Foreign Exchange Operations and the Federal Reserve

J. Alfred Broaddus, Jr., and Marvin Goodfriend

The operations of U.S. government agencies in foreign exchange markets are probably regarded as arcane by most Americans. These operations are, however, an important element of U.S. international economic policy. And from time to time they are highly visible to the public: for example, when the United States and other major industrial countries intervene jointly in the markets to influence exchange rates, or when they provide assistance to particular countries such as the substantial aid extended to Mexico in 1995.

The Gold Reserve Act of 1934 gives the Treasury primary responsibility for United States foreign exchange operations through its Exchange Stabilization Fund (ESF). Although the Federal Reserve (Fed) had been active in foreign exchange markets in the 1920s and early 1930s, its involvement ceased after 1934.[1] There was relatively little need for official U.S. foreign exchange operations in the early post-World War II period. Under the Bretton Woods arrangements of 1944, foreign governments assumed responsibility for fixing the value of their currencies against the dollar. For its part, the United States managed its monetary policy in accordance with the Gold Reserve Act so as to maintain the dollar's convertibility into gold at $35 an ounce.

U.S. authorities, however, were reluctant to pursue sufficiently tight monetary policy to protect the country's gold reserves following the resumption of

The authors are respectively president, and senior vice president and director of research. This article originally appeared in this Bank's *1995 Annual Report*. The article benefited greatly from stimulating discussions with Robert Hetzel and from presentations at the Macro Lunch Group of the Wharton School, University of Pennsylvania; the Masters of International Business Lecture Series, College of Business, University of South Carolina; and the Norwegian School of Management. Comments from Michael Dotsey, Thomas Humphrey, Robert King, Bennett McCallum, and Alan Stockman are greatly appreciated. The views expressed are the authors' and not necessarily those of the Federal Reserve System.

[1] See Chandler (1958) and Clarke (1967).

full convertibility among the major currencies in the late 1950s. And the Fed resumed foreign exchange operations in 1962, after a nearly 30-year hiatus, to supplement and substitute for monetary tightening in defense of the dollar. Although the Fed has consistently held that it has independent authority to undertake foreign exchange operations, in practice the Fed works closely with the Treasury in conducting them. Indeed, the Federal Open Market Committee's (FOMC's) foreign currency directive requires that these operations be conducted "in close and continuous consultation and cooperation with the United States Treasury."[2] So it seems fair to say that the Fed recognizes the Treasury's preeminence in foreign exchange policy.

The Treasury welcomed the Fed's renewed participation in large part because the Fed brought with it resources to supplement those of the ESF. In 1962 the Fed established reciprocal currency agreements—commonly called "swaps"—with nine central banks and the Bank for International Settlements. Further, in 1963, the Fed agreed to "warehouse" foreign currencies held by the ESF. The primary objective of these initiatives was to provide U.S. authorities with a supply of foreign currencies to buy back dollars in order to help protect U.S. gold reserves.[3]

FOMC discussions at the time made it clear that some Fed officials recognized how following the Treasury's lead in foreign exchange operations could compromise the Fed's independence in conducting monetary policy.[4] This risk did not present serious operational problems at the time, however, because the United States was committed to the Bretton Woods arrangements and monetary policy was committed to defending the dollar.[5] Thus, the Fed and the Treasury were working toward the same general objectives, and the Fed's independence was not a pressing issue in practice.

We argue below that subsequent developments have undermined the favorable conditions that enabled the Fed to participate in foreign exchange operations without compromising either its independence or its monetary policy goals. We make our case by developing several preliminary points. In Section 1 we explain how theoretical advances and practical experience in recent years teach that the Fed's longer-term low-inflation objective must be *credible* if the Fed is to pursue this objective efficiently via monetary policy. Moreover, the Fed's independence is the cornerstone of this credibility. In Section 2 we explain why Fed credibility based on independence is inherently fragile, and we

[2] See the discussion in Humpage (1994), pp. 3–4.

[3] Pauls (1990) details the evolution of U.S. exchange rate policy in the post-World War II period.

[4] See Hetzel (1996).

[5] That either a fixed exchange rate or a fixed gold price commitment requires monetary policy to be dedicated to that objective is emphasized, for example, by McCallum (1996b), Chapters 4 and 7.

emphasize the crucial importance of the Fed's off-budget status in supporting its independence.

We take up the role of the Fed in foreign exchange operations in Section 3, where we distinguish two broad types of official foreign exchange transactions: unsterilized and sterilized. As explained there, *unsterilized* transactions are essentially monetary policy actions and therefore are carried out independently by the Federal Reserve. Since *sterilized* transactions are *not* monetary policy actions, the Fed can acknowledge the Treasury's leadership regarding them without directly compromising its independence.

Evidence accumulated over the past two decades suggests, however, that sterilized intervention in exchange markets has at best only temporary effects on exchange rates and must be supported by monetary policy actions to have lasting effects. Consequently, the Fed's participation with the Treasury in sterilized operations creates confusion as to whether monetary policy is dedicated to the support of exchange rate or domestic objectives. Such confusion weakens the public's perception of the Fed's independence and undermines the credibility of the Fed's low-inflation goal.

In Section 4 we lay out in more detail the inherent contradictions for monetary policy that arise when the Fed follows the Treasury's lead on exchange rate policy. And we argue in Section 5 that the Fed's financing of even sterilized foreign exchange operations constitutes a misuse of the Fed's off-budget status that risks undermining the public's acceptance of the independence of the Fed. We believe that the best way to resolve the conflict between foreign exchange operations and monetary policy is for the Fed to disengage from foreign exchange operations completely. The concluding section summarizes our argument.

1. CREDIBILITY AND THE EFFECTIVENESS OF MONETARY POLICY

Numerous disinflations since the early 1980s have taught central bankers around the world that credibility—having a reputation for pursuing price level stability consistently and persistently—is the key to an effective anti-inflationary monetary policy.[6] We would even go so far as to say that the primary policy problem facing the Fed during this period has been the acquisition and maintenance of credibility for its commitment to low inflation—so much so that credibility concerns remain a motivating or restraining influence on monetary policy actions today, even though the Federal Reserve's low-inflation objective has nearly been achieved.

[6] See the accounts in Leiderman and Svensson (1995).

As it happens, the growing practical appreciation of the importance of credibility is supported by an improved scientific understanding associated with game theory and the rational expectations approach to monetary theory. In many ways, theory simply articulates what central bankers have learned from practical experience. Briefly, the theory recognizes that monetary policy involves continuous interaction between a central bank and the public that introduces a link, in the public's mind, between current policy and future policy actions. In the absence of credibility, expansionary current monetary policy tends to generate expectations of expansionary policy—and possibly excessively expansionary policy—in the future. Such expectations trigger aggressive wage and price increases that, in turn, neutralize the beneficial effects of the expansionary current policy. The result is higher inflation with little, if any, sustained increase in employment and output.

Theory supports the idea that the potential for future inflation, which can be thought of as a punishment imposed collectively by wage- and price-setters on a central bank, can discipline a central bank. In a reputational equilibrium, wage- and price-setters keep their part of an implicit bargain by not inflating as long as the central bank demonstrates its commitment to low inflation by eschewing excessively easy policy. A central bank may be said to have credibility when an implicit mutual understanding between the public and the central bank sustains a low-inflation equilibrium.[7]

The key point is that a low-inflation equilibrium sustained by central bank credibility is *fragile*. In such an equilibrium the public is very sensitive to any central bank departure from the behavior it has come to anticipate; this expected continued behavior, indeed, is the essence of the central bank's credibility. The public is particularly nervous about such departures when the central bank has acquired credibility only recently. But there is evidence that low-inflation equilibria sustained by credibility continue to be fragile even when a central bank's actions have repeatedly demonstrated its commitment to low inflation over a period of years.

The fragility of the Fed's credibility is evident in the behavior of long-term bond rates.[8] The real yield on the 30-year U.S. government bond probably moves within a range of 2 percentage points or so around 3 percent per year.[9] The remainder of the nominal long-term yield reflects inflation expectations. In the early 1960s, for example, when inflation averaged between 1 and 2 percent

[7] The introductory chapter in Persson and Tabellini (1994) contains a good survey of research on the role of credibility in monetary and fiscal policy. Barro and Gordon (1983), Cukierman (1992), and Sargent (1986) contain seminal analyses of credibility.

[8] Goodfriend (1993) and King (1995), for example, interpret movements in long-term bond rates as indicators of credibility for low inflation.

[9] Ireland's (1996) study of the ten-year bond rate provides some support for this view.

per year, the 30-year bond yielded roughly 4 percent.[10] In 1981, when the public's confidence in the Fed's commitment to controlling inflation was at its low point, the long-term bond yield reached nearly 15 percent. The rate stood at around 6 percent in late 1995, which indicated that the public expected about 3 percent inflation on average over the long term.

Doubts about a central bank's credibility often surface as "inflation scares" in the long-term bond market. Following a period of rising inflation in the late 1970s, for example, the 30-year rate jumped 2 percentage points in the first quarter of 1980, which signaled the most serious and sudden collapse of confidence in the Fed on record. The fragility of the Fed's credibility was apparent again in 1984 when the bond rate, after falling to about 10 percent in late 1982, registered another inflation scare by rising to around 13.5 percent, even though the Fed had by then brought actual inflation down from over 10 percent to around 4 percent.

The swings in the bond rate over the past two years have been less dramatic than in the early 1980s, but nonetheless substantial. Rising from a low of about 5.8 percent in October 1993, the bond rate peaked at around 8.2 percent in November 1994. We interpret that wide swing as evidence that the Fed's anti-inflationary credibility remains exceedingly brittle despite years of sustained progress in bringing the actual inflation rate down.

The fragile nature of the Fed's credibility imposes a number of costs on the economy. First, there is the direct cost of higher long-term interest rates with their negative effects on economic performance. Second, with inflation expectations higher than they should be, the Fed is left with the difficult choice of either accommodating these expectations and accepting higher rates of inflation or failing to accommodate them and risking negative short-term effects on real economic activity. Moreover, even hesitating to react can be costly because, by suggesting indifference, the Fed may encourage workers and firms to ask for wage and price increases to protect themselves from higher expected costs.

Finally—a related point—weak credibility makes it difficult for the Fed to respond when employment considerations call for an easing of policy, as they did in the 1990–91 recession and again in mid-1995. In such circumstances, the Fed must balance the desirable short-term effects of lower short-term rates against the risk of higher long-term rates.

2. FEDERAL RESERVE INDEPENDENCE

A number of prominent institutional mechanisms have been used to assist central banks in maintaining credibility for low-inflation objectives. Historically, a national commitment to a gold or silver standard—that is, a commitment to maintain a fixed currency price of gold or silver—was the most important. A

[10] See Salomon Brothers and Hutzler (1968).

second mechanism, more prominent in recent years, is for a country to commit to fix its exchange rate against the currency of a trading partner that credibly maintains the purchasing power of *its* currency. An important motivation for the establishment of the European Monetary System (EMS), for example, was the desire of some countries to import credibility for low inflation by pegging their currencies to the deutsche mark (D-mark). Difficulties with fixing exchange rates, including the near collapse of the EMS in the early 1990s, have led some countries to experiment recently with a third commitment device: inflation targets.[11] Finally, countries have relied on central bank independence to supplement one of the other mechanisms or to substitute for them.

Broadly speaking, central bank independence implies a separation of bank decisions from the regular decisions of the political system.[12] At a minimum, it means that a central bank is free to conduct monetary policy without interference from the Treasury. The degree of actual operational freedom enjoyed by an independent central bank, however, has varied widely depending on the circumstances. For instance, in the nineteenth century, when wide support for central bank independence first developed, independent central banks were narrowly constrained by national commitments to various commodity standards. Similarly, the Federal Reserve was established in 1913 as an independent central bank mandated by the Federal Reserve Act to stabilize financial markets while keeping the United States on the gold standard.

A central bank may be said to lack "goal independence" when its objective is given by legislative mandate; however, one can still speak of a central bank as having "instrument independence"—the freedom to use a short-term interest rate or other monetary policy instrument to achieve its mandated goals.[13] The Fed has had full instrument independence, except for the World War II years and the period from the end of the war to the 1951 Fed-Treasury Accord. During that time the Fed was obliged to maintain low interest rates on government securities to facilitate the Treasury's finances. The Accord reasserted the principle that monetary policy should be used for macroeconomic stabilization, the fiscal concerns of the Treasury notwithstanding. In terms of the above definitions, the Accord fully restored the Fed's instrument independence.[14]

The Accord did not give the Fed *goal* independence because monetary policy was still committed under the Bretton Woods arrangements to support the fixed dollar price of gold. When the Bretton Woods System collapsed

[11] Leiderman and Svensson (1995) and McCallum (1996a) contain accounts of the experience with inflation targets in a number of countries. For an empirical study of exchange rate credibility in the EMS, see Rose and Svensson (1994).

[12] This definition is from Hetzel (1990), p. 165.

[13] Fischer (1994), p. 292, distinguishes between goal and instrument independence.

[14] The Fed actually abandoned its short-term interest rate peg in 1947; it gave up its long-term rate peg in 1951. Stein (1969) contains a good discussion of developments leading up to the 1951 Fed-Treasury Accord.

in 1973, however, the national consensus on the proper goal for monetary policy collapsed with it, and the Fed has been operating without an explicit congressional mandate since then.[15] Thus, during this period the Fed has had goal independence by default, as it were, and this independence is now arguably the sole institutional mechanism supporting low inflation in the United States.

Independence and Credibility

A goal-independent Fed unrestrained by a legislative mandate is a particularly deficient mechanism for maintaining low inflation. The reason is that in this situation a low-inflation equilibrium must be supported entirely by credibility that the Fed creates for itself—credibility that is inherently fragile as discussed above. The unbridled discretion conferred on the Fed in this case only makes the acquisition and maintenance of credibility for low inflation more difficult. The Fed's goal independence gives other government entities strong incentives to attempt to influence its policies via such channels as congressional oversight hearings, appointments of Federal Reserve governors, proposed changes in the Fed's regulatory role, and so forth. Moreover, such attempts at influence can be of a conflicting nature, adding to the confusion. Knowing this, the public is rightly suspicious of any potential conflict between the Fed, the Treasury, and Congress. In this environment, any contact that Fed officials have with the rest of the government risks creating credibility problems for monetary policy.

At the same time—and paradoxically—central bank goal independence actually creates incentives for Fed officials to interact with the rest of the government.[16] The lack of clarity in the Fed's mandate necessitates deeper involvement in the legislative process by Fed officials who must see to it that proposed legislation does not compromise its monetary policy mission. Finally, the Fed's independence confers upon it a nonpartisan aura which leads others in government to seek its advice, certification, or arbitration in controversial policy disputes.

Financial Independence

In principle, a healthy democracy requires full public discussion of expenditures of public monies. The congressional appropriations process enables Congress to evaluate competing budgetary programs and to establish priorities for the allocation of public resources.

Congress has long recognized, however, that the pressure of budgetary politics could tempt future Congresses to press the Fed at least implicitly to

[15] It is true that the 1978 Humphrey-Hawkins law mandates the Fed to set monetary aggregate targets as guides to short-run policy. But the Humphrey-Hawkins law instructs the Fed to take account of so many potentially conflicting macroeconomic concerns in setting the targets that it has exercised little restraint on the Fed's freedom of action.

[16] See Bradsher (1995).

help finance federal expenditures through inflationary monetary policy. Consequently, the Fed has been made financially independent—its operations are funded from the interest payments on its portfolio of securities—and the Fed has wide discretion over the assets it holds. In short, the Fed is exempt from the congressional appropriations process in order to keep the political system from exploiting inflationary money creation. It is critically important that the Fed not misuse this exceptional "off-budget" status so as not to undermine public understanding of and support for its financial independence. This, in turn, requires the Fed to understand clearly what activities are and are not essential to its central banking mission.

3. THE ROLE OF THE FED IN FOREIGN EXCHANGE OPERATIONS

The points about credibility and independence developed above will serve as the basis for our assessment of the Fed's role in foreign exchange operations in what follows. Here we review the basic mechanics of foreign exchange operations. We begin by making the important distinction between unsterilized and sterilized transactions. Then we briefly discuss the means by which the Fed finances foreign exchange operations for its own account and warehouses foreign exchange for the ESF.[17] Our analysis identifies in a preliminary way the fundamental sources of conflict for monetary policy arising from the Fed's participation in foreign exchange operations.

Unsterilized and Sterilized Operations

The distinction between unsterilized and sterilized operations is straightforward: unsterilized transactions involve changes in the monetary base, and sterilized transactions do not. For example, the Fed could acquire foreign exchange in an *unsterilized* purchase using newly created base money: that is, bank reserves or currency. Such a transaction would be an expansionary monetary policy action because it would increase the monetary base.

A foreign exchange purchase would be *sterilized,* in contrast, if the Fed offset its effect on the base by selling an equivalent amount of dollar-denominated securities. Because the Fed controls the monetary base, it is in a position to determine whether a foreign exchange operation is sterilized or not. In practice, the Fed routinely sterilizes foreign exchange operations that it undertakes for its own account and for the ESF. In sterilized operations the current federal funds rate target (the key policy instrument indicating the current stance of monetary

[17] A detailed description of the mechanics of foreign exchange operations using T-accounts is found in Humpage (1994).

policy) is maintained. This point is important because it implies that—at least as a mechanical matter—the Fed can follow the Treasury's lead in sterilized foreign exchange operations without relinquishing control of monetary policy.

Nevertheless, sterilized foreign exchange operations, or "intervention," pose significant problems for the Fed. For the most part, economists agree that sterilized intervention by central banks in foreign exchange markets has no lasting effect on exchange rates.[18] In the absence of supporting monetary policy actions, sterilized interventions can influence exchange rates temporarily, especially when the interventions are unexpected. But obviously the ability of authorities to surprise markets is very limited. Sterilized intervention can be most effective when it signals a government's resolve to follow up with monetary or fiscal policy actions that will powerfully influence the exchange rate in the future.[19] Consequently, Fed participation in sterilized foreign exchange operations under the Treasury's leadership creates confusion as to whether monetary policy will support short-term exchange rate objectives or longer-term anti-inflationary objectives. Only occasionally will the monetary policy actions required to pursue these two objectives coincide.

This confusion is compounded by a lack of consistency in U.S. exchange rate policy in the post-1973 floating exchange rate regime. Officially, the objective of foreign exchange operations is to counter "disorderly market conditions," but that phrase has never been defined operationally. It was interpreted most narrowly in the first Reagan administration, when U.S. operations were minimal. It was interpreted broadly between 1977 and 1979 when the dollar was viewed as unacceptably low and again in 1985 when the dollar was unacceptably high. Intervention was undertaken in these periods to help push the dollar into an acceptable range. Extensive interventions were carried out in the years following the Louvre Accord of 1987 to help stabilize the exchange rate.[20]

Moreover, much U.S. intervention in recent years has been coordinated with foreign governments. The Group of Seven finance ministers and central bank governors meet regularly to discuss exchange rate objectives. The enormous publicity surrounding these discussions, designed to underscore international harmony on exchange rate policy, heightens uncertainty regarding whether the Fed will support sterilized operations with monetary policy actions. The widespread coverage of internationally "coordinated" foreign exchange operations is almost certainly harmful to the public's perception of the Fed's independence and thereby weakens the credibility of the Fed's low-inflation strategy.

---

[18] A representative survey of the academic literature on this point would include Bordo and Schwartz (1991), Edison (1993), and Obstfeld (1990), and references contained therein.

[19] See Mussa (1981).

[20] See Destler and Henning (1989), Funabashi (1989), and Pauls (1990) for discussions of U.S. exchange rate policy.

Financing Mechanisms

Federal Reserve acquisitions of foreign exchange are generally financed in one of three ways. If the FOMC approves, the Fed can acquire foreign exchange for its own account by creating additional bank reserves or currency—that is, via an unsterilized transaction. Sterilized acquisitions, on the other hand, are financed by selling Treasury securities from the Fed's portfolio. Finally, the Fed has the option of borrowing currencies from foreign central banks using reciprocal currency agreements—the so-called "swap" network. Swap facilities are, in effect, short-term lines of credit giving central banks access to one another's currencies. The facilities provide for the swap (simultaneous spot purchase and forward sale) of each other's currency by the Fed and the foreign central bank. Swaps typically are not accompanied by any change in monetary policy—in other words they are sterilized transactions.[21] The Fed holds foreign exchange in the form of short-term securities or interest-bearing deposits at foreign central banks, so that sterilized transactions amount to substituting foreign-currency-denominated interest-earning assets for dollar-denominated securities in the Fed's portfolio.

The Fed bears the exchange rate revaluation risk—as well as the credit risk—for any foreign-currency-denominated assets it holds for its own account. Since the Fed marks its foreign currency assets to market monthly, a depreciation of the foreign exchange value of the dollar, for instance, raises the dollar value of the Fed's foreign holdings. Any such gains or losses eventually show up as larger or smaller Fed payments to the Treasury after expenses.[22] Whenever the Fed disperses foreign exchange acquired through a swap, it bears the exchange risk involved in covering its forward commitment to reverse the swap.

