8th ed. New York: McGraw-Hill, 1970.

17. Thomson, Procter. "Variations on a Theme by Phillips." *American Economic Review* 46 (December 1956): 965-70.

18. Timberlake, Richard H. "A Reassessment of C. A. Phillips' Theory of Bank Credit." Unpublished Paper.

19. Torrens, Robert. *A Letter to the Right Honourable Lord Viscount Melbourne on the Causes of the Recent Derangement in the Money Market and on Bank Reform.* London: Longman, Rees, Orme, Brown, and Green, 1837.

THE EQUILIBRIUM APPROACH TO EXCHANGE RATES

*Alan C. Stockman**

1. Introduction

Media reports on foreign exchange rates are filled with discussions of "overvalued" or "undervalued" currencies. Stories in the financial press about changes in exchange rates frequently state that they affect international competitiveness and employment. The stories often discuss relations between exchange rates and the nation's trade deficit or the federal government's budget deficit. They often state that changes in the exchange rate hurt or benefit the economy, and sometimes discuss policy options available to the government.

Most of these stories are based on a particular disequilibrium theory of exchange rates that has come under increasing criticism in recent years. The disequilibrium theory conflicts with available evidence and an alternative equilibrium theory based on simple economic principles has been developed. The new theory has completely different implications and policy prescriptions than the earlier theory, which underlies most current public policy discussions. This article summarizes the basic elements of the equilibrium approach to exchange rate behavior and the evidence that conflicts with the older disequilibrium theory. It argues that the equilibrium approach to exchange rates is in better accord with this evidence. It concludes with a discussion of the implications of the equilibrium approach to exchange rates for economic policies.

2. Overview of the Issues

The main argument of the paper is the following. Economic theory predicts that real disturbances to supplies of goods or demands for goods cause changes in relative prices, including the "real exchange rate".[1] In a wide variety of circumstances, these changes in the real exchange rate are partly accomplished through changes in the nominal exchange rate. Repeated disturbances to supplies or demands thereby create a correlation between changes in real and nominal exchange rates. This correlation is consistent with equilibrium in the economy, in the sense that markets clear through price adjustments. This is the basis for the "equilibrium approach" to exchange rate changes, and it has several important implications about exchange rate changes. First, exchange rate changes are not "causes" of changes in relative prices, but part of the process through which the changes occur in equilibrium. Second, the question of whether a change in the exchange rate—or more general exchange rate volatility—is "good" or "bad" for the economy is not correctly posed because the exchange rate is an *endogenous* variable. The right question is whether the underlying disturbances to the economy are "good" or "bad," so (of course) the answer varies with the disturbance. Third, the correlation between nominal and real exchange rates is not exploitable by government policy in the sense that attempts by the government to affect the *real* exchange rate by changing the nominal exchange rate (e.g. through foreign exchange market intervention, a return to fixed exchange rates, or "target zones" for exchange rates) will fail. Fourth, there is no simple relation between changes in the exchange rate and changes in "international competitiveness" or employment. It is incorrect, according to the theory, to blame decreased "competitiveness" on the exchange rate. It is equally incorrect to expect that (by itself) an alternative exchange rate system such as fixed rather than floating exchange rates will affect

* Associate Professor of Economics, University of Rochester, and a Visiting Scholar at the Federal Reserve Bank of Richmond. The author wishes to thank Marvin Goodfriend, Thomas M. Humphrey, Anatoli Kuprianov and Torsten Persson for helpful comments.

[1] The real exchange rate is defined in this paper as the relative price of foreign goods in terms of domestic goods. This relative price is also known as the terms of trade. There are other definitions of the real exchange rate, involving relative prices of nontraded and traded goods. Equilibrium models of exchange rates with nontraded goods include Helpman and Razin (1982), Stockman (1983), Stockman and Dellas (1986), and Stulz (1986).

competitiveness. Fifth, there is no simple relation between the exchange rate and the balance of trade or the current account of the balance of payments.[2] Trade deficits do not "cause" currency depreciation, nor does currency depreciation by itself help reduce a trade deficit. Sixth, government budget deficits do not necessarily cause currency appreciation (even if they cause trade deficits). Finally, changes in exchange rates are not related in any simple manner to changes in international interest rate differentials (which may be affected by government budget deficits).

Many of these implications of the equilibrium approach may appear surprising. They conflict with claims that are commonly made in the financial press. But, according to the equilibrium view of exchange rates, many of the assumptions and statements commonly made in the media about exchange rates are simply wrong. This article will explain why.

Some of the propositions stated above may also appear at first to conflict with experience. But, this paper will argue, the experience that appears to conflict with these propositions is only selective. More generally, the evidence is consistent with the implications of the equilibrium approach and fails to support the older, alternative theory.

The alternative "disequilibrium" theories of the exchange rate are based on sluggish adjustment of nominal prices. According to the disequilibrium view, nominal disturbances can cause changes in real exchange rates: changes in nominal exchange rates are naturally translated into changes in real exchange rates because of slow prices adjustments. This view of exchange rate changes underlies most popular accounts of exchange rate changes and policy discussions in the media. It implies that the correlation between real and nominal exchange rate changes is exploitable by government policy (e.g. by establishing "target zones" for exchange rates or intervening in foreign exchange markets in some other manner). It implies that currencies may become "undervalued" or "overvalued" relative to equilibrium, and that these disequilibria affect international "competitiveness" in ways that are not justified by changes in comparative advantage (adjusted for government policies such as tariffs, regulations, etc.). Some versions of the disequilibrium approach also imply systematic relations between the exchange rate and the trade deficit (or the current account deficit), e.g. they imply that the current U.S. deficit will be reduced eventually by a fall in the value of the dollar, with a "hard landing" or "soft landing" occurring under various conditions that can perhaps be affected by government intervention in foreign exchange markets.

Econometric testing of these models is in its infancy, but there is some evidence that supports the equilibrium models. According to the disequilibrium approach, a change in the real exchange rate occurs in response to changes in the nominal exchange rate because of slow nominal price adjustment. But as prices eventually adjust toward their new equilibrium levels, the real exchange rate should adjust back toward its equilibrium value. Monetary disturbances, then, should create temporary movements in real exchange rates. Initial increases in the real exchange rate should be followed by decreases within a few years as nominal prices readjust to equilibrium.[3] According to many of the disequilibrium models such as Dornbusch (1976), monetary disturbances should also create temporary movements in nominal exchange rates.[4]

But statistical evidence indicates that changes in real exchange rates tend to be nearly permanent (on average), or to persist for very long periods of time. The evidence also indicates that changes in nominal exchange rates—even very short-term day-to-day changes—are largely permanent (statistically). This persistence is inconsistent with the view that nominal shocks, or even temporary real shocks, cause most of the important changes in exchange rates. Instead, it is consistent with the view that most changes in real exchange rates are due to real shocks with a large permanent component. Because changes in real and nominal exchange rates are very highly correlated and have similar variances, it is also consistent with the view that most changes in nominal exchange rates are due to largely permanent real disturbances.

This paper discusses the basics of the equilibrium models, their implications, and their relation to existing evidence.[5] Section 3 presents a simple model

[2] The current account equals the trade balance adjusted for any difference between exports and imports that can be paid for by income earned from ownership of foreign assets. For example, a country that is a net creditor earns income from loans it has made in the past, and could use this income to pay for a perpetual trade deficit. A country that did this would have a trade deficit but a balanced current account.

[3] Because nominal price sluggishness is also thought by many economists to be responsible for aggregate business fluctuations, the time involved for the real exchange rate to revert back to its equilibrium level following a disturbance should be similar to the time it takes for recovery from recessions. This argument suggests that the temporary changes in real exchange rates would tend to last, on average, no more than a few years.

[4] For further discussion, see Obstfeld and Stockman (1985).

[5] This paper bypasses a number of associated technical issues, such as the use of optimizing models or the introduction of money into the optimization process. Discussions of these technical issues are often confused with discussion of the basic economic points of the equilibrium models of exchange rates. There is no necessary reason to connect them, so the technical points are left aside here.

on which the remainder of the article builds. Some modifications of the mode are discussed in Section 4. Section 5 discusses some evidence on exchange rates, Section 6 discusses relations between the exchange rate, the balance of trade and some other economic variables, and Section 7 discusses some additional evidence about exchange rates. Finally, Section 8 concludes and raises some policy issues.

3. A Simple Model of Exchange Rates

This section will develop a simple core model of the exchange rate and discuss its properties. Subsequent sections will discuss some additional features that can be added to this model. The simplest model (from an example in Stockman, 1980) embodies the assumptions described below as A0-A6. The role of these assumptions is to clarify the exposition of the equilibrium approach to exchange rates. Most of these assumptions can be dropped without altering the main points of this article. One very important assumption that cannot be dropped without changing many of the results is discussed in Section 4.3. The first five assumptions are:

A0. There is only one period of time, so there is no borrowing or lending. (This assumption will be dropped in Section 6.)

A1. There are two countries, domestic and foreign, that are identical except for the differences spelled out in the other assumptions.

A2. There are two goods. The domestic country produces good X (only), while the foreign country produces only good Y. Output in each country is fixed each period (perfectly inelastic) due to fixed input supplies and technology. Both goods are perishable. There is perfect competition among producers.

A3. The two countries trade so that households can consume both goods. There are no barriers to trade, transportation costs, or transactions costs. Households in each country have the same tastes, expressed here as systems of indifference curves between X and Y (see Figure 1). Both goods are normal.[6]

A4. Households in the two countries are equally wealthy.[7]

The world supplies of X and Y can be divided by world population to obtain per capita supplies x^s and y^s, shown in Figure 1 along with some of the indifference curves.[8] Assumptions A3 and A4 state that households in both countries have the same tastes and resources. So all households will consume the same amounts of both goods. In equilibrium, each household consumes the quantities x^s and y^s, represented by point A in the figure. Because supplies of the goods are perfectly inelastic (i.e. completely insensitive to price changes), tastes for goods affect equilibrium prices but not quantities. The equilibrium *relative price* of the two goods is determined by the slope of the indifference curves at point A. In particular, the relative price of good Y in terms of good X, π_y, equals the absolute value of the inverse of the slope of the indifference curve passing through point A. Flatter indifference curves represent higher equilibrium relative prices of Y. Steeper indifference curves passing through point A represent lower relative prices of Y. The relative price of Y, π_y, is the *real exchange rate* (see footnote 1).

Nominal exchange rates become part of the model when money supplies and money demands are incorporated in the model. The nominal exchange rate is the price of foreign money—say pounds—measured in terms of domestic money—say dollars. Assumptions about the money supply and the demand for money in each country are required.

A5. The nominal supplies of domestic and foreign moneys, dollars and pounds, are denoted by M^s and M^{*s} and are fixed by the governments (or central banks) of the two countries.

A6. The demand for domestic money, dollars, is

(1a) $M^d/p_x = \alpha$

where M^d is the nominal quantity of dollars demanded, p_x is the nominal dollar price of good X, and α represents the real demand for

---

[6] A person's indifference curves describe his own tastes. Each curve shows the various combinations of goods that the person could consume without being either happier or less happy. Higher indifference curves represent greater happiness. A "normal good" is one that people want to buy more of (given its price) when their incomes rise.

[7] Assumption A4 simplifies the description of the model but is not essential. The assumption is useful in drawing Figure 1 because it implies that consumption in both countries can be represented by the same point in the figure.

[8] Assumption A1 implies that the two countries have equal populations. Denote these by N, so there are 2N people in the world. Let x^s be the (world) per capita supply of X, so total production of good X is $2Nx^s$. Similarly, total world supply of Y is $2Ny^s$, and y^s is the per capita supply.

dollars (in terms of good X), which is treated as exogenously fixed. Similarly, the demand for foreign money, pounds, is

(1b) $M^{*d}/p_y^* = \alpha^*$

where p_y^* is the nominal price of good Y measured in terms of pounds and α^* is the real demand for pounds, measured in terms of Y; α^* is also exogenously fixed.

In equilibrium, money demands and supplies must be equated. Setting $M^s = M^d$ and $M^{*s} = M^{*d}$ in (1) gives solutions for nominal export prices (or GDP deflators) p_x and p_y^*:

(2a) $p_x = M^s/\alpha$ and

(2b) $p_y^* = M^{*s}/\alpha^*$.

