1984

ECONOMIC REVIEW

FEDERAL RESERVE BANK
of CLEVELAND

FALL

Economic Review
Federal Reserve Bank of Cleveland
Fall 1984

Settlement risk is the risk that a bank will be unable to repay other banks for daylight credit, meaning the net amount (payments minus receipts) owed for a day's transactions on a payment system. Payments volumes have been growing rapidly on Fedwire and even more rapidly on privately operated systems. This growth highlights the need for management of settlement risk exposure. One vital ingredient of settlement risk management is the ability to control daylight credit extended or used by a bank across the various payment systems. Equally vital is the need for banks and their customers to recognize who is at risk in making and receiving large-dollar payments.

This article reconsiders the problem of a declining rate of return on capital. The rate of return for noncorporate businesses and corporations fell after 1965. A model designed to examine sources of the decline indicates that the major forces at work were inflationary distortions in the relative price of capital and falling capital productivity. Other contributors were a one-time-only price-cost distortion in the late 1960s and periodic cyclical weakness. None of these effects, perhaps excepting falling capital productivity, can be considered a permanent or long-term erosion in the profitability of capital.

Economic Review is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101. Telephone: 216/579-2000.

Editor: Pat Wren. Assistant editor: Meredith Holmes. Design: Jamie Feldman. Typesetting: Lucy Balazek.

ISSN 0013-0281

An assistant vice president with the Federal Reserve Bank of Cleveland, the author is responsible for Federal Open Market Committee briefings and other monetary policy advice. The author has benefited from the comments of Donald Hester, Ernest Patrikis, David Humphrey, James Hoehn, Roger Hinderliter, Kim Kowalewski, John Carlson, and Mark Sniderman.

Risk in Large-Dollar Transfer Systems

by E.J. Stevens

Risk in large-dollar transfer systems was an unfamiliar topic until quite recently. Even today, people who are not banking professionals probably have little notion of what these systems are, what risks they involve, and why those risks should be investigated.

In a nutshell, large-dollar transfer systems in the United States are telecommunications networks (currently including Fedwire, CHIPS, CashWire, and CHESS) almost entirely dedicated to same-day handling of multimillion dollar payments among banks (see box 1). These payments may be for a bank itself, as it buys and sells money, or for the accounts of the bank's customers, especially financial institutions, active in world money and capital markets, and nonfinancial corporations. The risk being discussed is settlement risk: that, at the end of a day, a participating bank will not be able to pay the net amount (payments minus receipts) it owes to the other banks on one of the systems for that day's transactions. How to deal with this risk may be viewed differently by banks, their customers, and the Federal Reserve System. Nev-

Box 1 Large-Dollar Transfer Systems

CashWire is operated by the Payment and Administrative Communication Corporation owned by a consortium of 180 U.S. banks. Service is currently provided to 17 of these banks, which send approximately 350 payments daily with an average value of about $700,000.

CHESS, or Clearing House Electronic Settlement System, is operated by the Chicago Clearing House Association with service available to members, who must be in the 7th Federal Reserve District. Service is currently provided to six banks, which send approximately 450 messages daily with an average value of about $1.0 million.

CHIPS, or Clearing House Interbank Payments System, is operated by the New York (City) Clearing House Association with service available to its members. Service is currently provided to over 120 institutions, which send approximately 75,000 payments daily with an average value of about $3.0 million.

Fedwire is operated by the 12 Federal Reserve Banks and their branches with service available to any depository institution with an account relationship with a Reserve Bank. Service is currently provided to over 7,000 institutions, which send approximately 150,000 payments daily with an average value of about $2.4 million.

ertheless, there is a common concern that the risk be recognized and methods of risk management be understood.[1] To that end, the Board of Governors of the Federal Reserve System has formally asked for industry comment on a number of proposed methods of risk management and control.[2]

The purpose of this article is to examine the concept of settlement risk (section II). This requires, in addition, a background description of the institutional basis for making large-dollar payments in the United States (section I) and an explanation of why concern about settlement risk has only recently come to the fore (section III).

I. Large-Dollar Payments and Settlement

Payment Systems

The nature of large-dollar transfer systems in the United States is best understood by contrasting them with small-dollar systems on the one hand and, on the other, with nonpayment message systems. Small-dollar transfer systems—cash, checks, automated clearinghouses (e.g., for direct deposit of Social Security or salary payments)—are familiar to most people. Whereas an average cash transaction is for considerably less than $100, and even the average check (including business checks) is for less than $1,000, the average wire transfer (the generic name for payments made on large-dollar systems) is in the million-dollar range. This is not to imply that wire transfers are somehow restricted to such large sums: small payments can be made as readily as large. However, routine wire transfer of small payments is usually un-economic because the value of gaining interest for a day or so by completing a same-day payment is not worth the additional cost of a wire transfer.[3]

Wire transfers typically are completed on a telecommunications system, although telephone and other devices can be employed. Fledgling debit card and home banking systems exist, some automated clearinghouse transactions are now made on a computer-to-computer basis, and electronic check collection is contemplated. Nonetheless, small payments typically are not completed on telecommunications systems.

Large-dollar transfer systems should also be contrasted to inter-bank message systems that, standing alone, do not necessarily effect payments. SWIFT and BankWire II, for example, are technologically sophisticated telecommunications systems; however, they are only capable of conveying messages. Those messages may be—frequently are—instructions authorizing a bank to remove funds from the account of one depositor and place them in the account of another. When both depositors maintain accounts at the bank receiving the message, only follow-up internal bookkeeping is required to complete the process. However, if the account to which the payment is directed is at any other bank, then no payment has been made until a follow-up interbank payment has been initiated via one of the large-dollar transfer systems. In this sense SWIFT and BankWire II, despite restricted access, standardized formats, and a sophisticated telecommunications vehicle for transmission, are closer to telex, telephone, and mail service than to large- or small-dollar transfer systems. What distinguishes large-dollar transfer systems from these message systems is the ability of one participant to transfer cash in the form of immediately available bank balances to any other participant by a single message. In the jargon of payments, these large-dollar transfer systems make payments because they include settlement—the irrevocable transfer of ownership of bank balances from one participant to another. This transfer is analogous to payment in cash in the form of legal tender currency and is brought about by the transfer of deposit balances at Federal Reserve Banks from the account of the paying bank to that of the receiving bank.

1. For example, see "Risks in Electronic Payments Systems," Report of the Risk Task Force, Association of Reserve City Bankers, October 1983.

2. See Board of Governors of the Federal Reserve System, "Proposals to Reduce Risk on Large-Dollar Transfer Systems," Docket No. R-0515.

3. Automated Fedwire payments, for example, involve a fee of 55¢ to each of the sending and receiving banks. The banks themselves would have to charge substantially more than this to cover the 55¢ fee plus inhouse costs.

Non-pecuniary costs may loom just as large. Specifically, it is sometimes argued that small-value wire transfers are more likely to go astray. A bank that fails to receive a transfer, standing to gain only the overnight return on the value of the payment, has little incentive to initiate a search for a missing $1,000 payment; other parties to the transaction may be relied on to do the job. The incentive is a thousand-fold greater on a million-dollar transfer, which is therefore likely, if lost, to be sought assiduously until found.

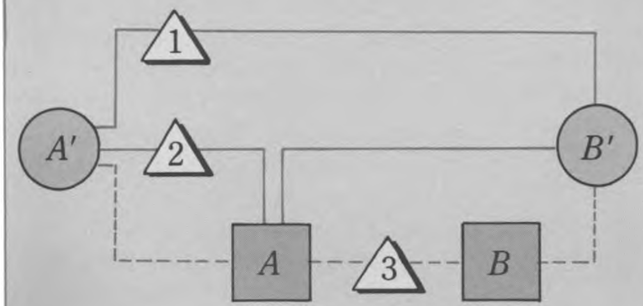
Settlement Risk

The risk that is the major concern in large-dollar transfer systems is settlement risk, reflecting the possibility that the recipient's bank will not be paid by the payor's bank. This risk should be distinguished from the more familiar risks confronted in ordinary payments, including the simple risk of nonpayment of a debt and the risk of being paid with a bad check (see box 2). Settlement risk is the risk that a bank, despite a paying customer's adequate balance or credit line, will be unable to cover payments it has initiated on various payment systems during the banking day.

The circumstances under which this might happen are extreme. Inability of a bank to settle would be the symptom not just of a bank failure but of an unexpected bank failure. Indeed, it might be argued that, at least for very large banks, such an event could not happen. After all, in the event of trouble, wouldn't the supervisory authorities be summoned immediately to arrange a last-resort loan from its Reserve Bank? In this way, every other participant in a large-dollar transfer system would be paid; no settlement failure would occur. This argument ignores two vital matters. First, as explained in the next paragraph, settlement risk would not disappear; it would simply be transformed into credit risk of the Federal Reserve or other creditors of the troubled institution. Second, as explained in section II, the supervisory authorities and the Federal Reserve might be well-advised to adopt a "hands-off" policy in the case of an incipient and isolated settlement failure, allowing other participants in a network to absorb the losses occasioned by settlement failure. Settlement risk is, therefore, the risk that no other bank, supervisory authority, or government lender of last resort will lend enough to a bank to cover its debit position vis-à-vis another institution before the end of the day.

Exposure to settlement risk may be borne by the recipient of a payment or by his bank, depending on which payment system is utilized. Settlement risk for Fedwire payments is borne by the Reserve Banks, which stand as the bankers (B in box 2) for the depository institutions receiving payments (B' in box 2). Notification of a payment sent by a Reserve Bank to a recipient bank is an irrevocable notice that immediately available funds have been credited to its account. If the Reserve Bank discovers at the end of the day that the paying bank has insufficient funds to cover its Fedwire payments, the Reserve Bank can-

Box 2 Three Kinds of Risk



A and *B* are banks.
A′ and *B′* are customers of those banks.

1. *Nonpayment*: *A′* owes *B′* money, but may fail to do anything about it. *B′* is exposed to a credit risk: *A′* may be a deadbeat.
2. *Bad Check*: *A′* pays *B′* by check, but has insufficient funds in his account with *A* to cover the check. *A′* gives the check to *B′*, who deposits it in *B*. *B* sends the check to *A* for payment, receiving provisional credit in *B*'s account at the Federal Reserve Bank, where *A*'s account is provisionally debited. *A* then finds *A′* has insufficient funds and returns the check to *B* via the Federal Reserve, which reverses the previous entries.
3. *Settlement Risk*: Upon instruction from *A′*, *A* wires funds to *B* for credit to *B′*. At the end of the day, *A* has insufficient funds to pay *B* the amount it owes for this (and other) payments, even after subtracting other payments made by *B* to *A*.

4. These matters are spelled out in the Federal Reserve's Regulation J, Section 210.36, Final Payment and Use of Funds: (a) Final Payment. A transfer item is finally paid when the transferee's Reserve Bank sends the transfer item or sends or telephones the advice of credit for the item to the transferee, whichever occurs first. (b) Right to Use Funds. Credit given by a Reserve Bank for a transfer of funds becomes available for use when the transfer item is finally paid, subject to the Reserve Bank's right to apply the transferred funds to an obligation owed to it by the transferee. See "Regulation J: Collection of Checks and Other Items and Wire Transfers of Funds," Federal Reserve Regulatory Service, vol. III, section 210, as amended, effective April 2, 1984, p. 7-043.

not recover the funds from the recipients.[4]

The location of settlement risk exposure in CHIPS is different from Fedwire. This large-dollar system is not a bank, but a clearinghouse with net settlement. Participating banks receive notices of payment during the day from each of the other banks and send notices of payments to the other banks. The clearinghouse records all of these payments, but no transfer of assets between participants takes place. At day's end, each bank then is in a net debit or credit position with respect to each other bank, reflecting the net balance of payments to, relative to payments from, each of those other banks. Of course, the sum of all these bilateral net debit and credit positions is always zero because what each bank owes is what each other bank is owed. This accounting identity makes clear why settlement of the day's payments can be such a simple matter. Each bank in net debit position with respect to the aggregate of all other participants might pay what it owes into a pool from which each bank in net credit position could be paid. In this way 120 participants would be able to settle 7,140 bilateral positions [$(120 \times 119) \div 2$] with a total of only 120 payments, some into and some out of the clearinghouse settlement account (see box 3).

The facts of CHIPS settlement are both more and less simple than these few sentences suggest. CHIPS participants include both settling and non-settling banks. Non-settling banks both send and receive payments in their own names. However, at the end of the day, they pay or receive the net balance due from or to them through one of the settling banks rather than directly with the settlement account at the Federal Reserve Bank of New York. A settling bank agrees to receive the net credits or pay the net debits attributable both to its own activity as well as to that of each of its non-settling respondents. Settlement then proceeds in several steps at the end of the day. First, net positions are reported so that each participant has an opportunity to verify all of its bilateral net debit and credit positions and its aggregate net debit or credit ("net net") position. Then, settling banks must indicate their readiness to settle, before which they have had an opportunity to confirm that their non-settling respondent participants in debit positions either have a sufficient balance with the settling bank, or are able to borrow a sufficient amount from the settling bank, to cover the debit. When all settling banks have signaled their readiness to settle, those in debit positions send the amounts they owe from their Federal Reserve Bank deposit accounts via Fedwire to the CHIPS settlement account at the Federal Reserve Bank of New York. When all payments are received, the balance in the settlement account is dispersed via Fedwire to the Federal Reserve Bank deposit accounts of the settling banks in credit positions.

Settlement risk in CHIPS, then, refers to the possibility of two different kinds of failure to pay. One would be the inability of a non-settling participant to cover its debit position, either by funding its account with the settling bank or by a loan from that bank. The other would be the inability of a settling bank to fund its account at the Federal Reserve Bank.

Locating the risk of loss in these hypothetical situations involves identifying how settlement is handled if one of the participants is unable to make settlement. CHIPS' Rule 13 specifies the series of steps for operating under such circumstances. First, more time can be granted, in effect extending the end of the day beyond the appointed settlement time (normally 5:45 pm). This extension may be granted by the clearinghouse because a settling bank notifies the clearinghouse of its unwillingness to settle for one of the non-settling par-

5. Deletion of transactions does not relieve the participant of its obligation to make those payments, according to Rule 13.

ticipants for whom it would otherwise be obligated to settle, or because a settling bank notifies the clearinghouse that it is unable to cover its own debit position by settlement time. The extra time would allow a non-settling participant to seek funding and/or an alternative settling bank, or would allow a settling participant to secure funds. If, despite extra time, a participant is still unable to settle, the day's transactions of that participant (both payments and receipts) can be deleted from the settlement.[5] This means that the aggregate net debit or credit position of each remaining participant will change by the amount of its bilateral net credit or debit for the day's transactions with the deleted institution. The remaining participants might then settle, if those with increased net debit positions have or can acquire sufficient funds to cover their new, larger settlement obligations. The deleted institution would then become the legal quarry of a host of unhappy customers, other banks and their customers, the clearinghouse, and regulatory authorities.

Box 3 The Benefits of a Net-Settlement System

Suppose 120 banks agree to send and receive payment messages among themselves during the day and to settle only the net amounts due to and due from one another at the end of the day, using deposit balances at Reserve Banks as the medium of settlement.

At the end of the day, each of the 120 banks either owes or is owed by each of the 119 other banks. Thus, there will be 7,140 (half of 120×119) pairs of net credit and debit positions, or bilateral net positions, to be settled. In the absence of a net-settlement system, settlement of these 7,140 positions would involve a payment by each of the debit position banks to each of the corresponding credit position banks.