The Exchange Stabilization Fund

As mentioned above, the Treasury conducts foreign exchange operations through its Exchange Stabilization Fund. When it was established by the Gold Reserve Act, the ESF was capitalized with $2 billion derived from the proceeds of the 1934 revaluation of the U.S. gold stock from $20.67 to $35 per ounce. Later, $1.8 billion was transferred from the ESF as partial payment on the U.S. subscription to the International Monetary Fund (IMF), which left $200 million as the remaining capital of the ESF. ESF capital has grown since then

[21] The Fed drew on its swap lines in the 1960s to protect the Treasury's gold stock by using the borrowed currencies to buy back dollar reserves from foreign central banks. These transactions effectively allowed the United States to assume a portion of other countries' devaluation risk. More recently, the United States has had sufficient foreign currency reserves and has not drawn on its swap lines.

[22] See the discussions in Goodfriend (1994) and Humpage (1994).

as a result of retained interest earnings, revaluations of gold, and profits on foreign exchange acquisitions.[23]

Since use of its funds is not subject to the appropriations process, the ESF provides the Treasury with a degree of flexibility and discretion in its foreign exchange operations. The ESF serves two broad purposes. First, it is used to intervene in foreign exchange markets to influence dollar exchange rates with major currencies such as the D-mark and the Japanese Yen. Second, the ESF makes loans to foreign governments—frequently to heavily indebted governments and often in association with IMF or other official assistance programs. Typically such loans are made to deal with a serious balance-of-payments problem or to assist a country managing its external debt. Often the currencies of recipient countries are not fully convertible or are of secondary importance.[24] The recent loans to Mexico are a prominent example of this type of assistance.

The ESF's capacity for purchasing foreign currencies is limited, however, as it has not received an appropriation from Congress since 1934. Apart from the retained earnings on its investments mentioned above, the ESF has been able to augment the resources at its disposal in three significant ways. First, Congress has authorized advancing to the ESF foreign currencies borrowed from the IMF. Second, the ESF receives the Special Drawing Rights (SDRs) allocated to the United States by the IMF.[25] Third, the Fed has provided the ESF with additional resources, either by helping to finance operations on its own account or by warehousing foreign exchange for the ESF. It was because the ESF's resources were limited that the Treasury encouraged the Fed in the early 1960s to participate for its own account in foreign currency operations and to warehouse foreign currencies. In 1990, the dollar value of U.S. net foreign currency balances (the sum of acquisitions on the Fed's and the ESF's accounts) exceeded $40 billion.[26] The FOMC authorized warehousing of ESF foreign currencies up to a limit of $15 billion in 1990.

Warehousing allows the ESF to finance purchases of foreign exchange in much the same way that securities dealers use repurchase agreements with banks to finance their portfolios. That is, warehousing allows the ESF to enlarge its portfolio of foreign-currency-denominated assets with funds borrowed from the Fed. Suppose, for example, that the ESF wishes to sell dollars for foreign exchange to depreciate the dollar but has inadequate resources to do so. The Fed can execute the transaction—warehouse the foreign exchange—by selling a Treasury security from its portfolio in the open market and using the proceeds

[23] U.S. Congress (1976), pp. 3–5.

[24] See U.S. Department of the Treasury (1991). U.S. Congress (1976) details ESF operations from 1968 to 1975. Todd (1992) presents a history of the ESF.

[25] SDRs are monetized by transferring them to the Fed.

[26] See Pauls (1990), pp. 894 and 904, and U.S. Congress (1976), pp. 3–5.

to acquire the foreign-currency-denominated securities on behalf of the ESF. Because the Fed executes the purchase of foreign exchange on behalf of the ESF, the latter remains exposed to the revaluation gains or losses on the foreign exchange warehoused. Interest earnings on the foreign currencies warehoused accrue to the Fed. Note that the warehousing operation amounts to a sterilized acquisition of foreign exchange.

Whether or not the Fed finances sterilized foreign exchange purchases for its own account, or warehouses foreign currencies for the ESF, a sale of Treasury securities to the public is the ultimate source of the funds. True, the securities involved are not newly issued; they are sold from the Fed's portfolio. The results, however, are equivalent in many ways to those of a new issue since the Fed simply returns to the Treasury all of the interest it receives on the Treasury securities that it holds, minus a small fraction that covers the Fed's operating expenses. The main difference between Fed financing and financing by the Treasury itself is that the former is arranged between Treasury and Fed officials without an explicit appropriation from Congress. A second difference is that Fed financing does not show up as a measured increase in the federal deficit, since it does not involve newly issued debt.

Although the Fed is the junior partner with the Treasury on foreign exchange policy, it is certainly an equal partner in terms of the resources provided. It is able to make these resources readily available without a congressional appropriation because its financial independence puts its open market operations in Treasury securities off-budget. The exchange operations arranged by the Treasury, however, not infrequently involve broader foreign relationships in ways that may be politically charged. Hence, the Fed's involvement, especially because it is outside the formal budget process, puts public support for its financial independence at risk, and with it, the credibility of its low-inflation policy.

4. THE CONFLICT BETWEEN EXCHANGE RATE POLICY AND MONETARY POLICY

The national commitment to the Bretton Woods arrangements minimized the risk of policy conflict between the Fed and the Treasury when the Fed resumed its participation in foreign exchange operations in the early 1960s. But the nation's unwillingness to support that commitment with sufficiently restrictive monetary policy led to the collapse of the fixed exchange rate system in 1973. Several years of sharply rising inflation followed. Despite this, Congress was unable to reach consensus on a new monetary policy mandate. Consequently, in 1979 the Fed asserted its own commitment to restore low inflation.

We believe that these developments have undermined the Fed's ability to participate in exchange rate policy without compromising its independence and its monetary policy goals. In particular, with the potential for Fed-Treasury policy conflicts now significantly enlarged, it is no longer possible for the Fed simply to follow the Treasury's lead on exchange rate policy without endangering its monetary policy credibility. This is true *even* in the case of sterilized interventions. Under the current arrangement, the Fed participates in sterilized operations without committing to support the operations with future monetary policy actions. This maintains the Fed's independence by keeping its options open. But such discretion increases the likelihood that particular operations may fail because the Fed is not willing to support them with monetary policy.

Failed foreign exchange operations are costly because they give the impression that the authorities are either unable or unwilling to achieve a prominent objective that they appear to be pursuing. For example, the failure of the June 24, 1994, intervention was reported in a front-page *New York Times* story carrying the headline: "16 Central Banks are Thwarted in Huge Effort to Prop Up Dollar."[27] Nor was attention to the event confined to major money centers. On the following day the *Richmond Times-Dispatch* reported the story with the front-page headline: "Effort to Bolster Dollar a Failure." Widely publicized policy failures undermine Fed credibility and thereby jeopardize the effectiveness of overall monetary policy.

We believe that, to best protect the credibility of its low-inflation goal and the independence of monetary policy more generally, the Fed should be separated completely from the Treasury's foreign exchange operations. In principle, the Fed could disengage unilaterally; however, there would be two major practical obstacles to such an action. The most serious obstacle is that the appointment process would make it difficult for the Fed to bind itself not to participate, since appointments to the Federal Reserve Board could be made on condition of cooperation with the Treasury. Congress might be able to block such conditions in the confirmation process in particular cases if it were so disposed, but legislation probably would be required to remove the Fed from exchange market intervention definitively.

The second main obstacle to unilateral disengagement is that it would deny the Treasury the benefit of the Fed's advice on foreign exchange intervention and the certification that goes with it. Here, though, the Fed cannot be indifferent to the use of its name in headlines that either box it in or harm its credibility. Moreover, the act of certification itself creates a perception of partisanship that erodes the value of that certification, even as it undermines the public's perception of the Fed's independence.

In these circumstances, it is natural to look for a middle-of-the-road solution to the problems presented by the Fed's involvement in exchange market operations. One might, for example, try to specify particular circumstances in

[27] See Friedman (1994).

which the Fed could participate. For instance, if the Fed routinely announced an inflation target, it could agree to help the Treasury intervene if the inflation rate were within a specified range of the target. Defining such conditions clearly, however, would be difficult, and this approach would leave the door open to many of the same problems the Fed faces currently.

5. THE CONFLICT BETWEEN FOREIGN EXCHANGE OPERATIONS AND THE FED'S FINANCIAL INDEPENDENCE

From the start, a major reason for the resumption of Federal Reserve foreign exchange operations in the 1960s was to make Fed resources available to the ESF. The Fed's financial independence gave it the discretion to allocate resources to foreign exchange operations without an explicit congressional appropriation. Apparently there was then little concern about misuse of the Fed's off-budget status because Fed financing of foreign exchange operations at the time seemed conformable with the nation's commitment to the Bretton Woods system. Such financing has become more problematic with the breakdown of the national consensus on monetary and exchange rate policy in the aftermath of the collapse of Bretton Woods.

Economists understand more clearly today than they did in the 1960s the distinction between Federal Reserve monetary policy and credit policy.[28] As pointed out in Section 3, sterilized foreign exchange operations are not monetary policy since they leave the monetary base and the federal funds interest rate target unchanged. Such operations do, however, constitute credit policy since they amount to a substitution of loans to foreign authorities for dollar-denominated securities in the Fed's portfolio. In effect, sterilized operations are extensions of Fed credit financed by selling Treasury debt from the Fed's portfolio. Such extensions of credit are clearly fiscal policy, not monetary policy.

The extension of credit by U.S. authorities involves both market and credit risk. Although the default or credit risk of the securities in which major foreign currency balances are held is negligible, the revaluation or market risk is considerable. Credit risk, however, can also be substantial when a loan is made to assist, say, a country managing its external debt or one with a serious balance-of-payments problem. Provisions can be made to take collateral if the borrowing country proves unable to make scheduled payments. But such provisions are not always feasible or entirely effective. When a borrowing country's financial problems prove persistent, the ESF and the Fed can be "taken out" by longer-term funding arranged through international organizations such as the

[28] This distinction is developed in Goodfriend and King (1990) and used in Goodfriend (1994).

IMF.[29] But to the extent that collateralization is incomplete or "take outs" are not arranged in advance or are uncertain, taxpayers are at risk. Thus, in their foreign exchange operations the Fed and the ESF assume risk—both market risk and credit risk—on behalf of the U.S. taxpayer.

The national decision to put funds at risk in foreign exchange operations is clearly an important fiscal policy matter. The presumption is that—as with any fiscal action—Congress should authorize the expenditure and explicitly appropriate the funds. Fed financing of foreign exchange operations through its own account and by warehousing funds for the ESF sidesteps congressional authorization and obscures the funding.

The Fed's financing of foreign exchange operations without explicit direction from Congress exposes it to potentially harsh criticism if an initiative goes badly. Unfavorable outcomes would obviously undermine public support for the Fed's financial independence. But there is a more subtle risk, even if foreign initiatives funded by the Fed go well. Some will ask whether, if Fed financing of credit extensions to foreigners is beneficial, it might also be desirable for the Fed to support worthy domestic objectives. Any attempt to exploit the Fed's financial independence in this manner would almost guarantee that its independence would be withdrawn over time.

Fed off-budget funding attracted substantial attention in the Mexican case in 1995, as indicated by a remarkable headline in *The New York Times*: "Clinton Offers $20 Billion to Mexico for Peso Rescue; Action Sidesteps Congress."[30] Should the Fed take comfort from the relative absence to date of significant negative repercussions from its involvement in this initiative? We think not. The publicity for the Mexican rescue put the Fed's off-budget funding powers on the radar screen, along with the potential risks described above. The Fed appeared to receive the implicit support of the congressional leadership in this instance, but Congress itself probably would not have voted to authorize the funds, and the public at large did not seem to favor such generous support for Mexico. Indeed, many Americans, including some prominent ones, viewed the transaction as a bailout of big investors. If, over time, developments in Mexico turn unfavorable, the result could be an erosion of public and congressional support for the Fed's financial independence.

[29] To the extent that the funds are provided by the United States in the first place, the possibility of such takeouts amounts to only a partial reduction of U.S. taxpayer risk. On some occasions when U.S. authorities have drawn and dispersed foreign currencies through the swap network, the U.S. Treasury has repaid the swap loans with foreign exchange borrowed on a long-term basis using so-called "Roosa," or "Carter," bonds. Such actions, however, only shift the market risk from short to long term. See U.S. Congress (1976), pp. 4, 5, and 40.

[30] See Sanger (1995). Folkerts-Landau and Ito et al. (1995) contains a thorough account of the Mexican peso crisis.

In brief, Congress deliberately placed the Fed outside the appropriations process in order to safeguard its independence. The Fed should not misuse its off-budget status to finance initiatives that are unrelated to monetary policy because there is very little to be gained and much to lose.

6. CONCLUSION

We have assessed the consequences of the Fed's participation in foreign exchange operations. Our analysis was based on the idea that central bank credibility for low inflation is the cornerstone of an effective monetary policy and that public support for Fed independence is the foundation of that credibility.

Distinguishing between sterilized and unsterilized foreign exchange operations, we recognized that as a mechanical matter the Fed can follow the Treasury's lead on sterilized operations without compromising its independence on monetary policy. There is little evidence, however, that sterilized intervention alone can have a sustained effect on the exchange rate. Thus, the Fed's participation in foreign exchange policy with the Treasury creates doubt about whether monetary policy will support domestic or external objectives, and this doubt undermines the credibility of the Fed's longer-term objective of reducing and ultimately eliminating inflation.

Although the Fed is the junior partner with the Treasury on foreign exchange operations, it has been an equal partner when it comes to providing the resources. The Fed can make these resources available without a congressional appropriation because its financial independence puts its open market operations off-budget. Foreign exchange operations initiated by the Treasury involve foreign relationships in ways that can be politically charged, especially when they involve direct loans to foreign governments. We think that Fed financing of such operations risks undermining public respect for its financial independence and with it the credibility of its longer-term price level stability objective.

We argued that central bank independence alone is an inherently fragile basis for the credibility of monetary policy. In view of that fragility, we recommended that the Fed be separated completely from foreign exchange operations. We did not argue that the nation should forsake official foreign exchange operations—only that the Fed, as an independent central bank, should not participate. The Treasury would be free to carry out sterilized operations. Having made this point, we acknowledged that it would be difficult for the Fed to disengage from foreign exchange operations unilaterally. Consequently, some sort of congressional legislation would probably be required to remove the Fed from foreign exchange operations permanently.

In our view, the problems created by the Fed's involvement in foreign exchange operations underscore the need for Congress to provide the Fed with a mandate for price level stability, recognizing a concern for the stabilization of employment and output. Such a mandate would constitute a long overdue replacement for the commitments made at Bretton Woods.[31] Moreover, firm congressional support is needed to strengthen the credibility of the Fed's anti-inflation strategy. By providing an overarching national goal for monetary policy once again, a price stability mandate would greatly reduce the risk of conflicts and credibility problems when the Fed works closely with the Treasury and other parts of the government.

REFERENCES

Barro, Robert, and David Gordon. "Rules, Discretion, and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics,* vol. 12 (July 1983), pp. 101–22.

Blinder, Alan. "Central Banking in Theory and Practice," The Marshall Lecture, Cambridge University. Federal Reserve Board, processed May 1995.

Bordo, Michael D., and Anna J. Schwartz. "What Has Foreign Exchange Market Intervention since the Plaza Agreement Accomplished?" *Open Economies Review,* vol. 2 (No. 1, 1991), pp. 39–64.

Bradsher, Keith. "Lately, A Much More Visible Fed Chief," *The New York Times*, January 17, 1995.

Chandler, Lester V. *Benjamin Strong, Central Banker*. Washington, D.C.: The Brookings Institution, 1958.

Clarke, Stephen V. O. *Central Bank Cooperation: 1924–31*. New York: Federal Reserve Bank of New York, 1967.

Cukierman, Alex. *Central Bank Strategy, Credibility, and Independence*. Cambridge, Mass.: MIT Press, 1992.

Destler, I. M., and C. Randall Henning. *Dollar Politics: Exchange Rate Policymaking in the United States*. Washington, D.C.: Institute for International Economics, 1989.

Edison, Hali J. "The Effectiveness of Central-Bank Intervention: A Survey of the Literature after 1982." Special Papers in International Economics, No. 18. Princeton: Princeton University, Department of Economics, International Finance Section, July 1993.

[31] Our conclusion that a central bank should have its goal legislatively mandated is also the recommendation, for example, of Blinder (1995), Lecture II, p. 16, Friedman (1962), pp. 224–43, and Fischer (1994), p. 316; although the suggested mandates differ from ours and from each other's in certain respects.

"Effort to Bolster Dollar a Failure," *Richmond-Times Dispatch*, June 25, 1994.

Fischer, Stanley. "Modern Central Banking," in Forrest Capie, Stanley Fischer, Charles Goodhart, and Norbert Schnadt, eds., *The Future of Central Banking: The Tercentenary Symposium of the Bank of England*. Cambridge: Cambridge University Press, 1994.

Folkerts-Landau, David, and Takatoshi Ito, et al., "Evolution of the Mexican Peso Crisis," in *International Capital Markets: Developments, Prospects, and Policy Issues*. Washington, D.C.: International Monetary Fund, 1995, pp. 53–69.

Friedman, Milton. "Should There Be an Independent Monetary Authority?" in *In Search of a Monetary Constitution*, Leland Yeager, ed., Cambridge, Mass.: Harvard University Press, 1962.

Friedman, Thomas. "16 Central Banks Are Thwarted in Huge Effort to Prop Up Dollar," *The New York Times*, June 25, 1994.

Funabashi, Yoichi. *Managing the Dollar: From the Plaza to the Louvre*, 2d ed. Institute for International Economics. Washington, D.C., 1989.

Goodfriend, Marvin. "Why We Need an 'Accord' for Federal Reserve Credit Policy," *Journal of Money, Credit, and Banking*, vol. 26 (August 1994), pp. 572–80.

__________. "Interest Rate Policy and the Inflation Scare Problem," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 79 (Winter 1993), pp. 1–24.

__________, and Robert King. "Financial Deregulation, Monetary Policy, and Central Banking," in W. Haraf and R. Kushmeider, eds., *Restructuring Banking and Financial Services in America*. Boston: Kluwer, 1990. Also in Federal Reserve Bank of Richmond *Economic Review*, vol. 74 (May/June 1988), pp. 3–22.

Hetzel, Robert. "Sterilized Foreign Exchange Intervention: The Fed Debate in the 1960s." Federal Reserve Bank of Richmond, 1996.

__________. "Central Banks' Independence in Historical Perspective," *Journal of Monetary Economics*, vol. 25 (January 1990), pp. 165–76.

Humpage, Owen F. "Institutional Aspects of U.S. Intervention," Federal Reserve Bank of Cleveland *Economic Review*, vol. 30 (Quarter 1, 1994), pp. 2–19.

Ireland, Peter N. Long-Term Interest Rates and Inflation: A Fisherian Approach," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 82 (Winter 1996), pp. 21-35.

King, Mervyn. "Credibility and Monetary Policy: Theory and Evidence," Bank of England *Quarterly Bulletin*, vol. 35 (February 1995), pp. 84–91.

Leiderman, Leonardo, and Lars Svensson, eds. *Inflation Targets*. London: Center for Economic Policy Research, 1995.

McCallum, Bennett. "Inflation Targeting in Canada, New Zealand, Sweden, the United Kingdom, and in General," forthcoming in the Bank of Japan Seventh International Conference, 1996a.

__________. *International Monetary Economics*. London: Oxford University Press, 1996b.

Mussa, Michael. "The Role of Intervention," Occasional Paper, No. 6. New York: Group of Thirty, 1981.

Obstfeld, Maurice. "The Effectiveness of Foreign-Exchange Intervention: Recent Experience, 1985–1988" in William H. Branson, Jacob A. Frenkel, and Morris Goldstein, eds., *International Policy Coordination and Exchange Rate Fluctuations*. National Bureau of Economic Research. Chicago: University of Chicago Press, 1990.

Pauls, Dianne. "U.S. Exchange Rate Policy: Bretton Woods to Present," *Federal Reserve Bulletin,* vol. 76 (November 1990), pp. 891–908.

Persson, Torsten, and Guido Tabellini, eds. *Monetary and Fiscal Policy, Vol. 1: Credibility, and Vol. 2: Politics*. Cambridge, Mass.: MIT Press, 1994.

Rose, Andrew, and Lars Svensson. "European Exchange Rate Credibility before the Fall," *European Economic Review,* vol. 38 (June 1994), pp. 1185–1216.

Salomon Brothers and Hutzler. *An Analytical Record of Yields and Yield Spreads: 1950–1968*. New York: Salomon Brothers and Hutzler, 1968.

Sanger, David. "Clinton Offers $20 Billion to Mexico for Peso Rescue; Action Sidesteps Congress," *The New York Times,* February 1, 1995.

Sargent, Thomas J. *Rational Expectations and Inflation*. New York: Harper and Row, 1986.

Stein, Herbert. "The Liberation of Monetary Policy," in *The Fiscal Revolution in America*. Chicago: University of Chicago Press, 1969.

Todd, Walker F. "Disorderly Markets: The Law, History, and Economics of the Exchange Stabilization Fund and U.S. Foreign Exchange Market Intervention," in George G. Kaufman, ed., *Research in Financial Services: Private and Public Policy,* Vol. 4. Greenwich, Conn.: JAI Press, 1992.

U.S. Congress, House of Representatives, Committee on the Budget. *Exchange Stabilization Fund*. Hearings before the Task Force on Tax Expenditures and Off-Budget Agencies of the Committee on the Budget, 94 Cong. 2 Sess. Washington: Government Printing Office, 1976.