The nominal exchange rate enters into the model because the *relative* price of Y in terms of X (which is minus the slope of the indifference curve passing through point A in Figure 1) is

(3) $\pi_y = ep_y^*/p_x$

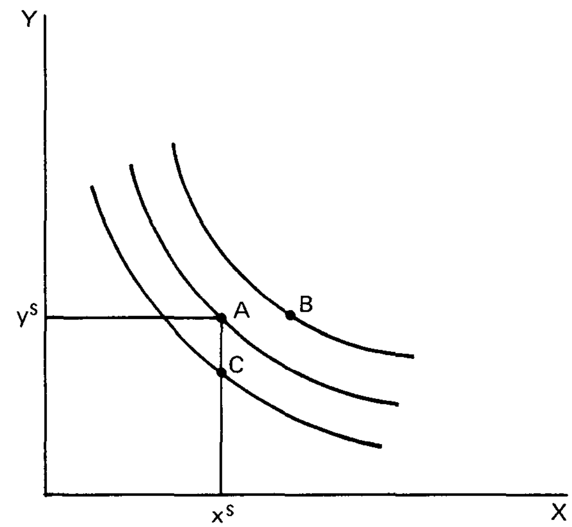
where e is the nominal exchange rate, i.e. the dollar price of one pound. Notice that the dollar price of the foreign good Y is given by arbitrage in goods markets at $p_x = ep_y^*$. Similarly, the pound price of the domestic good X is $p_x^* = p_x/e$. Substituting (2) into (3) gives an equation for the exchange rate:

(4) $$e = \frac{M^s \alpha^*}{M^{*s} \alpha} \pi_y .$$

This is the key equation determining the nominal exchange rate. The model can be modified and made more realistic in many ways, but some essential features of (4) will continue to describe exchange rates. This solution has several features, some of them more obvious than others. First, increasing the domestic money supply by k percent raises domestic prices by k percent and leads to a k percent rise in the exchange rate, which means a k percent *depreciation* of the dollar. Second, an increase in α, lowers domestic nominal prices and the nominal exchange rate (i.e. leads to dollar appreciation). Changes in foreign money supply or foreign money demand have the opposite effects on the nominal exchange rate.

A third key feature of (4) is that it involves the relative price, or real exchange rate, π_y. *Given* the nominal supplies of moneys, M^s and M^{*s} and *given* the real demands for moneys measured in terms of the goods produced in each country, α and α^*, an increase in the relative price of imports, π_y, raises the nominal exchange rate. Recall that an increase in π_y means a flattening of the indifference curve passing through the point in Figure 1 that corresponds to the (per capita) supplies of goods. There are two possible ways in which an increase in the relative price of imports can occur: a change in demand or a change in supply. (1) Demand may change because tastes change so that the indifference curve passing through point A becomes flatter. Or (2) the supplies of X or Y may change, so that the new supplies are represented by a point in Figure 1 at which the indifference curve is flatter, such as point B (resulting from a rise in the supply of X) or point C (resulting from a fall in the supply of Y).

Figure 1



When a change in supply or in demand occurs, it may affect foreign wealth, domestic wealth, or both. To determine the effects of a change in demand or supply, we must take into account its effects on wealth in each country. For example, suppose domestic output rises exogenously (because of an increase in domestic productivity). The domestic firms that produce the additional output may be owned entirely by people in the domestic country. Alternatively, if foreign households also own shares of stock in domestic firms then the rise in domestic output would also raise foreign wealth—because foreigners would share in the additional dividends or capital gains from shares of domestic firms. Even if *only* domestic households own domestic firms, an exogenous rise in domestic output will lower the relative price of the domestic good. If its price falls only a little then the domestic country will be wealthier than before—it has more goods to consume or sell. But if the price of domestic output falls very much then the domestic country will be less wealthy than before:

e.g. owning ten apples each worth one banana may be worse than owning eight apples each worth two bananas. In either case, foreign wealth rises because foreigners are able to buy domestic goods at a lower relative price. So, for a concrete discussion, we need to make an assumption about how changes in demand or supply affect the distribution of wealth. Tentatively we assume:

A7. People in both countries hold exactly the same fractions of their wealth in the stock of any firm (so foreigners own as much of domestic firms as domestic residents do, and the same applies to foreign firms).

Assumption A7 implies that a change in supply or demand for goods affects wealth by an equal amount in both countries, because shares of firms are equally owned by both countries. Then foreign and domestic wealth are equal *after* as well as before any change, so foreign and domestic consumption will be discussed in Section 4.1.

The effects on the exchange rate of changes in demands or supplies of goods can now be summarized. Consider in turn changes in each of x^s, y^s, tastes for goods, α, and α^*, holding money supplies and the other variables fixed.

(a) An increase in the supply of domestic goods raises (lowers) the relative price of foreign (domestic) goods and thereby depreciates the dollar (raises e). The physical quantity of exports also rises, as consumption of the good rises in both countries.[9] An observer, seeing that dollar depreciation is associated with a fall in the relative price of domestic exports and an increase in the volume of exports, might conclude that the domestic country had become "more competitive" as a result of the depreciation of the dollar. But this interpretation is confused. The change in the exchange rate does not *cause* changes in relative prices or the quantity of exports. The change in the exchange rate is itself a *result* of an underlying economic change which also affects other prices and quantities. The distinction is important not only for an accurate understanding of the economy but also for intelligent policy decisions. An observer who mistakenly believes that the "increase in competitiveness" (fall in the relative price of domestic exports) and increase in export volume was caused by a currency depreciation might be tempted to recommend that a further currency depreciation be engineered by increasing the domestic money supply or altering other policies so as to reduce domestic money demand. But, as noted in (d) below, these policy changes would affect the exchange rate without altering "competitiveness" or the quantity of exports.

(b) An increase in the supply of foreign goods lowers their relative price and appreciates domestic money (lowers e). The volume of domestic imports also rises. An observer, who witnesses a simultaneous dollar appreciation, decline in "competitiveness" in the sense of a rise in the relative price of domestic exportables in terms of foreign goods, and rise in the volume of imports, might mistakenly believe that the change in the exchange rate was the cause. He might recommend a rise in the money supply or other policies that reduce domestic money demand in order to mitigate or reverse the dollar's appreciation. But, while those policies may succeed in depreciating the dollar, they would fail to change relative prices (such as the real exchange rate) or the volume of imports.

(c) An increase in the demand for domestic goods and fall in the demand for foreign goods appreciates the dollar. (The demand for foreign goods falls because any change in the demand for domestic goods must be accompanied by a reduction in the demand for something else, given household budgets.) A shift in tastes away from foreign goods toward domestic goods is represented by a steepening of all the indifference curves, as shown in Figure 2. Given supplies of goods at point A, this implies a rise in the relative price of domestic goods.[10] This might be termed a fall in domestic "competitiveness" by some people, although the volumes of exports and imports would be unaffected if the change in tastes occurs in both countries equally (as assumption A3 states).[11] As before, it would be a mistake to conclude that the rise in the relative price of domestic goods was caused by the appreciation of the dollar. Instead, they are both results of an underlying change in demand.

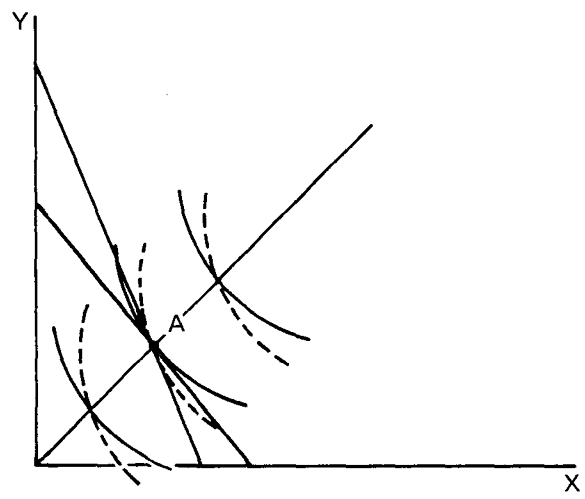
---

[9] In Figure 1, the increase in supply of domestic goods is represented by a shift from point A to point B. The original budget line of domestic (and foreign) households goes through point A and is tangent to the indifference curve touching point A. The new budget line goes through point B and is tangent to the indifference curve touching point B. The new, flatter, budget line represents a higher relative price of Y, the foreign good. Equation (4) implies that, because money supplies and money demands are unaffected, the exchange rate e rises, so the dollar depreciates. The quantity of domestic exports obviously rises: foreign households consume more of the domestic good (at point B) than before (at point A).

[10] In Figure 2, the indifference curve going through point A becomes steeper at that point due to the change in tastes. Assumption A7 implies that the budget lines of all (domestic and foreign) households continue to go through point A, but rotate so that they are tangent to the new indifference curve. So the relative price of the domestic good, X, rises. All households continue to consume at point A.

[11] Section 6.5 discusses a change in tastes in one country alone. In that case, volumes of exports and imports are affected. Also see Section 4.1.

Figure 2



(d) A rise in the domestic money supply or a fall in the domestic demand for money causes dollar depreciation. But relative prices and trade volumes are unaffected because nothing in Figure 1 changes.

It is not possible to discuss trade deficits with this model, because the model includes only a single time period. A dynamic model is required for analysis of such issues as the connections between exchange rates and trade imbalances, interest rates, international capital flows, and budget deficits. The model is expanded in Section 6 so that these issues can be discussed. But there are a number of other important points that can be made without the complications of a dynamic model.

4. Two Modifications of the Model

This section discusses two possible modifications of the model presented in Section 3. Section 4.1 contains a discussion that will be useful in Section 6; Section 4.2 develops a modification that will be used in Section 5. Section 4.3 discusses a very important assumption made in the equilibrium theory of exchange rates. Unlike the other assumptions of the model, it cannot be changed without altering many of the results.

4.1 Wealth Redistribution Effects Suppose assumption A7 is dropped. An alternative assumption is required to replace it. One alternative is that only domestic households own shares in domestic firms and only foreign households own shares in foreign firms. (This assumption leaves open the question of why households fail to achieve the gains that could be obtained, in terms of lower risk for the same return, by international portfolio diversification.) To keep the discussion simple and concrete, we add a stronger assumption than is necessary for the results. Assume A7 is replaced by the assumption

> A8. (i) Firms in each country are owned entirely by households in that country. (ii) The utility function is homothetic, i.e. if a person's income rises and the relative price of goods does not change, then the fraction of his income that he spends on each good does not change.[12]

Assumption A8 implies that changes in the international distribution of wealth can occur, but they do not affect the equilibrium relative price. If wealth is redistributed from the foreign to the domestic country, then the fall in foreign demand for each good is exactly offset by the rise in domestic demand for that good, leaving the total world demand (and the equilibrium relative price) unaffected. In the figures, A8 implies that all of the indifference curves have the same slope along a line coming out of the origin.

With assumption A8, the discussions above regarding changes in supplies of goods continue to apply, with one caveat: one country may end up wealthier—and so may consume more—than another.[13] This is illustrated in Figure 3. Assume there are N households in each country, so world population is 2N. World per capita output of the domestic good is x^s; its total output is $2Nx^s$. Each of the N domestic households owns $2x^s$ of the domestic goods before international trade takes place. An increase in domestic productivity raises total domestic output from $2Nx^s$ to $2N(x^s+\Delta)$. So the per capita supply of X rises from x^s (point A in Figure 3) to $x^s+\Delta$ (shown as point B). The budget line of a domestic household now goes through point G in Figure 3. Domestic households consume at point D and foreign households consume at point F. Average world consumption is at point B (as it must be, since total demand must equal total supply).

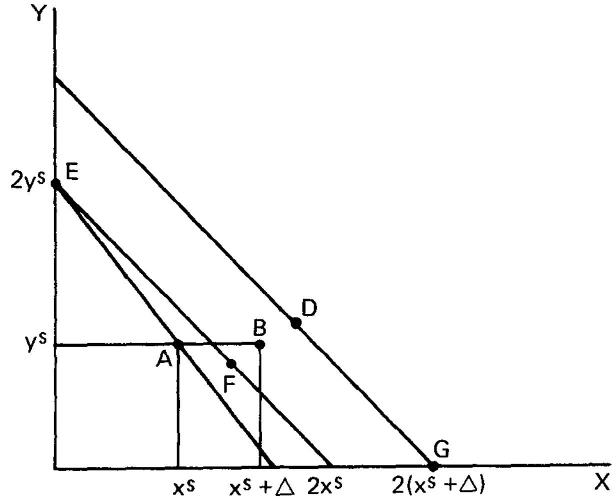
The discussion above regarding a change in demand for goods also requires only one modification:

---

[12] That is, the *relative* amounts of X and Y consumed depends on the relative price but not on income.

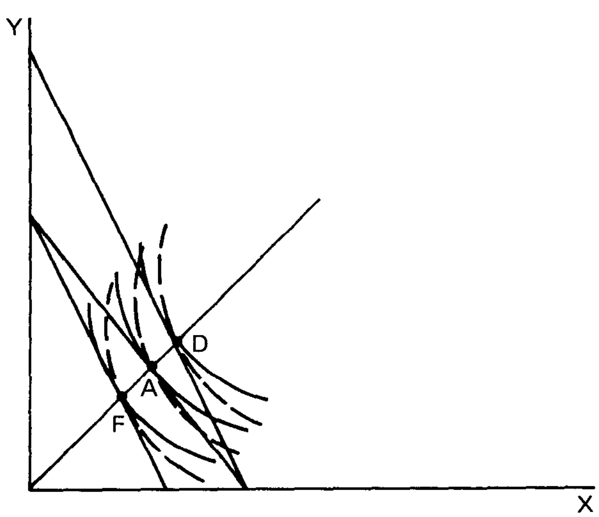
[13] An increase in the supply of domestic goods will raise exports, as before, but it is possible that the domestic country might reduce rather than increase its own consumption of the good. This can occur if the price of the domestic good falls sufficiently, as in Figure 6 below. If the utility function is Cobb-Douglas, i.e. if people always spend some fixed fraction of their incomes on each good, regardless of the relative price, then the countries end up equally wealthy after the change in domestic output, just as if assumption A7 rather than A8 had been invoked. In that case, budget lines for all households go through point B in Figure 1.

