THE TRADE THEORIST'S SACRED DIAGRAM: ITS ORIGIN AND EARLY DEVELOPMENT

Thomas M. Humphrey

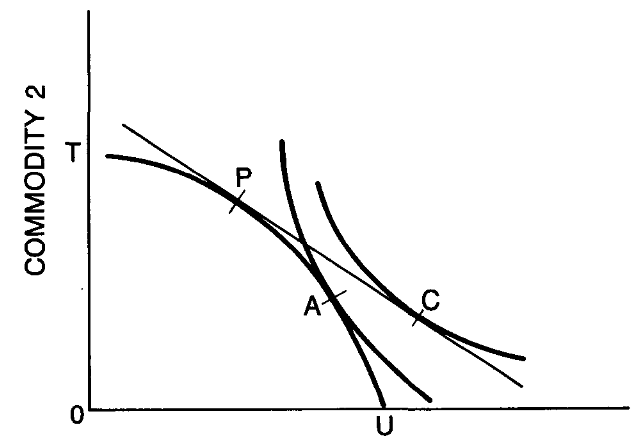
Introduction

In his celebrated 1945 essay on international trade under variable returns in a simple model[1] the noted Dutch economist Jan Tinbergen presented his version of what Robert Baldwin calls "the sacred diagram of the international trade economist" [1, p. 142]. Tinbergen used the diagram, which consists of a transformation or production possibility curve, taste indifference curves, and relative price or terms-of-trade lines, to show how a country gains from the opportunity to trade at a world price ratio different from the closed-economy one (see Figure 1). Given that opportunity, the country does two things. First, it produces the output mix that maximizes its national product valued at world prices. That is, it produces at the point of tangency of the production possibility curve and world price line. Then it trades along that line, exporting products in which it has a comparative cost advantage in exchange for imports of products in which it has a comparative disadvantage, until it reaches its point of maximum satisfaction on its highest attainable indifference curve. There it enjoys a bundle of goods that it could not produce or consume in isolation. Here is the economist's case for free trade captured in a single diagram.

That a simple geometrical diagram would become an icon is hardly surprising. Other economic diagrams have enjoyed that same distinction—the Keynesian cross, Marshallian sissors, Hicksian IS-LM, Edgeworth-Bowley box, Phillips curve, and Knightian circular flow are cases in point. What is surprising is how little has been written on the trade diagram's history. Few systematic surveys of that history exist; textbooks say little about it. Tinbergen himself said nothing about earlier versions of the diagram even though it was 38 years old at the time he presented it. Who invented the diagram? How was it initially received? Who exerted the greatest influence in getting it accepted into trade theory?

Figure 1

TINBERGEN'S DIAGRAM



Before trade the economy produces and consumes at A, the common point of tangency of transformation curve and indifference curve. Given the opportunity to trade at the world price ratio shown by the slope of line PC, it produces commodity bundle P which it then trades for bundle C to reach its point of maximum satisfaction C on its highest attainable indifference curve.

Source: Tinbergen [16, p. 129].

Today these issues still remain unresolved and one finds such writers as Samuelson, Baldwin, Maneschi and Thweatt disagreeing over whether Viner, Lerner, Haberler, or Barone contributed most to the diagram's development.[2] In an effort to rectify this situation and to provide some needed historical perspective, this article traces the evolution of the trade diagram from its 1907 origins to its presentation by Tinbergen in 1945 by which time it had already become the standard geometrical tool of the trade theorist. A word of explanation is in order, however. Today analysts put the diagram to many

[1] Tinbergen's essay, originally entitled "Professor Graham's Case for Protection," was reprinted in 1965 in a slightly abbreviated version as "International Trade Under Variable Returns in a Very Simple Model." See [16].

[2] See Maneschi and Thweatt [12, pp. 375-78] for a review of the controversy.

uses—to depict the effects of protection, of noneconomic objectives of tariffs, of domestic market distortions, and of growth on trade, to name just a few. Historically, however, economists chiefly employed it to illustrate trade equilibrium and the gains from trade in a fully competitive economy in which the balance of payments for simplicity consists of the balance of trade. Given this article's historical focus, it too concentrates on those traditional concerns.