With a net-settlement agreement, settlement requires only 120 payments, some into and some out of the settlement account at a Reserve Bank. Those owing more than they are owed (at least one bank) pay that "net-net" amount to the settlement account; those that are owed more than they owe are paid that "net-net" amount from the settlement account. When settlement is complete, the settlement account is exhausted because the sum of all banks' positions is precisely zero: what each bank owed was what another bank was owed.

The possibility of deleting payments helps to clarify the difference between provisional and final payment. Payments made via interbank net settlement networks for large-dollar payments are provisional until settlement is complete. This comes at the end of the day, when funding of the settlement account by net debit position banks is complete and payments have been sent from the settlement account to the banks in net credit positions. Between the morning opening of a network and evening settlement, payments are provisional in the sense that the receiving bank may not actually receive credit to its deposit account at a Reserve Bank. Those who use the proceeds of such payments without any other guarantee are exposed to the risk that the payment will not become final because of settlement failure.

Settlement risk is more narrowly focused in CashWire and CHESS. All participants in these systems are settling banks using their own accounts at Reserve Banks to pay to or receive from the settlement account. Also, at least as an interim matter, all CashWire participants guarantee irrevocable availability of funds to their receiving customers. Settlement risk is thus absorbed by the participating banks in CashWire, while it is shared with the customers of participating banks in CHIPS.

II. The Nature of Risk Exposure

Bilateral Risk Exposure

Identifying how losses might be realized in large-dollar transfer systems is not the same as evaluating the extent of risk exposure of any individual participant in those systems. Bilateral settlement risk exposure may be thought of as the expected value of the cost to a participant of a settlement failure by a par-

6. *The discussion here proceeds as though credit is being extended by the receiving bank to the paying bank, rather than to the customers of the receiving bank. This initial presumption is relaxed in the next section.*

ticular counterparty bank. Aggregate settlement risk exposure of a bank is the sum of the expected values of the cost of settlement failure evaluated across all counterparties in a large-dollar transfer system. Conceptually, then, a bank would measure its bilateral settlement risk exposure as the expected value of the unrecoverable portion of its bilateral settlement position (net credit) with respect to the i^{th} other bank. This risk exposure of one bank to another, ignoring any interdependence among all banks, has two ingredients.[6] One is the actual dollar value of the bilateral net credit extended by B to the i^{th} bank in a particular payment system during a day; the other is the expected value of the percentage of this position that will not be recovered.

Consider first the dollar value of the bilateral net credit extended by B to i. This amount will vary each day and during the day with the ebb and flow of payments to and from one another, and can be represented symbolically as $\$_{Bi}$.

Next, consider the unrecovered percentage of the bilateral net credit extended by one bank to another. That a bank is unable to settle at the end of a day would impose a cost on its net creditors in the form of interest foregone on the unsettled position for the period of time the position remained unsettled. Of course, this cost might be recovered, in effect writing an *ex post* loan and then being repaid, but costs of negotiating and litigation would remain. Alternatively, if the bank had failed, there might be some delay before its obligations were settled, involving both administrative and waiting costs. Finally, the unsettled position might turn out to be unrecoverable in part, in total, or (including litigation and waiting) more than total.

All of this potential write-down is what is imagined to underlie each participant's estimate or expectation of the percentage unrecovered of each net bilateral credit position linking it to another bank. Further, it seems plausible that these estimates are not made with certainty, but are better imagined as being distributed across a potentially wide spectrum of values, ranging from just above zero (complete recovery) to more than 100 percent (no recovery plus costs of waiting and litigation). The probability of any particular percentage, j, actually being realized from the i^{th} bank could then be stated as P_{ij}. A whole spectrum of such probabilities of percentage unrecovered would exist, with their distribution tightly clustered near zero for counterparty banks of good repute and long relationship, and less tightly clustered, more evenly spread above zero for less reputable, less well-known counterparties.

Combining the bilateral net credit position and the probability of percentage unrecovered, settlement-risk exposure of bank B to the i^{th} counterparty bank can be represented by the weighted average of all possible outcomes,

$$E(X^{Bi}) = \sum_{j>0}^{>1} P_{ij}\$_{Bi}.$$

Lest this characterization of risk exposure seem too remote from the real world, recognize that something like this risk-exposure calculation must be employed by any bank that makes loans. The expected return on a loan must incorporate both the explicit yield and the likelihood of costs on nonrepayment. In the large-dollar transfer system case, there is credit extended (the net bilateral credit position, $\$_{Bi}$) but no explicit yield because the credit is to be repaid before the end of the day and, with few exceptions, interest is charged only on loans with maturities of overnight or longer. For this reason the net bilateral credit position is often referred to as daylight credit extended by B to i, or as i's daylight overdraft with B.

Sharing

Risk exposure of B to i is affected by something in addition to the bilateral credit position and the creditworthiness of i as perceived by B. This additional matter involves the degree to which others share B's exposure to i. The more widely shared an exposure is, the better for B in two distinct ways. The first would be the extent to which funds, $\$_{Bi}$, are at risk. As in a case in which B is receiving payments from i for the accounts of a number of customers, B's exposure is reduced to the extent that its customers might both share litigation and collection costs as well as bear the burden of amounts that are ultimately unrecoverable. In a more limited way, some cost sharing might be possible among a number of banks all of whom were in net credit position when i failed to settle.

The second way would skew the spectrum of probabilities of percentage unrecovered toward zero. More widespread sharing of exposure may benefit B in that more numerous sources of independent scrutiny of i's creditworthiness may reduce the chances of i's pursuing imprudent courses of action. This is no more than the fooling-all-of-the-people-all-of-the-time impossibility theorem at work. In this application, the theorem reflects the increasing likelihood that (true) adverse information will become available to counterparties as the number of counterparties increases.

The power of information in constraining behavior and reducing risk exposure depends in part on who gets the information. A single customer of a bank, receiving adverse information about another bank from which payments might be received, can simply refuse to do business with customers of the suspect bank or request payment in some safer means. While this may be a prudent course of action for the payee, it has only limited power to force constructive change in the suspect bank, which may be able to ignore the qualms of one or a few isolated alarmists among its many customers, only some of whom may switch to a more sound bank. Less likely is a bank to be able to ignore the qualms of one or a few counterparty banks representing the interests of many customers. Banks are in a better position to profit from, and therefore would likely be more aggressive in seeking, information about the creditworthiness of counterparty banks. Similarly, the power of information to constrain risky behavior is likely to be greater when in the hands of a bank than in the hands of a customer. It is the credit judgments of other banks that provide the foundation of a bank's liquidity, determining its ability to obtain short-term funding in the interbank money market. Counterparty banks would be more likely to maintain a truly less risky posture, including smaller aggregate daylight overdrafts, when other banks have continuing daily incentive to discover their creditworthiness.

Systemic Risk

All of the foregoing discussion of risk exposure implicitly has assumed that a settlement failure is an isolated event; that losses because of a settlement failure would be small enough that banks could absorb the resulting unexpected loss of funds without themselves being unable to settle. Upon investigation, this is an overly strong assumption that ignores the interdependence of participants in a payment system and the resulting vulnerability of many banks directly or indirectly to a single counterparty's failure to settle.

Vulnerability to counterparty failure reflects both the speed with which a bank must gain access to cash and the likely scarcity of cash. An unexpected settlement failure would occur at the end of a day when most banks have finished matching their sources with uses of cash through purchases and sales of funds in the money market. Unexpected non-repayment of a daylight credit then removes one expected source of cash, which must be replaced before the close of Fedwire prevents further transfers of cash that day. Even though the Fedwire facility might be kept open beyond the normal 6 pm close if there were trouble, emergency financing arrangements would have to be made in an hour or so, or before the next morning.

Speedy access to cash might be sought in a bank's own balance at a Reserve Bank. This would require no action, because the cash already would be on deposit and, for reserve requirement purposes, the unexpected drain might be offset by larger cash holdings over future days. However, for many of the nation's largest money center banks, the size of daylight overdrafts is many times larger than their own or any counterparty bank's normal overnight reserve deposit position. For this reason, cash balances would be inadequate in preventing vulnerability to a counterparty failure. If insufficient cash is available in the bank's own account, then steps might be taken to sell liquid assets. However, many banks would have difficulty doing this because liquid assets would already have been serving as collateral for other purposes and therefore be unavailable. Their alternative would then be to borrow. Whether borrowing or, in exceptional cases, selling unpledged liquid assets, the challenge is both to find other banks with the required amount of cash as well as to convince them to make the cash available. Which of these hurdles—finding cash or acquiring it—would be the more troublesome is a debatable matter, for neither would be easy.

Finding banks with cash may be difficult, given the limited stock of cash on which the U.S. financial system operates. The relevant concept of cash is, again, in the form of deposit account balances at Federal Reserve Banks (because they can be transferred via Fedwire before the end of the day). Vault cash is unlikely to help, being of limited amount, cumbersome denomination, dispersed location, and probably sealed in time-locked vaults. Deposit balances at correspondent banks would be useful to a single bank, but their use would merely relocate the cash shortage to the correspondent bank. In this sense, the search for end-of-day cash is one of musical chairs in which the cash needs (players) are no larger than the available stock (chairs), but matching need with available stock will be accompanied by pandemonium, and the pandemonium may last longer than the few hours available to settle. Clearly, there is almost no excess cash in the system to be mobilized on short notice if by excess we mean excess reserve deposits. Excess reserves are widely distributed as "small change" across the entire banking system; if they were available to be mobilized into sellable quantities, the federal funds rate would already have induced their owners to bring them to market before the end of the day.

The total reserve deposits of banks with cash may be available to lend to those without cash to the extent that reduced holdings one day can be offset by enlarged holdings (or reduced requirements) later in a two-week reserve maintenance period or by carry-over of deficiencies into the next period. Further, just as one bank must cover an unrepaid daylight credit in the event of a settlement failure, so, too, other banks will have unexpected cash surpluses because of unsettled payments *to* the failed institution. On balance, the surpluses are less than the deficits only in the amount of the funding gap that triggered the initial bank failure. In general, then, sufficient cash will exist in the aggregate, but its redistribution must be accomplished very quickly, after the time at which most markets have begun to close, and subject to the perceived creditworthiness of deficit banks.

The difficulty arises, therefore, not just in the scarcity of total cash but in the distribution of unexpected cash deficits and surpluses across banks. If a bank with an unexpected deficit does not have established credit lines with a surplus bank or the surplus banks do not have management authorization to sell funds to the deficit bank, then a speedy redistribution of cash would be very difficult—perhaps impossible—by transactions among these banks late in the day in the money market. Intermediary banks, having authority to lend to the deficit bank and to whom the surplus banks will lend, would then be required to complete the process that would prevent a wave of systemic failures. The reliability of these market mechanisms would presumably depend importantly on the degree to which confidence in normal credit judgments about deficit banks could overcome a prudent inclination of surplus banks to seek safety in cash in the midst of waves of market talk occasioned by an unexpected bank failure.

Of course, this is just the setting—a "liquidity crisis"—in which a central bank lender of last resort might play a constructive role. Systemic risk—of a cascade of settlement failures—can be eliminated by isolating the initiating settlement failure, preventing it from forcing unexpected cash deficits on other banks, which in turn could force unexpected deficits on still other banks, and so on. The lender of last resort might isolate the problem in one of several ways. It might lend to a bank that would otherwise initiate a settlement failure, or to second-round banks with unexpected deficits produced by the settlement failure, or to banks that would be willing to lend to the second-round banks. However, the central bank may face constraints on its lending. For example, it may be prevented from making loans to foreign institutions or to banks without adequate collateral or to insolvent banks. Such constraints would influence which set of banks became the focus of last-resort-lending activity in isolating settlement failure and eliminating systemic failures. Beyond these institutional considerations, however, is an important issue to be faced in designing settlement-risk-management policy, involving the degree to which risk exposure is concentrated on banks in the payment system.

Concentration of Risk

There is a fundamental tension between two means of managing settlement risk in a payment system. One would seek to concentrate risk exposure on banks; the other would seek to distribute exposure more broadly over banks, their customers, and the lender of last resort.

Concentrating risk exposure on banks is promoted, as we have seen, when banks guarantee irrevocable availability of funds to their customers. The result should be incentives for banks to monitor carefully the condition of counterparty banks and to manage daylight credit extensions to those counterparties. At the same time, banks would have incentives to monitor their own market reputations and to protect those reputations through prudent banking practices, including management of their use of daylight credit.

Concentrating risk exposure could also be promoted by the lender of last resort through a "hands-off" attitude toward systemic risk. This would require a credible policy of unwillingness to be pressured into eleventh-hour lending even to otherwise sound institutions suffering the repercussions of a settlement failure. Knowledge that the lender of last resort would not intervene should create incentives for sound banks to manage carefully their bilateral net credit extensions across all payment networks as well as their net debit positions across networks. Scrupulous management of these daylight credit extensions would aim at narrowing the size of possible unexpected end-of-day financing needs to levels commensurate with access to cash available in the system to meet them.

This management might involve minimizing daylight credit extensions by careful control of the timing of payments relative to receipts, by lengthening the maturity of bank financing to reduce daily payments and receipts, and by increased holdings of cash to act as a buffer stock.

Distributing, rather than concentrating, risk exposure can be promoted if banks hold customers liable for funds provisionally credited to their accounts but not delivered because of settlement failure. Perhaps more important, distributing risk exposure can be promoted if the lender of last resort follows a policy of lending freely to banks in circumstances that might otherwise threaten systemic settlement failures.

There are important differences between concentrating and distributing risk as a means of managing settlement risks in a payment network, although both may be effective in achieving a low incidence of settlement failures. Concentration of exposure relies primarily on market pressures within the banking industry. Banks will wish to conserve their own liquidity by shunning normal reliance on volatile sources of funds, maintaining a good ability to borrow should unexpected settlement problems arise. This may also serve to economize on their daylight overdrafts by reducing the need to repay overnight funds each day before fresh funds have been received. Liquidity conservation will be encouraged by other banks that will be carefully managing daylight credit extensions and the quality of the banks to which such credit might be extended. Distributing exposure relies more heavily on bank customers and supervisory authorities to promote prudent banking practices, and ultimately on the intervention of the lender of last resort to prevent settlement failures. Clearly, distributing exposure creates a moral hazard, in the sense that a credible commitment by the lender of last resort to prevent systemic settlement failures will make redundant many management and monitoring efforts of banks. This will reduce the liquidity of the banking system and raise chances that imprudent practices will go undetected and uncorrected. In combination, both effects will increase the probability that the lender of last resort will have to intervene.

III. The Current Concern

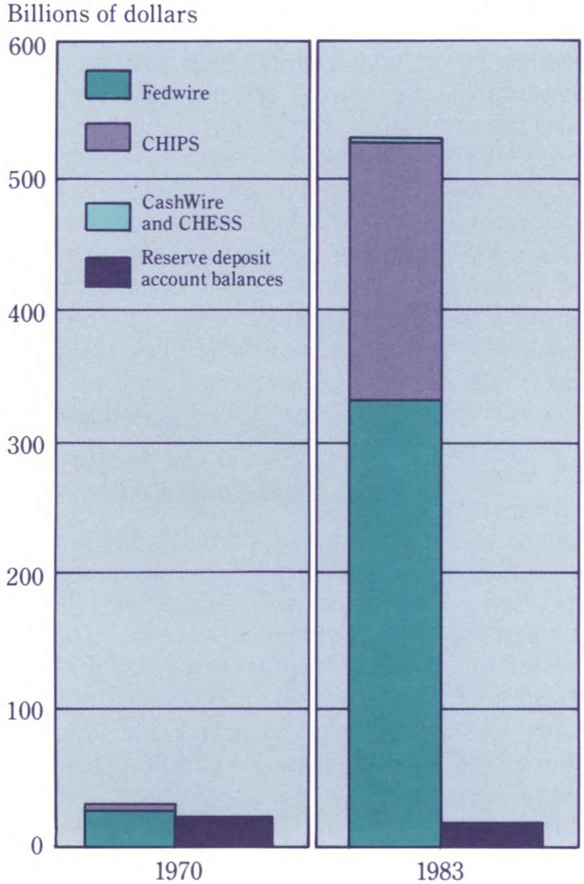
Both banks and the Federal Reserve are concerned about risk exposure on large-dollar transfer systems. What is not clear is whether this reflects a recent discovery of a longstanding and stable exposure, or an awareness of recent growth in that exposure. The evidence presented in what follows argues that it is the latter, the awareness of significant growth of settlement-risk exposure, that underlies the current concern.