Figure 3



volumes of exports and imports may be affected. If the demand for domestic goods rises (and the demand for foreign goods falls), then the rise in the relative price of domestic goods raises domestic wealth and reduces foreign wealth. This is illustrated in Figure 4. Initially, all domestic households consume at point A. The budget line going through point A is tangent to the indifference curve at that point. Then tastes change, and all indifference curves get steeper. In the new equilibrium, domestic households consume at point D and foreign households consume at point F. The volume of domestic exports falls and the volume of domestic imports rises. The fall in

Figure 4



exports would probably reinforce the views of someone who thought that the appreciation of domestic money caused the fall in competitiveness. But it would continue to be a mistake to think that the nominal exchange rate change caused the changes in the real exchange rate and the volumes of exports and imports: all are results of an underlying change in households' preferences for goods.[14]

4.2 An Alternative Specification of Money Demand

Suppose assumption A6, which specified that money demands are given by (1), is replaced by

A9. The demands for domestic and foreign money are given by

$$(1') \quad M^d/p_x = f(x^s) \quad \text{and} \quad M^{*d}/p_y^* = f^*(y^s).$$

This assumption states that real money demand in each country (in terms of that country's output good) is a function of the country's real income measured in the country's output good. A special case of (1') occurs if real money demands are given by

$$(5) \quad M^d/p_x = \alpha x^s \quad \text{and} \quad M^{*d}/p_y^* = \alpha^* y^s$$

so that money demand in each country is a function of that country's GDP (gross domestic product). Then α and α^* can be thought of as the inverses of the velocity of money in each country.

With assumption A9, equilibrium nominal prices and the equilibrium exchange rate are given by

$$(2') \quad p_x = M^s/f(x^s) \text{ and } p_y^* = M^{*s}/f^*(y^s),$$

and

$$(4') \quad e = \frac{M^s f^*(y^s)}{M^{*s} f(x^s)} \pi_y .$$

To determine the effects of changes in supplies or demands, we again invoke assumption A7 (rather than A8). Replacing the money demand specification (1) with (1') leaves the previous analyses of changes in money demands or supplies unaffected. The effects of changes in the demands for foreign versus domestic goods are also exactly the same as in the previous analyses. But the effects of changes in the *supplies* of goods are now more complicated.

An increase in the supply of domestic goods has two analytically separate effects. First, it raises π_y

---

[14] It might be more realistic to replace assumption A8 by the assumption that people in each country tend to buy relatively more of their own country's goods. Except under very peculiar conditions, the analyses in this article will continue to apply with few modifications. An exception is discussed in Section 6.7. Goodfriend (1979) addresses some related issues associated with wealth redistributions.

as before. Given p_x and p_y^*, (3) shows that this raises e, that is, it depreciates the dollar. This can be called the "relative price effect" of an increase in domestic output. The magnitude of the relative price effect (given the change in supply) is greater when the demand for the good is more inelastic, i.e. when the elasticity of substitution between foreign and domestic goods is smaller (see footnote 15). This occurs when the domestic and foreign goods are poor substitutes for each other. Second, an increase in domestic output raises the demand for money and, as (2′) shows, reduces the dollar price of domestic goods. Given the relative price π_y, this reduces the exchange rate e, that is, it appreciates the dollar. This can be called the "money-demand effect" of an increase in domestic output.

The "relative price effect" and the "money demand effect" push the nominal exchange rate in opposite directions in response to an increase in domestic output. Whether the exchange rate rises or falls depends on the relative sizes of these effects. The nominal exchange rate rises—as before—if and only if the relative price effect dominates the money demand effect, i.e. if and only if the inverse of the elasticity of substitution between foreign and domestic goods is smaller than the income elasticity of the demand for money.[15] In the special case of (5), the income elasticity of the demand for money is one.

Let k denote the income elasticity of money demand. Then the money demand effect alone implies that the exchange rate (and each domestic nominal price) falls k percent for each one percent rise in output. If foreign and domestic goods are sufficiently poor substitutes for each other, then the elasticity of substitution between the two goods will be less than 1/k. Then its inverse is larger than k, so a one percent rise in supply of the domestic good reduces its relative price by more than k percent. This effect alone raises the exchange rate by more than k percent. Combining these two effects, the exchange rate rises.

4.3 An Important Assumption The models described above have the essential feature that the demand for money in each country is fixed in terms of that country's output, as in (1), (1′), or the special case (5). Equation (5) implies that the nominal demand for money is proportional to nominal GDP. If, instead, the nominal demand for money were proportional to the nominal value of consumption (with the same factor of proportionality, α or α^*), then the demands for moneys would be

$$(5') \quad M^d = \alpha(p_x x^s + e p_y^* y^s) \quad \text{and}$$
$$M^{*d} = \alpha^*(p_x x^s/e + p_y^* y^s).$$

In this case, a change in the demand for goods—holding fixed money supplies and α and α^*—would alter π_y as before, but not the nominal exchange rate. Equations (5′) imply that $p_x x^s + e p_y^* y^s$ and $p_x x^s/e + p_y^* y^s = (p_x x^s + e p_y^* y^s)/e$ are both unaffected by the change in demand. Consequently, e is unaffected. So the change in the relative price π_y occurs through a change in p_x and p_y^*. For example, a shift in demand away from foreign goods and toward domestic goods lowers $\pi_y = e p_y^*/p_x$ by lowering p_y^* and raising p_x (while the weighted average of the two, $p_x x^s + e p_y^* y^s$, stays fixed). An increase in the supply of the domestic good now leaves the exchange rate unchanged. It raises π_y, the real exchange rate. But (5′) implies that $p_x x^s + e p_y^* y^s$ and e are unchanged, so p_y^* rises and p_x falls, with e unchanged. Evidently, a very important feature of the models in previous sections is that the demands for money in the two countries are appropriately expressed in "real" terms in terms of different bundles of goods. In other words, there are measures of "real" money demands in each country that are invariant to shifts in demand across goods or in supplies of goods, and these invariant measures of real money demands differ across countries. This issue seldom arises in macroeconomic discussions of other issues, but it is extremely important in the economics of exchange rates. The remainder of this article returns to the assumption A9. It is not at all unrealistic that money demands differ across countries in ways

---

[15] The income elasticity of money demand measures the degree to which people want to hold more money when their income rises. The elasticity of substitution between foreign and domestic goods measures the degree to which people are willing to substitute one of the goods for the other. The elasticity is larger as people are more willing to switch from one good to another as one of them becomes more expensive. The income elasticity of the demand for money is $k = x^s f'(x^s)/f(x^s)$, where f′ is the derivative of f. The elasticity of substitution is defined as minus the elasticity of x/y with respect to the relative price of x, along an indifference curve. So the elasticity of substitution is defined as

$$\epsilon \equiv -\frac{(p/ep^*)d(x/y)}{(x/y)d(p/ep^*)}.$$

Then, in response to a change in domestic output x, holding foreign output y fixed, the elasticity of the real exchange rate with respect to domestic output is

$$(x/\pi_y)d\pi_y/dx = 1/\epsilon,$$

and the elasticity of the nominal exchange rate with respect to domestic output is

$$(x/e)de/dx = (x/p)dp/dx + 1/\epsilon,$$

because (2′) implies that $dp^*/dx = 0$. But (2′) also implies that $(x/p)dp/dx = -k$. So

$$(x/e)de/dx = (1/\epsilon) - k.$$

similar to the assumptions made in earlier sections, such as (1'). Consumption bundles differ across countries particularly when allowance is made for nontraded goods and the nontraded components such as retail services, local inventories, transportation, etc., that are embedded in the retail prices of even ostensibly "traded" consumer goods.

5. Some Evidence on Actual Exchange Rates

At this point it is useful to view a plot of real and nominal exchange rates and other prices, as in Chart 1. The chart shows the nominal exchange rate e, the real exchange rate π_y, and the ratio of GNP deflators p_y^*/p_x, where p_y^* is the foreign GNP deflator and p_x is the US GNP deflator. The chart graphs quarterly data for Canada, Britain, and Germany (versus the United States) from the early 1970s when exchange rates were allowed to float. The qualitative features of the plot apply also to other pairs of countries with flexible exchange rates.

Notice that the nominal exchange rate and the real exchange rate move together fairly closely. Most variations in exchange rates—at least among countries with reasonably similar rates of inflation (e.g. OECD countries in the recent past)—are associated with roughly equal variations in the relative price of foreign and domestic goods. This implies that the main source of disturbances to exchange rates must be something—like the changes in supplies or demands for goods discussed above—that change the relative price, and not disturbances that affect only nominal variables (like changes in money demand or supply). Of course, much of macroeconomics is devoted to studying various possible effects of changes in money supply or demand on real variables such as output and relative prices. But these effects of monetary policy on real variables—if they are important—are temporary (or at least contain large temporary components). As we shall see, most of the evidence indicates that changes in nominal and real exchange rates are approximately (statistically) permanent, which is difficult to explain on the basis of temporary real effects of monetary disturbances. Another feature of Chart 1 is that the exchange rate varies much more than the ratio of nominal GNP deflators. (This feature also holds for other country pairs and time periods.) It is convenient to call this feature of the data the "excess variability of exchange rates," though this should not be presumed to imply that this variability is bad in any sense, or indicative of a problem with the operations of markets. It is simply a feature of the data whose interpretation is yet to be determined. This feature can easily be explained with the model from Section 3 above, consisting of equations (2), (3), and (4). Variations in supplies or demands for goods—holding M^s, M^{*s}, α, and α^* fixed—affect π_y but not p_x or p_y^*, so all changes in π_y occur through changes in the exchange rate. But the modified model from Section 4.2, consisting of equations (2'), (3), and (4') can explain the excess variability of exchange rates only under certain conditions. Shifts in demand between foreign and domestic goods change the exchange rate but not the ratio of nominal GDP deflators, so these shifts in demand can explain the excess variability of exchange rates without any additional assumptions. But shifts in supplies of goods only create excess variability in the exchange rate if the elasticity of substitution between foreign and domestic goods is ***smaller than the inverse of twice the income elasticity of money demand.***[16] A one percent rise in domestic output lowers the domestic nominal GNP deflator by k percent, where k is the income elasticity of money demand. If the elasticity of substitution in consumption is 1/k, then a one percent increase in domestic output reduces the new equilibrium relative price of domestic goods by k percent. Since p^* is unchanged, the k percent fall in p/ep^* occurs automatically by the k percent fall in p, without any change in the exchange rate. This explains why the direction of the exchange rate change depends upon whether the elasticity of substitution is larger or smaller than 1/k. Even if the elasticity is smaller than 1/k, in order to obtain a larger percentage change in the exchange rate than in the ratio of GNP deflators, it is necessary that the relative price effect not only be larger than the money demand effect (in order to counteract it completely), but more than double its size. So demand disturbances can clearly explain the excess variability of exchange rates with this model, but supply disturbances can do so only if the elasticity of substitution between foreign and domestic goods is particularly small.[17]

None of these results depend on whether assumption A7 or A8 is invoked. However, if both A7 and A8 are violated, then supply or demand changes affect the international distribution of wealth and alter relative prices. In that case, the exact conditions discussed here would have to be modified.