U.S. Department of the Treasury. *Exchange Stabilization Fund Annual Report*. Washington: U.S. Department of the Treasury, 1991.

Long-Term Interest Rates and Inflation: A Fisherian Approach

Peter N. Ireland

In recent years, Federal Reserve (Fed) policymakers have come to rely on long-term bond yields to measure the public's long-term inflationary expectations. The long-term bond rate plays a central role in Goodfriend's (1993) narrative account of Fed behavior, 1979–1992, which links policy-related movements in the federal funds rate to changes in the yield on long-term U.S. Treasury bonds. According to Goodfriend, Fed officials interpreted rapid increases in long-term bond rates as the product of rising inflationary expectations, reflecting a deterioration in the credibility of their fight against inflation. To restore that credibility, they responded by tightening monetary policy, that is, by raising the federal funds rate. Mehra (1995) presents statistical results that support Goodfriend's view. Using an econometric model, he demonstrates that changes in long-term bond rates help explain movements in the federal funds rate during the 1980s.

While these studies provide convincing evidence of a link between Fed policy and long-term bond rates, both start with the untested hypothesis that movements in such rates primarily reflect changes in long-term inflationary expectations. And while economic theory does identify expected inflation as one determinant of nominal bond yields, it suggests that there are other determinants as well. Using theory as a guide, this article seeks to measure the contribution each determinant makes in accounting for movements in long-term bond yields. By doing so, it attempts to judge the extent to which Fed policymakers are justified in using these bond yields as indicators of inflationary expectations.

The author thanks Bob Hetzel, Tom Humphrey, Yash Mehra, and Stacey Schreft for helpful comments and suggestions. The opinions expressed herein are the author's and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

Irving Fisher (1907) presents what is perhaps the most famous theory of nominal interest rate determination.[1] According to Fisher's theory of interest, movements in nominal bond yields originate in two sources: changes in real interest rates and changes in expected inflation. Thus, Fisher's theory provides a guide for investigating the extent to which long-term bond yields serve as reliable indicators of long-term inflationary expectations. Specifically, it implies that movements in long-term bond yields provide useful signals of changes in inflationary expectations if and only if their other determinant, the long-term real interest rate, is stable.

Although Fisher acknowledges the potential importance of risk in outlining his theory, he stops short of explicitly considering the effects of uncertainty in his graphical and mathematical treatment of interest rate determination. Recognizing that risk can play a key role in determining interest rates, and exploiting advances in mathematical economics made since Fisher's time, Lucas (1978) develops a model that extends the relationships obtained by Fisher to a setting where future economic magnitudes are uncertain.

In addition to real interest rates and expected inflation, Lucas's model identifies a third determinant of nominal bond yields: a risk premium that compensates investors for holding dollar-denominated bonds in a world of uncertainty. Thus, Lucas's model provides a more exhaustive set of conditions under which movements in long-term bond rates provide useful signals of changes in long-term inflationary expectations: it indicates that the long-term real interest rate must be stable and that the risk premium must be small.

This article draws on Fisherian theory to assess the practical usefulness of long-term bond yields as indicators of long-term inflationary expectations. It begins, in Section 1, by outlining Fisher's original theory of interest. It then shows, in Section 2, how Lucas's model generalizes the relationships derived by Fisher to account for the effects of uncertainty. Section 3 uses Lucas's model to decompose the nominal bond yield into its three components: the real interest rate, the risk premium, and the expected inflation rate. Section 4 applies this procedure to estimate the relative importance of the expected inflation component in explaining movements in long-term U.S. Treasury bond rates. Finally, Section 5 concludes the article.

1. FISHER'S THEORY OF INTEREST

To derive a relationship between the yield on a nominal bond and its determinants, Fisher (1907) considers the behavior of an investor in a simple model economy. The economy has two periods, labelled $t = 0$ and $t = 1$, and a single

[1] Although Fisher is usually identified as the inventor of this theory, Humphrey (1983) argues that its origins extend back to the eighteenth century writings of William Douglass.

consumption good. The consumption good sells for P_0 dollars in period $t = 0$ and is expected to sell for P_1^e dollars in period $t = 1$.

Fisher's investor chooses between two types of assets. The first asset, a nominal bond, costs the investor one dollar in period $t = 0$ and pays him a gross return of R dollars in period $t = 1$. The yield R on this nominal bond measures the economy's nominal interest rate. The second asset, a real bond, costs the investor one unit of the consumption good in period $t = 0$ and returns r units of the good in period $t = 1$. The gross yield r on this bond represents the economy's real interest rate.

In order to purchase a nominal bond in period $t = 0$, the investor must first acquire one dollar; he can do so by selling $1/P_0$ units of the consumption good. When it matures in period $t = 1$, the nominal bond returns R dollars, which the investor expects will buy R/P_1^e units of the good. Measured in terms of goods, therefore, the expected return on the nominal bond equals the investor's receipts, R/P_1^e, divided by his costs $1/P_0$. Letting $\pi^e = P_1^e/P_0$ denote the economy's expected gross rate of inflation, one can write this goods-denominated return as R/π^e.

In equilibrium, the goods-denominated returns on nominal and real bonds must be the same. For suppose the return R/π^e on the nominal bond were to exceed the return r on the real bond. Then every investor could profit by selling the real bond and using the proceeds to purchase the nominal bond. The resulting decrease in the demand for real bonds would raise the return r, while the increase in the demand for nominal bonds would depress the return R/π^e, until the two were brought back into equality. Similarly, any excess in the return r over R/π^e would be eliminated as investors attempted to sell nominal bonds and purchase real bonds. Thus, Fisher concludes that $R/\pi^e = r$ or, equivalently,

$$R = r\pi^e. \tag{1}$$

Fisher's equation (1) expresses the nominal interest rate R as the product of two terms: the real interest rate r and the expected inflation rate π^e. It therefore describes the circumstances under which the nominal bond yield serves as a reliable indicator of inflationary expectations. In particular, it implies that one can be sure that a movement in the nominal interest rate reflects an underlying change in inflationary expectations if and only if the real interest rate is stable.

The nominal bond in Fisher's model resembles a U.S. Treasury bond since, upon maturity, it returns a fixed number of dollars. Thus, the yield on Treasury bonds measures the economy's nominal interest rate R. Unfortunately, assets resembling Fisher's real bond do not currently trade in U.S. financial markets. As a result, it is not possible to directly observe the real interest rate r and then use equation (1) to determine the extent to which movements in Treasury bonds reflect movements in real interest rates rather than inflationary expectations.[2] However, Fisher's theory also links an economy's real interest rate to its growth rate of consumption. Hence, the theory suggests that the real interest rate may be observed indirectly using data on aggregate consumption.

To derive a relationship between the real rate of interest and the growth rate of consumption, Fisher returns to his model economy and uses a graph like that shown in Figure 1. The graph's horizontal axis measures consumption in period $t = 0$, and its vertical axis measures consumption in period $t = 1$.

Fisher's investor receives an income stream consisting of y_0 units of the consumption good in period $t = 0$ and y_1 units of the consumption good in period $t = 1$. He continues to trade in real bonds, which allow him to borrow or lend goods at the real interest rate r. In particular, if y_1 is large relative to y_0, the investor borrows by selling a real bond; this transaction gives him one more unit of the good in period $t = 0$ but requires him to repay r units of the good in period $t = 1$. Conversely, if y_1 is small relative to y_0, the investor lends by purchasing a real bond; this gives him one less unit of the good in period $t = 0$ but pays him a return of r units of the good in period $t = 1$. Thus, the real interest rate r serves as an intertemporal price; it measures the rate at which financial markets allow the investor to exchange goods in period $t = 1$ for goods in period $t = 0$. In Figure 1, the investor's budget line A, which passes through the income point (y_0, y_1), has slope r.

Fisher's investor has preferences over consumption in the two periods that may be described by the utility function

$$U(c_0, c_1) = \ln(c_0) + \beta \ln(c_1), \tag{2}$$

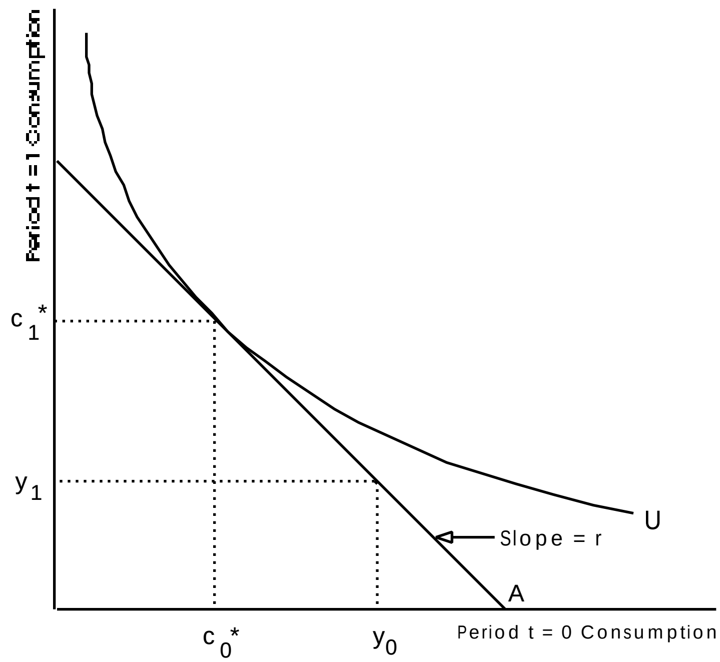
where c_0 denotes his consumption in period $t = 0$, c_1 denotes his consumption in period $t = 1$,ln is the natural logarithm, and the discount factor $\beta < 1$ implies that the investor receives greater utility from a given amount of consumption in period $t = 0$ than from the same amount of consumption in period $t = 1$.[3] In Figure 1, these preferences are represented by the indifference curve U, which traces out the set of all pairs (c_0, c_1) that yield the investor a constant level of utility as measured by equation (2).

The slope of the investor's indifference curve is determined by his marginal rate of intertemporal substitution, the rate at which he is willing to substitute consumption in period $t = 1$ for consumption in period $t = 0$, leaving his utility unchanged. Mathematically, the investor's marginal rate of intertemporal

[2] For exactly this reason, Hetzel (1992) proposes that the U.S. Treasury issue bonds paying a fixed return in terms of goods. Until Hetzel's proposal is implemented, however, only indirect measures of the real interest rate will exist.

[3] Although Fisher does not use a specific utility function to describe his investor's preferences, equation (2) helps to sharpen the implications of his theory by allowing the relationships shown in Figure 1 to be summarized mathematically by equations (3) and (4) below.

Figure 1 Fisher's Diagram



substitution equals the ratio of his marginal utility in period $t = 0$ to his marginal utility in period $t = 1$:

$$\frac{\partial U(c_0, c_1)/\partial c_0}{\partial U(c_0, c_1)/\partial c_1} = \frac{c_1}{\beta c_0}. \tag{3}$$

To maximize his utility, the investor chooses the consumption pair (c_0^*, c_1^*), where the budget line A is tangent to the indifference curve U. At (c_0^*, c_1^*), the slope of the budget line equals the slope of the indifference curve. The former is given by r; the latter is given by equation (3). Hence,

$$r = x/\beta, \tag{4}$$

where $x = c_1^*/c_0^*$ denotes the optimal growth rate of consumption.

Equation (4) shows how Fisher's theory implies that even when the real interest rate r cannot be directly observed, it can still be estimated by computing the growth rate x of aggregate consumption and dividing by the discount factor β. With this estimate in hand, one can use equation (1) to assess the usefulness of Treasury bond yields as indicators of expected inflation. Specifically, if the estimated real rate turns out to be fairly stable, then equation (1) implies that movements in Treasury bond yields primarily reflect changes in inflationary expectations.

2. LUCAS'S GENERALIZATION OF FISHERIAN THEORY

While Fisher recognized that the presence of risk may affect interest rates in important ways, he lacked the tools to incorporate uncertainty formally into his analysis and therefore assumed that his investor receives a perfectly known income stream and faces perfectly known prices and interest rates. More than seventy years later, advances in mathematical economics allowed Lucas (1978) successfully to generalize Fisher's theory to account for the effects of risk.

Lucas's model features an infinite number of periods, labelled $t = 0,1, 2, \ldots,$ and a single consumption good that sells for P_t dollars in period t. Lucas's investor receives an income stream consisting of y_t units of the consumption good in each period t and consumes c_t units of the good in each period t.

Lucas's investor, like Fisher's, trades in two types of assets. A nominal bond costs Lucas's investor one dollar in period t and returns R_t dollars in period $t+1$. Hence, R_t denotes the gross nominal interest rate between periods t and $t+1$. A real bond costs him one unit of the consumption good in period t and returns r_t units of the good in period $t+1$. Hence, r_t denotes the gross real interest rate between periods t and $t+1$. During each period t, the investor purchases B_t nominal bonds and b_t real bonds.

Unlike Fisher's investor, however, Lucas's investor may be uncertain about future prices, income, consumption, interest rates, and bond holdings. That is, he may not learn the exact values of P_t, y_t, c_t, R_t, r_t, B_t, and b_t until the beginning of period t; before then, he regards these variables as random.

As sources of funds during each period t, the investor has y_t units of the consumption good that he receives as income and $r_{t-1}b_{t-1}$ units of the consumption good that he receives as payoff from his maturing real bonds. He also has $R_{t-1}B_{t-1}$ dollars that he receives as payoff from his maturing nominal bonds; he can exchange these dollars for $R_{t-1}B_{t-1}/P_t$ units of the consumption good. As uses of funds, the investor has his consumption purchases, equal to c_t units of the good, and his bond purchases. His real bond purchases cost b_t units of the good, while his nominal bond purchases cost B_t/P_t units of the good. During period t, the investor's sources of funds must be sufficient to cover his uses of funds. Hence, he faces the budget constraint

$$y_t + r_{t-1}b_{t-1} + R_{t-1}B_{t-1}/P_t \geq c_t + b_t + B_t/P_t, \tag{5}$$

which is Lucas's analog to the budget line A in Fisher's Figure 1.

Lucas's investor chooses c_t, B_t, and b_t in each period t to maximize the utility function

$$E_t\left[\sum_{j=0}^{\infty}\beta^j \ln(c_{t+j})\right], \tag{6}$$

subject to the budget constraint (5), where E_t denotes the investor's expectation at the beginning of period t. Equation (6) simply generalizes Fisher's utility function (2) to Lucas's setting with an infinite number of periods and uncertainty. The solution to the investor's problem dictates that

$$1/r_t = \beta E_t\left[1/x_{t+1}\right] \tag{7}$$

and

$$1/R_t = \beta E_t\left[(1/x_{t+1})(1/\pi_{t+1})\right], \tag{8}$$

in each period $t = 0,1,2, \ldots$, where $x_{t+1} = c_{t+1}/c_t$ denotes the gross rate of consumption growth and $\pi_{t+1} = P_{t+1}/P_t$ denotes the gross rate of inflation between periods t and $t+1$.

Lucas's equation (7) generalizes Fisher's equation (4); it is analogous to the tangency between the investor's budget line and indifference curve shown in Figure 1. As in equation (3), the investor's marginal rate of intertemporal substitution is x_{t+1}/β. Hence, equation (7) shows that, under uncertainty, the investor chooses his consumption path so that the expected inverse of his marginal rate of intertemporal substitution equals the inverse of the real interest rate. Also, like Fisher's equation (4), Lucas's equation (7) suggests that while the real interest rate cannot be directly observed, it can still be estimated using data on aggregate consumption.

For any two random variables a and b,

$$E\left[ab\right] = Cov\left[a,b\right] + E\left[a\right]E\left[b\right], \tag{9}$$

where $Cov[a,b]$ denotes the covariance between a and b. Using this fact, one can rewrite equation (8) as

$$1/R_t = \beta Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right] + \beta E_t\left[1/x_{t+1}\right]E_t\left[1/\pi_{t+1}\right], \tag{10}$$

where Cov_t denotes the covariance based on the investor's period t information. In light of equation (7), equation (10) simplifies to

$$1/R_t = \beta Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right] + (1/r_t)E_t\left[1/\pi_{t+1}\right]. \tag{11}$$

Lucas's equation (11) generalizes Fisher's equation (1); it shows how, under uncertainty, the nominal interest rate R_t depends on the real interest rate r_t and the expected inflation term $E_t[1/\pi_{t+1}]$.

The covariance term in equation (11) captures the effect of risk on the nominal interest rate. It appears because random movements in inflation make the goods-denominated return on a nominal bond uncertain. To see this, recall from Section 1 that the return on the nominal bond, measured in terms of the consumption good, equals R_t/π_{t+1}. Since the inflation rate π_{t+1} remains unknown until period $t+1$, so too does R_t/π_{t+1}. Hence, random inflation makes the nominal bond a risky asset.

Equation (11) shows that inflation uncertainty may either increase or decrease the nominal interest rate, depending on whether the covariance term is negative or positive. In particular, inflation uncertainty increases the nominal interest rate if the covariance between $1/x_{t+1}$ and $1/\pi_{t+1}$ is negative, that is, if periods of low consumption growth coincide with periods of high inflation. In this case, high inflation erodes the nominal bond's return R_t/π_{t+1} precisely when the investor, suffering from low consumption growth, finds this loss most burdensome. Hence, the higher nominal yield R_t compensates the investor for this extra risk. Conversely, uncertainty decreases the nominal interest rate if the covariance term in equation (11) is positive, so that periods of high consumption growth coincide with periods of high inflation.

Thus, Lucas's model, like Fisher's, identifies real interest rates and expected inflation as two main determinants of nominal bond yields. Lucas's model goes beyond Fisher's, however, by identifying a third determinant: a risk premium, represented by the covariance term in equation (11), that compensates investors for holding dollar-denominated bonds in the presence of inflation uncertainty. According to Lucas's model, therefore, movements in long-term bond yields accurately reflect changes in expected inflation if and only if the real interest rate is stable and the risk premium is small.

3. DERIVING BOUNDS ON EXPECTED INFLATION

Lucas's equation (7), like Fisher's equation (4), suggests that the unobservable real interest rate can be estimated using data on aggregate consumption. Like the real interest rate, however, the risk premium component of nominal bond yields cannot be directly observed. Hence, without further manipulation, Lucas's equation (11) cannot be used to assess the extent to which movements in nominal bond yields reflect changes in expected inflation rather than changes in their other two components.

Fortunately, as shown by Smith (1993), Lucas's model also places bounds on the plausible size of the risk premium. These bounds, together with estimates of the real interest rate constructed from the consumption data, can be used to determine the extent to which movements in nominal bond yields reflect changes in inflationary expectations.[4]

[4] Smith (1993) takes the opposite approach: he uses the bounds on risk premia, along with estimates of expected inflation, to characterize the behavior of real interest rates.

Recall that the effects of risk enter into Lucas's equation (11) through the covariance term. Smith rewrites this term as

$$\beta Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right] = \beta\rho_t Std_t\left[1/x_{t+1}\right] Std_t\left[1/\pi_{t+1}\right], \tag{12}$$

where

$$\rho_t = Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right]/\left\{Std_t\left[1/x_{t+1}\right] Std_t\left[1/\pi_{t+1}\right]\right\} \tag{13}$$

denotes the correlation between $1/x_{t+1}$ and $1/\pi_{t+1}$ based on the investor's period t information and Std_t denotes the standard deviation based on period t information.

Equation (12) conveniently decomposes the covariance term into three components. The first component, the correlation coefficient ρ_t, can be negative or positive but must lie between -1 and 1. Hence, this component captures the fact that the covariance term may be of either sign. The second and third components, the standard deviations of $1/x_{t+1}$ and $1/\pi_{t+1}$, must be positive. Hence, these terms govern the absolute magnitude of the covariance term, regardless of its sign. Thus, equation (12) places bounds on the size of the covariance term:

$$\begin{aligned}\beta Std_t\left[1/x_{t+1}\right] Std_t\left[1/\pi_{t+1}\right] &\geq \beta Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right] \\ &\geq -\beta Std_t\left[1/x_{t+1}\right] Std_t\left[1/\pi_{t+1}\right],\end{aligned} \tag{14}$$

where the upper bound is attained in the extreme case where $\rho_t = 1$, the lower bound is attained at the opposite extreme where $\rho_t = -1$, and the tightness of the bounds depends on the size of the standard deviations.