Historical Evolution

Historically the diagram evolved through at least eight stages. Each stage saw a different innovator contribute to the diagram's development. Irving Fisher (1907) invented the diagram to illustrate a problem in capital theory, Enrico Barone (1908) extended it to international trade, and Allyn Young (1928) applied it to a hypothetical closed economy operating under constant, decreasing, and increasing returns. Gottfried Haberler (1930) introduced the strictly concave production frontier version into foreign trade theory. Jacob Viner (1931) added community indifference curves to Haberler's diagram and criticized the entire apparatus. Abba Lerner (1932) extended the diagram to the level of the aggregate world economy, Wassily Leontief (1933) applied it to two countries simultaneously, and Jan Tinbergen (1945) elegantly consolidated their results. Except for Young, each analyst used the diagram to emphasize the gains from trade. Of these analysts, it was Haberler and Leontief who had the greatest influence. It was they who convinced trade theorists to add the diagram to their analytical tool kit. What follows describes in chronological order the specific contributions of each of these pioneers.

Irving Fisher

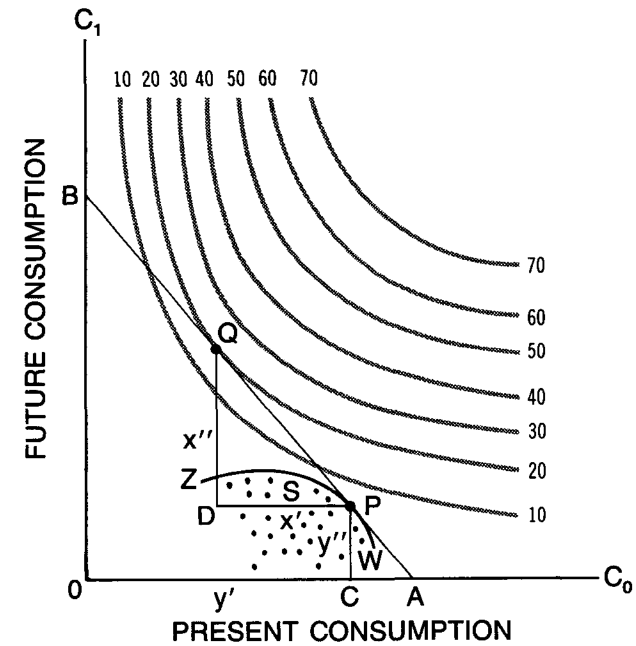
Francis Y. Edgeworth invented indifference curves in his *Mathematical Psychics* in 1881. Similarly, Vilfredo Pareto pioneered the use of transformation curves in his *Manuale di economia politica* in 1906. But Irving Fisher in his 1907 classic *The Rate of Interest* was the first to combine indifference and transformation curves together with market price lines in a single diagram and to use it to illustrate the gains from exchange (see Figure 2).

True, he applied his diagram to a problem in capital theory rather than to the theory of international trade. That is, he used it to depict an individual's optimum investment decision rather than a country's foreign trade equilibrium. But this difference is only superficial. Like trade theorists after him, Fisher used the diagram to demonstrate the gains from trade (albeit intertemporal rather than international). And like trade theorists, he showed the individual moving along the production possibility frontier to the highest attainable price line and then trading along that line to reach the point of maximum satisfaction. In terms of abstract economic logic, his demonstration matches that of the trade theorists. To Fisher, then, must go the credit for inventing the trade diagram.

His diagram appears on page 409 of *The Rate of Interest*.[3] The transformation or production possibility or (as Fisher called it) opportunity curve ZPW shows an individual's opportunity to transform present consumption (measured on the horizontal axis) into future consumption (measured on the

Figure 2

FISHER'S DIAGRAM



Given the interest rate implicit in the slope of line AB, an investor produces the two-period consumption bundle P having the highest present value. Then he trades that bundle for bundle Q by lending PD units of present consumption for DQ units of future consumption to reach his point of maximum satisfaction Q.

Source: Fisher [4, p. 409].