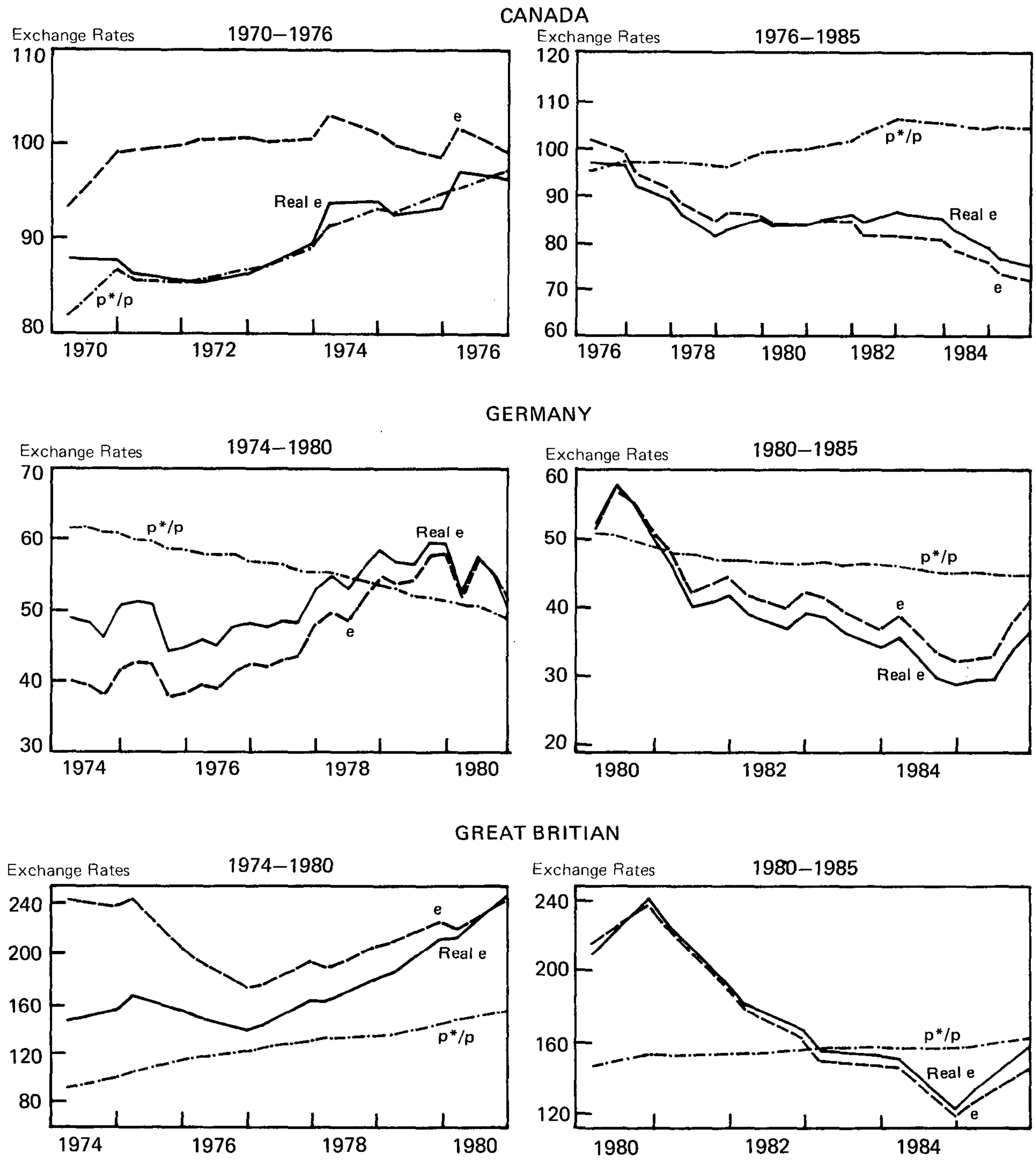
---

[16] A rise in domestic output by one percent lowers p by k percent, according to (2'), where k is the income elasticity of money demand. Footnote 15 implies that the percentage change in e exceeds k percent if and only if $(1/\epsilon)$-k > k, which requires that the elasticity of substitution is smaller than ½k.

[17] See Obstfeld and Stockman (1985). Stockman and Dellas (1986) discuss the issue in the context of a model that also includes nontraded goods.

Chart 1

RATIO OF GNP DEFLATORS, AND NOMINAL EXCHANGE RATES



6. The Exchange Rate and the Balance of Trade

If the model described in Section 4.2 (or the one from Section 3) is used to describe the world in each of a series of time periods, then it is possible to discuss the balance of trade, international capital flows, the effects of government budget deficits, and other related issues. This section discusses the operation of the model when nations are able to borrow or lend, i.e. to have trade deficits or surpluses. It then examines the relations between nominal and real exchange rates and the balance of trade in response to various exogenous disturbances.

Suppose there are two time periods rather than one. (The extension to more periods is straightforward.) The two-period intertemporal model can be described by repeating the model from Section 4.2 at each time period. Make assumptions A1, A2, A3, and A4. At each date there are fixed supplies of the domestic and foreign goods. The real exchange rate π_y is equal to (minus) the slope of the indifference curve passing through point A in Figure 1, just as before, at each date. Nominal prices and the exchange rate at each date are given by (2′) and (4′).

The equilibrium balance of trade, and the effects of various exogenous disturbances, depends on how the international distribution of wealth is affected by exogenous disturbances. (This issue also arose in the one-period models discussed in previous sections, but trade was always balanced in those models.) If a change in supply or in demand in the first period raises domestic wealth more than foreign wealth, then the domestic country will begin the second period with greater wealth than the foreign country. Assumption A4 (which postulated equal initial wealth) will not apply in the second period. If we make assumption A7 then both countries remain equally wealthy at all times. This corresponds to the model in Lucas (1982). On the other hand, if international trade in financial assets is limited in some effective way, then we may make assumption A8 and changes in supplies or demands may redistribute wealth, which corresponds to the model in Stockman (1980).

We adopt assumption A8 for the remainder of this section.[18] Then the relative price of the two goods is always the slope of the indifference curve passing through point A, but one country may consume more of both goods than the other, because (even if the countries begin with equal wealth) an exogenous disturbance may affect domestic and foreign wealth differently.

We now consider a series of exogenous disturbances, and in each case examine the effects on the real exchange rate, the nominal exchange rate, the balance of trade, and related variables.

6.1 A Permanent Increase in Domestic Productivity
If domestic output rises equally in both the first and second periods, then the relative price of the domestic good falls in both periods. The nominal exchange rate rises, i.e. the dollar depreciates, if the relative price effect dominates the money demand effect, as discussed in Section 4.2. Foreign wealth rises (as discussed in Section 4.1) because foreign households can import domestic goods at a lower relative price. Domestic wealth rises unless the fall in the relative price of the domestic good is very large. The case in which domestic wealth rises is illustrated in Figure 3, which describes both time periods (since they are the same). Whatever happens to the distribution of wealth and relative consumption levels, international trade is balanced.[19]

6.2 A Temporary Increase in Domestic Productivity
Suppose domestic output rises exogenously in the first period only. Then its relative price falls in the first period. Whether the nominal exchange rate rises or falls depends—as discussed in Section 4.2—on whether foreign and domestic goods are good or poor substitutes in consumption and on the income elasticity of the demand for money. If the goods are poor substitutes and/or the income elasticity of the demand for money is low, then the relative price effect of the change in output on the exchange rate dominates the money demand effect. Then the exchange rate rises (the dollar depreciates). Whether the domestic country has a balance of trade surplus or deficit in the first period also depends on the degree of substitutability of domestic and foreign goods. Suppose the goods are sufficiently good substitutes that a one percent increase in domestic output reduces its relative price by less than one percent as in Figure 3 (the elasticity of substitution is greater than one). Then the domestic country will have a balance of trade surplus in the first period, and the foreign country will have a deficit. The domestic trade surplus results because the temporary increase in domestic output raises domestic income

---

[18] Assumption A1 implies that households discount future utility at the same rate. The results in this section also assume additively separable utility in first- and second-period consumption with a time-invariant instantaneous utility function.

[19] The balanced-trade result is not robust to slight changes in the assumptions about tastes, but there is little theoretical presumption that the domestic country should have either a surplus or a deficit.

more than proportionally to foreign income. The first-period budget lines of both countries rotate as in Figure 3 because of the relative price change. The budget line of the domestic country rotates through point G in Figure 3 because the domestic people own the firms producing the domestic good. The foreign budget line rotates through point E, so the domestic budget line lies above the foreign budget line: the domestic country has greater income at date one. *If* it were *not* possible to borrow or lend, then the domestic country would consume at point D and the foreign country would consume at point F in Figure 3. In the second period, with output back to point A, both countries would consume at point A.

But it *is* possible to borrow and lend, i.e. it is possible to have a trade deficit or surplus. Both countries would like to save some income from period one for consumption in period two. But it is impossible for the world to save in this way because the goods are perishable. The domestic country sees a larger drop in its income and consumption from the first period to the second than does the foreign country. So there is a mutually advantageous trade: the domestic country will have a balance of trade surplus (lend to the foreign country) and the foreign country will have a trade deficit (and borrow). The equilibrium is shown in Figure 5. In the first period, the budget line of the domestic country shifts in while the budget line of the foreign country shifts out. Domestic households consume H in the first period while foreign households consume I. In the second period, this is reversed: the home country has a trade deficit (paid for by principal and interest received as foreigners pay off the loan) and the foreign country a trade surplus. Second-period domestic consumption is at point J while second-period foreign consumption is at point K.

If foreign and domestic goods are sufficiently poor substitutes that a one percent rise in domestic output reduces its relative price by more than one percent (the elasticity of substitution is less than one) then the situation described above is reversed: domestic income is lower than foreign income in the first period. This situation is illustrated in Figure 6. In the absence of borrowing and lending opportunities, domestic consumption would be at point D and foreign consumption would be at point F. With the opportunity to borrow or lend, the foreign country will have a trade surplus and the domestic country will have a trade deficit in the first period. Domestic households will consume at point H in the first period and foreign households will consume at point I. In the second period, domestic consumption is at point J and foreign consumption at point K.

Summing up: a temporary increase in domestic output causes, temporarily, real exchange rate depreciation (a fall in the relative price of domestic goods), and nominal exchange rate depreciation if the relative price effect dominates the money demand effect. This rise in the nominal exchange rate can

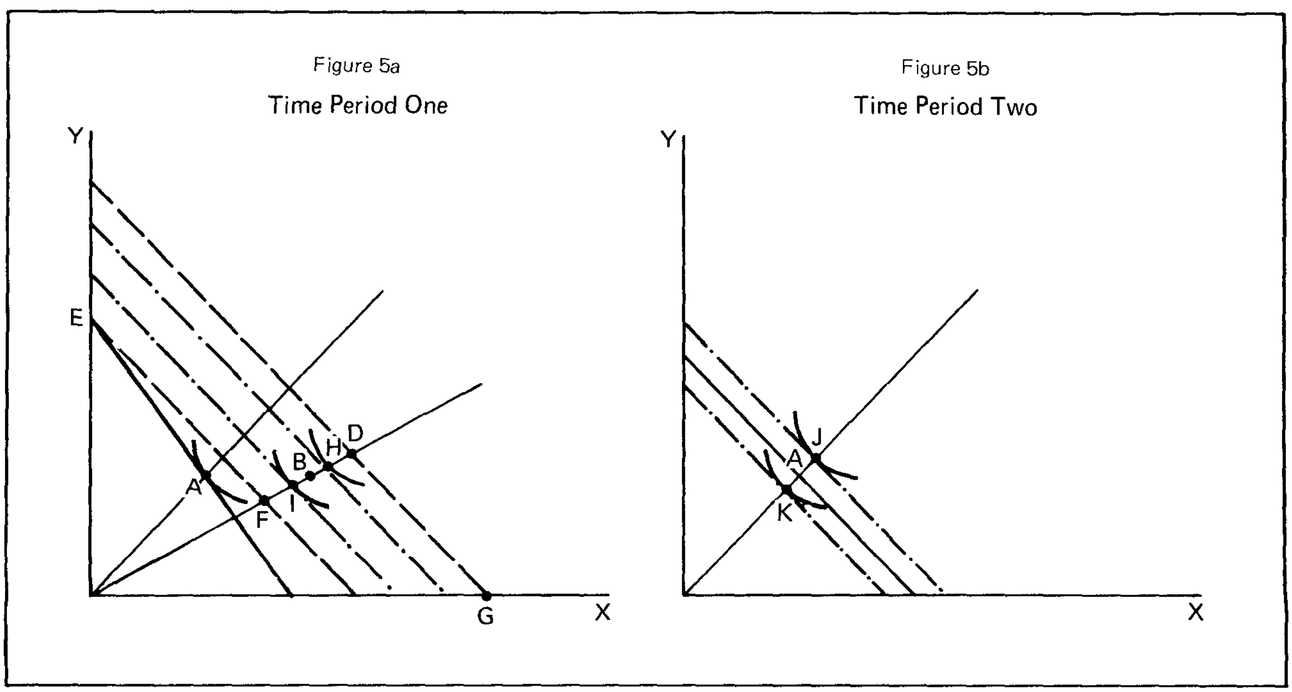
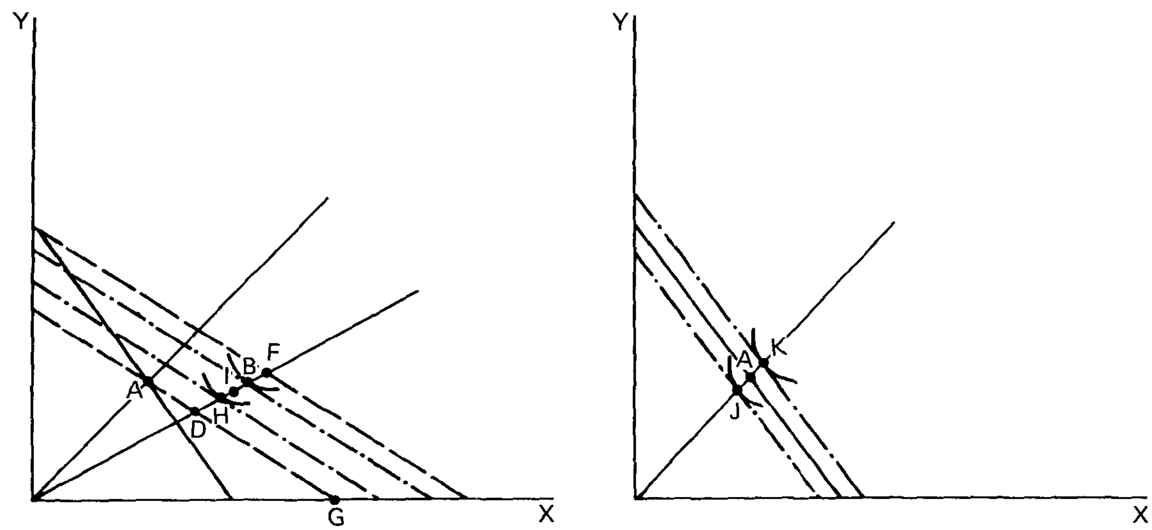