Evidence presented in the appendix justifies the additional assumption that inflation volatility in the United States is limited in the sense that the coefficient of variation of $1/\pi_{t+1}$ conditional on period t information is less than one:

$$Std_t\left[1/\pi_{t+1}\right]/E_t\left[1/\pi_{t+1}\right] \leq 1. \tag{15}$$

This assumption allows equation (14) to be rewritten

$$\begin{aligned}\beta Std_t\left[1/x_{t+1}\right] E_t\left[1/\pi_{t+1}\right] &\geq \beta Cov_t\left[(1/x_{t+1}),(1/\pi_{t+1})\right] \\ &\geq -\beta Std_t\left[1/x_{t+1}\right] E_t\left[1/\pi_{t+1}\right],\end{aligned} \tag{16}$$

which, along with equations (7) and (11), implies

$$\begin{aligned}\beta Std_t\left[1/x_{t+1}\right] E_t\left[1/\pi_{t+1}\right] &\geq 1/R_t - \beta E_t\left[1/x_{t+1}\right] E_t\left[1/\pi_{t+1}\right] \\ &\geq -\beta Std_t\left[1/x_{t+1}\right] E_t\left[1/\pi_{t+1}\right],\end{aligned} \tag{17}$$

or, equivalently,

$$\begin{aligned}\beta R_t\left\{E_t\left[1/x_{t+1}\right] + Std_t\left[1/x_{t+1}\right]\right\} &\geq 1/E_t\left[1/\pi_{t+1}\right] \\ &\geq \beta R_t\left\{E_t\left[1/x_{t+1}\right] - Std_t\left[1/x_{t+1}\right]\right\}.\end{aligned} \tag{18}$$

Since

$$1/E_t\left[1/\pi_{t+1}\right] \approx E_t\left[\pi_{t+1}\right], \tag{19}$$

equation (18) places bounds on the expected inflation component that is embedded in the nominal interest rate R_t. Again, these bounds arise because the covariance term in Lucas's equation (11) may be negative or positive and because the absolute magnitude of the covariance term depends on the standard deviation of $1/x_{t+1}$. In particular, the bounds will be tight if this standard deviation—and hence the magnitude of the risk premium—is small. Equation (18) also indicates that these bounds may be estimated using data on aggregate consumption.

Together, therefore, equations (7) and (18) show how one may use data on aggregate consumption to assess the usefulness of nominal bond yields as indicators of inflationary expectations. If the estimates provided by equation (7) show that the real interest rate is stable, and if the bounds provided by equation (18) indicate that the risk premium is small, then Lucas's model implies that most of the variation in the nominal bond yield reflects underlying changes in expected inflation.

4. ESTIMATING THE REAL INTEREST RATE AND BOUNDS ON EXPECTED INFLATION

In order to estimate the real interest rate and the bounds on expected inflation using equations (7) and (18), one must first obtain estimates of the quantities $E_t[1/x_{t+1}]$ and $Std_t[1/x_{t+1}]$. Suppose, in particular, that the evolution of $g_{t+1} = 1/x_{t+1}$, the inverse growth rate of aggregate consumption, is described by the linear time series model

$$g_{t+1} = \gamma + \Gamma(L)g_t + \epsilon_{t+1}, \tag{20}$$

where γ is a constant, $\Gamma(L) = \Gamma_0 + \Gamma_1 L + \Gamma_2 L^2 + ... + \Gamma_k L^k$ is a polynomial in the lag operator L, and ϵ_{t+1} is a random error that satisfies

$$E\left[\epsilon_{t+1}\right] = 0, Std\left[\epsilon_{t+1}\right] = \sigma, E_t\left[\epsilon_{t+1}\epsilon_{t-j}\right] = 0, E\left[\epsilon_{t+1}g_{t-j}\right] = 0 \tag{21}$$

for all $t = 0,1,2, \ . \ . \ .$ and $j = 0,1,2, \ . \ . \ . \ .$ One may then use estimates of γ, $\Gamma(L)$, and σ to compute $E_t[1/x_{t+1}] = E_t[g_{t+1}]$ and $Std_t[1/x_{t+1}] = Std_t[g_{t+1}]$ as

$$E_t\left[g_{t+1}\right] = \gamma + \Gamma(L)g_t \tag{22}$$

and

$$Std_t\left[g_{t+1}\right] = \sigma. \tag{23}$$

Here, as in Mehra (1995), the long-term nominal interest rate is measured by the yield on the ten-year U.S. Treasury bond. This choice for R_t identifies each period in Lucas's model as lasting ten years. In this case, $g_{t+1} = 1/x_{t+1}$

corresponds to the inverse ten-year growth rate of real aggregate consumption of nondurables and services in the United States, converted to per-capita terms by dividing by the size of the noninstitutional civilian population, ages 16 and over. The data are quarterly and run from 1959:1 through 1994:4.

Hansen and Hodrick (1980) note that using quarterly observations of ten-year consumption growth to estimate equation (20) by ordinary least squares yields consistent estimates of γ and the coefficients of $\Gamma(L)$. But since the sampling interval of one quarter is shorter than the model period of ten years, the least squares estimate of σ is biased. Thus, the results reported below are generated using the ordinary least squares estimates of γ and $\Gamma(L)$ and Hansen and Hodrick's consistent estimator of σ, modified as suggested by Newey and West (1987). The limited sample size and the extended length of the model period imply that only one lag of g_{t+1} can be included on the right-hand side of equation (20).

Finally, equation (18) indicates that the discount factor β determines the location of the bounds on expected inflation. Thus, β may be chosen so that the midpoint between the lower and upper bounds, averaged over the sample period, equals the actual inflation rate, averaged over the sample period. This procedure yields the estimate $\beta = 0.856$, which corresponds to an annual discount rate of about 1.5 percent.

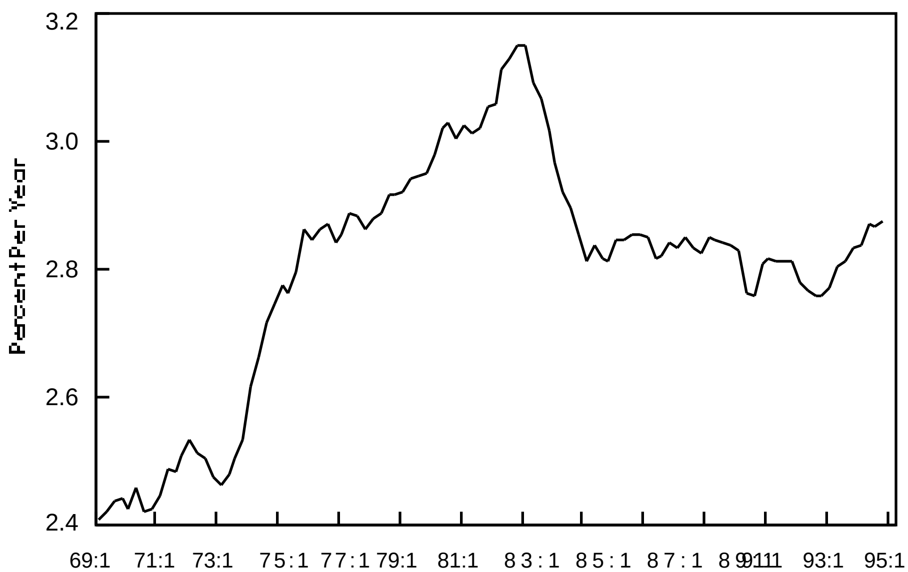
Figure 2 illustrates the behavior of the ten-year real interest rate, estimated using equations (7) and (22).[5] The real interest rate climbs steadily from 1969 until 1983 before falling sharply between 1983 and 1985. But despite these variations, the long-term real interest rate remains within a narrow, 75 basis point range throughout the entire 26-year period for which estimates are available. The average absolute single-quarter movement in the real interest rate is just two basis points; the largest absolute single-quarter move occurs in 1973:4, when the real interest rate increased by only eight basis points. Thus, Figure 2 suggests that the long-term interest rate in the United States is remarkably stable.

Figure 3 plots the bounds on ten-year expected inflation estimated using equations (18), (22), and (23). The bounds are very tight. Even at their widest, in 1981:4, they limit the expected rate of inflation to a 28 basis point band, implying that changes in the risk premium cannot account for movements in the ten-year bond rate larger than 28 basis points. Thus, Figure 3 suggests that the risk premium in the ten-year Treasury bond is very small.

Some intuition for the results shown in Figures 2 and 3 follows from equations (7), (18), (20), (22), and (23). Equation (22), along with equation (7), links variability in the long-term real interest rate to variability in the

[5] Although the sample used to estimate equation (20) extends back to 1959:1, the ten-year model period and the presence of one lag of g_{t+1} on the right-hand side imply that estimates of the real rate can only be constructed for the period beginning in 1969:1.

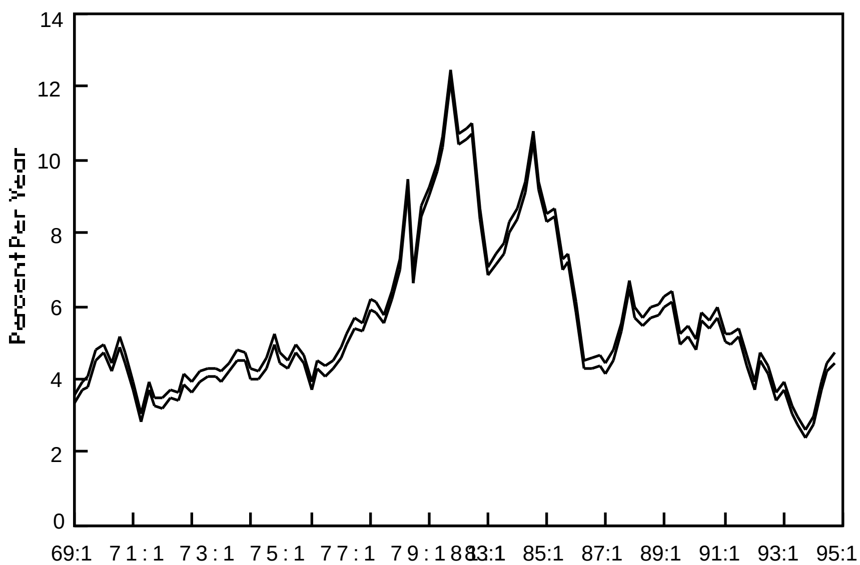
Figure 2 Ten-Year Real Interest Rate



predictable part of consumption growth, measured by $\gamma + \Gamma(L)g_t$ in equation (20). Equation (23), along with equation (18), links the size of the risk premium to variability in the unpredictable part of consumption growth, measured by ϵ_{t+1} in equation (20). In the U.S. data, aggregate consumption growth varies little over ten-year horizons. And since total consumption growth is quite stable, both of its components—the predictable and unpredictable parts—are quite stable as well. Thus, given the stability in aggregate consumption growth, Lucas's model implies that the long-term real interest rate must be quite stable and that the risk premium must be quite small.

According to Lucas's model, the stability of the real interest rate and the small size of the risk premium shown in Figures 2 and 3 imply that most of the variation in the ten-year Treasury bond rate reflects underlying changes in the third component, expected inflation. Indeed, as the largest quarterly real interest rate movement shown in Figure 2 is eight basis points, and as the bounds in Figure 3 are at most 28 basis points wide, the results suggest that any quarterly change in the ten-year bond rate in excess of 36 basis points almost certainly signals a change in inflationary expectations.

Figure 3 Bounds on Ten-Year Expected Inflation



5. CONCLUSION

Although Federal Reserve officials use the yield on long-term Treasury bonds to gauge the public's inflationary expectations, contemporary versions of Fisher's (1907) theory of interest suggest that variations in bond yields can originate in other sources as well. In particular, Lucas's (1978) model indicates that movements on long-term bond yields will accurately signal changes in long-term inflationary expectations if and only if long-term real interest rates are stable and risk premia are small.

Unfortunately, neither real interest rates nor risk premia can be directly observed. However, Lucas's model also shows how these unobservable components of nominal bond yields can be estimated using data on aggregate consumption.

This article lets Lucas's model guide an empirical investigation of the determinants of the ten-year U.S. Treasury bond yield. The results indicate that, indeed, the ten-year real interest rate is quite stable and the ten-year risk premium is quite small. Hence, according to Lucas's model, movements in the long-term bond rate primarily reflect changes in long-term inflationary expectations. Evidently, the Federal Reserve has strong justification for using long-term bond yields as indicators of expected inflation.

APPENDIX

The bounds on expected inflation given by equation (18) were derived in Section 3 under the extra assumption that equation (15) holds. Thus, this appendix provides some justification for (15).

Consider the following linear time series model for the inverse inflation rate $q_{t+1} = 1/\pi_{t+1}$:

$$q_{t+1} = \theta + \Theta(L)q_t + \eta_{t+1}, \tag{24}$$

where the random error η_{t+1} satisfies

$$E\left[\eta_{t+1}\right] = 0, Std\left[\eta_{t+1}\right] = \omega, E\left[\eta_{t+1}\eta_{t-j}\right] = 0, E\left[\eta_{t+1}q_{t-j}\right] = 0 \tag{25}$$

for all $t = 0,1,2, \ldots$ and $j = 0,1,2, \ldots$. One can use this model to estimate $E_t[1/\pi_{t+1}] = E_t[q_{t+1}]$ and $Std_t[1/\pi_{t+1}] = Std_t[q_{t+1}]$, just as equation (20) was used to estimate $E_t[1/x_{t+1}] = E_t[g_{t+1}]$ and $Std_t[1/x_{t+1}] = Std_t[g_{t+1}]$. In the U.S. data, q_{t+1} corresponds to the inverse ten-year growth rate of the price deflator for the aggregate consumption of nondurables and services.

Estimates of (24) using quarterly data from 1959:1 through 1994:4 reveal that $Std_t[q_{t+1}] = \omega = 0.0363$. The smallest estimate of $E_t[q_{t+1}]$ is 0.441, for 1969:1. Thus, for the entire sample period, estimates of $Std_t[1/\pi_{t+1}]/E_t[1/\pi_{t+1}]$ never exceed 0.0823, which suggests that the upper bound of unity imposed by equation (15) is an extremely conservative one.

REFERENCES

Fisher, Irving. *The Rate of Interest.* New York: MacMillan Company, 1907.

Goodfriend, Marvin. "Interest Rate Policy and the Inflation Scare Problem: 1979–1992," Federal Reserve Bank of Richmond *Economic Quarterly,* vol. 79 (Winter 1993), pp. 1–24.

Hansen, Lars Peter, and Robert J. Hodrick. "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy,* vol. 88 (October 1988), pp. 829–53.

Hetzel, Robert L. "Indexed Bonds as an Aid to Monetary Policy," Federal Reserve Bank of Richmond *Economic Review,* vol. 78 (January/February 1992), pp. 13–23.

Humphrey, Thomas M. "The Early History of the Real/Nominal Interest Rate Relationship," Federal Reserve Bank of Richmond *Economic Review,* vol. 69 (May/June 1983), pp. 2–10.

Lucas, Robert E., Jr. "Asset Prices in an Exchange Economy," *Econometrica,* vol. 46 (November 1978), pp. 1429–45.

Mehra, Yash P. "A Federal Funds Rate Equation," Working Paper 95-3. Richmond: Federal Reserve Bank of Richmond, May 1995.

Newey, Whitney K., and Kenneth D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* vol. 55 (May 1987), pp. 703–08.

Smith, Stephen D. "Easily Computable Bounds for Unobservable Real Rates." Manuscript. Federal Reserve Bank of Atlanta, August 1993.

The Early History of the Box Diagram

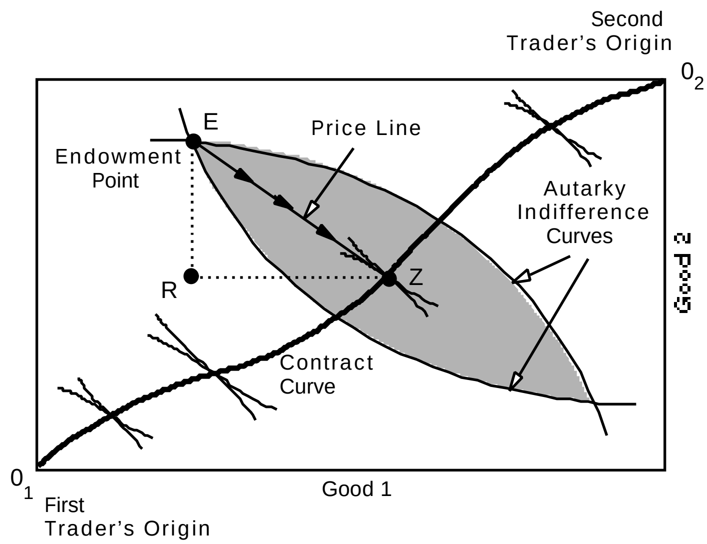
Thomas M. Humphrey

Economists hail it as "a powerful tool," "a work of genius," and "one of the most ingenious geometrical constructions ever devised in economics." It graces the pages of countless textbooks on price theory, welfare economics, and international trade. It is associated with some of the greatest advances ever made in economic theory. It elegantly depicts the two fundamental welfare theorems that are absolutely central to modern economics. In short, it ranks with the preeminent schematic devices of economics since it illuminates the most important ideas economists have to offer. It is none other than the celebrated box diagram used to illustrate efficiency in exchange and resource allocation in hypothetical two-agent, two-good, two-factor models of general economic equilibrium.

The box comes in two variants. The exchange version has dimensions determined by total available stocks of the two goods (see Figure 1). It incorporates traders' indifference maps, one with origin sited in the southwest corner and the other in the northeast corner. The box depicts opportunities for mutually beneficial trade. Thus a movement from initial endowment point *E* to point *Z* on the contract curve—a movement accomplished through a trade of *ER* units of the second good for *RZ* units of the first—benefits both traders simultaneously by putting them on higher indifference curves. In general, so long as the straight trading line *EZ*, whose slope measures the price of the first good in terms of the second, cuts the indifference curves of both parties at point *E*, it pays each to move along that line to the contract curve. Once on the contract curve, however, the potential for further mutually advantageous trades is at an end. Since the contract curve is the locus of indifference-curve tangency points, it follows that movements along the contract curve improve the welfare of one trader only by reducing that of the other.

For valuable comments on earlier drafts of this article, the author is indebted to D. P. O'Brien, J. Patrick Raines, Tibor Scitovsky, and especially to his Richmond Fed colleagues Peter Ireland, Jeff Lacker, Ned Prescott, and John Weinberg.

Figure 1 Exchange Box Diagram



Both traders prefer allocations in the shaded area to initial endowment *E*. Movement down the price vector connecting endowment point *E* to efficiency point *Z* on the contract curve puts both traders on higher indifference curves and thus makes them better off. At point *Z*, however, all potential mutually beneficial trades are at an end. Movements along the contract curve benefit one trader only at the cost of hurting the other.

The alternative production variant of the box depicts the fabrication of two goods from two factor inputs. It replaces indifference maps representing preference functions with isoquant maps representing production functions. It lets available factor quantities determine the dimensions of the box. Efficient factor allocations occur along the contract-curve efficiency locus. There, isoquants are tangent to each other such that the output of one good is maximized given the output of the other.

The chief appeal of the box diagram is its ability to explain much with little. A simple plane diagram, the box can, in Kelvin Lancaster's words, "show the interrelationships between no less than twelve economic variables" ([1957] 1969, p. 52). Moreover, it can do so without resort to algebra and calculus, techniques inaccessible to the mathematically untrained. Small wonder that economists extol the analytical and pedagogical properties of the box or that textbooks feature it as an expository device.

Where some textbooks go astray, however, is in their ahistoric presentation of the diagram. Typically, they say little or nothing about its origins and evolution. They simply present it as an accomplished fact without inquiring into its genealogy. A leading international trade theory textbook authored by Richard Caves and Ronald Jones (1981) provides a prime example. It attributes the box to no progenitor, not even to Francis Edgeworth or Arthur Bowley. The result is that the student is unaware of the circumstances prompting the diagram's development. He knows not who invented it, why it was invented, what problems it originally was designed to solve, or how it evolved under the impact of attempts to perfect it and extend its range of application. Nor can he appreciate the intellectual effort involved in its creation and refinement. Unaware of such matters, he may surmise that the diagram sprang fully developed from the brain of the latest theorist. Ahistoric textbooks indeed foster that very impression. Such are the hazards of disassociating an idea from its historical context and presenting it as a timeless truth.

Far from being timeless, the box diagram possesses a definite chronology. That chronology features some of the leading names in neoclassical and modern economics. Francis Edgeworth, Vilfredo Pareto, A.W. Bowley, Tibor Scitovsky, Wassily Leontief, Kenneth Arrow, Abba Lerner, Wolfgang Stolper, Paul Samuelson, T. M. Rybczynski, and Kelvin Lancaster all contributed to the diagram's development.

Edgeworth invented the exchange box in 1881. He used it to demonstrate the indeterminacy of isolated barter and the determinacy of competitive equilibrium. He showed that all final settlements are on the contract curve, that the competitive equilibrium is one such settlement, and that the contract curve shrinks to the competitive equilibrium as the number of traders increases. Pareto in 1906 demonstrated his celebrated optimality criterion with the aid of the box. Bowley in 1924 generalized Edgeworth's work with his notion of the bargaining locus. Scitovsky in 1941 employed the box to formulate his famous double-bribe test of increased efficiency. Leontief coordinated, consolidated, and clarified the earlier accomplishments in his 1946 rehabilitation of the exchange box. In so doing, he paved the way for the post-war popularity of the diagram. Following hard on Leontief's heels, Arrow in 1951 employed the concepts of convex sets and supporting hyperplanes to analyze the problem of corner solutions on the boundary of the box. And Samuelson in 1952 employed the box to investigate how international transfers affect the terms of trade.

When the foregoing contributions threatened to exhaust the analytical potential of the exchange box, economists turned to the alternative production version. Already, Lerner had presented the first production box in a pioneering 1933 paper whose publication unfortunately was delayed for nineteen years. In the meantime, Stolper and Samuelson published the first production box diagram to appear in print. As employed by them in 1941, by Rybczynski in 1954, and by Lancaster in 1957, the production diagram proved indispensable to the derivation and illumination of certain core propositions of the emerging Heckscher-Ohlin theory of international trade.