Degrees of settlement-risk exposure cannot be measured. Creditworthiness cannot be modeled by an exact science; the expected value of the unrecovered percentage of a bilateral net credit position necessarily lies in the eye and mind of its beholder. Bilateral and aggregate net credit and debit positions on a payment system during a day may be recorded by individual banks, but have only recently begun to be recorded by operators of the payment systems. Therefore, any historical evaluation of exposure must rely on inferences drawn from indirect evidence bearing on creditworthiness, bilateral and aggregate net credit and debit positions of individual banks, and systemic risk.

Individual Bank Risk

The dollar volume of daylight credit extended apparently has been growing rapidly since at least 1970. The dollar volume of large-dollar payments grew at a 24 percent compound annual rate from 1970 through 1983, almost 2½ times the rate of growth of the value of national income and output over the same period. All but 3 percentage points of this volume expansion was in the number of transactions rather than in their average value. This was quite different from national income and output, where inflation-adjusted output grew only 3 percent, while the average value (price) of output grew 7 percent. In itself, growth in the number and value of transactions does not necessarily require growth in daylight credit extended. It is possible that, with increasingly careful control of the timing of payments relative to receipts, banks could have accommodated a larger value of payments with no increase in daylight credit—that is, without using larger bilateral net credit extensions during a day to make payments.

Fig. 1 Large-Dollar Transfers Made or Settled through Reserve Deposit Accounts



SOURCE: Board of Governors of the Federal Reserve System.

Sequencing—delaying payments until receipts arrive—is an alternative to daylight overdrafts in covering payments. It is an approximation of a "realtime" accounting system in which payments could only be made if a sufficient balance were on hand at the instant the payment were ordered. Payments accounting is not typically "realtime," but "batched" overnight. Daylight overdrafts arise if payments are ordered before the receipts come in that will leave a positive cash balance. In principle, the total stock of cash in the banking system could be as little as a single dollar, as long as that dollar were used and reused over 500 billion times during the day. In practice, with a total stock of deposit balances at Federal Reserve Banks of a little more than $20 billion, the average dollar would have to turn over only about 26 times daily to be used and reused in completing over $500 billion of large-dollar transfers. This is up from only 1½ times daily as recently as 1970 (see figure 1). Nonetheless, the feat of turning over the stock of cash 26 times daily would indeed be prodigious if each payment were made against an adequate cash balance in "realtime," so that dollars were actually used and reused 26 times. Sequencing would require that initial payments be made up to the limit of the opening cash balance, but further payments be delayed until sufficient receipts accumulated to fund them. Alternatively, the feat of 26 times per day turnover is not at all prodigious if daylight overdrafts can be used. Payments far in excess of opening cash balances could be made with the expectation of covering them with offsetting receipts later in the day before settlement.

Increasingly powerful computer and telecommunications technology might have made it possible for banks to sequence payments carefully as the volume of transfers grew,

7. *The Federal Reserve Bank of New York maintains an on-line monitor of positions of Edge Act customers and restricts their daylight overdrafts. Starting in 1984, an* ex post *monitor was put in place for all Fedwire users.*

thereby avoiding increased daylight overdrafts. Corporate cash management became a sophisticated science during this period. Indeed, the spread of cash management had as a by-product some of the rapid growth in wire transfer volume. Moreover, computerized telecommunications for wire transfer, integrated with realtime accounting systems, did make it possible to build into banks' systems preset limits on daylight credit extensions to customers. This had a counterpart in the "store and forward" environment of the CHIPS system. Banks could store payment messages in the system until such time as they were prepared to authorize a payment, when the payment would be released and the transaction completed. Beyond this, however, there is little to suggest the application of sophisticated technology to the problem of sequencing inter-bank payments traffic to enable banks or networks to control their net bilateral or aggregate daylight credit or debit positions.

Lack of evidence of attempts to sequence inter-bank wire payments is not surprising. Investing resources in managing the sequencing of payments would be likely only if there were some clear incentive to do so. In fact, incentives to manage daylight credit positions have been weak in the large-dollar transfer systems.

Fedwire, until recently, had no systemwide mechanism for monitoring daylight credit extended to banks.[7] With no restrictions on their daylight overdrafts, efficient management of commercial bank operations had every incentive to use large daylight overdrafts to reduce other costs. For example, large banks routinely finance themselves by buying overnight funds from many other banks and financial institutions (aggregating over $100 billion daily in mid-1984). Just as routinely, the banks can prepare the necessary list of repayment messages at the end of a day for automated telecommunications transmittal on the following day. The first job each morning, before competing uses of people and equipment intervened, would be to activate the system to transmit the string of repayments. The result would be efficient work flow and a routine drawing down of the bank's Federal Reserve deposit account balance at the opening of business each morning. If overnight sources of funds normally exceeded reserve deposit balances (reserve deposits of *all* banks amounted to about $20 billion in mid-1984), then the result would be an immediate daylight overdraft that would last until fresh overnight funding had been bought and received.

The incentive structure in CHIPS has been somewhat different. Settling banks have had an interest in the amount of daylight credit provisionally extended to their respondent banks participating in CHIPS. Indeed, the "store and forward" mode of operation made it possible to set dollar limits on the net credit extended to a customer by a bank, whether the customer were a large nonfinancial corporation making trade payments or a respondent bank paying for money or foreign exchange market purchases. Another indication of settling participants' interest in controlling risk exposure may be found in the change in CHIPS finality from 1:00 pm of the next day, first to 10:00 am of the next day (in 1979) and then to 6:30 pm of the same day (in 1981). Shortening this time gap had the effect of substantially reducing the duration of settlement-risk exposures.

These efforts by settling banks to control their own potential credit-risk exposure to customers might also be viewed as efforts to control settlement risk. Any limitation on credit extended through payments made for a customer is also a limit on daylight indebtedness of the paying bank to other participants in the network. Until recently, however, these incentives were not reflected in any network limit on bilateral credit or debit positions, or in a bank's ability to set a limit on bilateral or aggregate credit extended on the system.

CashWire and CHESS do have both bilateral net credit limits and sender net debit caps, but these cannot be attributed to any incentives operating within the group of participants, because they are an interim requirement of the Federal Reserve for net settlement of the networks. CashWire also requires customer guarantees.

The conclusion that one draws is that the bilateral and aggregate net credit positions generated on large-dollar payment systems have been growing rapidly over the years since at least 1970. Payments volume grew rapidly; the cash position of the banking system using these payment networks did not grow at all; incentives for participating banks to sequence their own transactions to avoid reliance on daylight credit were weak, at best. If risk exposures were not growing, it could only have been because the creditworthiness of participants was improving enough to offset growing daylight credit usage.

The creditworthiness of system participants probably was not improving. In addition to a general weakening in capital positions of the U.S. banking system, two specific developments suggest this conclusion. One has to do with the range of institutions eligible for direct access to Fedwire and the second with the burgeoning role of foreign participants in the daily flow of inter-bank large-dollar payments.

Prior to 1980, access to Fedwire was restricted to member banks, a (declining) subset of depository institutions that included all national banks plus those state banks that chose to become members. With passage of the Depository Institutions Deregulation and Monetary Control Act of 1980, access was broadened to include all commercial banks, thrift institutions, and credit unions eligible for federal deposit insurance. Without arguing that any particular newly eligible institutions were less creditworthy than any member banks, it is still possible to argue that Federal Reserve risk exposure increased. More institutions, supervised by a more diverse set of regulatory authorities, entering new markets and (for some) entering the twilight of viable operation during 1981–82, surely could have reduced the average creditworthiness of the set of institutions eligible to use Fedwire. How much Federal Reserve risk exposure actually increased as a result is impossible to quantify. Actually, it may not have been a substantial increase. Most of the newly eligible institutions did not avail themselves of access to Fedwire or any of the other large-dollar payment systems.

A more important increase in risk exposure was the result of growing international interbank volume. The 1970s first saw U.S. banks follow U.S. multinational corporations abroad and then become immersed in recycling petro dollars through the world banking system. CHIPS dollar volume was growing at a 35 percent annual rate starting in 1970. Between mid-1978, when current data reporting began, and the end of 1983, U.S. banks' own claims on foreigners were growing at a 30 percent annual rate; overnight Eurodollar deposit holdings were growing at a 50 percent annual rate starting in 1977. The burgeoning daily flow of payments through CHIPS involving New York-based subsidiaries of foreign banks was a likely source of increased risk exposure. Foreign-owned institutions handling the transactions of foreign-based banking establishments subject to unfamiliar legal and regulatory systems necessarily injected uncertainty and new risks into banking relationships.

Systemic Risk

The conclusion to be drawn must be that, both on account of growing daylight credit extensions and on account of the widening set of banking institutions whose creditworthiness is relevant, individual risk exposures of participants in large-dollar transfer systems has been growing. Systemic risk exposure is not necessarily driven by these individual exposures, however. Mechanisms to isolate the impact of a settlement failure might have been strengthened or weakened, offsetting or augmenting the effects of growing daylight credit extensions and the changing creditworthiness of individual banks. Nonetheless, it would appear that systemic risk has also been growing in recent years.

Systemic risk is absent from Fedwire because of the finality of Fedwire payments. Since 1970, the percentage of the dollar volume of large-dollar-value payments made on Fedwire has declined from 88 percent to 54 percent, suggesting substantial growth of systemic risk exposure. Moreover, the shift from next-day to same-day CHIPS settlement completed in 1981 might be interpreted as a further increase in systemic exposure in one important sense. Participants did reduce the duration of their bilateral exposures to settlement risk, and the shift to same-day settlement was designed for that purpose. However, systemic risk exposure may have been increased because of the shortened time available during which banks, in the event of a settlement failure, could develop alternative financing arrangements. This would be the case if a participant's ultimate inability to settle were known or strongly suspected before normal settlement hour, so that the period between the end of a day's activity and settlement could be used by otherwise sound banks to find alternative funds in the cash markets.

Whether individual bank participants would have recognized the potentially offsetting increase in systemic risk exposure when their individual exposures were reduced cannot be determined. Their own systemic risk exposure would depend as well on the extent to which Federal Reserve Banks' lending could be expected to isolate settlement failure and control systemic risk.

That systemic risk has increased seems clear. Whether the exposure is that of private financial institutions and their customers or that of the Federal Reserve Banks is not so clear. The distribution could only be determined by experience with failures, which does not exist and no one wants, or by a binding commitment by the Federal Reserve about how it would act in the event of a failure, which is fraught with difficulty.

A lender of last resort could state its commitment not to lend in the event of a settlement failure, thereby attempting to concentrate systemic risk exposure squarely on private institutions. That such a commitment could be credible is itself incredible, given the panic-avoiding objective of a central bank lender of last resort. Ambiguity might be the most that could be gained from such an approach, creating a minimal sense of systemic risk exposure in private institution managements and some minimal risk-controlling behavior. On the other hand, a stated commitment to isolate all institutions from the impacts of a settlement failure, while more credible than a pledge of inaction, carries with it the moral hazard that private institutions will have weak incentives to consider systemic risk in their control practices.

Recent Proposals

Silence about settlement-risk exposure was a reasonable Federal Reserve policy when cash balances equaled or exceeded all of a day's payments on Fedwire, when Fedwire was the dominant large-dollar transfer system, and when supervisory oversight of member banks provided a monitor of the creditworthiness of all participants in Fedwire. Recently, however, risk exposure has grown rapidly, both individual and systemic. Uniform procedures for examination have been adopted by the Federal Financial Institution Examination Council that seek to assure, among other things, that financial institution managers are aware of settlement-risk exposure. Beyond that, the Federal Reserve has sought comment on three principal risk-management devices (receiver guarantees, bilateral net credit limits, and sender net debit caps) that might be used singly or in combination by network participants, by network managements, and/or by regulators seeking to assure prudent settlement-risk management.

The Federal Reserve clearly has a large stake in this. Its own bilateral credit risk has been growing rapidly, unshared with its depositor banks because of Fedwire finality of payment. Its systemic risk exposure has also grown rapidly because its lender-of-last-resort function implies an obligation to isolate the impacts of a settlement failure. A prudent central bank, no less than a prudent private bank, must manage its risk exposure.

IV. Conclusion

Settlement risk is real, although actual settlement failures have been nonexistent. Records of bilateral daylight credit positions are beginning to be kept, but evaluating risk exposures will always be judgmental. This is because the creditworthiness of "borrowers" is a subjective judgment and because the incidence of systemic risk depends on the reaction of the lender of last resort to an actual settlement failure. It seems clear, nonetheless, that both individual and systemic settlement risk exposure have increased rapidly in recent years. This has prompted the current interest of bankers and the Federal Reserve in assuring prudent management of settlement-risk exposure by private financial institutions and their customers and by the Federal Reserve.

Two matters seem central to prudent management of settlement-risk exposures. One is the ability of participants to maintain limits on their own bilateral net credit and aggregate debit positions. This involves both information systems that track those positions and control over their size. Improved position-monitoring capabilities are being developed by the respective large-dollar payment systems that may provide the necessary information; proposals are being considered that would result in debit and credit limits, either self-imposed or derived from uniform rules.

The other matter is the recognition of settlement-risk exposure. Supervisory examination may assure that banks have settlement-risk-management procedures in place, but risk unperceived will go unmanaged. Ambiguity about who is at risk in making large-dollar payments clouds this perception. The Federal Reserve is at risk in extending daylight credit on Fedwire and must manage its exposure. In the absence of receiver guarantees, banks and their customers are both at risk in making payments on net settlement systems. Whether either party, but especially customers, clearly perceives the full extent of its exposure is hard to determine. Systemic risk exposure cannot be unambiguously located, but must be managed by the Federal Reserve jointly to protect the resilience of the banking system to an unexpected settlement failure and the discipline of a grudging lender of last resort.

Economic advisor Roger H. Hinderliter researches capital investment and business finance for the Federal Reserve Bank of Cleveland.

Sources of Change in Rates of Return on Capital: 1952-82

by Roger H. Hinderliter

Business earnings are the reward to capital from current production and sale of goods and services. As such, they reflect the performance of existing capital in an uncertain economic environment. Business earnings also are a major influence on investment decisions of firms and on the willingness of individuals and institutions to help finance investment by purchasing debt or equity issued by firms. Funds raised by businesses in capital markets are invested in new capital assets in expectation of future earnings from the assets' productive use. If past expectations were unfulfilled in current earnings, further investment would become less attractive, and, in time, output growth, creation of new jobs, and living standards might suffer. Thus, while measuring the performance of existing capital assets, business earnings also might be an important indicator of the nation's future economic health.

Beginning in the late 1960s, the economy of the United States was beset by a number of related problems threatening future prosperity. These problems included falling labor productivity growth, accelerating inflation, sharp increases in energy costs, and an apparent decline in relative business earnings. Earnings are measured in relative terms as the rate of return on capital or capital's share of nominal output.[1] In studying the performance of corporate business, economists generally have concluded that relative earnings declined sharply in the late 1960s and, if not falling further, at least remained low during the 1970s. Declining relative earnings are not unambiguously a signal of weakness and disappointed expectations. They could reflect, for example, the disappearance in the 1960s of capital scarcity accumulated during the Depression and war years. Although economists agree that the rate of return and the capital share fell, there is little agreement over the sources and implications of the decline.