[3] Fisher also used the diagram in his *The Theory of Interest* (1930). On Fisher's diagram see Hirshleifer [9, pp. 330-32] and Samuelson [15, pp. 29-33].

vertical axis) by investing in real capital projects. The concave shape of the curve represents diminishing returns to investment as the sacrifice of more and more units of consumption today yields smaller and smaller increments to consumption tomorrow.

The set of convex iso-desirability curves (as Fisher called them) labeled 10, 20, 30, etc. constitute the individual's indifference map. Each curve shows alternative combinations of present and future consumption that yield equal satisfaction. Higher curves represent higher levels of satisfaction. Finally, the interest line AB shows the opportunity to convert P dollars of present consumption into Q dollars of future consumption by lending at the market rate of interest shown by the slope of the line. In other words, one can lend as well as invest.

Fisher explains that the individual, if deprived of the opportunity to lend on the money market, would choose the two-period consumption combination shown by the common point of tangency of indifference curve and production possibility curve (point S).[4] This is analogous to the trade diagram's closed-economy equilibrium production and consumption point.

Given the opportunity to lend at the going rate of interest, however, the individual equates that rate with the marginal rate of return on real investment by moving along the production frontier to point P on the highest attainable interest line AB. That is, he chooses the two-period consumption bundle having the highest present value calculated at the market interest rate shown by the slope of AB. Then he trades along that line, lending PD ($= x'$) dollars of current consumption in exchange for DQ ($= x''$) dollars of future consumption, to reach a point of maximum satisfaction Q. In short, given the opportunity to trade at a market price, the individual produces the bundle of goods having the highest market value and then trades it for a preferred bundle lying beyond the production frontier. But this is exactly what a fully competitive open national economy does when given the opportunity to trade at world prices.

Modern users of the trade diagram note that international equilibrium requires the world price ratio be such as to balance trade across nations. In other words, the desired exports of one nation must at the equilibrium price ratio equal the desired imports of another and vice versa. Fisher argued the same about the equilibrium rate of interest. That rate, he said, equates the desired lending of one individual with the desired borrowing of another—that is, it ensures that the legs of the trade triangle PDQ are equal in length but opposite in sign across lenders and borrowers. Thus Fisher did more than specify trade equilibrium conditions for a single individual facing a given market rate. He also specified the market equilibrium conditions that determine that rate. True, he did not show such conditions in his diagram. That is, he did not extend it to the two-person case. But he stated how it could be done. His work presaged later uses of the diagram to depict world trade equilibrium in the two-country case.

[4] Fisher omits the relevant indifference curve to avoid cluttering the diagram.

Enrico Barone

If Fisher was the first to use the diagram to show the gains from *intertemporal* trade, then Enrico Barone, the Italian mathematical economist and author of the famous article on "The Ministry of Production in the Collectivist State," was the first to use it to depict the gains from *international* trade.[5] In a long footnote in the 1908 edition of his *Principi di economia politica*, he presented a diagram showing pre- and post-trade equilibrium positions for a single national economy that produces and consumes two goods A and B (see Figure 3). His diagram, like Fisher's, consists of three types of curves.

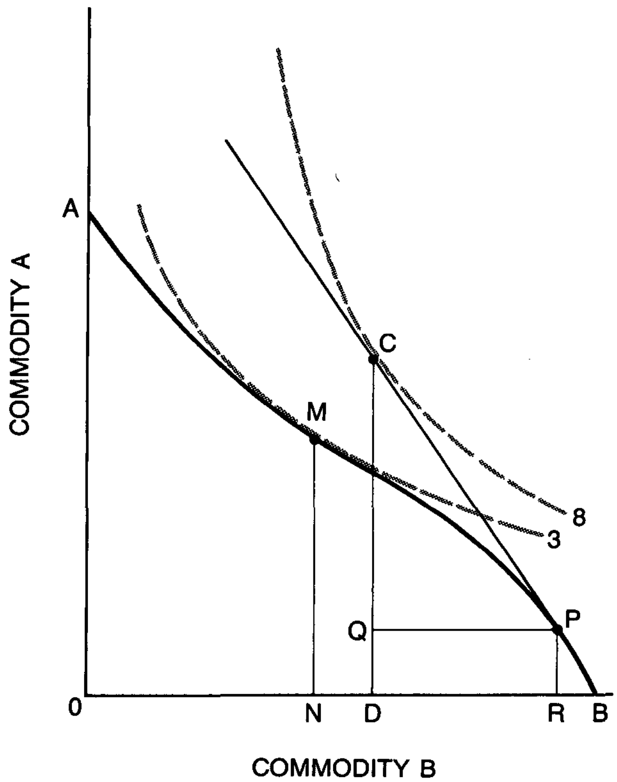
His "production indifference" or transformation curve AB shows the maximum alternative combinations of the two goods the economy can produce from available resources. Its nonlinear curved shape indicates that production takes place under conditions of nonconstant costs. The slope of the curve at any point M represents what Barone called "comparative cost," or the ratio of the marginal costs of production.