The paragraphs below attempt to trace this evolution and to identify specific contributions to it. Besides unearthing lost or forgotten insights, such an exercise may serve as a partial antidote to the textbooks' ahistorical treatment of the diagram. One conclusion emerges: namely that the box hardly developed autonomously. Rather it evolved in a two-way interaction with its applications. Thus an unsolved puzzle in microeconomics prompted the invention of the box—a prime example of a seemingly intractable problem inducing the very tool required for its solution. The resulting availability of the diagram then spurred economists to find new applications for it. These new uses in turn triggered modifications of the diagram. Applications were both cause and effect of the diagram's development.

1. FRANCIS Y. EDGEWORTH

The box diagram makes its first appearance on pages 28 and 113 of Francis Edgeworth's 1881 *Mathematical Psychics*. Motivated by a problem in microeconomic theory, Edgeworth invented the diagram and its constituent indifference-map and contract-curve components to solve the problem.

Edgeworth's predecessors had long known that equilibrium price in isolated, two-party exchange is indeterminate. They also understood that equilibrium between numerous buyers and sellers operating in competitive markets is determinate. But they had been unable to reconcile the two results. They could not show rigorously how increasing numbers lead to price determinacy.

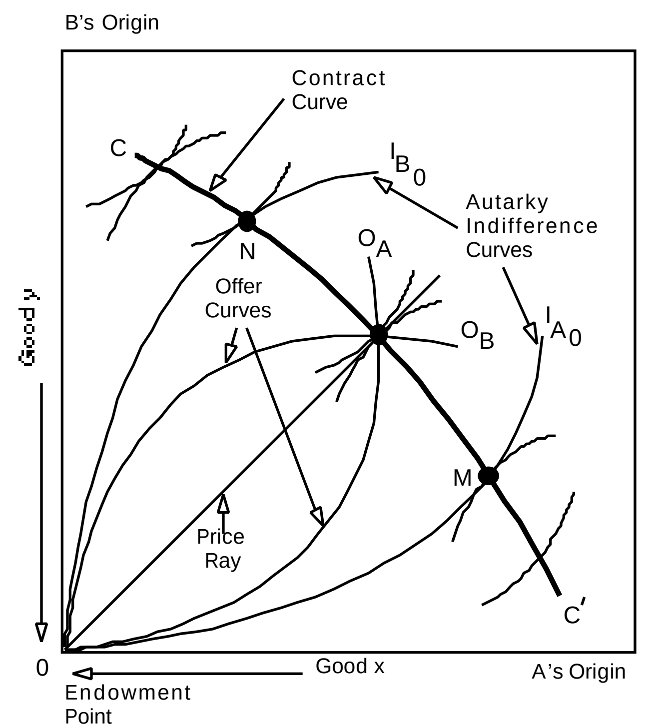
This task Edgeworth sought to accomplish. Using the box diagram, he established (1) that final outcomes must be on the contract curve, (2) that the contract curve shrinks as the number of competitors increases, (3) that competitive equilibrium is one point on the contract curve, and therefore (4) that as the number of competitors increases without limit the contract curve shrinks to a single point, namely the competitive equilibrium.[1] Here was his rationale for inventing the diagram.

Edgeworth's Invention and its Components

Edgeworth's diagram depicts two isolated individuals, *A* and *B*, trading fixed stocks of two goods, *x* and *y*, whose quantities determine the dimensions of the box (see Figure 2). Individual *A* initially holds the entire stock of good *x* and individual *B* the entire stock of good *y*. Superimposing indifference maps on the box, Edgeworth sites the origin of *A*'s map in the lower right corner and the origin of *B*'s map in the upper left corner. This arrangement fixes point 0

[1] In the case of multiple competitive equilibria, the contract curve shrinks not to one but to several points. Edgeworth recognized such a possibility. But he tended to focus on the case of singular rather than multiple equilibrium. See Newman (1990, p. 261).

Figure 2 Edgeworth's Version of the Box



Edgeworth's original diagram depicts three components: (1) autarky indifference curves going through the endowment point and defining the cigar-shaped area of mutually advantageous trades; (2) offer curves or sets of points of tangency of indifference curves and the price ray as it swings about the endowment point; and (3) the contract curve, whose relevant segment lies between the autarky indifference curves.

in the lower left corner as the endowment point of both individuals. That is, the length of the lower horizontal axis measured from right to left indicates the amount of good *x* held by *A* just as the length of the left vertical axis measured from top to bottom indicates the amount of good *y* held by *B*. Since *A* holds no *y* nor *B* any *x*, these axes establish the endowment point.

From the indifference curves radiating outward from their respective origins, Edgeworth selects one particular curve for each trader, namely the curves passing through the endowment point. These curves show alternative combinations of goods that yield the same satisfaction as the endowment bundle. They indicate the level of utility each person would enjoy if he consumed his endowment bundle and refrained from exchange. They also trace out the zone of mutually beneficial exchanges that make both traders better off than they would be under autarky.

Next, Edgeworth draws in the contract curve CC' along which indifference curves are tangent such that one trader cannot occupy a higher indifference curve unless the other is forced to occupy a lower one. Especially significant is the portion of the contract curve bounded by the autarky indifference curves. Since traders require that potential exchanges make them at least as well off as they would be under autarky, they will never voluntarily agree to trades outside those bounds. It follows that the relevant segment of the contract curve lies in the lens-shaped area between the indifference curves going through the endowment point.

Finally, Edgeworth sketches traders' reciprocal demand schedules or offer curves. These curves apply to the special case where the two traders act as representative price-takers operating on opposite sides of a competitive market. Offer curves show how much each trader is willing to exchange at all possible prices. Edgeworth of course did not invent such curves. That honor goes to Alfred Marshall. But he was the first to derive them as the locus of points of tangency of indifference curves and the price ray as it pivots about the endowment point. He likewise was the first to explain that each point on an offer curve represents an outcome of constrained utility maximization in which the commodity price ratio, or slope of the price ray, equals the ratio of marginal utilities, or slope of the indifference curves.

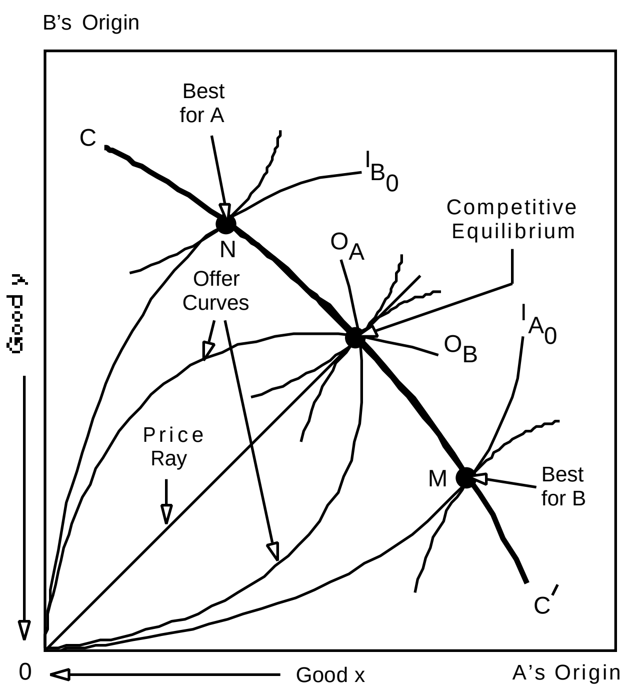
Exploiting Potential Mutual Gains from Exchange

Having derived the exchange box and its constituent components, Edgeworth employed it to illuminate five basic propositions. His first proposition states that final settlements must be on the contract curve. At any other point, both parties could make themselves better off by renegotiation. Consider any point lying off the contract curve. Going through that point are intersecting indifference curves enclosing a cigar-shaped area that spells unexploited potential mutual gains from exchange. Traders will not let such opportunities go unrealized. Instead, they will exploit them until they reach the contract curve where indifference curves are tangent and further mutual gains are at an end.

Efficiency of Competitive Equilibrium

Edgeworth's second proposition refers to the efficiency of competitive equilibrium. It states that such equilibrium is always on the contract curve. The reason? Competition establishes a common, market-clearing price ratio. Competitive price-takers independently respond to that ratio by trading at the point where each supplies the quantity the other demands and vice versa. That is, price-takers operate at the point where their offer curves intersect (see

Figure 3 Competitive Equilibrium versus the Indeterminacy of Two-Party Barter



Bargaining between two isolated traders can lead to an outcome anywhere on the segment of the contract curve between the autarky indifference curves. By contrast, the competitive equilibrium is uniquely determined at the intersection of the price ray and offer curves. There the traders' indifference curves are tangent to the common price ray and thus to each other.

Figure 3). At this point, indifference curves are tangent to the common price ray emanating from the endowment point and thus are tangent to each other. Since such tangencies occur only on the contract curve, it follows that competitive equilibrium is on that same curve.

Indeterminacy of Isolated Two-Party Barter

Edgeworth's third proposition refers to the indeterminacy of isolated two-party exchange. In such bilateral monopoly situations, it is impossible to determine, from indifference maps and endowments alone, the precise price-quantity equilibrium that will emerge. All one can say is that equilibrium must lie on the segment of the contract curve between the autarky indifference curves. But which one of the infinity of possible equilibria will prevail will depend upon considerations external to Edgeworth's model, namely the relative bargaining skills and strengths of the traders as well as the strategies and tactics they employ. Economists traditionally have had little to say about such matters.[2] They cannot confidently predict any unique outcome. The precise ingredients of shrewd, effective bargaining remain subtle, elusive, and obscure. Still, economists can note that the gain one bargainer gets from exchange is limited only by the other's effort to get the best for himself. Final settlement will be near point *N* if *A* is the superior bargainer. It will be nearer to point *M* if *B* has the bargaining advantage.

Neither outcome, Edgeworth noted, necessarily coincides with the point of maximum aggregate welfare on the contract curve. There the sum of the traders' satisfactions is at its peak. Identifying this unique maximum point of course requires that utility be cardinally measurable and comparable across individuals—properties Edgeworth thought utility possessed. It was on these grounds that he advanced his famous principle of arbitration. Compulsory arbitration, he argued, could do what unrestricted bargaining could not do. By imposing the utilitarian sum-of-satisfactions solution on the bilateral monopolists, arbitration would yield a determinate, socially optimum outcome. Conversely, in the absence of such arbitration indeterminacy would continue to characterize the isolated two-party case.

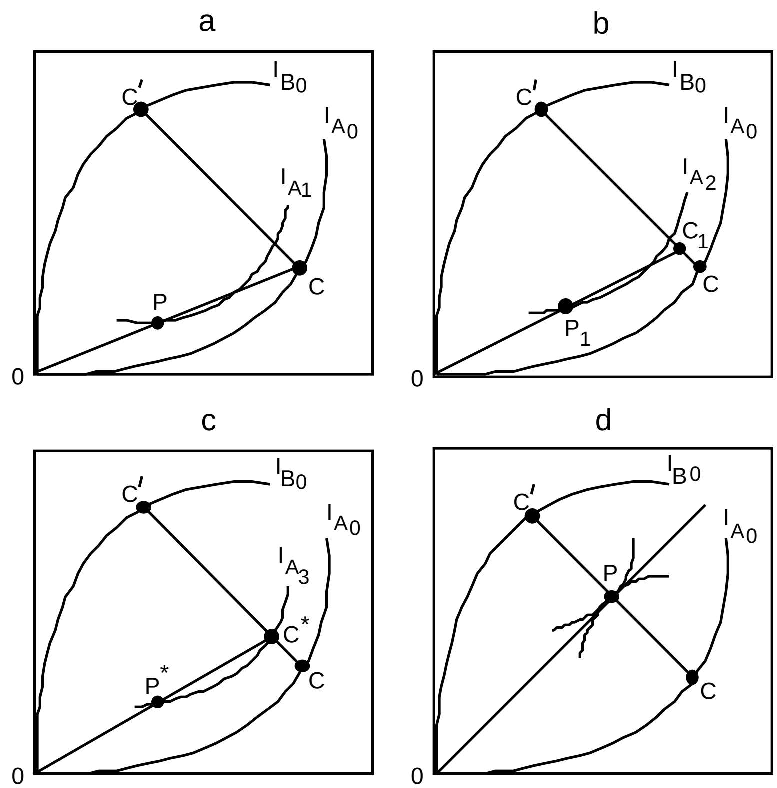
Recontracting and the Role of Numbers

Edgeworth's fourth proposition, his recontracting theorem, refers to the role of numbers in reducing indeterminacy. It states that as the number of traders gets large, the contract curve shrinks to a single point, the competitive equilibrium.

Edgeworth sketches a proof on pages 35–37 of his *Mathematical Psychics* (see Creedy [1992], pp. 158–65, for a particularly clear and insightful interpretation). He starts with the two-person case in which party *A* provisionally contracts with party *B* to reach point *C* on *A*'s indifference curve I_{A_0} (see Figure 4a). He then introduces a new pair of traders identical to the first pair. This

[2] They have said something, however. John Nash (1950) thought the bargainers might agree to maximize the multiplicative product of their respective utility gains from trade (the excess of post-trade over autarky levels of satisfaction). John Harsanyi (1956) showed that the Danish economist Frederick Zeuthen (1930) had proposed essentially the same solution two decades before Nash. Ariel Rubinstein (1982) showed that the Nash solution is the outcome of a noncooperative, offer-counteroffer game. John Creedy (1992, pp. 193–99) suggested a variant of the Nash solution, namely the maximization of a geometrical weighted average of the traders' utility gains, with the weights measuring the relative bargaining powers of the two parties. These solutions establish unique potential agreement points on the contract curve. Since it is unlikely that the bargainers would always agree to go to such proposed points, however, indeterminacy remains.

Figure 4 Edgeworth's Recontracting Process



Recontracting plus convexity of indifference curves implies that the contract curve shrinks as traders become more numerous. With a single pair of traders, the contract curve is *CC*' as shown in panel a. Adding another pair shrinks the curve by the amount *CC** at both ends (panel c). When the number of pairs gets very large, the curve shrinks to the competitive equilibrium (panel d).

maneuver allows him to use the same box diagram to deal with four parties. It permits him to represent the preferences of each of the *A*s (and the *B*s) with a single indifference map.

It also means that the agreement reached at point *C* cannot be final. For the two *A*s can now ignore one of the *B*s and deal with the other at point *C*. When they split the resulting bundle equally among themselves, they will each reach the half-way point *P* on the trade vector 0*C*. That point, because of the convexity of indifference curves, is on a higher such curve than before. Thus the *A*s are better off, their trading partner *B* is just as well off as initially, and the excluded *B* is at his endowment point.

In retaliation, the excluded *B* then underbids his competitor by offering the *A*s a trade on better terms at point C_1 (see Figure 4b). The *A*s, by accepting, can share the resulting bundle among themselves to each attain point P_1 on a still higher indifference curve than before. Repeated recontracting brings the parties to point C^* (see Figure 4c). There the *A*s are indifferent between (1) trading with both *B*s at C^* and (2) dealing with just one *B* at C^* and splitting the resulting bundle at point P^*. Either option puts them on the same indifference curve.

There being no advantage to choosing option 2 over option 1, the *A*s will trade with the two *B*s at point C^*. The result is that recontracting, which began at point *C*, ends at point C^*. Adding a trader to each side of the market shrinks the contract curve by the amount CC^*. The same logic of course applies to point C' at the other end of the contract curve. Recontracting initiated there shrinks the contract curve inward as each of the *A*s continually underbids the other to attract the business of the *B*s. In other words, the contract curve shrinks at both ends.

Although two pairs of traders shrink the range of indeterminacy, they hardly eliminate it. To reduce it further, Edgeworth adds a third pair. Doing so gives room for two *B*s to underbid the third for the patronage of the *A*s. Dealing with the two *B*s, the three *A*s each can reach a point *P* two-thirds the distance from the origin to any point *C* on the contract curve. Final settlement occurs when point *C* shrinks inward sufficiently to lie on the same *A*-indifference curve as point *P*. The same reasoning holds for the other end of the contract curve, which of course shrinks too.

Let the number of pairs of traders *N* grow without limit. Then point *P*, which according to Edgeworth is $(N - 1)/N$ times the distance from the origin to point *C*, converges on that latter point. Expressed geometrically, final settlement in the large-numbers case occurs where an *A*-indifference curve is tangent to a ray from the origin (see Figure 4d). The same holds true for an indifference curve of the *B*s. The result is that both indifference curves are tangent to the same ray and thus to each other just as in the competitive equilibrium. Large numbers shrink the contract curve to the point of competitive equilibrium.

Monopoly Pricing—An Exception to Edgeworth's Rule?

Finally, Edgeworth considered a case that apparently violated his postulate that final settlements lie on the contract curve. That case has two bargainers agreeing on price but making no agreement on the quantities to be traded. An extreme example confronts a representative competitive price-taker with a monopolistic price-maker. The monopolist is of the simple, or non-price-discriminating, variety. He sets a single price for all units exchanged and leaves the competitor free to determine how much he (the competitor) wants to trade at that price along his offer curve.

Let *A* be the representative competitor and *B* the monopolist. If *B*'s monopoly power is absolute, he will set the single price that puts him on his highest attainable indifference curve given *A*'s offer curve (see Figure 5). That is, he chooses the price that takes him to point *Q*, where his indifference curve just touches *A*'s offer curve. Of course, if his monopoly power is less than absolute, his fear of losing *A*'s patronage to potential rival traders may induce him to charge the slightly lower price shown by the slope of ray 0*q*. In any case, the result is that trade takes place at a point like *Q* (or *q*) on *A*'s offer curve rather than on the contract curve. Here is an apparent exception to the rule that final settlements tend to be efficient.

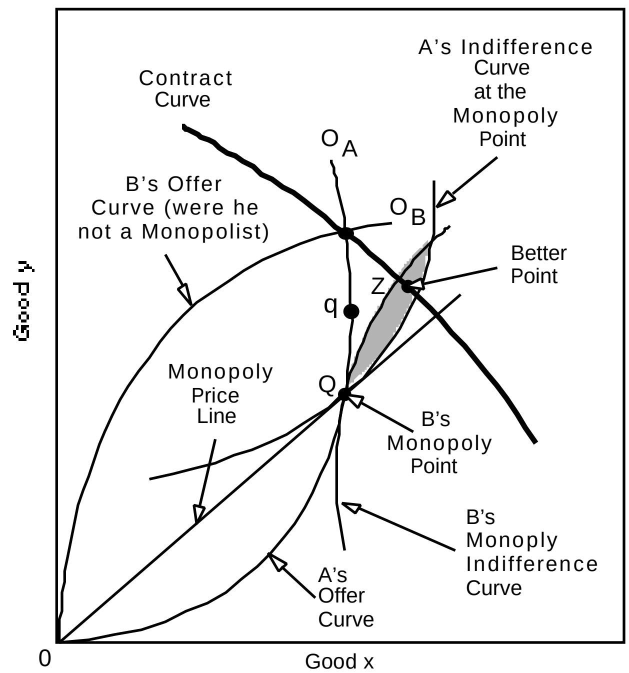
Edgeworth was quick to point out, however, that the exception stems from the assumption that the parties contract over price alone. Were they to contract over quantity as well, they both could move advantageously to the contract curve. Thus Edgeworth questioned the validity of the assumption. To him, rational behavior required that parties bargain over both price and quantity dimensions of a deal, especially when it was to their mutual advantage to do so.

In illustration, Edgeworth referred again to monopoly point *Q* reached through a price-only contract. From that point, superior outcomes are possible in the sense that both parties can move to higher indifference curves than those crossing through *Q*. Edgeworth realized, however, that such improved positions would never be attained by new price settings alone. For, given that the monopolist is constrained by the competitor's offer curve, any change in the slope of the price line 0*Q* would make him (the monopolist) worse off than he is at *Q* and for that reason would be resisted. But mutually beneficial positions could be reached if the competitor somehow could be induced to leave his offer curve. Such an inducement could take the form of a new contract specifying quantity as well as price.

For example, monopolist *B* might dictate terms corresponding to point *Z*, thus improving his own welfare. He would lower the price against himself in exchange for a more-than-compensating rise in quantity traded. And he would do so confident that *A* would gladly agree to the larger trade volume in return for the guarantee of a lower price. In other words, *A* would concur with any price-quantity package moving him to an indifference curve higher than the one he would otherwise occupy at point *Q*. And if such a negotiated package fell short of the contract curve, the parties could renegotiate other packages until they finally arrived there.[3]

[3] Tibor Scitovsky, in his classic 1942 article "A Reconsideration of the Theory of Tariffs," showed that the parties could reach point *Z* by an alternative route. Competitor *A* could bribe monopolist *B* to act as a competitor operating on his own offer curve. The bribe, paid in *A*'s own good, would result in a rightward shift of the endowment point and its attendant offer curves by the amount of the payment. So shifted, the offer curves would intersect at point *Z*. The monopolist would gain from the bribe and the price-taker would gain from the lower, competitive price. Edgeworth, however, said nothing of this scheme.

Figure 5 Monopoly Outcomes



Monopolist *B* sets the price that puts him on his highest attainable indifference curve given the offer curve of the competitive price-taker *A*. The monopolist goes to tangency point *Q*. There, however, *A*'s indifference curve is tangent to the price ray and not to *B*'s indifference curve. The resulting intersecting indifference curves create a lens-shaped area of unexploited mutually beneficial exchanges. If the parties could agree to let monopolist *B* set both price and quantity, they could move to efficient point *Z* where both are better off than at point *Q*.