1. Business earnings may be measured relatively with respect to the value of total output (capital's share of the proceeds from production and sales); the replacement value (current cost) of the capital stock (the rate of return on capital); or net worth (the rate of return on equity). The rate of return on equity is important only if the division of earnings between equity holders and debt holders is an issue. Thus, while the rate of return on equity may be a critical measure for evaluating the performance of an individual firm, it is less useful in evaluating the overall performance of capital assets. The other two ratios are closely linked, and most efforts to track the earnings problem have concentrated on them.

2. This brief review of past studies highlights fundamental issues. A more thorough evaluation of the literature is provided by Scanlon (1981).

In an early attempt to track the earnings performance of corporations, Okun and Perry (1970) attributed the initial decline in the capital share, between 1965 and 1969, to a rise in labor cost relative to output price and to poor labor productivity growth. These forces were viewed from the perspective of 1970 as transitory. Somewhat later, Nordhaus (1974) concluded that a longer-term, more permanent decline in the capital share had occurred. He suggested the decline reflected the effects of higher levels of investment achieved after World War II. This investment was made possible by a lower cost of capital associated with declining corporate tax rates and a falling risk premium on corporate equity as the Depression became a more distant memory.

Nordhaus' conclusion was supported by additional research results. Kopcke (1978) estimated a "normal" rate of return from a productivity model based on a Cobb-Douglas production function and also found that a longer-term decline was indicated. Lovell (1978) examined a number of relative earnings concepts and, after accounting for cyclical fluctuations and productivity growth, concluded that a downward trend characterized earnings patterns. Other studies, however, produced contradictory results. Working with gross and net (of depreciation) rates of return on capital, Feldstein and Summers (1977) found lower earnings in the 1970s probably resulted from a variety of factors—wage-price controls, rising inflation, and larger fluctuations in capacity utilization—none of which would confirm a long-term downward trend. Runyon (1979) compared the approach of Nordhaus with that of Feldstein and Summers and also concluded that a secular decline in relative business earnings was not supported.[2]

In short, the decline of relative business earnings may have been the result of temporary reversible disturbances without lasting effect on investment, a sustainable downward trend possibly constraining investment over a longer period, or a complex mixture of short-run and long-run factors with equally mixed implications for investment. The strong recovery in investment that began in 1983 is an encouraging sign of an improved investment environment. Many industries, even manufacturing industries widely regarded as inefficient high-cost producers, now are making or planning large investments in technologically advanced capital assets. A prolonged and successful investment program depends heavily on whether business earnings justify the effort over a longer period of time. Thus, any judgment on the promise of the current recovery depends on unraveling the mysteries surrounding earnings patterns in the past. This study approaches that task in five stages.

A method of evaluating earnings performance is proposed in section I. The rate of return on capital and capital's share of nominal output are joined together and the resulting expression defining the rate of return is decomposed into a *margin* effect (which captures interactions among the output price, the replacement price of capital, and unit costs) and a *multiplier* effect (which captures interactions between capacity utilization and capital productivity).

In section II, estimates of before- and after-tax net rates of return and their margin and multiplier components are presented for total business (corporate and noncorporate businesses). One benefit of a total business framework is that GNP, the normal unit of account in discussions of economic performance, becomes the relevant measure of output in the calculations. Thus, the estimates can be related to other issues—the productivity slowdown, economic growth potential, and the transmission of policy actions, for example—that have an economy-wide focus. Moreover, while the noncorporate sector has always been much smaller than the corporate sector, it still is a significant force in creating jobs and controlling capital. Because most studies of business earnings have focused on the larger corporate sector, it is unclear to what extent noncorporate businesses shared in any earnings difficulties that emerged in the postwar period. Noncorporate business has undergone a striking metamorphosis over these years. It has been transformed from a largely agrarian sector to a largely service-producing sector, and these structural dynamics are one element affecting rates of return.[3]

The before- and after-tax rates of return are re-estimated in section III under the counterfactual hypothesis of a constant relative price of capital. Eisner (1980) argued that relative price changes during inflationary periods create capital gains for owners of existing capital assets, and these are appropriately treated as income. Rates of return measured with respect to replacement value may understate earnings performance, because the relative price of capital does change, especially when land is included in the capital stock, and gains typically are not included in income.[4]

A model for evaluating changes in rates of return that attempts to identify and quantify specific sources of change is proposed in section IV. Apart from relative price increases and cyclical variation in capital utilization, slower capital productivity growth and a one-time distortion of the relationship between output price and unit costs in the late 1960s, as first noted by Okun and Perry, contributed to the fall in observed rates of return. The extent to which these forces represent a permanent erosion of business earnings is considered in section V. The emphasis in this study is not one of providing better numerical estimates of rates of return. Although there are interesting differences between the estimates developed here covering total business and prior estimates of corporate rates of return, judgments about the significance of these differences are hazardous and necessarily speculative. The overall pattern of movements in rates of return found here is quite similar to evidence from past studies. The major departure of this study is the attempt to identify and quantify the sources of change in this pattern of behavior.

I. The Rate of Return on Capital

The rate of return on capital is a measure of relative business earnings derived from capital budgeting theory. In principle, it compares the discounted present value of the expected future earnings stream to the current cost of capital assets. Thus, the rate of return is a useful guide in choosing among investment projects. Those projects that offer the highest rates of return are implemented, and those projects with relatively low rates of return are discarded.

The rate of return also is useful as an indicator of the performance of existing capital. As such, it represents an average rather than a marginal rate and is defined as current nominal business earnings divided by the replacement value of the capital stock. This measure can be defined before or after taxes, and on a net or gross basis depending on whether a fully depreciated or an undepreciated capital stock is entered in the denominator. The choice between net and gross capital stocks is governed by the way capital affects output. If capital becomes less efficient in production over time, if it gradually wears out year-by-year, the net concept is appropriate and, correspondingly, earnings are measured net of depreciation or capital consumption.

Measurement of the average rate of return on capital rests on a capital maintenance theory of income. Nominal earnings in the numerator are the residual gain from the current year's production and sale of goods and services, after deducting labor and other costs and providing for the maintenance of capital assets at their replacement value. Nominal earnings include profits and rents (with appropriate capital consumption and inventory valuation adjustments) of the corporate and noncorporate business sectors. Noncorporate profits are estimated from total noncorporate earnings by assuming the owners of these businesses are paid the average wage in their industry after correcting for the effects of cyclical and secular changes in labor market

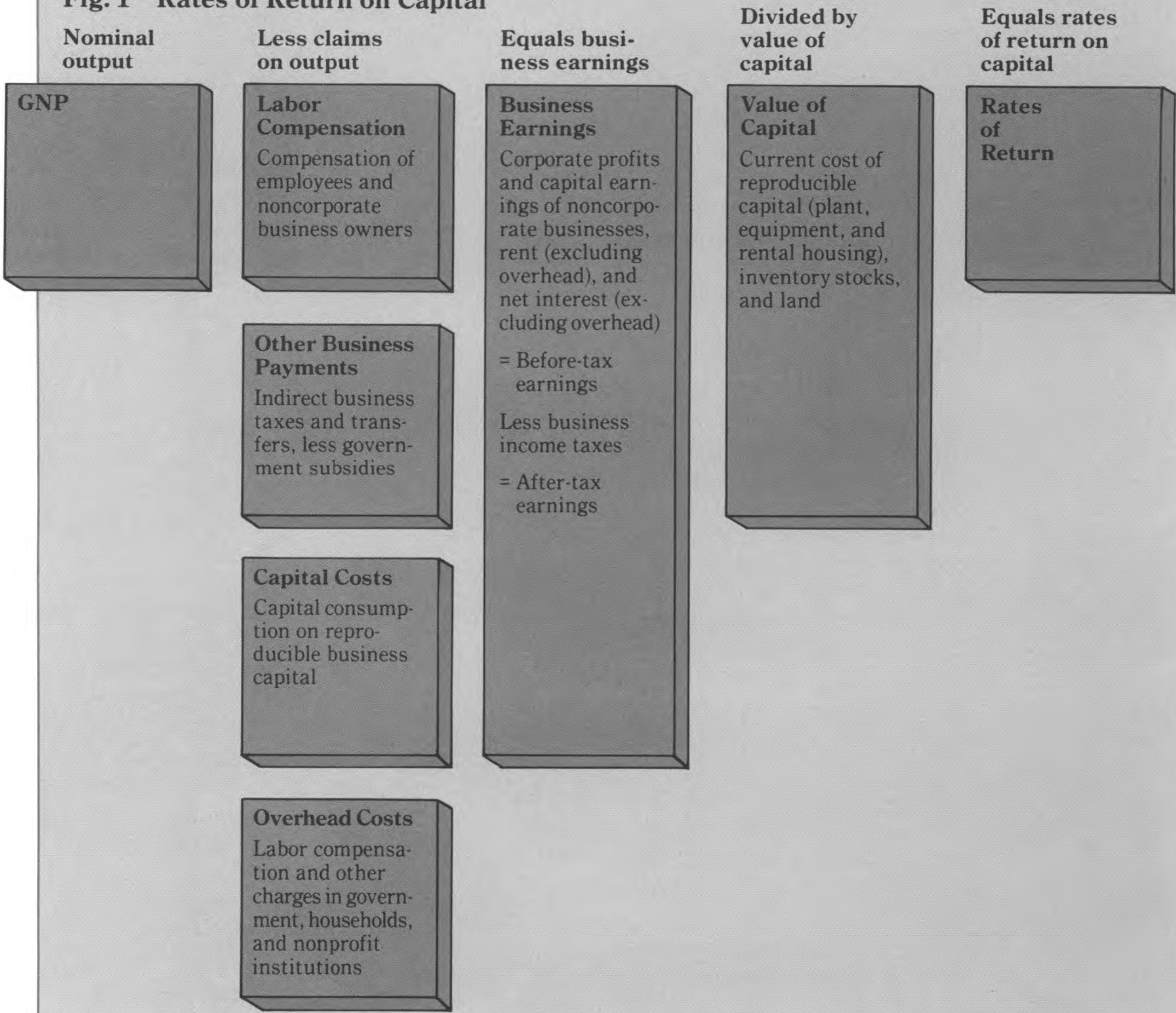
3. Feldstein and Summers (1977) speculated that countervailing differences in tax treatment and risk made the noncorporate rate of return indeterminate relative to the corporate rate in theory. In one empirical study that included noncorporate business (the nonfarm component), Bosworth (1982) resolved the indeterminacy in favor of lower noncorporate rates of return (increasingly so in the 1970s). However, Bosworth's estimates of the capital component of total noncorporate earnings (capital plus labor) seem to have placed a disproportionate share of cyclical and secular weakness in the 1970s on capital earnings (see below, appendix 1).

4. Holland and Myers (1979) include capital gains (or losses) in estimates of after-tax rates of return. They find that individual years can be greatly affected, but over a longer period the average rate of return is little changed by the capital gains calculation. This is to be expected if relative price movements are such that gains and losses roughly balance over time. The estimates by Holland and Myers excluded land, which severely limits capital-gains potential.

conditions. The correction (detailed in appendix 1) is designed to avoid placing the entire burden of cyclical and secular weakness in the economy on capital earnings. In addition to rents that are embedded in profit figures, rental income exclusive of rents on owner-occupied nonfarm housing is included in the numerator of the rate-of-return ratio. Because the division of gains between equity and debt holders is not an issue in rate-of-return calculations, net interest payments are added to nominal earnings.[5]

The nominal earnings stream—profits plus interest plus rent—is a broad-based measure of the returns to business capital. The denominator of the rate of return ratio accordingly is defined to include the broad set of capital assets generating this stream. The relevant

Fig. 1 Rates of Return on Capital



5. See, for example, Walton (1981). Lovell (1978) suggests that including interest also may help correct for the financial distortions of inflation, which otherwise tend to understate earnings by raising interest charges without a corresponding reduction in the real value of outstanding debt.

capital stock is the sum of business inventories, plant and equipment, tenant-occupied residential housing, and land used for business purposes. Among these components, only land presents difficult measurement problems. In past studies of business earnings, a variety of simplifying assumptions have been employed to deal with these problems, among them ignoring land altogether. Here, the basic land series for rate of return calculations are the market value estimates from the flow-of-funds balance sheets. Real value of land stocks and the corresponding price indexes, which are needed for margin and multiplier calculations, are estimated from the sketchy data available on agricultural land, land transition ratios, and land ownership (see appendix 2). The resulting estimates suggest a nearly constant real value of business land, produced by a reduction of farm land at about 0.5 percent per year and an increase in nonfarm business land at about the same rate between 1952 and 1982. Land prices increased rapidly, especially in the 1970s.

A schematic view of the derivation of rates of return is shown in figure 1. Starting from nominal GNP, labor and other superior claims are deducted, leaving business earnings as the residual claim. The ratio of this earnings stream (before or after taxes) to the value of the capital stock determines rates of return. Formally, the definition of the before-tax rate of return can be stated as follows:

$$(1) \quad R_t = \frac{E_t}{V_t} = \frac{CPFT_t + NCPFT_t + NINT_t + RENT_t}{KPR_t(INV_t + FIX_t + LAND_t)},$$

where

E = nominal business earnings—sum of corporate profits, noncorporate profits, net interest, and rent,

V = replacement value of capital—(weighted average) capital price index times sum of real values of inventories, fixed reproducible capital, and land.

The alternative measure of relative business earnings examined in the studies cited above is defined as the earnings stream divided by nominal output or GNP. This is the share of nominal output generated during the year claimed by the business owners of capital. It is related to the rate of return on capital in the following manner:

$$(2) \quad \frac{E_t}{V_t} = \frac{E_t}{GNP_t} \times \frac{GNP_t}{V_t}.$$

The rate of return is equal to the capital share times the ratio of nominal output to the value of the capital stock. If the components of equation (2) are rearranged slightly, the rate of return becomes:

$$(3) \quad \frac{E_t}{V_t} = \frac{E_t}{Q_t} \times \frac{Q_t}{V_t} = \left(\frac{QPR_t - UCS_t}{KPR_t}\right) \times (QQ_t \times QK_t),$$

where

QPR, KPR = price indexes of output (real GNP) and real capital stock, respectively,

UCS = total unit cost of producing output,

QQ = ratio of actual output to potential output,

QK = ratio of potential output to real capital stock.

The first component on the right-hand side of the final expression in equation (3) is the *earnings margin*. This margin includes nominal variables relating the spread between output price and unit costs to the replacement price of capital. Thus, the price of capital relative to the price of output is an important determinant of the margin effect on the rate of return. Unit costs are divided into four groups (as identified in figure 1) for before-tax estimates of the rate of return. These are labor compensation, other business payments, capital costs, and overhead costs. Labor compensation is the payment for labor services in the corporate and noncorporate business sectors. Other business payments include indirect business taxes and business transfer payments. Capital costs are the depreciation charges (capital consumption with capital consumption adjustment) on the fixed reproducible portion of the business capital stock. Finally, overhead costs include compensation paid in government, households, and nonprofit institutions, and rents, interest, and other charges on owner-occupied residential housing and the fixed capital of nonprofit institutions. Government, households, and nonprofit institutions produce no earnings, but they perform functions necessary for the ongoing activities of earnings centers. Government provides the legal system, national defense, and rules and regulations for society's organization and well-being. The household is the fundamental organizational unit of society, performing many functions that support the earnings centers. Nonprofit institutions likewise perform a variety of needed functions definitionally outside the domain of earnings centers. Hence, they are treated as the overhead component of the national economy. Business income taxes, of course, are deducted along with the above costs to calculate the after-tax rate of return.