Figure 6a

Time Period One

Figure 6b

Time Period Two



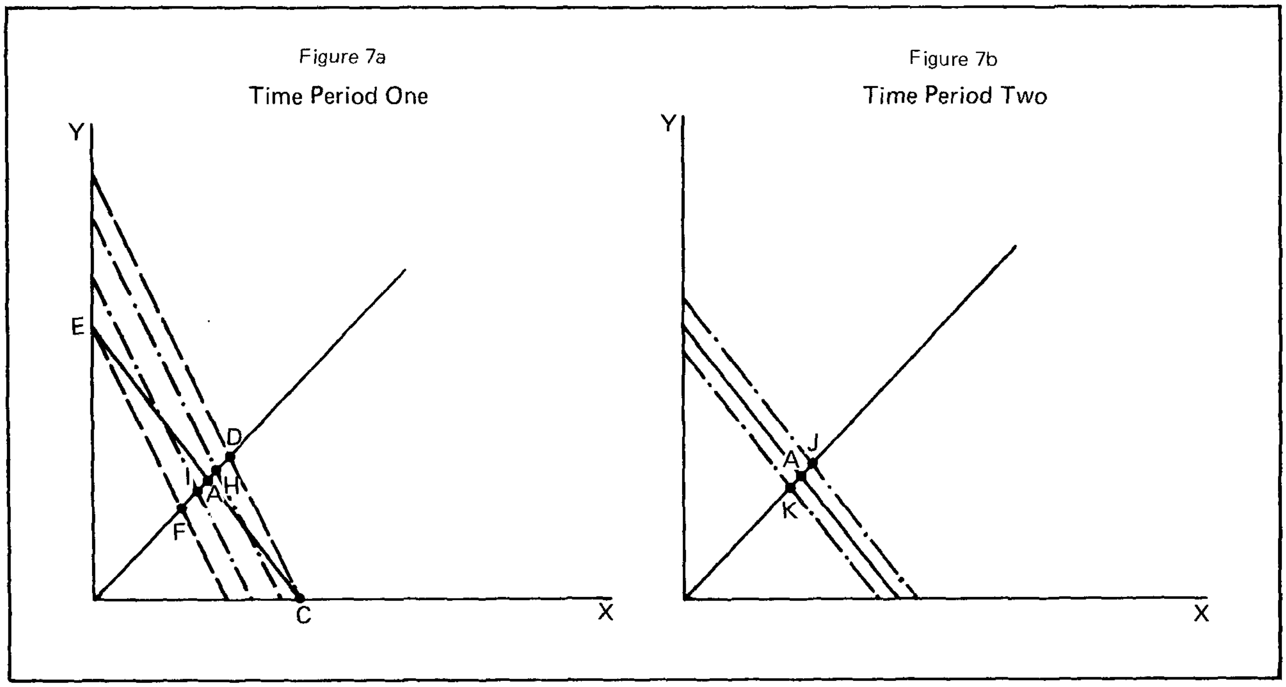
be accompanied by either a trade surplus or a trade deficit. Trade deficits and exchange rate depreciation do not necessarily go together.

6.3 A Temporary Increase in Demand for Domestic Goods Suppose the demand for domestic goods rises in the first period because of a temporary change in tastes. (A change in government spending—another reason for a change in demand—could be modeled as a change in supply.) Indifference curves in the first period shift so that they are steeper than before at every point. Figure 7 illustrates the equilibrium after the shift in indifference curves. Without the shift, equilibrium consumption for each country would have been at point A. Point A still shows the per capita supplies of goods, but the increase in the relative price of domestic goods—due to the increase in demand—raises domestic income and reduces foreign income. The domestic country's budget line rotates through point C and the foreign country's budget line rotates through point E. If borrowing and lending were not possible, the domestic households would consume at point D while foreign households would consume at point F.

But borrowing and lending is possible. The domestic country has temporarily high income and would like to save some of it; the foreign country has temporarily low consumption and would like to borrow. So the domestic country has a trade surplus and the foreign country has a trade deficit. In the first period, the domestic country consumes at point H while the foreign country consumes at point I. In the second period, the domestic country consumes at point J and the foreign country at point K. The temporary trade surplus in the domestic country is associated with real and nominal appreciation, i.e. the relative price of the domestic good rises and the nominal exchange rate falls (domestic money appreciates).

If there had been a temporary fall (rather than rise) in demand for the domestic good, this would have created a temporary real and nominal depreciation and a (temporary) trade deficit. In this case, depreciation and trade deficits go together, and as time passes the domestic currency appreciates while the deficit is eliminated. Despite this relation between currency depreciation and the trade deficit, it would be incorrect to say that the depreciation caused the deficit (or vice versa). Both were results of the underlying change in demand for goods. It would also be impossible for government policy to reduce the trade deficit by monetary policies or similar attempts to stabilize the nominal exchange rate.

6.4 An Expected Future Increase in Demand for Domestic Goods Suppose the increase in demand for domestic goods—discussed in Section 6.3—occurs in the second period rather than the first. Suppose it was also expected (in the first period) to occur. Figure 7 will again illustrate the equilibrium *with an*



important modification: the panel labeled "period one" in Figure 5 will apply to period two, while the panel labeled "period two" will apply to period one. In the first period there is no exogenous change in demand or supply. But the expectation of a future increase in demand for the domestic good raises expected future domestic income. Similarly, the change in demand lowers expected future foreign income. The domestic country will want to borrow in the first period while the foreign country will want to lend. That is, the domestic country will have a trade deficit in the first period (and consume at point J) and the foreign country will have a trade surplus (and consume at point K). But relative prices and the nominal exchange rate will be unaffected by expectations of the future. In the second period, domestic real and nominal appreciation will accompany a domestic trade surplus. Second period domestic (foreign) consumption is at point H (point I) in Figure 7.

If the model were modified in some realistic ways, the real and nominal exchange rates would change in the first period. The expectation of an increase in the relative price of the domestic good in the future would tend to increase its price now (e.g. if it can be stored over time, or if households can substitute consumption of the domestic good now—while it is still cheaper—for consumption of the good later when it costs more). This increase in the relative price of the domestic good would occur partly through a fall in the nominal exchange rate in the first period (just as if the original change in demand had occurred in the first period). With this modification of the model, the first-period trade deficit would be associated with real and nominal appreciation. The size of the first-period appreciation would depend on the degree to which suppliers and demanders can substitute goods over time.

A second modification would reinforce the nominal (though not the real) appreciation associated with the first-period trade deficit. An expected fall in the future nominal exchange rate (dollar appreciation) makes dollars less costly to hold now. If the demand for money were sensitive to its holding cost (the nominal interest rate), then the first-period real demand for dollars would rise by an amount that depends on the interest-elasticity of money demand. This would reduce the nominal exchange rate (and all nominal prices) in the first period, and reinforce the nominal appreciation associated with the trade surplus. Comparing the results in Sections 6.2, 6.3, and 6.4, it is clear that a trade deficit can be associated with either real and nominal depreciation or real and nominal appreciation, depending on the original disturbance (and, in some cases, on the magnitudes of certain parameters).

6.5 An Increase in Demand by the Domestic Country Only In the examples of changes in demand discussed above, households in both countries change their tastes. Suppose, instead, that only the domestic household increases its demand for the domestic good, due to a temporary change in tastes in the first period. As in the case of a worldwide change in tastes (Section 6.3), the relative price of the domestic good rises in the first period. This occurs through a fall in the nominal exchange rate. So the domestic country experiences real and nominal appreciation in the first period. But, in contrast to the results of Section 6.3, the domestic country can experience either a trade deficit or a trade surplus. Whether the real and nominal appreciation is accompanied by a surplus or deficit depends on which of two effects dominates. On the one hand, the rise in the relative price of domestic exports in the first period creates a temporary increase in domestic real income and a temporary decrease in foreign real income (as in Figure 7). As in Section 6.3, this tends to create a domestic trade surplus in the first period. But there is now another force that may tend to create a trade deficit. If the change in tastes by domestic households represents an increased demand for domestic goods in the first period at the expense of *all* other goods, including foreign goods in the first period and both goods in the second period, then domestic demand for *both* goods in the second period falls. The decrease in demand for second-period goods tends to create a domestic trade deficit in the first period. As a result, the domestic country can have either a trade deficit or surplus to accompany its real and nominal appreciation.[20]

6.6 *A Domestic Government Budget Deficit* Suppose the government of the domestic country cuts nondistorting (lump sum) taxes in the first period without changing government spending in either period, (i.e. the government makes lump sum transfers to domestic households, financed by borrowing). The government raises nondistorting taxes in the second period to pay off principal and interest on the debt. The "Ricardian-equivalence proposition" (Barro, 1981) states that under certain conditions the deficit will not affect interest rates or consumption.[21] Under those conditions, people save the entire tax cut, buy the bonds issued by the government, and use the interest on the bonds to pay the higher future taxes. Among the conditions for Ricardian equivalence in this model are that households fully anticipate the higher second-period taxes, and view those taxes as a liability with present value equal to the current tax cut. In that case, households do not gain wealth from the tax cut because liabilities rise as much as current taxes fall. Under the conditions for Ricardian equivalence, an increase in the government budget deficit has no effect on the real or nominal exchange rate or on the trade balance.

A more interesting case arises when the conditions for Ricardian equivalence are violated. To simplify matters, assume that households are shortsighted: in the first period they entirely ignore the higher taxes that will be imposed in the second period. Assume that households ignore the future taxes because they fail to understand that the government must raise future taxes to pay the additional interest (and principal, in this model) generated by the debt issued in the first period. Then the deficit makes domestic households feel wealthier, because they get the current tax cut but ignore the higher future taxes.

Under these assumptions, domestic households will spend part of the tax cut and save the rest for future spending. In the new equilibrium, both foreign and domestic households buy the debt issued by the domestic government. Because money supplies and money demands are unchanged, p and p* are unaffected by the deficit.[22] The interest rate rises because the increase in the quantity of loans demanded by the government exceeds the increase in the quantity of loans supplied by domestic households who save part of the tax cut. That is, the increase in demand for goods in the first period raises the relative price of first-period goods in terms of second-period goods. This relative price is just the real interest rate (plus one). So the higher government budget deficit raises the real interest rate. In addition, the budget deficit causes a trade deficit, because domestic households use the tax cut to buy more imports and to buy more domestic goods (that would otherwise have been exported).

But the budget deficit does not cause a change in either the real or nominal exchange rate, under assumption A8. Domestic households raise demands for *both* goods in the first period in such a way that their relative price is unaffected. Because p and p* are also unaffected, so is the nominal exchange rate.

---

[20] A borderline case occurs with time-separable Cobb-Douglas utility (an elasticity of substitution equal to one), in which case trade is balanced each period.

[21] Roughly, those conditions are: perfect capital markets, a long planning horizon for households, rational expectations, and nondistorting taxes.

[22] If the demand for money depended on the nominal interest rate, then the increase in the interest rate would reduce money demand in both countries, as world interest rates rise. Then p and p* would both fall. If they fell by the same percentage, then the implications for the exchange rate would be the same as if p and p* were both fixed.

The equilibrium is illustrated in Figure 8. The tax cut makes domestic households feel wealthier and raises domestic demand for goods to point B. Then world demand for first-period goods exceeds supply. The real interest rate rises to induce increased saving (lower demand for first-period goods). As all households reduce demand for goods in the first period, an equilibrium is reached at which domestic households (who feel wealthier than foreign households) consume at point D and foreign households consume at point F. The domestic country is borrowing to consume more than point A in the first period. When the domestic country repays the foreign country in period two, domestic consumption is at point J and foreign consumption is at point K.