In short, Edgeworth showed that agreements fixing both price and quantity inevitably lead to the contract curve. By contrast, agreements limited to establishing price alone may, under certain circumstances, lead only to the offer curve. But he insisted that rational agents have an incentive to choose the former agreements over the latter. Thus all final settlements tend to be on the contract curve.

Appraisal

Edgeworth's contribution must be judged one of the greatest virtuoso performances in the history of economics. Going beyond the mere creation of the box diagram itself, he invented its principal components, the indifference map and the contract curve. True, he did not invent offer curves. But he did give the earliest demonstration of their derivation from the underlying indifference contours and price ray. Moreover, in showing that offer curves intersect at the contract curve, he was the first to use them to demonstrate the efficiency of competitive equilibrium.

Edgeworth's work is remarkable in another respect. His five propositions essentially point the way to all of modern economics. One finds in them both a treatment of competitive equilibrium and its efficiency in exhausting the gains from trade and a framing of the problems that arise when perfect competition ceases to prevail. These problems arguably constitute the fundamental motivation for the development of game theory.

Indeed, Edgeworth himself contributed to this development by anticipating key game-theoretic ideas. He demonstrated that final allocations must lie on the segment of the contract curve spanning the indifference curves going through the endowment point. In so doing, he identified what game theorists some seventy-five years later were to call the *core* of the economy. And, in illustrating that the contract curve shrinks to a single point, he showed how the core behaves as its agents increase in number. Finally, his recontracting theory foreshadowed the game-theoretic notion that no coalition of traders can block the emergence of competitive equilibrium. Mark Blaug (1986, p. 70) said it all when he described Edgeworth's theory of the core as "his most beautiful contribution."

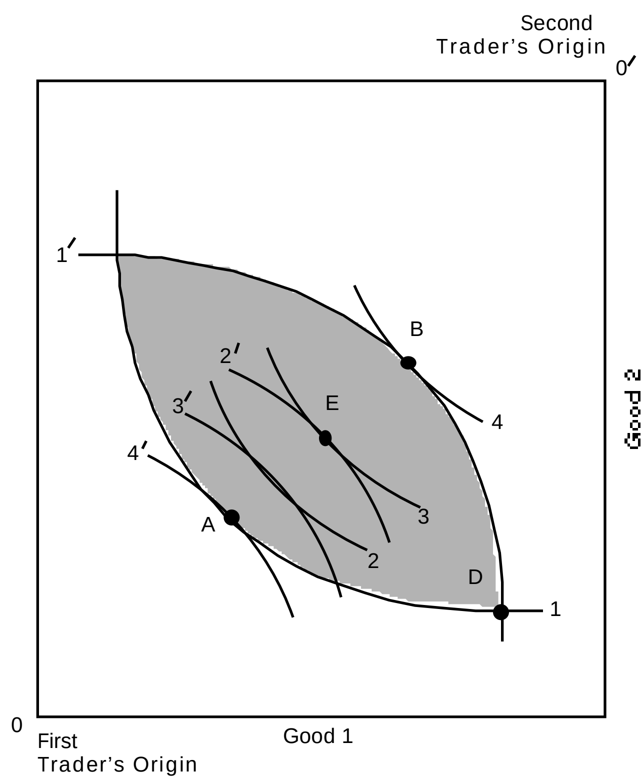
2. VILFREDO PARETO

The next to present the box diagram was Vilfredo Pareto, who did so in his 1906 *Manuale d'economie politica*. Pareto's work obviously owes much to Edgeworth. Indeed, commentators including Maffeo Pantaleoni (1923, p. 584) and John Creedy (1980, p. 272) have stressed that very point. But Pareto also modified Edgeworth's work in at least two key respects.

For one thing, he presented the box in its now-conventional form. That is, he located the origins of the indifference maps in the southwest and northeast corners, respectively, rather than in the other two corners as Edgeworth had done (see Figure 6). The result was that the succession of indifference-curve tangency points—Pareto did not draw the efficiency locus—sloped upward from left to right rather than downward as in Edgeworth's version.

Second and more important was Pareto's interpretation of the welfare implications of the box. Unlike Edgeworth, who believed that interpersonal comparisons of utility make it possible in principle to identify a unique point of maximum aggregate welfare on the contract curve, Pareto denied that such comparisons could be made and indeed refused to make them.

Figure 6 Pareto's Diagram



Movements from point *D* to any point within the lens-shaped area are unambiguously welfare-improving since both parties gain and neither lose. But movements across tangency points *A*, *E*, and *B* are welfare-ambiguous and defy comparison since one party gains while the other loses.

Pareto Optimality

Accordingly, he held that only outcomes involving gains for some and losses for none are unambiguously welfare-improving just as outcomes involving gains for none and losses for some are unambiguously welfare-decreasing. By contrast, outcomes involving gains for some and losses for others are ambiguous. They cannot be judged in terms of quantitative utility comparisons. The inadmissibility of interpersonal comparisons of utility (or "ophelimity" as Pareto termed the utility concept) foils their evaluation.

It follows that movements from points like *D*, where indifference curves cross, to points like *A*, *E*, and *B*, where the curves are tangent, constitute

Pareto-superior moves. They put at least one party on a higher indifference curve and none on a lower one. But movements across successive tangency points like *A*, *E*, and *B*, involving as they do higher curves for one person and lower curves for the other, defy comparison. An infinity of such Pareto-optimal points exists, none of which can be judged superior to the others.

In short, there is no single point of maximum welfare, Edgeworth's claim to the contrary notwithstanding. All one can say is that points off the tangency locus are economically inefficient since everyone could gain by moving to a point at which no mutually advantageous reallocations are possible. Likewise, points on the locus are economically efficient in the sense that no reallocation could improve the position of both parties. Edgeworth's notion of a unique welfare optimum gave way to Pareto's notion of an infinity of noncomparable optima.

3. ARTHUR W. BOWLEY

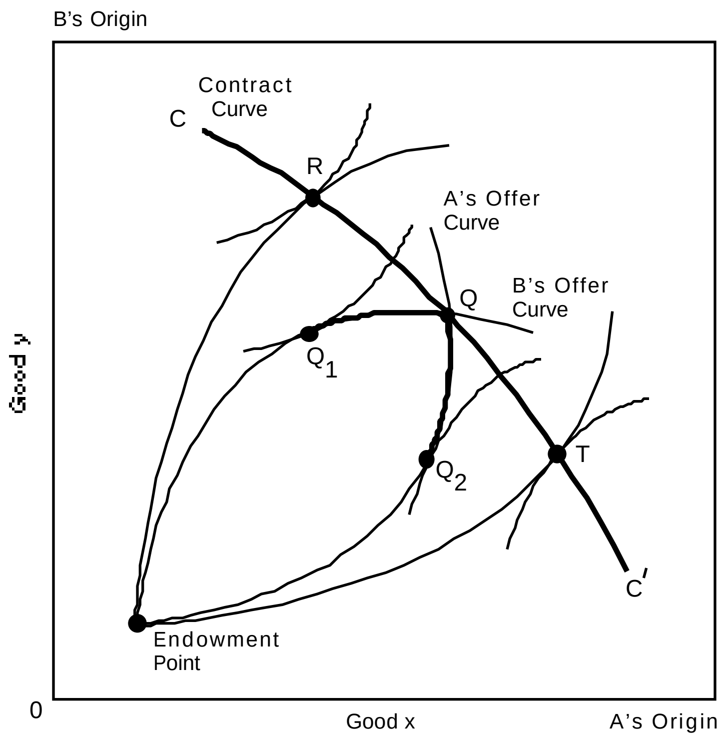
After Pareto's *Manuale*, fully eighteen years elapsed before the box diagram made its next appearance in A.W. Bowley's famous 1924 *Mathematical Groundwork of Economics*. Inspired by Edgeworth and Pareto, Bowley generalized and extended their work in three ways. First, he replaced their assumption that each hypothetical trader initially holds the entire stock of one good and none of the other. He replaced it with the alternative assumption that each trader initially holds some of both goods. The result was to fix the endowment point in the interior of the box rather that at one of its corners (see Figure 7). Bowley's innovation is conventional practice today.

Bargaining Locus

Second, he supplemented Edgeworth's analysis of bilateral monopoly with his concept of the bargaining locus. In defining that locus, which consists of the offer-curve segments Q_1QQ_2, Bowley argued as follows. If the two parties contract over price alone, equilibrium may well be on the offer curves rather than on the contract curve. The party possessing the superior bargaining power will set the price and leave the other free to determine the trade volume at that price along his offer curve. Accordingly, the outcome will be somewhere on the price-taker's offer curve.

Suppose *B* is the price-maker whose bargaining superiority is absolute. He will set the price to reach point Q_2 where his highest attainable indifference curve just touches *A*'s offer curve. But if his bargaining superiority is somewhat weakened by the countervailing bargaining skills of *A*, he will be forced to shade his price downward and occupy a position on *A*'s offer curve in the direction of competitive point *Q*. These considerations trace out the lower Q_2Q segment of the bargaining locus.

Figure 7 Bowley's Version of the Box



Bargaining between price-making, price-taking traders establishes the curve Q_1QQ_2 as the locus of final outcomes. Final settlement occurs on the upper or lower segment depending upon whether *A* or *B* is the dominant bargainer. If both parties possess equal and offsetting monopoly power, final settlement occurs at *Q*.

Similarly, if *A* is the price-maker, trade will occur at the point of intersection of the price ray he sets and *B*'s offer curve. Trader *A* will aim at reaching point Q_1, where the offer curve is tangent to his highest attainable indifference curve. But if *A*'s bargaining power is less than absolute, he may be forced to lower the price against himself and thus move to a point on *B*'s offer curve to the right of point Q_1. These considerations establish the upper Q_1Q segment of the bargaining locus.

The upshot is that if either one trader or the other sets the price, trade occurs at some point on the combined upper and lower segments of the offer curves between points Q_1 and Q_2.[4] With the single exception of point Q, where equal and offsetting bargaining power yields the competitive equilibrium, all these points are off the contract curve. Thus Bowley confirms Edgeworth's contention that when price-maker confronts price-taker over price alone the outcome is rarely efficient.

Trading at Disequilibrium Prices

Finally, Bowley advanced an alternative to Edgeworth's treatment of how the economy converges to its core. As mentioned above, Edgeworth, in considering such convergence, ruled out trading at disequilibrium prices. For him, contracts become binding and exchanges occur only at final equilibrium prices corresponding to points on the contract curve. Disequilibrium contracts he treated as tentative, provisional, non-binding, and subject to revision until the equilibrium contract emerged.

By contrast, Bowley permitted exchanges to take place at disequilibrium prices. He envisioned traders moving across a succession of intermediate positions in the lens-shaped area enclosed by indifference curves emanating from the endowment point. From each such intermediate trading position, they would move to a subsequent, Pareto-improving one changing the price as they went. They would continue in this fashion until they reached the core. The resulting path to equilibrium is described by a broken, or segmented, price line and final settlement can occur anywhere on the section *RT* of the contract curve.

For all its apparent realism, however, Bowley's analysis comes at a high cost. It greatly complicates the diagram. Each disequilibrium trade means a new allocation of goods such that the endowment point shifts continually. Since offer curves emanate from such endowment points, a new set of offer curves has to be drawn at each stage of the process. The result is to clutter the diagram unduly. For this reason, Edgeworth's simplification seems superior pedagogically to Bowley's treatment.

4. TIBOR SCITOVSKY

In the twenty-two years following the publication of Bowley's *Mathematical Groundwork*, the exchange box virtually disappeared from the literature. It surfaced briefly in 1941 when Tibor Scitovsky employed it to expose a flaw

[4] This result holds even when both bargainers, after agreeing on a noncompetitive price, treat it as given and act as price-takers operating on their respective offer curves. In this special case, the price ray will cut the respective offer curves at different points. One party, in other words, will wish to trade a larger quantity at the bargained price than will the other. Here, the smaller quantity will be the one actually traded. The outcome will be exactly the same as if one party unilaterally set the price (see Scitovsky 1951, p. 418).

in compensation tests of increased efficiency. Nicholas Kaldor and John R. Hicks had proposed such tests to circumvent Pareto's prohibition banning the evaluation of changes favoring some people while hurting others. Applied to such situations, the compensation test was supposed to reveal whether a change from one non-optimal state to another was, on balance, welfare-improving if some gained and some lost. The change was said to pass the test if the gainers could fully compensate the losers and still be better off.

But Scitovsky noted a paradox. The test might reveal both states to be superior to each other. Observe a change-induced reallocation from goods-bundle A to bundle B (see Figure 8). Let points A and A' have the same vertical height with the same being true of points B and B'. Then compensation can be represented as a quantity of the horizontally measured good alone. Gainer J (whose indifference map originates at the lower left) could fully compensate loser I (whose indifference map originates in the upper right) by an amount $B'B$ and still be better off. The hypothetical transfer would leave him occupying a higher indifference curve than the one going through his initial position A. Similarly, in a reverse transition from B to A, individual I could bribe individual J by an amount AA' and still be better off than at B. The test would reveal allocation B as preferred to allocation A. Once at B, however, the same test would reveal A as the superior allocation.

Double-Bribe Criterion

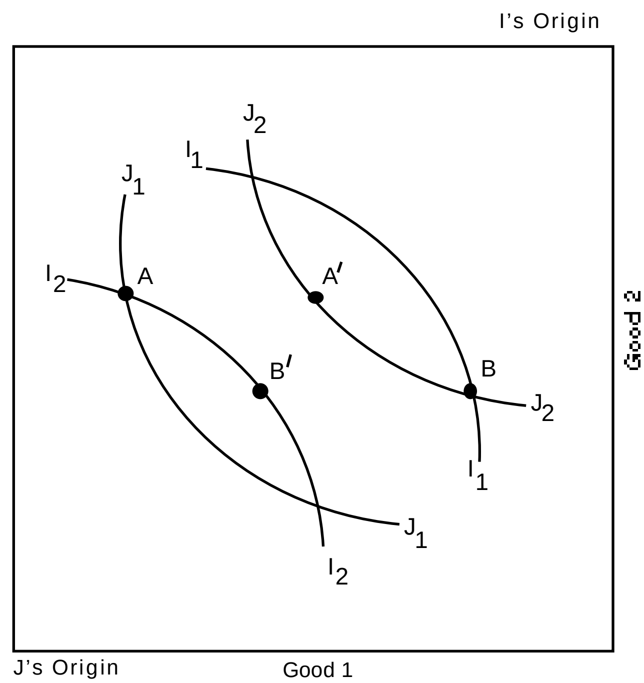
To avoid such contradictions, Scitovsky proposed a double test. Situation B is preferred to situation A if the gainers from the change can profitably compensate the losers, or bribe them to accept it, while the potential losers cannot profitably bribe the gainers to oppose the change.

Scitovsky's double-bribe criterion impressed economists far more than did the box diagram he used to exposit it. For that reason, his paper served merely to interrupt rather than to halt the diagram's pre-World War II lapse into obscurity. That lapse persisted for five more years.

5. WASSILY LEONTIEF

Then came Wassily Leontief's 1946 *Journal of Political Economy* article on "The Pure Theory of the Guaranteed Annual Wage Contract." Employing perhaps the most elaborate version of the exchange box to be found in the scholarly literature of the time, Leontief summarized, consolidated, and clarified all earlier work. He spelled out such notions as the lens-shaped zone of mutually advantageous trades, the contract curve, offer curves, the competitive and simple monopoly (price-maker, price-taker) outcomes and their welfare implications with a lucidity and elegance unmatched in earlier work. In so doing, he reawakened economists to the power and subtlety of the diagram and thus initiated its post-war revival.

Figure 8 Scitovsky's Paradox

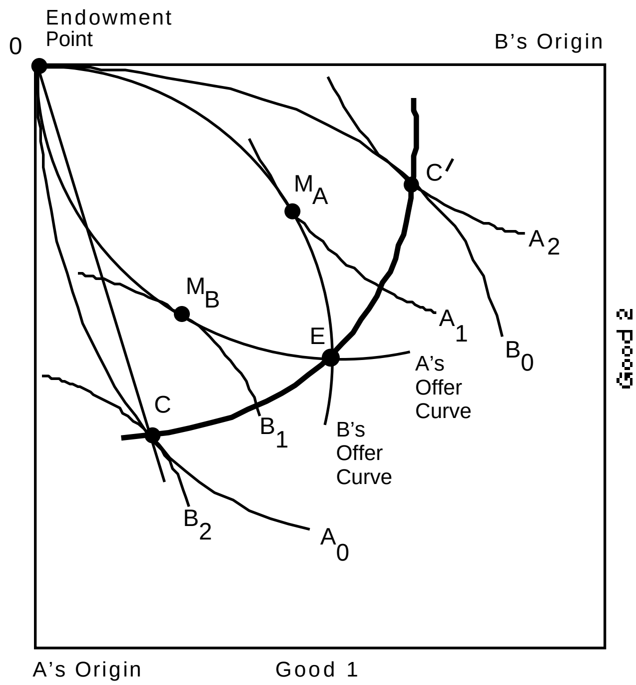


Paradoxically, the Kaldor-Hicks compensation test may justify both a move from situation *A* to situation *B* and a reverse move from *B* back to *A*. In the move from *A* to *B*, agent *J* could compensate agent *I* by the amount *BB*' and still be better off. He would still occupy a higher indifference curve than at *A*. Contrariwise, in the reverse move from *B* to *A*, agent *I* could compensate agent *J* by the amount *AA*' and still be better off than at *B*.

Perfectly Discriminating Monopoly

Leontief's main contribution, however, was to specify exactly how a dominant bargainer might extract for himself all the potential gains from trade. Let that bargainer present his passive counterpart with an all-or-nothing, take-it-or-leave-it option to trade the entire fixed bundle *C* at a fixed price equal to the slope of ray 0*C* (see Figure 9). The passive party either accepts the option or rejects it and remains at his endowment point. Since the option leaves him no worse off than does the autarky outcome, he accepts it. The resulting settlement is at one end of the core, namely at the extreme that yields the dominant party all the gains from exchange.

Figure 9 Perfectly Discriminating Monopolist



Trader *B*, a perfectly discriminating monopolist, captures all the potential gains from trade for himself by going to point *C* on his trading partner *A*'s autarky indifference curve A_0. He presents *A* with an all-or-nothing, take-it-or-leave-it option to trade the fixed bundle *C* at the fixed price denoted by the slope of ray 0*C*. Alternatively, *B* achieves the same result by moving down *A*'s autarky indifference curve, charging the highest price he can get for each successive unit of trade until he arrives at point *C*. Unlike the simple monopoly outcome M_B, the discriminating monopoly outcome is on the contract curve and therefore is efficient.

Leontief further noted that all-or-nothing option contracts are equivalent to perfect price discrimination. With price discrimination, the dominant trader moves along the autarky indifference curve of the passive trader. He does so by charging the highest price he can get for each successive unit of trade—that is, the highest price his partner is willing to pay rather than do without the unit—until he (the dominant trader) reaches the core at a point most favorable to himself. The result, in terms of the distribution of the gains from trade, is clearly the same as that achieved by the take-it-or-leave-it option.

In stressing this point, Leontief also emphasized that price discrimination, because it leads to the contract curve, is economically efficient. Like perfect competition, it wastes no resources. In this respect, the discriminating monopoly outcome is preferable to the simple monopoly one.

The significance of Leontief's contribution was this. Edgeworth and Bowley had stated that final settlement might occur at either extreme of the core. But they had failed to identify such outcomes with all-or-nothing options and discriminatory pricing. Leontief did so and established once and for all the exact price-quantity agreements that produce such outcomes.

6. OTHER POST-WAR CONTRIBUTIONS: KENNETH ARROW AND PAUL SAMUELSON

Leontief's rehabilitation of the exchange box contributed greatly to its popularity in the late 1940s and early 1950s. Extensions and generalizations followed when Kenneth Arrow and Paul Samuelson found imaginative new uses for the box.

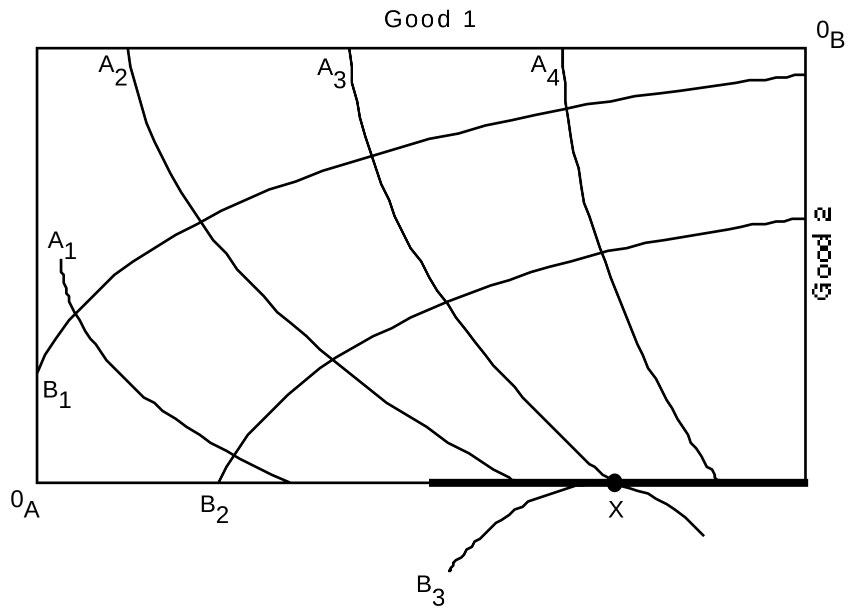
Arrow, in his 1951 essay "An Extension of the Basic Theorems of Classical Welfare Economics," did at least three things. First, he introduced modern set-theoretic concepts into the box. He interpreted the relevant regions of indifference maps as convex consumption sets and price or budget lines as their supporting hyperplanes. Doing so allowed him to replace local or first-order optimality criteria—the familiar marginal conditions—with global criteria.