The second component on the right-hand side of equation (3) is the *earnings multiplier*. The earnings multiplier includes real variables, measuring capacity utilization and capital productivity. These components, which together measure the efficiency of the capital stock in producing output, essentially determine how a given earnings margin will be transformed into a rate of return. The ratio of actual output to potential output (the *QQ* ratio) captures the effects of capacity utilization on the rate of return. High capacity utilization (larger values of the *QQ* ratio) supports a higher rate of return by stretching the margin over a more efficient volume of output. The ratio of potential output to the real capital stock (the *QK* ratio) is a measure of capital productivity. Higher capital productivity boosts earnings by squeezing larger potential (full-capacity) volumes out of a given capital stock.[6]

II. Tracking the Rate of Return

Before-tax (*BXR*) and after-tax (*AXR*) average net rates of return on capital, calculated from the definition of equation (1) for the total business sector, are plotted in figure 2.[7] The estimates cover the period from 1952, when any distortions from World War II and its aftermath might be expected to have dissipated, to 1982, the latest year for which complete and compatible earnings and capital stock data were available. This is a period of rich economic experience. It includes episodes of stable economic growth and episodes of accelerating inflation; structural changes in output and business organization and cyclical swings of major proportions; wage-price controls in the early 1970s and periodic tax changes, the most sweeping of which came in 1981.

In the early part of the period, both before-tax and after-tax estimates were relatively stable. *BXR* declined slightly, and *AXR* changed even less, except for cyclical fluctu-

6. The QK ratio is a hybrid measure in that it relates potential output to actual capital. Perhaps a more appropriate measure of long-term capital productivity would be potential output divided by potential (cyclically adjusted) capital. This would require a separate estimate of potential capital, one consistent with the chosen measure of potential output. Deviation of actual capital from potential capital then would become a third component of the earnings multiplier. No significant cyclical variation in the QK ratio was detected, however, and this refinement was not made.

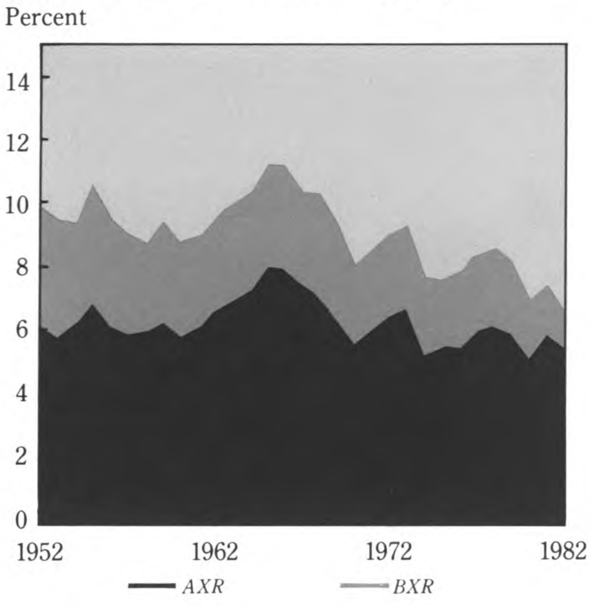
7. The major source of data used in this study is the national income and product accounts. The accounts provide estimates of earnings, prices, costs, output, and inventory stocks. Fixed reproducible capital stocks are from the Commerce Department's tangible wealth estimates. Land stocks and prices are estimated from several underlying data sources, as described in appendix 2. Potential output is from Clark (1983). All data are mid-year values. Capital-stock data reported end-of-year are shifted to mid-year by simple averaging.

8. Kopcke (1978) and other researchers argue that after-tax estimates of the rate of return are more revealing measures of performance. Because taxes are levied on reported earnings rather than economic earnings (with capital consumption and inventory valuation adjustments), effective tax rates increase during a period of accelerating inflation, possibly distorting before-tax estimates relative to their after-tax counterparts. When interest was included with earnings, however, AXR and BXR generally moved together, and AXR fell only marginally faster than BXR between 1965 and 1978.

ations produced by recessions in 1954, 1958, and 1960. In the early 1960s both measures rose sharply, reaching 30-year highs of 11.2 percent (*BXR*) and 8.0 percent (*AXR*) in 1965. After 1965, a sustained erosion in both measures is clearly indicated. Cycle peaks in 1973, 1978, and 1981 were progressively lower than previous peaks. In 1981 *BXR* was 7.5 percent, about two-thirds of its mid-1960s record high, while *AXR* had declined to 5.8 percent. In the recession year of 1982, *BXR* fell to a historic low of 6.5 percent. The after-tax measure recorded its record low (5.0 percent) in 1980, prior to the tax legislation of 1981. Indeed, *AXR* increased relative to *BXR* prior to 1965 and after 1978, especially in 1982 when the effects of the Economic Recovery Tax Act of 1981 became significant.[8]

Comparison of the rates of return shown in figure 2 with the results of other studies is hazardous because of substantial differences in methods of variable estimation. With this caveat in mind, it can be noted that the patterns of change in *BXR* and *AXR* are similar to those found by other researchers. The bulge in the mid-1960s and the decline thereafter are pieces of the earnings puzzle common to virtually all estimates of the rate of return. It also appears that the decline is somewhat less pronounced (at least through the late 1970s) in the estimates presented here. For example, Feldstein and Summer's (1977) before-tax estimates of the net corporate rate of return are higher than *BXR* throughout the 1950s and 1960s, by 2.4 percentage points near the peak (1964–66). This gap diminishes in the 1970s, and, in several years, particularly 1973–75, *BXR* exceeds Feldstein and Summer's corporate rate. Although these comparisons are highly provisional, it does not appear that a general deterioration of noncorporate rates of return relative to corporate rates in the 1970s, noted by Bosworth (1982), is supported here.

Fig. 2 Rates of Return on Capital



Noncorporate business was in sharp transition throughout the postwar period, a transition with roots reaching back into the Depression years of the 1930s. In 1952, 39 percent of all noncorporate entrepreneurs operated agricultural businesses, and 19 percent were in the service and finance industries. By 1982, the composition was reversed. Over the same period, farm earnings (labor plus capital) fell from about one-third to about one-fifth of total noncorporate earnings. Thus, by the 1970s the marginal farm businesses in the noncorporate sector, probably characterized by low rates of return on capital, were substantially reduced in number and influence on earnings, and services had supplanted agriculture as the dominant force in the noncorporate business structure.

Structural transition in the noncorporate sector probably contributed stability to capital earnings in the sector. Certainly, noncorporate businesses performed less well than corporations in the 1950s and 1960s, because their capital earnings did not match the corporate pace. In the early 1950s (1952–54), noncorporate business profits were about 43 percent of corporate profits. By the late 1960s (1967–69), the profit proportion had declined to 24 percent, but it fell only slightly in the 1970s, to 22 percent by the end of the decade (1977–79). More stable profits coupled with traditionally slower growth of capital (at replacement value) would have supported noncorporate rates of return in the 1970s. In 1980–82, however, the noncorporate sector collapsed. Even with considerable allowance being made to insulate capital earnings from slack labor market conditions, the profit proportion fell sharply in these years, to about 9 percent; hence, the extremely low rates of return of this period apparently were concentrated in the noncorporate sector.[9]

Following the definition of equation (3), the rate of return on capital can be decomposed into the margin effect, which accounts for the influence of prices and unit costs, and the multiplier effect, which accounts for the influence of capacity utilization and capital productivity. Before-tax (*BXM*) and after-tax (*AXM*) earnings margins are shown in figure 3. The earnings multiplier (*MULT*), cycling about the productivity ratio (*QK*) as capacity utilization rises and falls, is shown in figure 4.

There can be little doubt that a long-term margin squeeze was a substantial contributor to falling rates of return. Even in 1965, the record year for rates of return, the before-tax margin fell short of previous peaks, and subsequent peaks continued to exhibit deterioration. The after-tax margin reached an all-time high in 1965, but steadily declined thereafter. The sharpest declines in both margins occurred in the late 1960s (1965–70).

Contributions to rates of return from the earnings multiplier can be separated into three periods. The first, from 1952 to 1964, was a period of rapid growth in the multiplier, produced by sharply rising capital productivity. Although relatively low capacity utilization limited the effect on rates of return in many of these years, the rapid growth offset much of the erosion of margins and paved the way for record rates of return in the mid-1960s. In the late 1960s the multiplier plateaued at a high level. Capacity utilization was quite high, but capital productivity was not growing. The multiplier in the 1970s was characterized by irregular growth in capital productivity and large cyclical swings in capacity utilization.

III. Constant Relative Price Estimates

The persistent decline in both before- and after-tax margins, which would virtually foreclose all doubt about trend movements in rates of return were it not for the partially offsetting increases in the mutliplier, raises an important question about rate-of-return measurement. Rates of return are based on replacement value of the capital stock, and earnings margins correspondingly are measured relative to the replacement price of capital. If the capital price increases faster than the output price, such that the relative price of capital increases, should account be taken of the capital gains accruing to the owners of capital assets?

Two types of capital gains can be identified. Nominal capital gains on inventories and fixed reproducible capital are removed from earnings by the inventory valuation and capital consumption adjustments. These adjustments are necessary because otherwise earnings would be distorted by amounts properly reflecting current economic costs of capital maintenance—replacement of inventories, plant, and equipment. During inflationary periods both the inventory valuation adjustment and the capital consumption adjustment lower reported earnings, often by sizable amounts. In 1979–81, for example, the peak of the last inflationary spiral, the two adjustments lowered business earnings by an average of $67 billion a year. This produced a before-tax rate of return that averaged 1.5 percentage points less than would have been estimated without the adjustments.

Price increases need not be and usually are not equally distributed among all price indexes. Although nominal gains are excluded from earnings, many researchers contend that relative price changes create real capital gains, which augment earnings.[10] Thus, if the replace-

9. If no adjustment for labor market conditions had been made (that is, if noncorporate business owners were paid the average wage without qualification), the profit proportion in the late 1970s would have been about 12 percent rather than 22 percent, suggesting a deterioration as noted by Bosworth (1982). Noncorporate profits and rates of return in 1980–82 then would have been negative.

10. See, for example, Eisner (1980) and Cagan and Lipsey (1978).

ment price of capital rises faster than the output price, the relative price change creates a real capital gain for the business owners of capital assets. This gain simply reflects the fact that existing capital has risen in value, and because the owners of capital have use of higher-valued assets (at no additional cost), allowance should be made in earnings. Conversely, a decline in the relative price of capital results in a capital loss and a reduction in earnings. With no accounting for real capital gains, rates of return and earnings margins almost certainly would exhibit a downward trend during periods of prolonged accelerating inflation such as the 1970s. Replacement prices of fixed reproducible capital, and especially land prices, tend to outpace the general price level under such conditions. If relative price movements are the major factor depressing estimated rates of return through the margin effect, it is at least arguable that no meaningful decline occurred.

Fig. 3 Earnings Margins

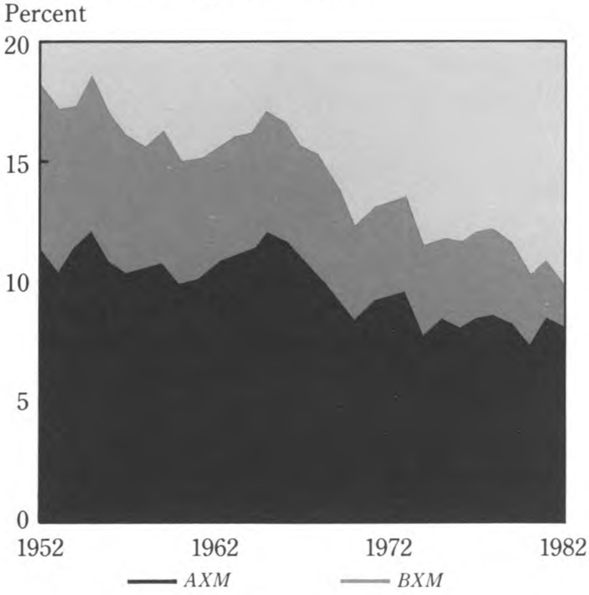
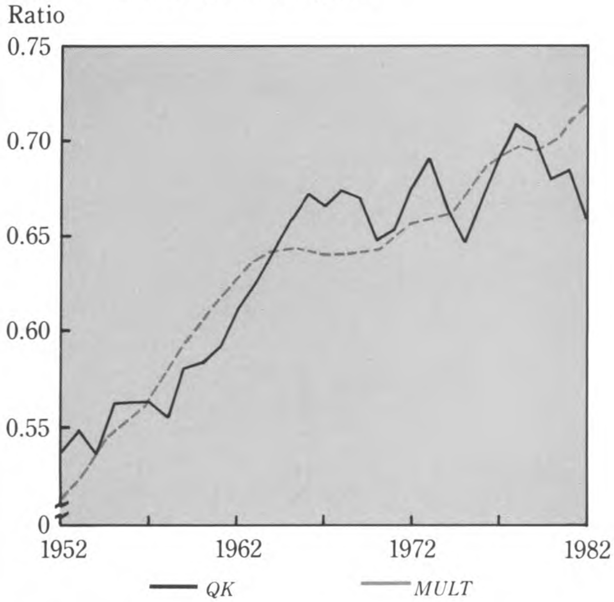


Fig. 4 Earnings Multiplier



A number of difficult problems arise in estimating real capital gains. Usually, no sale of capital assets takes place so a true picture of capital gains is unavailable from market data. Eisner (1980) develops a revaluation approach to estimating real capital gains. The carryover capital stock (that portion of the stock in use from one period to the next) is valued first at actual replacement prices and then at a simulated price that is limited to the increase in the general price level. The difference between the first and second of these formulations is an estimate of real capital gains. Holland and Myers (1979) estimate corporate rates of return that include real capital gains from revaluations of reproducible capital stocks. These gains raise rates of return in the early 1950s (increasing the after-tax rate of return by an average of 2.5 percentage points between 1951 and 1956), but have less effect thereafter. If revaluations were included in the rate of return estimates computed here for the total business sector, both before-tax and after-tax estimates would be higher in the 1970s (through 1979) by an average of 2.5 percentage points.

The larger gains in the 1970s found here, compared with Holland and Myers' estimates (through 1976), reflect the importance of land in computing capital gains. In the 1970s, when gains were accruing to both land and reproducible capital, the gains on land accounted for about 72 percent of the total. However, the movements in the relative price of reproducible capital imply substantially smaller gains in the 1950s, and substantially larger gains in the 1970s, than those estimated by Holland and Myers. Apart from differences in the relevant business sectors, and therefore in the measurement of the variables used to estimate capital gains, differences between output prices used in revaluation and methods of estimating gains on inventory stocks seem to account for these results. Eisner's (1980) net revaluations of reproducible capital were generally larger in the 1970s (through 1977) than those estimated here.