The real and nominal exchange rates could change if domestic and foreign preferences differed. If domestic households had a preference for domestic goods (and vice versa), then the relative price of the domestic good would rise in the first period. Given p and p*, this rise in p/ep* would occur through a fall in e. So if households in each country have a relative preference for their own country's good, then an increase in the domestic government's budget deficit would raise interest rates, cause a domestic trade deficit, and lead to real and nominal appreciation.[23]

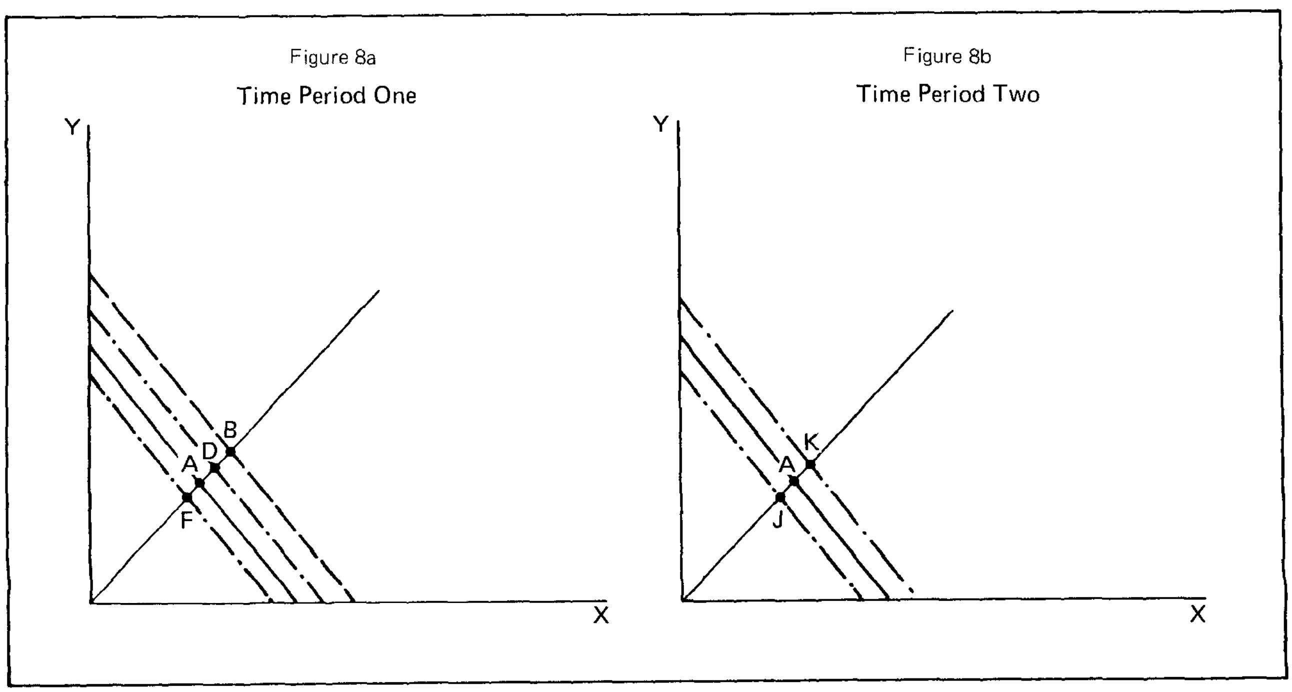
[23] Note that this result has nothing to do with the issue of whether foreign and domestic assets are good (or perfect) substitutes or not, or with the effect of a budget deficit on *relative* interest rates across countries.

6.7 A Shift in Desired Asset Holding It is frequently stated that a change in the preferences of investors to hold interest-bearing assets denominated in dollars or pounds affects the exchange rate. If these assets are not perfect substitutes, it is reasonable to assume that households' demand for each type of asset rises with its own rate of returns and falls with the rate of return on the other type of asset.

Begin with an initial equilibrium in which interest rates in the two countries are the same. Then suppose that foreign households change their preferences for assets in the first period: they wish to hold more assets denominated in pounds and fewer denominated in dollars. As foreigners attempt to buy pound-denominated assets and sell dollar-denominated assets, the relative price of these assets changes. In the new equilibrium, the interest rate on dollar-denominated assets is higher and the interest rate on pound-denominated assets is lower. These interest rates must change until people are willing to hold the existing asset supplies. Because this shift in preferences for assets does not increase or decrease the demands for either *good* or for either *money*, the real and nominal exchange rates are left unchanged.[24]

If foreign and domestic assets are imperfect substitutes then the effect of a budget deficit differs

[24] If money demands depend on interest rates then nominal prices p and p*, and the nominal exchange rate, e, may be affected by the change in asset demands. But—as long as demands for or supplies of goods are unaffected—the *real* exchange rate is unaffected.



slightly from the analysis in Section 6.6. The domestic government is assumed to issue dollar-denominated debt when it has a budget deficit. This increase in the supply of dollar assets lowers the relative price of those assets in terms of other assets, i.e. the domestic interest rate rises relative to the foreign interest rate. In this case, a domestic government budget deficit raises the interest differential between dollar- and pound-denominated assets (and, as before, causes a trade deficit). However, under assumption A8 the real and nominal exchange rates remain unaffected. It is only if tastes differ across countries, with households in each country having a relative preference for their own country's goods, that the domestic country experiences real and nominal appreciation.

7. Additional Evidence and Issues

The typical behavior of real and nominal exchange rates was graphed in Chart 1. Statistical evidence indicates that changes in nominal exchange rates and real exchange rates tend not to be followed quickly by other changes that either reinforce or reverse the original change. The evidence shows the changes in real and nominal exchange rates are either statistically permanent (in the sense that, on average, they are not reversed or reinforced), or highly persistent in the sense that the exchange rate takes a long time to begin returning toward its original level.[25] Huizinga (1987) finds evidence that the real exchange rate begins to reverse its previous changes only after four to seven years. His evidence covers a period of only twelve years; studies over longer time periods sometimes find even larger amounts of persistence, and the uncertainty in statistical estimation is large enough that, with a few exceptions, the evidence is consistent with completely permanent changes in the real exchange rate. The evidence similarly indicates that changes in the nominal exchange rate are either permanent or highly persistent. As argued in footnote 3, this degree of persistence appears to be too large to explain on the basis of disequilibrium models that postulate sticky nominal prices. Many macroeconomists believe that sticky nominal prices play a major role in business cycles (though there are clearly controversies about this). The length of time over which the economy recovers from recessions would provide a rough estimate of the time it takes the overall price level to adjust to its new equilibrium following a disturbance. This estimate would suggest a period of two to three years. In fact, because there are many reasons for business cycles to persist once they have begun, two to three years is probably an upper bound. Disequilibrium theories of exchange rates, based on sticky nominal goods prices, predict that real and nominal exchange rates should return toward their equilibrium levels when nominal goods prices do. This means that they predict systematic changes in real and nominal exchange rates that are not found in the data. The equilibrium theory of exchange rates, on the other hand, is consistent with this evidence if the underlying disturbances to the economy are permanent or highly persistent.

Evidence from the forward exchange market also suggests that changes in exchange rates are expected to be roughly permanent, or highly persistent. Many foreign currencies are traded like commodities on organized futures markets and on forward markets. The futures prices and forward exchange rates move roughly the same amount as spot exchange rates do. While the forward exchange rate may contain a risk premium and so deviate from the market's expectation of the future nominal exchange rate, that premium is unlikely to move systematically so as to mask any expected changes in exchange rates. So available data indicate that people expect changes in exchange rates to be highly persistent rather than temporary as the disequilibrium theories imply. This finding of persistence is inconsistent with the disequilibrium models of exchange rates, but is consistent with equilibrium models that incorporate permanent (or highly persistent) real disturbances. A recent study by Campbell and Clarida (1987) also shows that there is little evidence of any relation between exchange rate changes and real interest rate differentials across countries of the kind that many disequilibrium models predict. Finally, there is only a little evidence to support the contention that government budget deficits per se cause exchange rate changes of the kind predicted by the disequilibrium models or the equilibrium model of Section 6.6, though there is some evidence that variables such as government purchases affect exchange rates as the equilibrium models might suggest (Evans, 1986).[26]

Major questions remain unanswered by current research. Attempts to explain exchange rates empirically using economic "fundamentals," i.e. variables predicted by a theory to have important effects, have

[25] Papers that have documented these facts include (among many others) Roll (1979), Adler and Lehmann (1983), Meese and Rogoff (1983a, b, and 1985), Wasserfallen and Zimmerman (1985), Hsieh (1985), Hakkio (1986), and Huizinga (1987).

[26] Feldstein (1986) argues that budget deficits affect exchange rates. See also Stockman's comments (1986). Evans (1986) presents evidence that government spending rather than deficits affects exchange rates.

generally performed poorly (see, e.g. Meese and Rogoff, 1983a). But the equilibrium approach to exchange rates suggests that the trade balance, output, and other "fundamental" economic variables are not systematically related to the exchange rate in any particular direction, as explained in Section 6. Whether a trade deficit, or increase in domestic output, is associated with depreciation or appreciation depends, according to the theory, on the underlying disturbance. But if real disturbances cause changes in nominal and real exchange rates, then what *are* these disturbances? Can we identify specific examples of underlying changes in technology, tastes, etc. that cause exchange rate changes? While similar questions also remain unanswered for other economic phenomena such as changes in stock prices or business cycle phenomena, further attempts to identify the important exogenous disturbances seems essential.

Another unresolved question involves the explanation for a different fact: the variability of real exchange rates has been much greater when a country adopts a policy of floating nominal exchange rates than when it pegs (fixes) its nominal exchange rate (as under the old Bretton-Woods system that preceded widespread floating beginning in the 1970s). While the explanation is straightforward from the viewpoint of the disequilibrium models, any explanation consistent with an equilibrium model must be more subtle. Indeed, this evidence is sometime cited in support of the disequilibrium models and as contradicting the equilibrium models (e.g. by Mussa, 1987). There are many conditions—not all very realistic—that the economy must meet for the nominal exchange rate system to be totally irrelevant for real exchange rates.[27] One condition requires that all other government policies, including tariffs and quotes on international trade, restrictions on international financial markets, and fiscal policies, are the same under both exchange rate systems. If they are not, then the behavior of real exchange rates may differ under the two systems even if the equilibrium models are right. These issues are currently unresolved.

8. Policy Implications

Clearly the equilibrium theory of exchange rates has radically different policy implications than do disequilibrium theories.[28] First, the government cannot affect the real exchange rate simply by changing the nominal exchange rate, e.g. with policies such as foreign exchange market intervention, target zones, etc. Policies like "talking down (or up) the dollar" may affect the nominal exchange rate because they signal a willingness to pursue policies that affect it; they affect the *real* exchange rate only if they signal a willingness to pursue policies that affect *it*. Unfortunately, those policies generally include protectionist measures that reduce overall economic welfare.

Second, the equilibrium models imply that changes in the exchange rate do not "cause" or "reduce" inflation. Clearly, the exchange rate is an endogenous variable. Moreover, if most changes in exchange rates among countries with similar inflation rates are due to real disturbances to supplies of goods or demands for goods, then changes in the exchange rate may not even be particularly good *signals* of inflation. Exchange rate changes would not be particularly helpful in formulating monetary policies designed to maintain price stability or low inflation.

Third, the choice of fixed versus flexible exchange rates is, by itself, not important for real exchange rates, the trade balance, etc. The choice of an exchange rate system can then be made on the basis of whether one system provides more discipline to policymakers, or whether one would force a country to maintain a higher (or lower) inflation rate than it would like. Similarly, foreign exchange market intervention, "target zones" for exchange rates, and similar policy proposals should be judged on two main criteria: (i) how they would affect inflation, and (ii) how they would affect government incentives to pursue other policies.

Fourth, and perhaps most important, the government should not invoke protectionist restrictions on trade in goods or financial assets as a response to changes in exchange rates. "Undervalued" or "overvalued" currencies are not the issue; exchange rates are only reflections of underlying market conditions and government policies. Variability of exchange rates is no more inherently undesirable than variability in a person's mood throughout a day, and both reflect underlying conditions and policies. The main contribution of the equilibrium theory of exchange rates is to suggest an explanation for exchange rate behavior that is consistent with the notion that markets work reasonably well if they are permitted to. If so, the theory can help us avoid the substitution of folly for wiser policies.

---

[27] Stockman (1983) discusses these conditions.

[28] Most of the research in this area has concentrated attention on positive economics rather than on policy. Additional papers that have used equilibrium models or ideas from them include Helpman (1981), Helpman and Razin (1982, 1984), Hsieh (1982), Sachs (1983), Stockman (1985), Stockman and Hernandez (1987), Stockman and Svensson (1987), Stulz (1986), and Svensson (1985). Other discussions of these ideas can be found in Krueger (1983) and Obstfeld and Stockman (1985); a related discussion appears in Friedman (1953).

References

Adler, Michael, and Bruce Lehmann. "Deviations from Purchasing Power Parity in the Long Run." *Journal of Finance* 38 (December 1983): 1471-87.

Barro, Robert. "Public Debt and Taxes." In *Money, Expectations, and Business Cycles.* New York: Academic Press, 1981.

Campbell, John, and Richard Clarida. "The Dollar and Real Interest Rates." Carnegie-Rochester Conference Series on Public Policy. Edited by Karl Brunner and Allan H. Meltzer. Amsterdam: North-Holland, 1987. Forthcoming.

Dornbusch, Rudiger. "Expectations and Exchange Rate Dynamics." *Journal of Political Economy* 84 (December 1976): 1161-74.

Evans, Paul. "Is the Dollar High Because of Large Budget Deficits?" *Journal of Monetary Economics* 18 (November 1986): 227-50.

Fama, Eugene. "Forward and Spot Exchange Rates." *Journal of Monetary Economics* 14 (November 1984): 319-38.

Frenkel, Jacob, and Michael Mussa. "Asset Markets, Exchange Rates, and the Balance of Payments: The Reformulation of Doctrine." In *Handbook of International Economics,* Vol. 2, edited by R. Jones and P. Kenen. Amsterdam: North-Holland, 1985.

Friedman, Milton. "The Case for Flexible Exchange Rates." In *Essays in Positive Economics,* edited by M. Friedman. Chicago: University of Chicago Press, 1953.

Goodfriend, Marvin. "Exchange Rate Policy and the Dual Role of Exchange Rate Movements in International Adjustments." Federal Reserve Bank of Richmond, *Economic Review* 65 (September/October 1979): 16-26.