Second, he employed the foregoing concepts to establish the two fundamental theorems of welfare economics. Theorem one states that every competitive equilibrium, because it occurs at a point where each agent maximizes his satisfaction given the level of satisfaction of the other, is a Pareto optimum. Theorem two states that every Pareto optimum, because it can be supported by a price vector that equates supply and demand, is a competitive equilibrium. Arrow demonstrated that both theorems hold for the standard case where indifference-curve tangencies occur in the interior of the box.

Third, he analyzed boundary optima in which interior tangencies give way to corner solutions on the edges of the box. His analysis yielded a positive and a negative result. The positive result was that competitive equilibria retain their optimality properties even when they occur on the borders of the box. His negative result was that, without extra assumptions, there may be Pareto optimal points on the boundaries that cannot possibly be equilibrium allocations (see Figure 10).

Consider point *X*. There agent *A*'s downward-sloping indifference curve meets the corresponding curve of agent *B* at its peak. Clearly this is a Pareto-efficient allocation since each agent is on his highest attainable indifference curve given the curve of the other. Nevertheless, this optimum cannot sustain an equilibrium. For given the flatness of *B*'s curve at its peak, the tangent price vector that separates the two indifference curves at *X* is necessarily a

Figure 10 Arrow's Exceptional Case



Pareto-efficient solution *X* contradicts the notion that optimality guarantees a competitive equilibrium. For the horizontal, tangent price line separating the indifference curves at *X* induces agent *B* to maximize his satisfaction by remaining at that point. Contrariwise, it induces agent *A* to maximize his satisfaction by moving as far to the right as possible. The upshot is that *A* and *B* seek incompatible allocations and the market fails to clear.

horizontal line coinciding with the lower edge of the box. Its slope implies a zero relative price that induces the agents to register incompatible claims. Given the zero price, *B* maximizes his utility by remaining at point *X*. By contrast, *A* maximizes his utility by moving rightward as far as possible along the price line, reaching ever-higher indifference curves as he goes.

The upshot is that *A* and *B* seek inconsistent allocations at the prices implied by corner-solution *X* and so the market fails to clear. Students refer to this curiosum as Arrow's Exceptional Case. It violates the theorem that every Pareto optimum guarantees a competitive equilibrium.[5]

[5] The theorem holds, however, when both indifference curves possess negative slopes at boundary optima. In such cases, a downward-sloping, tangent price line can always be fitted between the curves. Its slope represents the market-clearing price ratio that induces both parties to go to the optimum point.

Samuelson, in his classic 1952 *Economic Journal* article on "The Transfer Problem and Transport Costs," used the exchange box to determine if a transfer payment made by Europe to America would worsen or improve Europe's terms of trade. According to him, the transfer shifts the endowment point to the left and with it the offer curves and terms-of-trade ray that intersect at world trade equilibrium (see Figure 11). But whether the new ray is less or more steeply sloped than the old depends on the relative marginal propensities to consume Europe's export good, clothing, in both countries. If the transfer reduces Europe's clothing consumption more than it expands America's, the result is an excess world supply of clothing whose relative price must therefore fall. The terms of trade will turn against Europe. On the other hand, if the transfer-induced fall in Europe's demand for its exportable good, clothing, is less than the rise in America's demand for that same good, the resulting excess world demand for clothing will bid up its relative price. Europe's terms of trade will improve. The slope of the terms-of-trade ray can become either flatter or steeper. It all depends on the relative propensities to consume.

These extensions, however, brought the evolution of the exchange box to a halt. For the combined contributions of Leontief, Arrow, and Samuelson had virtually exhausted the analytical potential of the diagram and left it with little new to do. True, it maintained its popularity in the textbooks. But it was clear to all that the exchange box had seen its heyday. By the mid-1950s, its main use was to illustrate established ideas rather than to generate new ones.[6] Not so the alternative production variant, however. Economists were increasingly finding new applications for that version of the box.

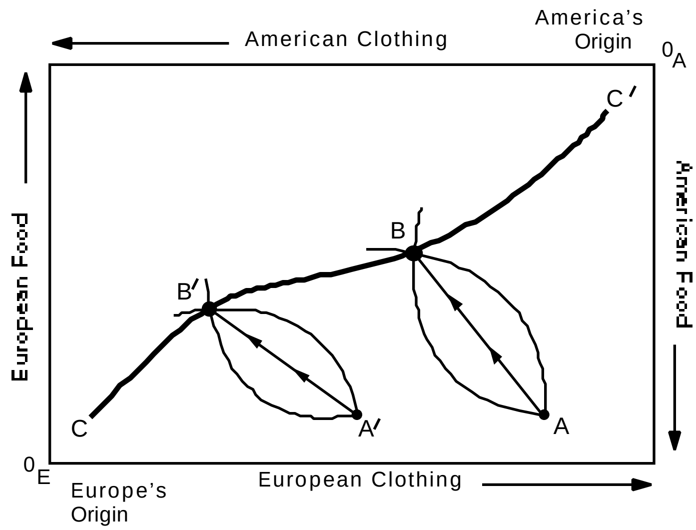
7. ABBA LERNER

Already, in December 1933, Abba Lerner had drawn perhaps the earliest version of the production box. He presented it in a term paper on factor-price equalization which he wrote for Lionel Robbins's seminar at the London School of Economics.

Lerner's diagram superimposes isoquant, or production indifference, maps of two industries fully employing two factor inputs whose fixed quantities determine the dimensions of the box (see Figure 12). Each isoquant shows alternative factor combinations capable of producing a given level of output. Any point in the box represents a particular allocation of the two factors between

[6] This situation, however, proved to be temporary. Unforeseen at the time was the post-1970 resurrection of the box to depict Kenneth Arrow's notion of insurance as trade in state-contingent commodities (see Duffie and Sonnenschein [1989], pp. 584–86, and Niehans [1990], pp. 493–95). By exchanging such commodities, agents could in principle profitably insure themselves against the risks of unfavorable states of the world. In so doing, they could reach the contract curve where the allocation of risk-bearing is optimal.

Figure 11 The Transfer Problem



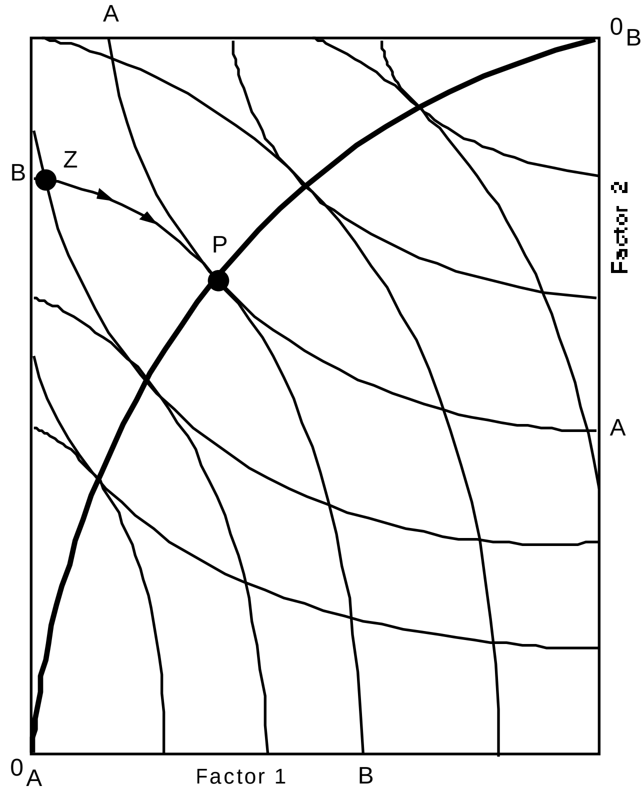
Before transfer, Europe and America trade along the terms-of-trade vector *AB* linking the autarky endowment point *A* to the free-trade equilibrium point *B*. When Europe pays to America a tribute of *AA*' units of cloth, the endowment point shifts horizontally leftward from *A* to *A*'. The corresponding new trade equilibrium point is *B*'. Whether the post-transfer terms-of-trade vector *A*'*B*' is steeper or flatter than the old one depends on the countries' marginal propensities to spend the proceeds of the tribute on cloth versus food. Either result is possible. Thus a transfer can cause the paying country's terms of trade to deteriorate, improve, or remain unchanged.

the production of the two goods. Isoquants going through such points show quantities of both goods produced with this factor allocation.

Regarding factor allocations off the locus of isoquant tangency points, Lerner notes that they are technologically inefficient. They squander scarce resources. They leave room for at least one industry, via mere reallocation of existing inputs, to increase its output with no loss of output of the other. Thus, starting from point *Z*, one can, by moving along isoquant *B* until it reaches tangency with isoquant *A*, increase the output of good *A* with no decrease in the output of good *B*. No such feat is possible on the efficiency locus itself, however. There, one industry's expansion spells the other's contraction. There, factor allocations are technologically efficient in the sense that they maximize the output of one good given the output of the other.

Efficiency, then, requires producers to operate on the contract curve. And, according to Lerner, competition and factor mobility together ensure that they

Figure 12 Lerner's Production Box Diagram



At inefficient point *Z*, good *A*'s output can be increased with no loss in good *B*'s output. Producers simply reallocate factor inputs to *A*-production so as to move along the given *B* isoquant cutting successively higher *A* isoquants in the process. At efficient point *P* on the contract curve, however, the output of one good can be increased only by decreasing the output of the other.

will do so. Competition forces producers to hire both factors until the ratio of their marginal products, represented by the slopes of isoquants, equals the ratio of their prices, represented by the slope of a relative factor-price line. And factor mobility dictates that resource prices, and so their ratio, are the same in both industries. Consequently, both industries operate at a point where their isoquants are tangent to a common factor-price-ratio line and thus are tangent to each other. Such points lie on the contract curve.

Delayed Publication

Unfortunately, economists had to wait for nineteen years to see Lerner's pioneering diagram. Tibor Scitovsky, in his essay on "Lerner's Contributions to Economics," tells why. In 1948 and 1949, Paul Samuelson published his celebrated proof of the factor-price-equalization theorem. Robbins, upon reading Samuelson's papers, recalled Lerner's 1933 term paper on the same subject. Robbins still had a copy of the paper in his files. Upon his urging, Lerner published the manuscript without alteration as the 1952 *Economica* piece "Factor Prices and International Trade."

As to why Lerner neglected to publish the paper in 1934, Tibor Scitovsky recounted a story he heard in 1935 when he was one of Lerner's students. Evidently, Lerner had given his only corrected copy to another student to be typewritten for submission to a scholarly journal. But the student lost the paper on a London bus and was unable to retrieve it. Lerner, who was busy writing other papers at the time, could not find the time to reproduce the lost manuscript. The resulting delay made Lerner's pathbreaking work and its innovative diagram seem less-than-novel when they finally appeared. In any case, it was not from Lerner but from Wolfgang Stolper and Paul Samuelson that the economics profession first learned of the production box.

8. WOLFGANG STOLPER AND PAUL SAMUELSON

Wolfgang Stolper and Paul Samuelson published the first production box diagram to appear in print. It features prominently in their 1941 *Review of Economic Studies* article on "Protection and Real Wages." Of the two authors, Stolper (1994, p. 339) credits Samuelson with the idea of using the box. In any case, they applied it to derive their famous theorem according to which free trade benefits the relatively plentiful factor and hurts the relatively scarce one while protective tariffs do the opposite.

Stolper-Samuelson Theorem

The Stolper-Samuelson theorem rests on two propositions. First, compared with autarky, free trade raises the price of the relatively abundant factor and lowers the price of the relatively scarce one. Conversely, trade restriction raises the scarce factor's price and lowers the plentiful factor's. Second, it follows that a tariff-induced restriction of trade may benefit labor in countries where labor is the scarcer factor. In such countries, a tariff may raise real wages and increase labor's real income both absolutely and relatively as a percent of the national income.

Stolper and Samuelson reached these conclusions via the following route. Suppose in the absence of trade a country produces wheat and watches with a fixed factor endowment consisting of much capital and little labor (see Figure 13). Measure capital on the horizontal axes of the box and labor on the vertical ones. The box, being wider than it is tall, indicates a high ratio of capital to labor and thus identifies capital as the relatively plentiful factor and labor as the relatively scarce one.

Next assume that, at any given factor-price ratio, wheat production requires a higher ratio of capital to labor than does watch production. Wheat, in other words, is capital intensive and watches are labor intensive. The slopes of labor-to-capital factor-proportion rays going through any point on the contract curve show as much. Those rays are steeper for watches than for wheat. Moreover, the contract curve lies everywhere below the diagonal of the box. Were factor intensities the same in both industries, the contract curve would coincide with the diagonal. And were factor-intensity reversals to occur, the contract curve would cross the diagonal. Neither possibility is allowed. Both are ruled out by assumption.

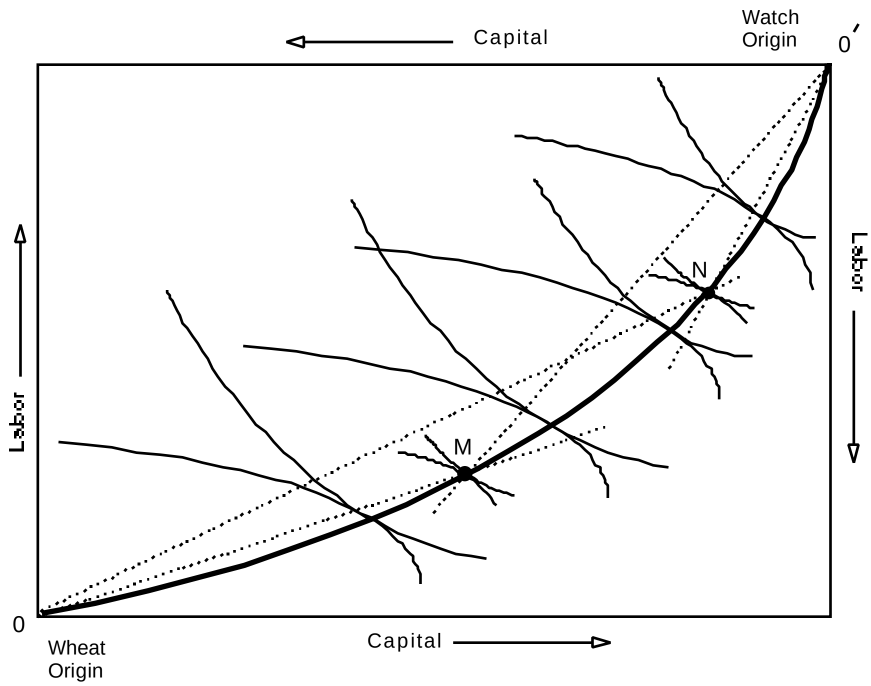
Initially, in the absence of trade, the country produces and consumes at point *M* on the contract curve. Wheat, embodying relatively large amounts of relatively cheap and plentiful capital, is the low-cost good. Conversely, watches, embodying much scarce and hence relatively dear labor, constitute the high-cost good. Given that the opposite conditions prevail in the rest of the world, the result is that wheat is cheaper in terms of watches at home than abroad.

Free Trade Helps the Plentiful Factor

When trade opens up, foreigners will import the home country's cheap wheat and home residents will import foreigners' cheap watches. The consequent increased demand for the home country's wheat and the decreased demand for its watches bids up the domestic price of wheat relative to the price of watches. The resulting price rise induces wheat producers to expand by hiring capital and labor from watch producers so as to move to free-trade point *N*. But the contracting watch industry, being labor-intensive, releases relatively little capital and relatively much labor compared to the ratio in which the capital-intensive wheat industry wants to absorb those factors. The ensuing labor surplus and capital shortage bids wages down and capital rentals up. The lower wages and higher rentals in turn induce both industries to substitute cheaper labor for dearer capital. The upshot is that the labor-to-capital ratio rises in both industries. And it does so even as the overall economy-wide endowment ratio shown by the slope of the diagonal stays unchanged. In terms of the diagram, both factor-proportion rays through point *N* are steeper than those through point *M*.

With less capital working with each unit of labor in both industries, the marginal product of labor falls and the marginal product of capital rises. Under competitive conditions, those marginal products constitute factor real rewards

Figure 13 Stolper-Samuelson Theorem



Free trade moves the capital-rich economy from autarky point *M* to free-trade point *N*. The dashed rays show that the labor-to-capital ratio rises in both industries. With more labor working with each unit of capital in each industry, the marginal productivity of capital and hence its real return rises while the marginal productivity of labor and hence its real wage falls. Free trade helps the plentiful factor and hurts the scarce one. Conversely, a protectionist move from point *N* to point *M* helps the scarce factor and hurts the plentiful one.

which factor mobility equalizes across industries. It therefore follows that real wages fall and real rentals rise expressed in terms of either good. Indeed, the flatter common slope of the isoquants at point *N* than at point *M* signifies as much. Those slopes indicate the rise in capital's and the fall in labor's real return. They show that free trade benefits the country's abundant capital factor and hurts its scarce labor one.

Protection Helps the Scarce Factor

Conversely, protection does the opposite. It raises the relative price and thus stimulates the output of import-competing watches at the expense of wheat production. In so doing, protection moves the domestic product-mix and its associated interindustry factor allocation from free-trade point *N* toward autarky point *M*. To induce the expanding watch industry to absorb factors in the proportion released by the contracting wheat industry, rentals must fall relative to wages. The consequent fall in capital's relative price encourages both sectors to adopt more capital-intensive techniques. The result is a rise in the capital-to-labor ratio in both industries as shown by the flatter slope of the rays going through *M* than through *N*. With more capital working with each unit of labor in both sectors, the marginal product of labor rises and the marginal product of capital falls. With factor real rewards equal to marginal products, the real wage of scarce labor rises while the real rental of abundant capital falls, as shown by the steeper common slope of the isoquants at *M* than at *N*. In short, import tariffs raise real wages and lower real rentals when the import-competing sector is more labor-intensive than the export sector. Protection benefits the scarce factor and hurts the plentiful one.

Evaluation

The Stolper-Samuelson paper is a milestone in the history of the box diagram and the evolution of trade theory. It crystallized certain components of the emerging Heckscher-Ohlin theory of international trade into a two-good, two-factor general equilibrium model. It then condensed that model into a simple box diagram capable of showing how commercial policy affects distributive shares. In so doing, it demonstrated the box's power in handling a large number of interrelated variables and thus established it as the standard tool of trade theory. Once established, the box proved indispensable in the derivation of such key trade propositions as the factor-price-equalization, Heckscher-Ohlin, and Rybczynski theorems.

Most important, the box diagram, in Stolper's and Samuelson's hands, taught that informal intuition on trade issues could be misleading. Before Stolper and Samuelson, most economists believed instinctively that free trade benefits all factor inputs. In demonstrating rigorously that such was not necessarily the case, Stolper and Samuelson made economists more cautious in discussing the benefits of trade. Thereafter, economists would acknowledge possible losses to the scarce factor in movements to free trade. But they would insist, on the grounds that trade benefits the country as a whole, that the gains of the abundant factor exceed the scarce factor's losses. Citing Scitovsky, they would argue that the abundant factor could in principle compensate the scarce factor for its losses and still be better off whereas the scarce factor would be unable to profitably bribe the abundant factor to oppose free trade.

9. T. M. RYBCZYNSKI

Stolper and Samuelson had used the box to link trade- or tariff-induced changes in commodity prices to changes in factor prices. They had shown how a product price increase causes a more-than-proportional rise in one factor's real reward while lowering the reward of the other. By contrast, T. M. Rybczynski in 1955 used the box to link changes in factor endowments to changes in commodity outputs. He showed that when one factor increases in quantity (product prices held constant), it causes a more-than-proportional increase in the output of one good and an absolute fall in the output of the other. Here was a startling revelation. Before Rybczynski, most economists felt that an increase in the endowment of one non-specific factor would lead to a rise in the output of all goods.

Rybczynski's demonstration goes as follows (see Figure 14). Let the country's initial factor endowment be that indicated by the dimensions of box *ABCD*. The economy initially produces at point *P* on the contract curve. The slope of the factor-intensity ray emanating from the wheat origin *A*, being flatter than its counterpart originating from the watch origin *C*, identifies wheat as the capital-intensive good and watches as the labor-intensive one.