Fig. 5 Constant-Relative-Price Rates of Return on Capital

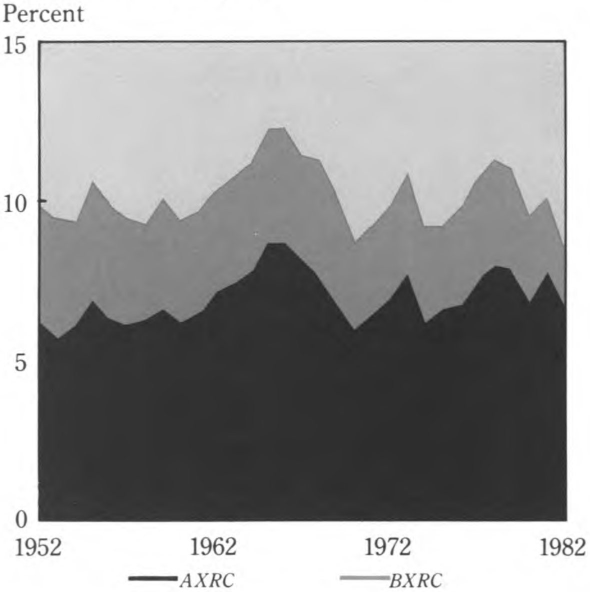


Table 1 Capital Productivity Growth
Average rate in percent per annum

Period	*QK* (potential output per unit of capital)	Potential output	Capital stock
1952–64	1.88 (0.53)	3.61 (0.10)	1.70 (0.47)
1965–70	0.01 (0.29)	3.44 (0.09)	3.43 (0.29)
1971–82	0.95 (0.71)	3.48 (0.10)	2.51 (0.71)

NOTE: Standard deviations are in parentheses.

Scanlon (1980) points out that an alternative approach to revaluation is to assume that capital and output prices move together, thus entirely submerging the capital gains question. This assumption will not do, of course, as long as the relative price of capital clearly is not constant. It may be made operational, as a counterfactual experiment, by limiting capital price increases to increases in the output price, thus creating a synthetic capital price that remains constant relative to the output price. Variation in rates of return and earnings margins that remain must result from forces other than changes in relative prices.

Before-tax (*BXRC*) and after-tax (*AXRC*) rates of return on capital estimated with a constant relative price of capital are shown in figure 5. As expected, constant-relative-price rates of return are higher than the variable price estimates for most of the period 1952–82. The gap widens in the 1970s, indicating that failure to take account of relative price changes and potential capital gains that accompany them is an important force behind the apparent decline in rates of return after 1965. Wider cycles are evident in *BXRC* in the 1970s, but the level is about the same as the 1950s. Indeed, the bulge in rates of return during the 1960s stands out in these estimates. The after-tax measure, *AXRC*, rises in the 1970s despite greater cyclical variations.

IV. Sources of Change in Rates of Return

As outlined above, estimating the pattern of long-term change in the rate of return on capital generally has been approached by searching for statistically significant trends in models that also standardize for cyclical variation. Results from these efforts have been mixed and have revealed little about underlying sources of change. The trend coefficients have had no clear interpretation, and sorting out different short-run trends within a longer period often has been an arbitrary process.

Advancing beyond simple trend/cycle models depends on an ability to assign trend effects in some systematic way to the margin and multiplier components of the rate of return. This, in turn, depends on how output price is determined as a markup over unit costs, which governs long-term change in the margin, and on the capital-output relationships that characterize long-term growth paths. Markup rules and capital-output relationships can be modeled explicitly with considerable rigor.[11] In this study, where isolating the sources of change in rates of return is of paramount interest, an approach retaining the simplicity of trend/cycle models, though enhancing interpretability, is followed.

The behavior of the *QK* ratio shown in figure 4 suggests three different ***regimes*** of capital productivity growth: 1952–64, 1965–70, and 1971–82. These regimes are identified in table 1. It is evident that slower capital productivity growth in the periods 1965–70 and 1971–82 is associated with faster growth in capital stock, the denominator of the productivity ratio, in those periods compared with the 1950s and early 1960s. Potential output grew at about the same rate in all three periods, despite the acceleration in capital-stock growth. Other researchers, using different methodologies, have found broadly similar behavior in the relationship between output and capital. Bosworth (1982) estimated a capital output ratio adjusted for cyclical change, and found it rising in the late 1960s and 1970s (implying the inverse, comparable with the *QK* ratio, fell). Bernstein (1983) noted a flat capital-output ratio after 1965 for nonfinancial corporations, while Baxter (1978) measured a declining output-equipment ratio (without adjustment for the cycle) in manufacturing from 1965 on. The major result—faster capital-stock growth after 1965 did not produce commensurate increases in potential output—is the same, which had to depress rates of return on capital.[12]

If changes in the rate of growth of capital productivity identify periods of different underlying trend growth, rates of return can be estimated from a linear spline model (see table 2).[13] Trend coefficients in the model are interpreted as long-term growth rates, depending on the rate of capital productivity growth in each period, and the margin ***drift*** that characterizes the behavior of earnings margins in each period (see appendix 3). Margin drift accounts for long-term growth in the earnings margin as determined by the price markup rule that best approximates pricing strategy for the aggregate economy. In general, margin drift is expected to be greater than or equal to zero. Thus, estimated trend coefficients are expected to be greater than or equal to observed capital productivity growth in each period.

Estimates of trend coefficients for all rate of return concepts employed here are shown in table 2. The coefficients displayed are maximum likelihood estimates with first- and second-order autocorrelation corrections. The model appears to fit the data veı y well. All individual coefficient estimates are significantly different from zero at the 1 percent level. Estimates of β_4 suggest considerable cyclical sensitivity in rates of return. The elastic response corresponds to the direct effect from changing output volume (unitary elasticity) and the indirect effect of the cycle on prices and costs in the earnings margin (elastic). An increase in the *QQ* ratio from recession levels (95 percent) to boom levels (105 percent) increases rates of return from 30 percent (*AXR*) to 33 percent (*BXRC*).

11. See, for example, Nordhaus (1974) and Kopcke (1978).

12. The causal factors in the capital productivity decline are not known with any precision. Bernstein (1980) suggests "willful" and "inescapable" errors in capital-stock management (associated with increasing attention given to short-run profit performance in the first instance, and energy price increases and pollution abatement requirements in the second) produced ". . . the wrong stock in the wrong place at the wrong time." Baxter (1978) further speculates that falling research and development expenditures and heightened emphasis on replacement investment may have aided the decline. No single factor seems capable of accounting for the timing of the sharp decline in 1965 (prior to energy problems and mandated capital expenditures) and the partial reversal of capital productivity growth after 1970. This question, part of the wider mystery surrounding the behavior of productivity, is left to future research.

13. For a description of spline functions, see Poirier (1976).

14. If margin drift were approximately constant (not necessarily zero), the sum of the trend shift coefficients (β_2, β_3) should not differ significantly from the differential in capital productivity growth rates between 1952-64 and 1971-82 (appropriately signed). Based on restricted regression F-tests, the difference is statistically significant (at 5 percent) in variable-relative-price equations, but not in constant-relative-price equations.

The major points of interest are the trend, coefficients (β_1, β_2, β_3), and their sums, which are estimates of underlying growth in rates of return for the periods 1952-64, 1965-70, and 1971-82. In 1952-64, the estimates of β_1 imply rate-of-return growth ranged from about 1.6 percent a year (*BXR*) to about 3.5 percent a year (*AXRC*). The sharp downward shift (β_2) in the late 1960s produced significantly negative growth rates ($\beta_1 + \beta_2$) in all measures of the rate of return in 1965-70. The shift was reversed (β_3) in the 1970s, but in 1971-82 only constant-relative-price rates of return grew at a pace significantly different from zero ($\beta_1 + \beta_2 + \beta_3$).

The trend coefficients quantify the combined effects of capital productivity growth and margin drift on rates of return. Although the trend shifts were controlled by observed changes in productivity growth, coefficient estimates also reveal the importance of margin drift in determining the course of rates of return. Quite different growth patterns are implied by the coefficients estimated from variable-relative-price and constant-relative-price measures of the rate of return. To disentangle the productivity and drift components, the estimated trend coefficients can be compared with mean values of productivity growth rates (table 1). That is, the trend coefficient estimates can be tested against the hypothesis of zero margin drift.

The rate of growth of *BXR* was significantly lower than productivity growth alone would suggest in each of the three periods, implying negative margin drift over the entire 30 years. Setting aside 1965-70 for the moment, the appearance of negative drift in these estimates reflects only the increasing relative price of capital in both 1952-64 and 1971-82, with a steeper increase, of course, in the latter period. When holding the relative price constant (*BXRC*), estimated growth rates in 1952-64 and 1971-82 were significantly greater than could be accounted for by productivity growth alone, indicating positive margin drift and a pricing strategy that does not exclude rapidly rising unit costs from the markup function. However, the difference between growth rates in 1952-64 and 1971-82 is accounted for by the capital productivity growth differential between the two periods. Thus, margin

Table 2 Maximum Likelihood Estimates of Rate of Return Growth

From the process described in appendix 3, the rate of return on capital over three regimes of capital productivity growth may be expressed as follows:

$$R_t = R_0 \prod_{i=1}^{3} (1 + r_i)^t QQ_t^{\lambda} \epsilon,$$

where

t = time,
r = $r_M + r_{QK}$ (sum of growth rates of earnings margin and capital productivity ratio),
i = index denoting three periods of capital productivity growth.

The parameters of this equation—R_0, r_i, and λ—are estimated from a linear spline function:

$$\ln R_t = \beta_0 + \beta_1 T + \beta_2 TS + \beta_3 TSS + \beta_4 \ln QQ_t + \mu,$$

where

T = linear trend (T = 1, 2, . . . , 31),
TS = shift in trend at 1965 ($TS = T$-12),
TSS = shift in trend at 1971 ($TSS = T$-18),
μ = random error,

and

$\beta_0 = \ln R_0$,
$\beta_1 = \ln(1 + r_1)$,
β_2 = shift in $1 + r$ at 1965,
$\beta_1 + \beta_2 = \ln(1 + r_2)$,
β_3 = shift in $1 + r$ at 1971,
$\beta_1 + \beta_2 + \beta_3 = \ln(1 + r_3)$,
$\beta_4 = \lambda$.

drift in *BXRC* was positive, which is acceptable from the standpoint of pricing strategy, and also was approximately constant from 1952–64 to 1971–82.[14] Similar results are implied by the after-tax trend coefficients. Estimated growth rates of after-tax rates of return were higher than their before-tax counterparts in both 1952–64 and 1971–82, although the differential narrowed slightly in the 1970s. Declining tax costs in the 1950s at least kept *AXR* growing apace of *BXRC*, but this correspondence was broken in the 1970s.

The sharp downward shift and the resulting negative growth rates in all measures of the rate of return in 1965–70 cannot be explained by a combination of changing capital productivity growth and constant margin drift. Although capital productivity growth was essentially zero, it did not become negative, on average, for the period. Yet even constant-relative-price rates of return exhibited substantial negative growth rates, suggesting a restructuring of the spreads between output price and unit costs depressed earnings.

The contribution of price and unit costs to changes in earnings spreads is illustrated in table 3. The average annual percentage change in price is shown in row 1. Share-weighted changes in unit costs and earnings spreads then represent the claims on the price increase. Clearly, the relationships between price and unit costs have not been constant in the postwar period. Price increases averaged 1.93 percent in 1952–64 and accelerated to 7.06 percent in 1971–82. Claims represented by other business payments (*UBC*) and overhead costs (*UOC*) declined, while capital costs (*UKC*) increased. These changes were relatively steady transformations. Unit labor compensation (*ULC*), however, rose sharply from 39 percent of the price increase in 1952–64, to 53.7 percent in 1965–70, before falling back to 47.2 percent of the price increase in 1971–82. Given changes in other unit costs, the reduction in labor's claim on price increases in 1971–82 was sufficient to restore the before-tax spread to the relative position in 1952–64. (About 12.8 percent of the average price increase went into the before-tax earnings spread in each period.) After-tax spreads were relatively lower in 1971–82 because unit-tax-cost reductions of the 1950s were not duplicated in the 1970s. Thus, the late 1960s was a period when accelerating labor costs were not marked up sufficiently to maintain a spread sufficient to cover other costs and provide earnings as well.[15]

Coefficient	Rate of return: *BXR*	*BXRC*	*AXR*	*AXRC*
β_0	−2.4445* (0.0136)	−2.4643* (0.0202)	−2.9279* (0.0235)	−2.9464* (0.0185)
β_1	0.0158* (0.0016)	0.0255* (0.0024)	0.0255* (0.0028)	0.0352* (0.0022)
β_2	−0.0602* (0.0045)	−0.0731* (0.0066)	−0.0735* (0.0078)	−0.0873* (0.0061)
$\beta_1 + \beta_2$	−0.0445* (0.0032)	−0.0480* (0.0047)	−0.0480* (0.0056)	−0.0525* (0.0044)
β_3	0.0433* (0.0050)	0.0685* (0.0073)	0.0540* (0.0087)	0.0810* (0.0068)
$\beta_1 + \beta_2 + \beta_3$	−0.0012 (0.0021)	0.0205* (0.0029)	0.0060 (0.0036)	0.0285* (0.0029)
β_4	2.7749* (0.1917)	2.8292* (0.2539)	2.6137* (0.3245)	2.7753* (0.2573)
$\bar{R}^2$	0.995	0.898	0.986	0.989
ρ^a	−0.586	−0.123	−0.818	−0.636
SE	0.032	0.037	0.050	0.045

NOTE: Standard errors are in parentheses.
* = Significant at 0.01.
a. ρ = Sum of first- and second-order autocorrelation coefficients.

V. Summing Up

Rates of return on capital from the early 1950s through the early 1980s were influenced by four sets of factors: changes in capital productivity growth; changes in the relative price of capital; changes in margin drift associated with disturbances in the price-markup function; and changes in capacity utilization.

15. *The early stages of the labor productivity slowdown accounted for only a fraction of the increase in unit labor compensation relative to output price in 1965–70 (Okun and Perry 1970). Relative increases in the compensation rate, perhaps induced by the long duration of high capacity utilization, accounted for the remainder. Wage-price controls, followed by deep recessions in 1973–75 and 1981–82, could have curtailed relative increases in the compensation rate and partially offset even greater deterioration of labor productivity in 1971–82.*

An idea of the relative importance of these factors can be obtained by asking what a particular rate of return might have been in 1982 had it grown from its initial value along alternative growth paths where the separate effects are controlled. This experiment is illustrated in table 4 for the before-tax rate of return. The first entry projects the rate of return along a driftless and cycle-free path defined by average capital productivity growth in 1952-64 extended through 1982. The second entry allows average productivity growth to change as observed in the three sub-periods considered in the analysis, while maintaining the zero drift and cycle assumptions of the first entry. In the third entry, margin drift is added but the relative price of capital is held constant. This incorporates the legacy of the markup failure in the late 1960s. The fourth entry allows for a variable relative price and thus illustrates the effects of inflation-induced increases in the relative price of capital on the rate of return. Finally, in the fifth entry, the estimate of *BXR* in 1982 picks up the cyclical and random fluctuations of that year.

Table 3 Price, Unit Cost, and Earnings Spreads

Spread component	1952-64 Rate of increase, percent per year	1952-64 Percent of price increase	1965-70 Rate of increase, percent per year	1965-70 Percent of price increase	1971-82 Rate of increase, percent per year	1971-82 Percent of price increase
QPR	1.93	—	3.87	—	7.06	—
ULC	0.75	39.0	2.08	53.7	3.33	47.2
UBC	0.19	9.9	0.35	8.9	0.46	6.5
UKC	0.13	6.5	0.35	9.1	0.80	11.3
UOC	0.61	31.8	1.15	29.6	1.57	22.3
BXSPRD	0.25	12.7	-0.06	-1.5	0.90	12.8
UXC	-0.02	-1.0	0.02	0.4	0.12	1.7
AXSPRD	0.27	13.7	-0.08	-1.9	0.78	11.0

NOTE: Unit costs and earnings spreads are share-weighted average increases, where weights are the ratio of the individual cost or spread component to price.