Hakkio, Craig. "Does the Exchange Rate Follow a Random Walk? A Monte Carlo Study of Four Tests for a Random Walk." *Journal of International Economics* 20 (June 1986): 221-30.

Helpman, Elhanan. "An Exploration in the Theory of Exchange Rate Regimes." *Journal of Political Economy* 89 (October 1981): 865-90.

Helpman, Elhanan, and Assaf Razin. "Dynamics of a Floating Exchange Rate Regime." *Journal of Political Economy* 90 (August 1982): 728-54.

________. "The Roles of Savings and Investment in Exchange Rate Determination under Alternative Monetary Mechanisms." *Journal of Monetary Economics* 13 (May 1984): 307-26.

Hsieh, David. "The Determination of the Real Exchange Rate: The Productivity Approach." *Journal of International Economics* 12 (May 1982): 355-62.

________. "The Statistical Properties of Daily Foreign Exchange Rates." University of Chicago, 1985. (Processed.)

Huizinga, John. "An Empirical Investigation of the Long-Run Behavior of Real Exchange Rates." Carnegie-Rochester Conference Series on Public Policy. Edited by Karl Brunner and Allan H. Meltzer. Amsterdam: North-Holland, 1987. Forthcoming.

Krueger, Anne O. *Exchange Rate Determination.* Cambridge Surveys of Economic Literature. Cambridge: Cambridge University Press, 1983.

Lucas, Robert. "Interest Rates and Currency Prices in a Two-Country World." *Journal of Monetary Economics* 10 (November 1982): 335-60.

Meese, Richard, and Kenneth Rogoff. "Empirical Exchange Rate Models of the Seventies: Are Any Fit to Survive?" *Journal of International Economics* 14 (February 1983): 3-24 (a).

________. "The Out-of-Sample Failure of Empirical Exchange Rate Models: Sampling Error or Misspecification?" In *Exchange Rates and International Macroeconomics,* edited by J. Frenkel. Chicago: University of Chicago Press, 1983 (b).

________. "Was It Real? The Exchange Rate—Interest Differential Relation, 1973-1984." NBER Working Paper No. 1732, October 1985.

Mussa, Michael. "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates." In *Real Business Cycles, Real Exchange Rates and Actual Policies.* Carnegie-Rochester Conference Series on Public Policy, Vol. 25, edited by Karl Brunner and Allan H. Meltzer. Amsterdam: North-Holland, 1987, pp.117-214.

Obstfeld, Maurice, and Alan Stockman. "Exchange Rate Dynamics." In *Handbook of International Economics,* Vol. 2, edited by R. Jones and P. Kenen. Amsterdam: North-Holland, 1985.

Roll, Richard. "Violations of Purchasing Power Parity and Their Implications for Efficient International Commodity Markets." In *International Finance and Trade,* Vol. 1, edited by M. Sarnat and G. P. Szego. Cambridge, MA: Ballinger, 1979.

Sachs, Jeffery. "Aspects of the Current Account Behavior of OECD Economies." In *Recent Issues in the Theory of Flexible Exchange Rates,* edited by E. Claassen and P. Salin. Amsterdam: North-Holland, 1983.

Stockman, Alan. "A Theory of Exchange Rate Determination." *Journal of Political Economy* 88 (August 1980): 673-98.

________. "Real Exchange Rates under Alternative Nominal Exchange Rate Systems." *Journal of International Money and Finance* 2 (August 1983): 147-66.

________. "Recent Issues in the Theory of Flexible Exchange Rates: A Review Article." *Journal of Money, Credit and Banking* 17 (August 1985): 401-410.

________. "'Comment' on 'The Budget Deficit and the Dollar' by Martin Feldstein." In *NBER Macroeconomics Annual,* edited by Stanley Fischer, 1986.

Stockman, Alan, and Harris Dellas. "International Portfolio Nondiversification and Exchange Rate Variability." University of Rochester, 1986. (Processed.)

Stockman, Alan, and Alejandro Hernandez. "Exchange Controls, Capital Controls, and International Financial Markets." *American Economic Review* (1987). Forthcoming.

Stockman, Alan, and Lars Svensson. "Capital Flows, Investment, and Exchange Rates." *Journal of Monetary Economics* 19 (March 1987): 171-202.

Stulz, Rene. "An Equilibrium Model of Exchange Rate Determination and Asset Pricing with Non-traded Goods and Imperfect Information." Ohio State University Working Paper, 1986.

Svensson, Lars. "Currency Prices, Terms of Trade, and Interest Rates: A General Equilibrium Asset-Pricing, Cash-in-Advance Approach." *Journal of International Economics* 18 (February 1985): 17-41.

Wasserfallen, Walter, and Heinz Zimmerman. "The Behavior of Intradaily Exchange Rates." *Journal of Banking and Finance* 9 (March 1985): 55-72.

IPC OR TOTAL DEPOSITS? THERE IS A DIFFERENCE!

Donald L. Welker

"This probably sounds like a basic question, but. . . ." Some variation of this introduction often is a prelude to a discussion of how to report bank concentration for bank merger or bank holding company application purposes. Other than applications to form one-bank holding companies, most applications to acquire banks or bank holding companies require information on market concentration. The prospective applicant usually knows about such things as market tables and Herfindahl-Hirschman Indices. The question is, should the market table be constructed from total deposits or IPC deposits?

Tactful attempts to explain that the Federal Reserve System prefers total deposits for purposes of competitive analysis tend to provoke the objection that "other agencies" emphasize IPC deposits. The caller is referring, of course, to the U. S. Department of Justice, the Office of the Comptroller of the Currency (OCC)[1] and the Federal Deposit Insurance Corporation (FDIC).

This article attempts to clarify the distinction between IPC deposits and total deposits. Then it will show the effect of using the alternative deposit definitions to measure concentration in selected Fifth District banking markets. The expanding role of thrift institutions as competitors of banks also will be discussed.

Deposits of Individuals, Partnerships and Corporations (IPC Deposits)

Normally the largest subset of a bank's deposits, this IPC category represents exactly what the name signifies. Most of the locally limited customers who provide a basis for the concept of a banking market are included here, although a large percentage of IPC deposits may be held by customers with access to national markets.

The most commonly used source of deposit information for specific banking markets is the Summary of Deposit data published annually by the FDIC. (This information is included in a publication entitled *Data Book-Operating Banks and Branches.)* One computes total IPC deposits for each institution by combining the two classifications of IPC Transaction Accounts and IPC Nontransaction Accounts for each geographic location.

Total Deposits

In addition to IPC deposits, total deposits encompass a variety of bank creditors who may not be effectively restricted to the local banking market. An important group of depositors, duly reported in the Summary of Deposits, are those holding "public funds" including federal, state and municipal governments. The deposits of these public bodies are often characterized as "political" deposits.

A reason for excluding governmental units from local banking markets is that they may have access to a national funds market. In practice, however, numerous state and local laws limit political deposits to the taxing jurisdiction and thus to specific banking markets. By contrast, large corporations often have far greater access to national markets through use of cash management services.

Other non-IPC categories not listed separately in the Summary of Deposits include deposits of foreign governments, commercial bank deposits, and certified and officers checks. Bank deposits are the major item in this group. While banks occasionally maintain correspondent relationships with competitors, self-interest determines that most accounts will be maintained with correspondent banks located outside the respondents' markets.

Basis for Determining Market Structure

As mentioned in the introduction, the Fed traditionally favors total deposits[2] when evaluating

Josephine O. Hawkins provided expert research assistance.

[1] Since 1985, the OCC has incorporated a "Quick Check Merger Screen" in its application process which defers to Federal Reserve market definitions. IPC deposit information must be included, however, as a part of all applications which fail to pass the initial screen for material competitive issues.

[2] A study prepared at the Board in 1965 based on data from the Distribution of Bank Deposits by Counties and Standard Metropolitan Areas for 1956 and 1960 concluded that concentration ratios computed from IPC deposits produced ". . . essentially the same results" as concentration ratios derived from total deposits [Flechsig, 1965].

structural relationships whereas the Department of Justice and other bank regulatory agencies prefer to use IPC deposits. This distinction may be more apparent than real in terms of practical results. As an example, the following section will show that in the top ten markets in the Fifth District concentrated markets remain concentrated whether classified by total deposits or IPC deposits. Unconcentrated markets on the basis of total deposits do not become concentrated when limited to IPC deposits.

The trend to include all or a portion of the deposits held by thrift institutions in banking markets, however, has the potential to modify some relationships as thrifts evolve toward becoming full competitors of banks. Correspondent banking currently is not a routine function of thrift institutions. Nor have thrifts developed the capital structures which would facilitate the ability to compete aggressively for public funds despite the removal of some legal barriers to such deposits in recent years. In fact, the Federal Home Loan Bank Board (FHLBB) does not even report IPC deposits for savings and loan associations. Any market table constructed from publicly available data must perforce focus on total deposits at thrifts.

Results in the Fifth District's Top Ten Markets

Non-IPC deposits are a comparatively small but material part of large banking markets in this District. Within a narrowly defined product definition limited to commercial banks, non-IPC deposits range from a low of 4.3 percent in the unconcentrated Washington, D. C., market to a high of 15.0 percent in the concentrated Richmond, Virginia, area with a weighted average for the ten markets of 7.7 percent (Table 1).

Recalling that thrifts report only total deposits, it follows that expansion of the product market to include thrifts would tend to reduce the relative significance of non-IPC deposits. Non-IPCs as a percent of aggregate bank and thrift deposits in the top ten markets range from 2.4 to 11.0 percent with a mean of 4.7 percent. Washington again has the smallest proportion with only 2.4 percent, but the greatest percentage of non-IPCs is now identified with the Winston-Salem, North Carolina, market at 11.0 percent (Table 2).

Table 1

TOP TEN BANKING MARKETS
FIFTH DISTRICT
June 30, 1985

(Dollar amounts in thousands)

	Total Bank Deposits	Total Bank IPC Deposits	Non-IPC Deposits as a Percent of Total Deposits
Washington, D.C.	$22,172,280	$21,210,219	4.34
Baltimore, Maryland	11,547,840	10,608,132	8.14
Charlotte, North Carolina	5,266,793	4,811,986	8.64
Richmond, Virginia	5,067,217	4,304,988	15.04
Norfolk-Portsmouth, Virginia	3,682,253	3,379,413	8.22
Winston-Salem, North Carolina	2,596,404	2,214,065	14.73
Raleigh, North Carolina	2,202,738	2,026,739	7.99
Columbia, South Carolina	1,930,330	1,685,142	12.70
Charleston, West Virginia	1,880,521	1,764,152	6.19
Greenville, South Carolina	1,429,134	1,333,277	6.71
Total	$57,775,510	$53,338,113	7.68

Table 2

TOP TEN BANKING MARKETS

FIFTH DISTRICT

June 30, 1985

(Dollar amounts in thousands)

	BANKS AND THRIFTS		
	Total Deposits	Total IPC Deposits	Non-IPC Deposits as a Percent of Total Deposits
Washington, D.C.	$39,947,208	$38,985,147	2.41
Baltimore, Maryland	19,536,585	18,596,877	4.81
Charlotte, North Carolina	6,817,605	6,362,798	6.67
Richmond, Virginia	7,529,874	6,767,645	10.12
Norfolk-Portsmouth, Virginia	6,349,866	6,047,026	4.77
Winston-Salem, North Carolina	3,476,383	3,094,044	11.00
Raleigh, North Carolina	2,986,878	2,810,879	5.89
Columbia, South Carolina	3,142,144	2,896,956	7.80
Charleston, West Virginia	2,241,979	2,125,610	5.19
Greenville, South Carolina	2,841,265	2,745,408	3.37
Total	$94,869,787	$90,432,390	4.68

The market tables confirm that alignment of market structure often is not affected by the use of IPC deposits as an alternative to total deposits. But there are exceptions. For example, consider the Richmond, Virginia, market when all thrift deposits are included (Table 3). Here the four largest institutions are commercial banks. Now refer to Table 4 where the Richmond bank/thrift market structure is determined by total IPC deposits. Under this alternative, the first and second ranked banks in the area have swapped places and the four largest depository institutions now include a savings and loan association.

One usually constructs market tables for the purpose of measuring concentration in terms of deposit concentration ratios and the Herfindahl-Hirschman Index (HHI). The HHI may be defined simply as the sum of the squares of the respective market shares of all participants in the market. For example, to determine the contribution to the HHI by a bank with 12 percent of the deposits in a given market, simply multiply .12 times .12 times 10,000 which equals 144. Then add the comparable data computed for all other banks in the market to obtain the HHI. (See Tables 3 and 4 for practical illustrations of the technique.) Following the U. S. Department of Justice's publication in 1982 of its Merger Guidelines based on the HHI, this statistic has become a widely accepted measure of concentration. Justice's guidelines for bank acquisition permit an increase of 200 in a concentrated market's HHI which is equivalent to combining two banks with respective market shares of 10.0 percent.