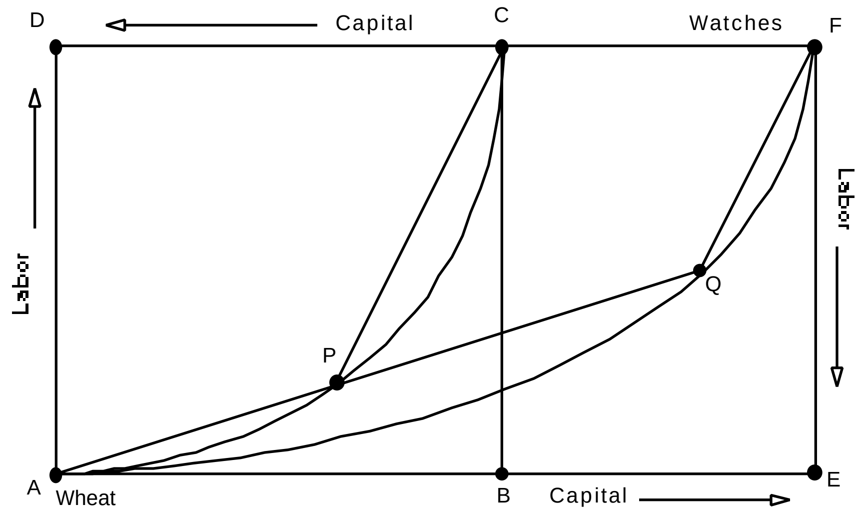
Rybczynski Theorem

Now assume that the economy's capital endowment expands by the amount *BE* while its labor endowment remains unchanged. The result is that the box annexes the new rectangle *BEFC*. How does the capital accumulation and the corresponding expansion of the box affect the output-mix of wheat and watches? Rybczynski's assumption of constant commodity prices provides the answer. Such constancy holds for small open economies taking their prices as given exogenously from the closed world economy.

Constant commodity prices imply constant factor prices. And with linear homogeneous production functions, constant factor prices imply unchanged factor proportions in both industries. Point *Q* in the new box satisfies that latter criterion. Only at that point are the capital-to-labor ratios (as shown by the slopes of the factor-intensity rays) the same as they were at point *P* in the old box. Thus the new equilibrium factor allocation must be *Q*. This new allocation, however, sees more labor and capital devoted to wheat production and less of both to watch production. The result is that wheat production expands and watch production contracts. Here is the famous Rybczynski theorem: Let one factor increase while the other stays constant. Then output of the good intensive in the increased factor will, at constant commodity prices, increase in absolute amount. Conversely, output of the other decreases absolutely.

The reasoning is straightforward. The expanding factor must be absorbed in producing the good using it intensively. To keep factor proportions fixed, as implied by the assumption of constant commodity prices, the expanding

Figure 14 Rybczynski Theorem



At constant commodity prices, a rise in the country's capital stock with no growth in its labor force raises the output of capital-intensive wheat and lowers the output of labor-intensive watches. Why? Because fixed commodity prices imply fixed factor prices which imply unchanged factor proportions in both industries. Thus as wheat output expands to absorb the extra capital, it requires extra labor to keep factor proportions unchanged. The only source of this extra labor is the watch industry, which therefore must contract. We go from point *P* to point *Q* with the slopes of the factor proportion rays remaining unchanged throughout.

industry must hire the non-increasing factor too. The only source of this factor is the other industry, which therefore must contract. Once again, the box diagram had rendered a seemingly counterintuitive proposition transparent.

10. KELVIN LANCASTER

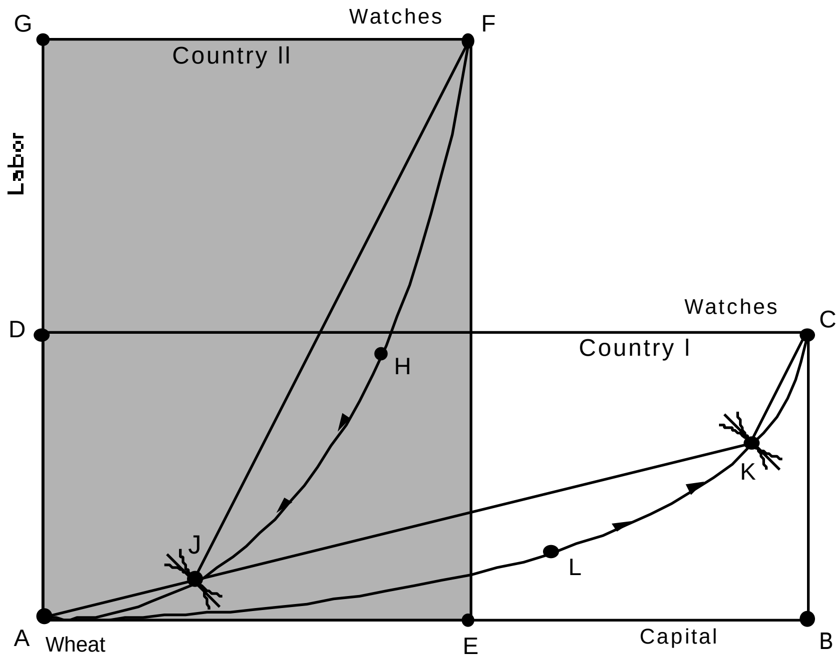
The box diagrams of Lerner, Stolper-Samuelson, and Rybczynski referred to a single country only. As such, they were hardly equipped to accommodate two-country models of international trade. The emerging Heckscher-Ohlin model was a prime example of such a model. True, the above-mentioned writers had introduced some Heckscher-Ohlin components into their single-country diagrams. But the list of included components was incomplete. Exposition of the full model required boxes referring to at least two countries.

Credit for developing the two-country box in its Heckscher-Ohlin form goes to Kelvin Lancaster.[7] His diagram, as presented in his 1957 *Economica* article on "The Heckscher-Ohlin Trade Model: A Geometric Treatment," embodies the standard features of that two-by-two-by-two model. Two countries produce two goods from two factor inputs. The countries are incompletely specialized. They produce both goods before and after trade. One good is always more capital-intensive than the other. Factor endowments differ across countries. Full employment prevails as does perfect competition in product and factor markets. Both countries share the same linear homogeneous production technology exhibiting constant returns to scale. Such technology ensures that factor marginal productivities are determined by factor-input ratios and not by scale of output.

The diagram incorporating these assumptions superimposes production boxes representing the countries' different factor endowments (see Figure 15). Capital is measured horizontally, labor vertically. The wide box *ABCD* identifies country *I* as the relatively capital-abundant nation. Similarly, the tall box *AEFG* specifies country *II* as the relatively labor-plentiful nation. Lancaster assumes that wheat production is always capital-intensive and watch production labor-intensive in both countries. The contract curves indicate as much. They lie below the diagonals of the boxes. Thus as one moves along a contract curve from left to right, the capital-to-labor ratio declines in response to a rising rental-to-wage ratio. But, at any given factor- price ratio, the capital-to-labor ratio is always higher in wheat than in watches. Were such not the case, the contract curves would either coincide with the diagonal or cross it.

[7] Even before Lancaster, Jan Tinbergen (1954, p. 137) had presented an alternative version of the two-country box. But his diagram, unlike Lancaster's, maps production possibility curves into commodity space. Depicting global trade equilibrium, he fits an equilibrium world price line between the production transformation curves and consumption indifference maps of the two countries. Both countries produce at the common point of tangency of their respective transformation curves and the price line. Then they trade along that line, each exporting its comparative advantage good and importing its comparative disadvantage one, until they reach the point of maximum satisfaction on their highest attainable indifference curves. In this way, trade enables both to consume beyond their transformation curves.

Figure 15 Heckscher-Ohlin Theorem and Factor-Price Equalization



Free trade moves countries *I* and *II* from autarky production points *L* and *H* to post-trade points *K* and *J*. Capital-rich country *I* produces more capital-intensive wheat and fewer labor-intensive watches. Labor-rich country *II* does the opposite. The slopes of the isoquants are the same at both post-trade points. This implies equal labor marginal productivities and equal capital marginal productivities in both countries. Since trade equalizes product prices worldwide and factor prices equal product prices times factor marginal productivities, it follows that trade equalizes factor prices too.

Heckscher-Ohlin Theorem

Having constructed the diagram, Lancaster used it to demonstrate the celebrated Heckscher-Ohlin and factor-price-equalization theorems. These theorems, together with their companion Stolper-Samuelson and Rybczynski postulates, constitute the core propositions of Heckscher-Ohlin trade theory. The Heckscher-Ohlin theorem predicts that each country will export the good intensive in its abundant factor and import the good intensive in its scarce factor. And the factor-price-equalization theorem says that free trade in commodities equalizes factor prices worldwide just as unrestricted factor mobility would do. The box diagram clarifies the underlying logic.

Initially, in the absence of trade, the countries operate in isolation at points *L* and *H* on their respective contract curves. At those autarky points, factor prices and factor combinations used to produce each good differ across the two countries as do product prices. Wheat, the capital-intensive good, is cheapest in terms of watches in capital-rich country *I*. Conversely, watches, the labor-intensive good, are cheapest in terms of wheat in labor-abundant country *II*.

When trade opens up, country *I* produces more of its export good, wheat, and fewer import-competing watches. The country moves along its contract curve to the free-trade point *K*. There, *I*'s relative commodity prices, or terms of trade, are the same as those abroad such that no incentive remains for further expansion of trade. At point *K* the rays *AK* and *CK*, whose slopes represent the factor proportions employed in *I*'s wheat and watch industries, respectively, intersect as required by the full-employment assumption.

Lancaster proves that the corresponding free-trade point for country *II* is *J*. The reason is simple. Free trade equalizes the ratio of commodity prices worldwide. In equilibrium, that ratio equals the marginal rate of factor substitution which equals the ratio of factor prices. Relative factor prices in turn uniquely determine factor input ratios in production functions exhibiting constant returns to scale. With both countries facing the same relative factor prices and sharing the same production functions, it follows that both must use the same factor input ratios too. Geometrically, the capital-to-labor factor-proportion rays going through country *II*'s free-trade point must have the same slopes as those intersecting at country *I*'s free-trade point. Such indeed is the case. Ray *AJ* is identical to ray *AK*. And ray *FJ* is parallel to ray *CK*. So country *II*'s free-trade point *J* corresponds to country *I*'s free-trade point *K*. Point *J* is the only point on *II*'s contract curve cut by factor-intensity rays of the same slope as those going through point *K* on *I*'s contract curve.

Having established the corresponding free-trade points, Lancaster required one final step to complete his demonstration. He took advantage of the property that linear homogeneous production functions allow output to be measured as the distance along any ray from the origin. He employed this property to compare the post-trade product mixes in the two countries. Country *I* produces more wheat ($AK > AJ$) and fewer watches ($CK < FJ$) than does country *II*. Thus country *I*'s product mix is heavily weighted toward wheat and country *II*'s toward watches. Here is Lancaster's demonstration of the Heckscher-Ohlin theorem: each country produces (and exports) relatively more of the good intensive in its abundant factor.

Factor-Price-Equalization Theorem

As for absolute factor-price equalization, Lancaster offered the following demonstration. Observe the tangent isoquants at the free-trade equilibrium points *J* and *K*. Constant-returns-to-scale considerations dictate that these isoquants, lying as they do on identical or parallel factor-proportion rays, possess the same slopes at one equilibrium point as they do at the other. But these slopes represent the ratios of factor marginal productivities which, as noted above, free trade equalizes across countries. Indeed, Lancaster shows that a stronger condition holds. When the two countries share the same linear production technology, free trade equalizes absolute as well as relative marginal productivities. Each factor's individual marginal productivity is the same in both nations.

Two additional steps complete the argument. The first cites the law-of-one-price notion that free trade renders the price of any traded commodity everywhere the same. The second refers to the competitive equilibrium condition that the price of any factor equals its marginal productivity multiplied by commodity price. Since trade equalizes commodity prices and marginal productivities worldwide, it equalizes their multiplicative product, factor prices, as well.

Lancaster's demonstration appeared at a time when other scholars were contributing to production-box analysis. Complementing his work were Kurt Savosnick's 1958 derivation of production possibility curves within the confines of the box and Ronald Jones's 1956 use of the diagram to examine the effects of factor-intensity reversals. These applications would have delighted Edgeworth. Apparently there was no end to what his invention could accomplish.

11. CONCLUSION: HOW THE BOX EVOLVED

The history of the box diagram reveals how analytical tools evolve in a complex interaction with their uses (see Koopmans [1957] 1991, pp. 169–71, for the definitive statement of this thesis). In this interaction, tools play a double role of servant and guide. As servants, they help solve problems motivating their invention. As guides, they alert their users to other problems solvable with their aid. Certainly Edgeworth regarded the box as servant when he invented it to demonstrate gains from exchange and to resolve the puzzle of how increasing numbers lead to the competitive equilibrium.

Once invented, however, the exchange box took on the status of guide. Its very existence made economists aware of other phenomena potentially seeking its application. In short order, it was employed to explain the rationale of such things as Pareto optimality, simple and discriminatory monopoly pricing, compensation tests, the transfer problem, and corner solutions. All were manifestations of the drive to generalize the exchange box and extend its range of application.

This same drive produced the alternative production box. Here a simple analogy sufficed. Economists saw how the exchange box depicted the allocation of fixed stocks of goods between the utilities of two individuals. They quickly realized that an analogous version could depict the allocation of a fixed stock of factor inputs between the production of two goods. Thus was born the production box, whose initial use was to devise rules for optimal factor allocation. Once available, however, the production box spurred economists to find new applications for it. Chief among these applications was the Heckscher-Ohlin theory of international trade. Accordingly, the box was deployed to derive, prove, or illustrate the key propositions of that theory.

The above experience contradicts Tjalling Koopmans's ([1957] 1991, p. 175) contention that diagrams, though a powerful aid to intuition and exposition, are nevertheless no match for higher mathematics in rigorous economic analysis. For the box diagram is surely an exception to that rule. In support of his allegation, Koopmans cites (1) the unreliability of diagrams as guides to reasoning, (2) their confining effect on the choice of problems studied, and (3) their inability to handle problems of more than two dimensions.

While these charges may stand up in a general comparison of diagrammatic and mathematical techniques, the box diagram itself pleads innocent to them. Far from being unreliable, it proved to be a highly accurate tool in the hands of economists ranging from Edgeworth to Samuelson. So accurate was it, in fact, that successive users found little need to modify it substantially. Far from being confining, it freed its users to attack long-unsolved problems such as how free trade affects the absolute and relative income shares of factor inputs. Its limited dimensionality likewise proved to be no handicap. On the contrary, Lancaster noted that the diagram could show the interrelationships between no less than twelve economic variables. Edgeworth likewise found the diagram's two-dimensionality no bar to analyzing what happens when unlimited pairs of traders are introduced into the model. In short, the history shows that this simple geometrical diagram, in terms of its ability to yield penetrating insights into problems of economic theory, has been a powerful mathematical tool in its own right.

Of course the box, like any diagram, cannot handle all problems. Far from it. Nobody would deny, for example, the diagram's insufficiency to represent infinite-horizon models involving infinite-dimensional commodity space. Such complex models are beyond the capacity of the box. Rather the box's strength lies in depicting simple general equilibrium models. As these models are extremely useful, so too is the diagram that embodies them.

In any case, it was the box diagram itself, more than any accompanying mathematics, that captured the attention of the economics profession. The result was that the box became a fixture of trade and welfare theory and a commonplace of textbooks. The survival of the concept testifies to its continued usefulness. Even today, if one wishes to understand the sources of and gains from exchange as well as the optimality of competitive equilibrium and the logic of efficient resource allocation, one can do no better than study the diagram.

REFERENCES

Arrow, Kenneth. "An Extension of the Basic Theorems of Classical Welfare Economics," in Jerzy Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1951, pp. 507–32, reprinted in Kenneth Arrow, *Collected Papers of Kenneth Arrow,* Vol. 2, *General Equilibrium*. Cambridge, Mass.: Harvard University Press, 1983, pp. 13–45.

Blaug, Mark. *Great Economists before Keynes*. Cambridge: Cambridge University Press, 1986.

Bowley, Arthur W. *The Mathematical Groundwork of Economics*. Oxford: Clarendon Press, 1924, reprinted, New York: Augustus M. Kelley, 1965.

Caves, Richard E., and Ronald W. Jones. *World Trade and Payments: An Introduction*. 3d ed. Boston: Little, Brown and Company, 1981.

Creedy, John. *Demand and Exchange in Economic Analysis: A History from Cournot to Marshall*. Aldershot, U.K.: Edward Elgar, 1992.

__________. "Some Recent Interpretations of *Mathematical Psychics,*" *History of Political Economy,* vol. 12 (Summer 1980), pp. 267–76.

Duffie, Darrell, and Hugo Sonnenschein. "Arrow and General Equilibrium Theory," *Journal of Economic Literature,* vol. 27 (June 1989), pp. 565–98.

Edgeworth, Francis Y. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: Kegan Paul, 1881, reprinted, New York: Augustus M. Kelley, 1967.

Harsanyi, John C. "Approaches to the Bargaining Problem before and after the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories," *Econometrica*, vol. 24 (April 1956), pp. 144–57.

Jones, Ronald W. "Factor Proportions and the Heckscher-Ohlin Theorem," *Review of Economic Studies,* vol. 24 (October 1956), pp. 1–10.

Koopmans, Tjalling C. *Three Essays on the State of Economic Science*. New York: McGraw-Hill, 1957, reprinted, Fairfield, N.J.: Augustus M. Kelley, 1991.

Lancaster, Kelvin. "The Heckscher-Ohlin Trade Model: A Geometric Treatment," *Economica*, vol. 24 (February 1957), pp. 19–39, reprinted in Jagdish Bhagwati, ed., *International Trade: Selected Readings*. Baltimore: Penguin Books, 1969, pp. 49–74.

Leontief, Wassily W. "The Pure Theory of the Guaranteed Annual Wage Contract," *Journal of Political Economy,* vol. 54 (February 1946), pp. 76–79, reprinted in Wassily Leontief, *Essays in Economics,* Vol 1. White Plains: International Arts and Sciences Press, Inc., 1966, pp. 108–15.

Lerner, Abba P. "Factor Prices and International Trade," *Economica,* vol. 19 (February 1952), pp. 1–16.

Nash, John F. "The Bargaining Problem," *Econometrica,* vol. 18 (April 1950), pp. 155–62.

Newman, Peter. "The Great Barter Controversy," in J. K. Whitaker, ed., *Centenary Essays on Alfred Marshall*. Cambridge: Cambridge University Press, 1990, pp. 258–77.

Niehans, Jurg. *A History of Economic Theory: Classic Contributions, 1720–1980*. Baltimore: Johns Hopkins University Press, 1990.

Pantaleoni, Maffeo. "Pareto: An Obituary," *Economic Journal,* vol. 33 (December 1923), pp. 582–90.

Pareto, Vilfredo. *Manuale d'economie politica*. Milan: Societa editrice libraria, 1906. French translation and revision: *Manuel d'economie politique*. Paris, 1909. English translation: *Manual of Political Economy,* trans. Ann S. Schwier. New York: Augustus M. Kelley, 1971.

Rubinstein, Ariel. "Perfect Equilibrium in a Bargaining Model," *Econometrica,* vol. 50 (January 1982), pp. 97–109.

Rybczynski, T. M. "Factor Endowment and Relative Commodity Prices," *Economica,* vol. 22 (November 1955), pp. 336–41, reprinted in Richard E. Caves and Harry G. Johnson, eds., AEA *Readings in International Economics*. Homewood, Ill.: Richard D. Irwin, 1968, pp. 72–77.

Samuelson, Paul A. "The Transfer Problem and Transport Costs: The Terms of Trade when Impediments Are Absent," *Economic Journal,* vol. 62 (June 1952), pp. 278–304, reprinted in Joseph E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson,* Vol. 2. Cambridge, Mass.: MIT Press, 1966, pp. 985–1011.

Savosnick, Kurt M. "The Box Diagram and the Production Possibility Curve," *Ekonomisk Tidskrift,* vol. 60 (September 1958), pp. 183–97.

Scitovsky, Tibor. "Lerner's Contribution to Economics," *Journal of Economic Literature,* vol. 22 (December 1984), pp. 1547–71.

__________. "A Reconsideration of the Theory of Tariffs," *Review of Economic Studies,* vol. 9 (Summer 1942), pp. 89–110, reprinted in Tibor Scitovsky, *Papers on Welfare and Growth*. Stanford: Stanford University Press, 1964, pp. 139–66.

__________. "A Note on Welfare Propositions in Economics," *Review of Economic Studies,* vol. 9 (November 1941), pp. 77–88, reprinted in Tibor Scitovsky, *Papers on Welfare and Growth*. Stanford: Stanford University Press, 1964, pp. 123–38.

__________. *Welfare and Competition*. Chicago: R. D. Irwin, 1951.

Stolper, Wolfgang F. "Afterthoughts on 'Protection and Real Wages'," in Alan V. Deardorff and Robert M. Stern, eds., *The Stolper-Samuelson Theorem: A Golden Jubilee*. Ann Arbor: University of Michigan Press, 1994.

__________, and Paul A. Samuelson. "Protection and Real Wages," *Review of Economic Studies*, vol. 9 (November 1941), pp. 58–73, reprinted in Jagdish Bhagwati, ed., *International Trade: Selected Readings*. Baltimore: Penguin Books, 1969, pp. 245–68.

Tinbergen, Jan. *International Economic Integration.* Amsterdam: Elsevier, 1954.

Zeuthen, Frederick L. *Problems of Monopoly and Economics*. London: Routledge & Kegan Paul, 1930.