Table 4 Before-tax Rates of Return in 1982
For alternative growth assumptions

Growth path	Rate of return, percent
1. Defined by average capital productivity growth 1952-64, extended through 1982	15.5
2. Defined by average capital productivity growth 1952-64, 1965-70, and 1971-82	12.4
3. Defined by growth coefficients in column 2, table 2	11.3
4. Defined by growth coefficients in column 1, table 2	8.0
5. Before-tax rate, 1982	6.5

It would be stretching this example beyond its bounds to claim that a 15.5 percent before-tax rate of return on capital was an attainable goal in 1982. Nevertheless, the rates illustrated in table 4 indicate the nature and importance of the changes that distinguished later years from earlier years in the postwar period. The difference between the rate of return in line 1 and line 5 is -9 percentage points, split about equally between margin effects and multiplier effects. The smaller margin and multiplier effects, dealing with the legacy of an already reserved change in price-cost relationships and cyclical fluctuations in capacity utilization, clearly are transitory. Of the larger effects, the margin component reflects a permanent erosion in the rate of return only if the inflationary experience of the 1970s is a permanent feature of the U.S. economic environment. Even under inflationary conditions, if capital gains on the existing capital stock are counted in business earnings, the erosion is offset when these gains are valued on the same terms as cash earnings. In a disinflationary environment differences associated with changes in the relative price of capital are narrowed. Declining capital productivity is more difficult to explain. Because the decline in productivity was associated with an acceleration in capital stock growth after 1965, achieving capital sufficiency or related explanations seem doubtful. The puzzle to solve here is one of determining how the U.S. economy managed to wring so much output growth out of so little capital stock growth in the 1950s and early 1960s.

Appendix 1 Estimating the Capital Component of Noncorporate Earnings

In the national income and product accounts, noncorporate earnings are not divided into labor and capital components. The entry in the accounts for *proprietors' income* is simply the owners' total earnings and therefore includes both a return for their labor services and a return on the capital invested in the businesses. To estimate the labor-capital division of earnings, the noncorporate business sector could be given the same return on capital as corporations, thus estimating labor compensation as a residual. Alternatively, owners of the noncorporate businesses could be paid the average wage prevailing in their industry, thus estimating business earnings on capital as a residual. Here, the second option is the relevant one. If owners of noncorporate businesses are to be paid the average wage, they are compensated partly as managers, partly as skilled workers, and partly as unskilled workers, depending on the characteristics of the work force embodied in each industry's average wage.

To extract the labor component from total noncorporate earnings, some account must be taken of cyclical and secular changes affecting labor markets. In the corporate sector workers are laid off in periods of slack demand to reduce or limit the increase of the wage bill. Wage rates, however, are affected less by falling demand and usually continue to rise during recessions. Controlling labor costs through layoffs distributes the burden of recession to labor income as well as capital income.

Noncorporate businesses have similar control of their hired labor force, but not of their own amalgamated labor services. Owners of noncorporate businesses do not lose their jobs during recessions, except those whose firms fail. Consequently, paying the owners the average wage during recessions would mean that cyclical weakness in the economy falls most heavily on their capital earnings. To avoid this incidence, owners of noncorporate enterprises must take a cut in labor income when economic activity and labor market conditions weaken. On the other hand, they may be paid an "entrepreneurial bonus" during boom periods when labor markets are tight and labor services in short supply.

A second problem encountered in estimating the division of total noncorporate earnings into labor and capital components is associated with secular changes in labor markets. The *natural rate of unemployment*, or the *nonaccelerating inflation rate of unemployment*, has risen since the 1950s, and most of the increase occurred in the late 1960s and 1970s.

One estimate, constructed by Clark (1983, table 5) for modeling potential GNP, shows this underlying unemployment rate rising from 4.5 percent in 1954 to 7 percent in 1978, before easing slightly in the early 1980s. As long-term labor market conditions deteriorated, the number of full-time partners and proprietors in the noncorporate sector (reported in the national income and product accounts) reacted in a curious way. Noncorporate business owners declined from 1952 to 1967 at about 2 percent a year, largely as the result of an exodus from farming that began in the 1930s. Between 1967 and 1972, the number of full-time noncorporate owners was roughly constant, increasing at about 1.5 percent a year thereafter. This change in noncorporate business formation and retention slowed the decline in relative (to employment in private industry) noncorporate ownership from about 3.3 percent a year to about 0.5 percent a year on average beginning in the late 1960s.

One reason why the noncorporate sector turned around as it did may be that increasing long-term weakness in labor markets discouraged people from a career of working for wages. When job availability is diminished on a long-term basis, the noncorporate sector may serve as the employer of last resort. This suggests, however, that labor earnings potential in the noncorporate sector may not be as great as the average wage implies. Certainly, the notion of hidden unemployment in the noncorporate sector is not new, even applied to highly developed economies.[a] If owners of noncorporate enterprises were paid the average wage during periods when long-term labor market conditions are deteriorating, then just as in the case of cyclical fluctuations the burden of limited earnings capacity would fall most heavily on the capital component.

To adjust for the effects of cyclical and secular distortions of the composition of noncorporate earnings, we estimated a simple interactive model of the ratio of full-time owners of noncorporate businesses to full-time employees in private industry. Using maximum likelihood techniques, this employment ratio was regressed on a time trend and measures of long- and short-term labor market conditions:

$$\text{(A.1)}\quad \ln RFTE = \underset{(0.1681)}{-0.0551} - \underset{(0.0059)}{0.0837T} + \underset{(0.0052)}{0.0517LT} - \underset{(0.1618)}{1.2392L} + \underset{(0.0121)}{0.1090S} + e.$$

$\bar{R}_2 = 0.981 \quad \rho = 0.682 \quad SE = 0.014$
(Standard errors are in parentheses.)

where

$RFTE$ = partners and proprietors devoting substantially full time to their businesses divided by full-time equivalent employees in private industry,

L = measure of long-term labor market conditions: natural rate of unemployment divided by minimum natural rate achieved in period 1952–82,

S = measure of short-term labor market conditions: difference between actual and natural rates of unemployment divided by natural rate.

In this model the sum ($-0.0551 - 1.2392L$) defines the intercept for each level of the natural rate of unemployment. The trend component ($-0.0837T$) captures the long-term decline in the full-time equivalents ratio, and the interaction term ($0.0517LT$) retards the trend rate of decline as the natural rate rises. The effects of short-term cyclical fluctuations are captured by ($0.1090S$).

To estimate the labor component of total noncorporate earnings, an adjusted full-

a. See Phillips (1962). It appears that one force behind the turnabout was growing numbers of women and the young entering the noncorporate sector, two groups whose employment opportunities might be more severely curtailed as the natural rate rises (Fain 1980).

employment equivalents ratio, *ARFTE*, was calculated by setting the actual rate of unemployment equal to the natural rate, and both equal to the minimum natural rate. That is, all labor market fluctuations about the minimum natural rate were removed.[b] The labor component of total noncorporate earnings was calculated from *ARFTE* as follows:

$$NCLC = (ARFTE \times EPI \times W)w,$$

where

$NCLC$ = noncorporate labor composition,
EPI = full-time equivalent employees in private industry,
W = average wage, weighted by distribution of partners and proprietors by industry,
w = markup on wages for other labor compensation (pension and profit sharing).

When *ARFTE* exceeds *RFTE*, the owners of noncorporate enterprises receive the entrepreneurial bonus (that is, they earn more than the average wage for their labor services); when *ARFTE* is less than *RFTE*, their wages are cut. In the 1970s, the implied wage cut progressively widened, and by 1982 it amounted to nearly 30 percent of the average wage. Even with this allowance for changes in labor market conditions, capital earnings in the noncorporate sector declined sharply relative to labor earnings in the early 1980s. Capital earnings generally exceeded 30 percent of total noncorporate earnings prior to 1979. This proportion fell to an average of 13 percent between 1980 and 1982, and, in the recession year of 1982, capital earnings were about 3 percent of total noncorporate earnings.

A final problem of estimating the division of total noncorporate earnings into labor and capital components relates to taxes. To calculate the after-tax rate of return, taxes also must be apportioned to labor and capital earnings. The first step here was to estimate the tax liability on total noncorporate earnings and rent. This was done by computing the ratio of total noncorporate earnings plus rent to personal income and multiplying the result by personal taxes paid (federal, state, and local).[c] The resulting tax liability then was distributed to capital (including rents) and labor according to their proportions in the aggregate. The effective tax rate on capital earnings was 12.4 percent in 1952 and was relatively constant until the late 1960s. Thereafter, the effective tax rate rose to about 20 percent in the early 1980s. For years when comparisons can be made, these tax rates are similar to estimates made by other researchers (Kahn 1964).

Appendix 2 Estimates of Land Prices and Real Values

One of the most challenging problems in constructing estimates of the rate of return on capital is to derive adequate measures of land. Land is an integral part of the capital stock, along with inventories and fixed reproducible capital, and business earnings must be evaluated relative to its value as well as the value of the reproducible assets. Nominal land value is the product of market price (which is equal to the replacement price for nonreproducible assets such as land) and the quantity of land in use. The analysis of rates of return developed here requires estimates not only of nominal land value, for computing rates of return, but a separation of this value into price and real value (quantity), for computing the margin and multiplier effects on rates of return.

Detailed statistical information on land prices is restricted to farm land (Goldsmith 1982). Past studies of the rate of return filled the information gap in a variety of ways: by ignoring land altogether (Lovell 1978); by linking nominal land value to the value of structures on the land (Feldstein and Summers 1977, following Goldsmith 1962 and Den-

b. This admittedly is somewhat arbitrary. It suggests that any deviation from the period minimum natural rate, if uncorrected, distorts the labor capital division of noncorporate earnings, but that the minimum determines an appropriate base line division. Several alternatives were explored, including replacing the minimum with a trend value of the natural rate, and making no adjustment at all. In 1982, the estimate of the before-tax rate of return was 6.5 percent. If a trend value of the natural rate had been used, the rate would have been lower, 6.0 percent; if no adjustment had been made, the estimate would have been 5.4 percent.

c. In the tax calculations income measures were converted to taxable income equivalents. The relationship between personal income in the national income and product accounts and taxable income (adjusted gross income) is illustrated by Hinrichs (1975). Noncorporate capital earnings and rent were approximated on an AGI basis by removing the capital consumption and inventory valuation adjustments and all imputations made in the national income accounts.

d. As an example of this calculation, in 1980, 710,390 private single-family housing units were authorized, which also means 710,390 structures were authorized, mostly for owner occupancy. At the same time, 53,768 two-family units were authorized, or 26,884 structures, half of which were presumably for rental. Three or more family units contributed 39,895 structures, virtually all for rental. Thus, there were 777,169 structures authorized, 53,337 of which were for rental or business use, a ratio of 6.9 percent. See U.S. Department of Commerce, Bureau of the Census, Housing Units Authorized by Building Permits and Public Contracts: Annual 1980, *July 1981, table 2, p. 4.*

ison 1974); by using book-value (historical price) data from income-tax records (Holland and Myers 1979); and by combining income-tax data, benchmark estimates from property-tax-assessment records, and price distribution assumptions such as uniform land prices across sectors (Fraumeni and Jorgenson 1980). The few attempts to estimate land prices outside the farm sector (Milgram 1973) have not produced continuing time series beyond the limits of the study.

Market value estimates of farm and nonfarm business land recently have been developed in the Board of Governors (1983) flow-of-funds national balance sheets. These data provide the variables necessary for estimating rates of return and are the starting point for separating values into price and real value components. Acreage in farm production is monitored and reported by the U.S. Department of Agriculture (1983). The data on acreage allow an average dollar price per acre to be computed from the market value estimates of farm land in the flow-of-funds balance sheets. Thus, real value of the farm land (1972 dollars) and a simple price-relative index can be computed. The price index increases from 0.321 in 1952 to 3.396 in 1982, an annual rate of increase of about 8 percent, compared with about 4 percent for the GNP deflator. The most rapid increase, of course, took place in the 1970s. Between 1970 and 1981, the price index on farm land rose by about 14 percent a year; in 1982, the price index declined. Real value of the farm land stock declined by about 0.5 percent a year between 1952 and 1982.

As indicated by the variety of approaches used in other studies to estimate land values, only a rough-and-ready approximation of nonfarm land price and quantity components can be developed. In this study, it is assumed that all changes in nonfarm business land use were transfers from agricultural use. This is approximately true, judging from the little information available on land transition. An Agriculture Department photographic interpretation study of land transition in 53 rapidly growing counties throughout the United States between 1961 and 1970 indicated that over 90 percent of the land entering nonfarm business use, including forests, was withdrawn from agricultural use (Zeimetz 1976). It is not certain, of course, that data drawn from such a small sample is representative of the United States as a whole, but it is likely that land transition has been concentrated in fast-growth areas.

From transition matrixes compiled from the photographic evidence, it appears that the transition rate from agricultural to commercial and industrial use was about 20 percent; that is, for every 100 acres withdrawn from agriculture, 20 acres was shifted into commercial and industrial use. The highest transition rate (58 percent) was into residential use and another 18 percent of the loss in agricultural land was diverted into transportation and recreation uses. Commercial and industrial businesses include a large proportion of total business enterprises, but do not exhaust the total. Some portion of the land transferred to transportation and recreation uses may represent a business gain, but the largest exclusion is the rental component of the residential transfer.

An examination of housing units authorized since the mid-1960s suggests that an average of about 6.5 percent of new structures was multi-family dwellings.[d] Multi-unit structures naturally are more land-intensive than single-unit structures. Two-family structures may use only marginally more land, while large apartment buildings can use several times the land of the typical single-family home. There is no way of precisely determining intensity differentials, but an intensity coefficient of 2 was used here. This assumes two-family structures are equal to single-family structures in land use, three- and four-family structures are twice as land-intensive, and five- or more family structures are three times as land-intensive. Thus, about 13 percent of the land transferred into residential use should be counted in the business sector. This raises the transition ratio from 20 percent to 28 percent, and implies a total transition to nonfarm business use of about 47 million acres between 1952 and 1982.

A survey of private landownership in 1978 indicated that nonfarm land in some form of business organization (proprietorship, partnership, or corporation) amounted to about 30 percent of the land in farm use (see USDA 1979). From the farm acreage data cited above, this would imply about 312 million acres in nonfarm business use in 1978. Using the transition ratio of 28 percent and the estimates of changes in farm acreage, nonfarm business land acreage can be estimated for 1952–82. From the market-value balance-sheet estimates, the price index and real value components also can be constructed. In 1972, the estimated average price per acre of nonfarm business land was $886.70, nearly 5 times the average price of farm land. The price index of nonfarm business land increased from 0.416 in 1952 to 2.779 in 1982, implying an annual rate of increase of about 6.5 percent. The real value of nonfarm business land increased by about 0.5 percent a year. Although estimates of nonfarm business land prices are considerably higher than farm land prices, the rate of increase is less rapid, though still faster than the rate of increase in the GNP deflator. This is reasonable if improvements to the land limit the price inflation.

Appendix 3 A Technical Note on Margin Drift

Consider rate-of-return behavior as a growth process:

$$(A.2)\quad R_t = R_0(1+r)^t QQ_t^\lambda \epsilon,$$

R_t = rate of return in current period,
R_0 = rate of return in base period,
r = "normalized" (invariant with respect to cyclical and random disturbances) growth rate,
$QQ_t^\lambda \epsilon$ = allowance for cyclical and random movements in R ($\epsilon > 0$ fluctuates randomly about 1).