As depicted in Table 5, calculation of the HHI on the basis of IPC deposits will reduce the indicated levels of concentration for the first nine markets in the District by amounts ranging from just one for Baltimore, Maryland, to 498 for the Winston-Salem, North Carolina, market. Note, however, that the HHI for the Greenville, South Carolina, market actually registered an increase of 44. By contrast, the ten-market average change in the HHI was a decrease of 78. This means that, on the average, two banks with respective market shares of 6.24 percent could merge in the composite market measured by IPC deposits without exceeding the HHI for the market based on total deposits.

Adding thrift deposits to the markets reduces absolute levels of concentration, but deletion of

Table 3

RICHMOND, VA, RMA BANK/THRIFT MARKET

June 30, 1985

(Dollar amounts in thousands)

Rank	Bank	Total Deposits	Percent of Total Deposits in Market	Herfindahl-Hirschman Index	Cumulative Herfindahl-Hirschman Index
1	United Virginia Bank	$1,372,240	18.22	332.11	332.11
2	Bank of Virginia	1,216,014	16.15	260.80	592.91
3	Sovran Bank, NA	1,142,387	15.17	230.17	823.08
4	Central Fidelity Bank	529,363	7.03	49.42	872.50
5	Heritage S&LA	525,600	6.98	48.72	921.23
6	Investors S&LA	355,135	4.72	22.24	943.47
7	Virginia FS&LA	346,580	4.60	21.19	964.66
8	Dominion Bank of Richmond, NA	296,630	3.94	15.52	980.17
9	Franklin FS&LA	277,946	3.69	13.63	993.80
10	Southern Bank	249,016	3.31	10.94	1004.74
11	Citizens S&LA, FA	190,365	2.53	6.39	1011.13
12	Security FS&LA	189,627	2.52	6.34	1017.47
13	First Virginia Bank—Colonial	173,566	2.31	5.31	1022.78
14	Colonial S&LA	136,807	1.82	3.30	1026.08
15	Lincoln S&LA	132,456	1.76	3.09	1029.18
16	Cardinal S&LA	103,226	1.37	1.88	1031.06
17	Pioneer FS&LA	52,624	0.70	0.49	1031.55
18	Virginia First Savings, FSB	52,592	0.70	0.49	1032.03
19	Consolidated Bank & Trust Co	43,205	0.57	0.33	1032.36
20	Dominion FS&LA	41,988	0.56	0.31	1032.67
21	First FSB of Virginia	33,233	0.44	0.19	1032.87
22	Bay Savings Bank, FSB	24,478	0.33	0.11	1032.97
23	Virginia Capital Bank	21,301	0.28	0.08	1033.05
24	The Suburban Bank	11,600	0.15	0.02	1033.08
25	Union Bank & Trust Co	5,447	0.07	0.01	1033.08
26	Peoples Bank of Virginia	4,177	0.06	0.00	1033.09
27	First National Bank, Louisville	2,271	0.03	0.00	1033.09
	Total Market	$7,529,874	100.00	1033.09	1033.09

Notes: The three bank concentration ratio is 49.54 percent.

The four bank concentration ratio is 56.57 percent.

THRIFT DEPOSITS WEIGHTED AT 100.00 PERCENT

Table 4

RICHMOND, VA, RMA BANK/THRIFT MARKET[1]

June 30, 1985

(Dollar amounts in thousands)

Rank	Bank	Total IPC Deposits	Percent of Total Deposits in Market	Herfindahl-Hirschman Index	Cumulative Herfindahl-Hirschman Index
1	Bank of Virginia	$1,154,202	17.05	290.86	290.86
2	United Virginia Bank	1,122,280	16.58	275.00	565.86
3	Sovran Bank, NA	871,753	12.88	165.92	731.78
4	Heritage S&LA	525,600	7.77	60.32	792.10
5	Central Fidelity Bank	413,535	6.11	37.34	829.44
6	Investors S&LA	355,135	5.25	27.54	856.98
7	Virginia FS&LA	346,580	5.12	26.23	883.20
8	Franklin FS&LA	277,946	4.11	16.87	900.07
9	Dominion Bank of Richmond, NA	249,197	3.68	13.56	913.63
10	Southern Bank	245,152	3.62	13.12	926.75
11	Citizens S&LA, FA	190,365	2.81	7.91	934.66
12	Security FS&LA	189,627	2.80	7.85	942.51
13	First Virginia Bank—Colonial	168,413	2.49	6.19	948.71
14	Colonial S&LA	136,807	2.02	4.09	952.79
15	Lincoln S&LA	132,456	1.96	3.83	956.62
16	Cardinal S&LA	103,226	1.53	2.33	958.95
17	Pioneer FS&LA	52,624	0.78	0.60	959.55
18	Virginia First Savings, FSB	52,592	0.78	0.60	960.16
19	Dominion FS&LA	41,988	0.62	0.38	960.54
20	Consolidated Bank & Trust Co	38,600	0.57	0.33	960.87
21	First FSB of Virginia	33,233	0.49	0.24	961.11
22	Bay Savings Bank, FSB	24,478	0.36	0.13	961.24
23	Virginia Capital Bank	21,128	0.31	0.10	961.34
24	The Suburban Bank	11,261	0.17	0.03	961.36
25	Union Bank & Trust Co	5,447	0.08	0.01	961.37
26	Peoples Bank of Virginia	3,949	0.06	0.00	961.37
27	First National Bank, Louisville	71	0.00	0.00	961.37
	Total Market	$6,767,645	100.00	961.37	961.37

Notes: The three bank concentration ratio is 46.52 percent.

The four bank concentration ratio is 54.29 percent.

THRIFT DEPOSITS WEIGHTED AT 100.00 PERCENT

[1] Total IPC deposits for banks and total deposits for thrifts.

Table 5

TOP TEN BANKING MARKETS
FIFTH DISTRICT
June 30, 1985

	HHI Based on Total Bank Deposits	HHI Based on Total Bank IPC Deposits	Change	Percent of Change
Washington, D.C.	816	807	-9	-1.10
Baltimore, Maryland	1254	1253	-1	-0.08
Charlotte, North Carolina	3126	3003	-123	-3.93
Richmond, Virginia	1998	1983	-15	-0.75
Norfolk-Portsmouth, Virginia	2270	2210	-60	-2.64
Winston-Salem, North Carolina	4969	4471	-498	-10.02
Raleigh, North Carolina	1481	1451	-30	-2.03
Columbia, South Carolina	1905	1871	-34	-1.78
Charleston, West Virginia	1430	1380	-50	-3.50
Greenville, South Carolina	1475	1519	44	2.98
Average Change			-77.6	-3.74

non-IPC deposits yields changes in the HHI comparable to results already observed when IPC deposits are considered for banks only. IPCs reduce the ten-market average HHI by 76 when thrifts are added to the product market compared with a reduction of 78 in the HHI when the market is restricted to banks. This average includes reductions in HHIs for specific markets ranging from 6 in the Washington market to 437 for Winston-Salem. Greenville again represents an exception with an increase in the HHI of 52 (Table 6).

It is widely recognized that thrifts may not be fully comparable to commercial banks in all respects despite the enactment in recent years of legislation which enables thrifts to accept demand deposits (NOW accounts) and grant commercial loans. Others suggest that one hundred percent of thrift deposits is the relevant standard because thrifts have the potential to become full competitors of banks. The Board of Governors' pragmatic approach to this reality usually has been to permit the inclusion of 50 percent of the deposits held by thrifts for the purpose of determining concentration in a banking market. On the other hand, the U. S. Department of Justice elects to calculate separate indices for "wholesale" and "retail" markets. Justice includes one hundred percent of thrift deposits in the retail market, while only twenty percent of thrift deposits are added to the wholesale market.

Table 7 demonstrates the effect of weighting thrift deposits at 50 percent in the District's largest markets. This approach produces the greatest variation in the HHI when IPC deposits are compared with total deposits. The mean reduction in HHI after removing non-IPC deposits from the market is 96 under this alternative. The increase in concentration for the Greenville, South Carolina, market due to using IPC deposits shows the risks inherent in making sweeping generalizations about banking markets. Banks in the market hold approximately 50.3 percent of total bank/thrift deposits, but only 48.6 percent of total IPC deposits.The smaller banks in the market apparently have managed to attract a disproportionately large share of non-IPC deposits. The first and second largest depository institutions in the market are thrifts. These two organizations hold 43.2 percent of total deposits and 44.7 percent of total IPC deposits.

Conclusion

Analysts usually include at least a portion of thrift deposits when measuring banking market structure. The only thrift deposit category currently reported

Table 6

TOP TEN BANKING MARKETS
FIFTH DISTRICT
June 30, 1985

	HHI Based on Total Deposits of Banks and Thrifts	HHI Based on Total IPC Deposits of Banks and Thrifts	Change	Percent of Change
Washington, D.C.	371	365	−6	−1.62
Baltimore, Maryland	522	501	−21	−4.02
Charlotte, North Carolina	1946	1810	−136	−6.99
Richmond, Virginia	1033	961	−72	−6.97
Norfolk-Portmouth, Virginia	1038	993	−45	−4.34
Winston-Salem, North Carolina	2948	2511	−437	−14.82
Raleigh, North Carolina	1017	993	−24	−2.36
Columbia, South Carolina	1062	1036	−26	−2.45
Charleston, West Virginia	1112	1069	−43	−3.87
Greenville, South Carolina	1324	1376	52	−3.93
Average Change			−75.8	6.13

Table 7

TOP TEN BANKING MARKETS
FIFTH DISTRICT
June 30, 1985

	HHI Based on Total Bank Deposits and 50 Percent of Thrift Deposits	HHI Based on Total Bank IPC Deposits and 50 Percent of Thrift Deposits	Change	Percent of Change
Washington, D.C.	466	453	−13	−2.79
Baltimore, Maryland	725	699	−26	3.59
Charlotte, North Carolina	2401	2257	−144	−6.00
Richmond, Virginia	1339	1258	−81	−6.05
Norfolk-Portsmouth, Virginia	1333	1261	−72	−5.40
Winston-Salem, North Carolina	3691	3187	−504	−13.65
Raleigh, North Carolina	1138	1099	−39	−3.43
Columbia, South Carolina	1235	1173	−62	−5.02
Charleston, West Virginia	1221	1170	−51	−4.18
Greenville, South Carolina	1082	1110	28	2.59
Average Change			−96.4	−6.59

by geographic location, however, is total deposits. This constitutes a persuasive reason for continuing to evaluate market concentration on the basis of total deposits despite the attraction of IPC deposits. Combining total deposits of thrifts with total IPC deposits of banks may overemphasize the market concentration attributed to thrift institutions. Proponents of thrifts as full competitors of banks do not attempt to claim that thrift deposits should be weighted more heavily than deposits held by commercial banks when assessing competitive relationships.

Our review of large banking markets in the Fifth Federal Reserve District tends to confirm that non-IPC deposits are more significant relative to the structure of some markets than for others. Whenever HHI statistics for banking markets begin to approach the critical range as determined by the Merger Guidelines, both applicants and bank regulatory agencies may find it constructive to review the market in terms of alternative deposit definitions as well as to explore the underlying causes of those differences.

References

Daskin, A. J. "Aggregate Concentration and Geographic Diversification in U.S. Commercial Banking, 1970-1982." Journal of Economics and Business 37 (August 1985): 237-51.

Decision Research Sciences, Inc. *1985-1986 Branch Directory and Summary of Deposits.* Blue Bell, PA., 1985.

Federal Deposit Insurance Corporation. *Data Book—Operating Banks and Branches.* June 30, 1985.

Flechsig, Theodore G. *Banking Market Structure & Performance in Metropolitan Areas.* Washington: Board of Governors of the Federal Reserve System, 1965.

Gilbert, R. A. "Bank Market Structure and Competition." *Journal of Money, Credit and Banking* 16 (November 1984): 617-60.

Keely C. and Gary C. Zimmerman. "Determining Geographic Markets for Deposit Competition in Banking." Federal Reserve Bank of San Francisco, *Economic Review* (Summer 1985), pp. 25-45.

Mote, Larry R. "Competition in banking: What is known? What is the Evidence?" Federal Reserve Bank of Chicago, *Business Conditions* (February 1967), pp. 7-16.

Rand McNally & Company. *Commercial Atlas and Marketing Guide.* Chicago: Rand McNally & Company, 1985.

Tayloe Murphy Institute. "Branch Deposits in Financial Institutions in Virginia." University of Virginia: The Colgate Darden Graduate School of Business Administration. June 30, 1985.

Watro, Paul R. "Thrifts and the Competitive Analysis of Bank Mergers." Federal Reserve Bank of Cleveland, *Economic Review* (Winter 1983), pp. 13-32.

Welker, Donald L. "Thrift Competition: Does it Matter?" Federal Reserve Bank of Richmond, *Economic Review* 72 (January-February 1986): 2-10.

Whitesell, William E. "The Bank Merger Act of 1966: Past, Present and Prospects." Federal Reserve Bank of Philadelphia, *Business Review* (November 1968), pp. 3-9.

Yanni, Joseph A. "Managing the Competitive Factors in a Merger." *Issues in Bank Regulation* 7 (August 1983): 7-10.

Yesley, Joel M. "Defining the Product Market in Commercial Banking." Federal Reserve Bank of Cleveland, *Economic Review* (June-July 1972), pp. 17-31.