From the definition of equation (3) in the text, the rate of return also is expressed as the product of the earnings margin (M) and the earnings multiplier (m), and the rate of growth (r) can be viewed as the sum of the growth rates of these two components.

$$(A.3)\quad r = r_M + r_m = r_M + r_{QQ} + r_{QK}.$$

In the sense of a normalized long-term growth rate, the middle term in equation (A.3) is zero. The capacity ratio rises and falls as the economy moves through expansion and recession phases of the business cycle (these effects on the rate of return are captured by QQ^λ in equation (A.2)), but does not grow over time. In the long term, only contributions from the margin and capital productivity influence rate-of-return growth. This can be shown directly as follows: Let R_0 be defined such that $QQ_0 = 1$ (i.e., the initial value is a full-employment level). From equations (3) and (4) in the text, the normalized growth rate of the rate of return must be:

$$(A.4)\quad (1+r)^t = \frac{M_t QQ_t QK_t}{R_0 QQ^\lambda \epsilon}.$$

The earnings margin and capital productivity have grown from the base period at normalized rates of r_M and r_{QK}, respectively. In addition, the margin fluctuates cyclically and randomly, while capital productivity fluctuates randomly about its normalized growth path. Thus,

$$(A.5)\quad (1+r)^t = \frac{[M_0(1+r_M)^t QQ_t^\alpha u] QQ_t [QK_0(1+r_{QK})^t w]}{R_0 QQ_t^\lambda \epsilon} = (1+r_M)^t(1+r_{QK})^t,$$

where

$\epsilon = uw,$
$\lambda = (1+\alpha), \alpha > 0.$

Thus,

$$(A.6)\quad r = r_M + r_{QK} + r_M r_{QK},$$

where

$r_M r_{QK}$ = small second-order effects.

The question now becomes one of determining the nature of long-term growth in the earnings margin and capital productivity. Once cyclical and random movements in the margin are accounted for, a nonzero *drift* may remain, depending on the price markup rule that best approximates pricing strategy for the aggregate economy. To illustrate margin drift, consider the case where the relative price of capital is held constant (as in *BXRC* and *AXRC*), and output price is determined as a markup on a subset of total unit costs (for example, unit labor costs). Unit costs are normalized for cyclical and random variation in the markup function, and the earnings margin corresponding to the growth rate r_M would be defined as follows:

$$\text{(A.7)}\quad M_t = \frac{(1+p)(1+g_i)^t U_{io} - (1+g_i)^t U_{io} - (1+g_e)^t U_{eo}}{(1+p)(1+g_i)^t U_{io}},$$

where

p = price markup,
U_i = unit costs included in markup function (subscript 0 denotes base period),
g_i = growth rate of included unit costs,
U_e = normalized unit costs excluded from markup function,
g_e = growth rate of excluded unit costs.

The continuous change in the normal margin with respect to time is given by

$$\text{(A.8)}\quad \frac{dM}{dt} = \left[\frac{-U_{eo}}{(1+p)U_{io}}\right] \times [\ln(1+g_e) - \ln(1+g_i)] \times \left[\frac{1+g_e}{1+g_i}\right]^t.$$

If growth rates of unit costs and the price markup are constant, the following possibilities characterize change in the normal margin:

(a) $\frac{dM}{dt} = 0$, if $g_e = g_i$.

(b) $\frac{dM}{dt} > 0$, if $g_e < g_i$, $\frac{dM}{dt} \rightarrow 0$, as t increases.

(c) $\frac{dM}{dt} < 0$, if $g_e > g_i$, $\frac{dM}{dt} \rightarrow -00$, as t increases.

In case (a), no drift in the earnings margin is indicated; hence, the normalized growth rate (r_M) would be zero. In case (b), however, a positive drift is indicated, gradually becoming smaller over time. When the difference between growth rates of included and excluded unit cost is fairly small, margin drift is also small but narrows very gradually (i.e., approximately constant for relatively long periods). Finally, if those unit costs excluded from the markup function grow faster than included unit costs, as in case (c), the normal margin would continuously deteriorate over time. This could not be descriptive of pricing strategy over any extended period, but it could describe a transition period of price-cost realignment. Negative drift would be possible in variable relative-capital-price margins, even if $g_i > g_e$. Increases in the replacement price of capital pull the margin down; the faster the capital price rises relative to the output price, the more likely the possibility of negative margin drift.

References

Baxter, John D. "What's Behind Falling Productivity of Equipment?" *Iron Age*, December 11, 1978, pp. 27-30.

Bernstein, Peter L. "Capital Stock and Management Decisions," *Journal of Post Keynesian Economics*, Fall 1983, pp. 20-38.

Board of Governors, Federal Reserve System. *Balance Sheets for the U.S. Economy 1945-82*. Washington, DC: October 1983.

Bosworth, Barry P. "Capital Formation and Economic Policy," *Brookings Papers on Economic Activity*, 2:1982, pp. 273-326.

Cagan, Phillip, and Robert E. Lipsey. *The Financial Effects of Inflation*. Cambridge, MA: Ballinger Publishing Company, 1978.

Clark, Peter K. "A Kalman Filtering Approach to the Estimation of Potential GNP," Processed. Revised, November 1983.

Denison, Edward F. *Accounting for United States Economic Growth 1929-1969*. Washington, DC: Brookings Institution, 1974.

Eisner, Robert. "Capital Gains and Income: Real Changes in Value of Capital in the United States, 1946-77," in Dan Usher, Ed., *The Measurement of Capital*. Chicago: University of Chicago Press, 1980, pp. 175-342.

Fain, T. Scott. "Self-employed Americans: Their Number Has Increased," *Monthly Labor Review*, November 1980, pp. 3-8.

Feldstein, Martin S., and Lawrence Summers. "Is the Rate of Profit Falling?" *Brookings Papers on Economic Activity*, 1:1977, pp. 211-27.

Fraumeni, Barbara M., and Dale W. Jorgenson. "The Role of Capital in U.S. Economic Growth, 1948-1976," in George M. Von Furstenberg, Ed., *Capital, Efficiency, and Growth*. Cambridge, MA: Ballinger Publishing Company, 1980, pp. 9-250.

Goldsmith, Raymond W. *The National Wealth of the United States in the Postwar Period*. Princeton: Princeton University Press, 1962.

______. *The National Balance Sheet of the United States 1953-1980*. Chicago: University of Chicago Press, 1982.

Hinrichs, John C. "The Relationship between Personal Income and Taxable Income," *Survey of Current Business*, February 1975.

Holland, Daniel M., and Stewart C. Myers. "Trends in Corporate Profitability and Capital Costs," in Robert Lindsey, Ed., *The Nation's Capital Needs: Three Studies*. New York: Committee for Economic Development, 1979, pp. 103-88.

Kahn, C. Harry. *Business and Professional Income Under the Personal Income Tax*. Princeton, NJ: Princeton University Press, 1964.

Kopcke, Richard W. "The Decline in Corporate Profitability," *New England Economic Review*, Federal Reserve Bank of Boston, May/June 1978, pp. 36-60.

Lovell, Michael C. "The Profit Picture: Trends and Cycles," *Brookings Papers on Economic Activity*, 3:1978, pp. 769-88.

Milgram, Grace. "Estimates of the Value of Land in the United States Held by Various Sectors of the Economy, Annually, 1952 to 1968," in Raymond W. Goldsmith, Ed., *Institutional Investors and Corporate Stock—A Background Study*, New York: Columbia University Press, 1973, pp. 343-77.

Nordhaus, William D. "The Falling Share of Profits," *Brookings Papers on Economic Activity*, 1:1974, pp. 169-208.

Okun, Arthur M., and George L. Perry. "Notes and Numbers on the Profits Squeeze," *Brookings Papers on Economic Activity*, 3:1970, pp. 466-72.

Phillips, Joseph D. *The Self-employed in the United States*. Urbana, IL: University of Illinois Press, 1962.

Poirier, Dale J. *The Econometrics of Structural Change*. Amsterdam: North-Holland Publishing Company, 1976.

Runyon, Herbert. "Profits: A Declining Share to Capital?" *Business Economics*, vol. XIV, no. 4 (September 1979), pp. 85-94.

Scanlon, Martha S. "Postwar Trends in Corporate Rates of Return," *Public Policy and Capital Formation*. Board of Governors of the Federal Reserve System, April 1981, pp. 75-87.

U.S. Department of Agriculture, *Who Owns the Land? A Preliminary Report of a U.S. Landownership Survey*, September 1979.

U.S. Department of Agriculture, *Economic Indicators of the Farm Sector*. Income and Balance Sheet Statistics, 1983.

Walton, John. "Capital Maintenance and the Measurement of Corporate Income," *Review of Income and Wealth*, series 27, no. 2 (June 1981), pp. 109-35.

Zeimetz, Kathryn A., et al. *Dynamics of Land Use in Fast Growth Areas*, U.S. Department of Agriculture, April 1976.

Working Paper Review

This working paper complements James G. Hoehn and William C. Gruben, with Thomas B. Fomby, Some Time Series Methods of Forecasting the Texas Economy, *Working Paper 8402, Federal Reserve Bank of Dallas, May 1984. James G. Hoehn is an economist with the Federal Reserve Bank of Cleveland.*

James G. Hoehn
A Regional Economic Forecasting Procedure Applied to Texas
Working Paper 8402.
September 1984. 45 pp. Bibliography.

In this paper economist James G. Hoehn proposes and implements a relatively simple method for building a multivariate autoregressive forecasting model for regional economic time series. The method used is a time series approach requiring little *a priori* or theoretical knowledge, as in the building of structural econometric models. In this way, the method is similar to the so-called vector autoregression (VAR) models of Anderson and Kuprianov and Lupoletti.[1] However, the way variables are chosen to be included and the way relationships are estimated involve more hypothesis tests and less prior knowledge.

We applied the method to the problem of forecasting quarterly growth rates of seven Texas variables that were seasonally adjusted. Each of the seven variables was related to its own past two observations in a simple regression equation to determine whether lagged values would aid forecasts. Then, two lagged growth rates of other regional and national variables were added to the equation to construct *Granger causality* tests. Such tests suggested whether inclusion of the other regional and national variables was likely to improve forecasts. After the series of causality tests was completed, a list of variables was suggested as likely candidates for inclusion in the forecasting equation for each Texas variable. This list was trimmed, and lag specifications were chosen, using the criteria of minimizing the standard error of the equation and the principle of parsimony. The resulting equations, although chosen on statistical grounds alone, conformed reasonably well to intuitions about the regional economy and its relations with the national economy. Of 14 national variables tried, inclusion of 4 appeared to capture most of the information—the index of leading indicators, the index of coincident indicators, the producer price index, and the federal funds rate.

Out of sample, the model was found to forecast consistently better (lower root mean square error) than benchmark univariate autoregressive integrated moving average (ARIMA) models, sometimes substantially better. In some cases, the improvement was statistically significant at the 0.05 level, despite the shortness of the forecasting sample, according to a test adapted from Ashley, Granger, and Schmalensee.[2] The model was reasonably stable as re-estimated over the out-of-sample period, and its forecasts could not be systematically improved by combining them with ARIMA forecasts.

The results of the study are subject to a number of caveats common to statistical studies. Nevertheless, results suggest that the proposed forecasting procedure can provide systematically better forecasts than univariate ARIMAs. Such systematically better forecasts apparently have not been delivered by the complex simultaneous equation models or by other time series methods. The procedure proposed requires only ordinary least squares regressions. A by-product of model building is insight into the regional economy and the strength of various leading relationships in the data. The method can easily be applied to other regional economies, and we have begun building a model for Ohio.

1. *Anderson, Paul A. "Help for the Regional Forecaster: Vector Autoregression,"* Quarterly Review, *Federal Reserve Bank of Minneapolis, vol. 3, no. 3 (Summer 1979), pp. 2-7; and Kuprianov, Anatoli, and William Lupoletti. "The Economic Outlook for Fifth District States in 1984: Forecasts from Vector Autoregression Models,"* Economic Review, *Federal Reserve Bank of Richmond, vol. 70/1 (January/February 1984), pp. 12-23.*

2. *Ashley, Richard A., C.W.J. Granger, and R. Schmalensee. "Advertising and Aggregate Consumption: An Analysis of Causality,"* Econometrica, *vol. 48, no. 5 (1980), pp. 1149-67.*

The working paper series is published by the Research Department of the Federal Reserve Bank of Cleveland to stimulate discussion and critical comment. Copies of working papers, either future or past issues, are available through our Public Information Department, Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101; (216) 579-2048.

Working Paper Series

8101
The Welfare Implications of Alternative Unemployment Insurance Plans
Mark S. Sniderman
April 1981, 20 pp.

8201
Multibank Holding Company Organizational Structure and Performance
Gary Whalen
March 1982, 30 pp.

8202
Stability in a Model of Staggered-Reserve Accounting
Michael L. Bagshaw and William T. Gavin
August 1982, 27 pp.

8203
A Micro View of the Transactions Money Market
Mark A. Zupan
September 1982, 31 pp.

8301
Non-Nested Specification Tests and the Intermediate Target for Monetary Policy
Mitsuru Toida and William T. Gavin
June 1983, 14 pp.

8302
Holding Company Organizational Form and Efficiency
Gary Whalen
July 1983, 20 pp.

8303
Extension of Granger Causality in Multivariate Time Series Models
Michael L. Bagshaw
August 1983, 11 pp.

8304
Forecasting the Money Supply in Time Series Models
Michael L. Bagshaw and William T. Gavin
December 1983, 23 pp.

8401
Economic Estimates of Urban Infrastructure Needs
Paul Gary Wyckoff
June 1984, 36 pp.

8402
A Regional Economic Forecasting Procedure Applied to Texas
James G. Hoehn
September 1984, 45 pp.

8403
Forecasting Using Contemporaneous Correlations
Michael L. Bagshaw
September 1984, 17 pp.

8404
Dollar Intervention and the Deutschemark-Dollar Exchange Rate: A Daily Time-Series Model
Owen F. Humpage
September 1984, 28 pp.

8405
Velocity: A Multivariate Time-Series Approach
Michael L. Bagshaw and William T. Gavin
November 1984, 29 pp.

8406
Monetary Policy and Real Interest Rates: New Evidence from the Money Stock Announcements
William T. Gavin and Nicholas V. Karamouzis
December 1984, 43 pp.

8407
Monetary Policy Regimes: A Synthesis of the Monetary Control and Rational Expectations Literatures
James G. Hoehn
December 1984, 60 pp.

Economic Review *is published quarterly by the Research Department of the Federal Reserve Bank of Cleveland. Copies of the issues listed here are available through our Public Information Department, 216/579-2048.*

Economic Review

Winter 1983

"Location and Reinvestment: The Youngstown Steel District"
Robert H. Schnorbus

"Thrifts and the Competitive Analysis of Bank Mergers"
Paul R. Watro

Working Paper Review:
Stability in a Model of Staggered-Reserve Accounting
Michael L. Bagshaw and William T. Gavin

Spring 1983

"Money Demand: Cash Management and Deregulation"
John B. Carlson

"Divisia Monetary Aggregates: Would They Be More Palatable than the Traditional Simple-Sum Stews?"
Mark A. Zupan

Fall 1983

"Plant Closings and Worker Dislocation"
Daniel A. Littman and Myung-Hoon Lee

"Prevailing Wage Laws, the Federal Reserve, and the Service Contract Act"
Mark S. Sniderman

Working Paper Review:
Forecasting the Money Supply in Time Series Models
Michael L. Bagshaw and William T. Gavin

Winter 1984

"Reflections on Money and Inflation"
William T. Gavin

"The Outlook for Inflation"
K. J. Kowalewski and Michael F. Bryan

Spring 1984

"Estimating Infrastructure Needs: Methods and Controversies"
Paul Gary Wyckoff

"Nonbanking Operations of Bank Holding Companies"
Gary Whalen

Summer 1984

"The Implementation of Industrial Policy"
Daniel A. Littman

"Voluntary Export Restraints: The Cost of Building Walls"
Michael F. Bryan and Owen F. Humpage

Working Paper Review:
Dollar Intervention and the Deutschemark-Dollar Exchange Rate: A Daily Time-Series Model
Owen F. Humpage