The curves bearing the numbers 3 and 8 are two of a set of community taste indifference curves that represent demand conditions in the economy. Each curve shows alternative commodity bundles yielding equal satisfaction. Higher curves represent higher levels of satisfaction as indicated by the higher numbers they bear. Finally, the curve PC is the world price line whose slope indicates the relative cost of obtaining goods A and B on the world market.

Before trade, the country produces and consumes at the autarky equilibrium point M characterized by the common tangency of production possibility and taste indifference curves. The slope of that tangent represents the domestic pre-trade price ratio and indicates that the country has a comparative cost advantage over the rest of the world in the production of good B.

[5] What follows draws heavily on Maneschi and Thweatt [12].

Figure 3
BARONE'S DIAGRAM



Given the opportunity to trade along the world price line PC, the country shifts production from autarky point M to specialization point P. Then it trades commodity bundle P for bundle C by exporting PQ of B for QC of A to reach its point of maximum satisfaction C.

Source: Maneschi and Thweatt [12, p. 381].

When trade opens up at the world price ratio given by the slope of line PC, the country exploits its comparative advantage by moving to production point P where the ratio of domestic marginal costs equals the world price ratio and GNP valued at world prices is maximized. In short, the country produces at the point of tangency of the transformation curve and the (highest attainable) world price line. Then it trades along that line, exporting PQ of good B in exchange for imports of QC of good A, until it reaches the point of maximum satisfaction C. By taking advantage of trade, it separates its production and consumption points and consumes beyond its transformation curve.

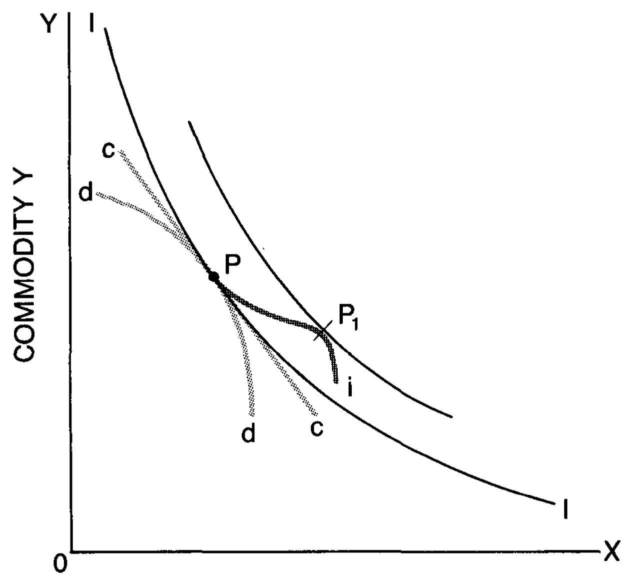
Here are all the elements found in modern versions of the diagram— the three-curve apparatus, the gap between autarky (closed economy) and world prices that makes trade feasible, the movement to the specialization point of maximum-value output, the post-trade separation of production and consumption points, and the trade triangle that reconciles those points. All this was a brilliant performance that should have made Barone the leading figure in the diagram's development. Such, however, was not the case. For all its brilliance, his contribution went largely unnoticed and consequently had no discernible influence on the work of his contemporaries and immediate successors. Barone himself may have been partly responsible for this state of affairs. By burying his diagram in a footnote of the 1908 *Principi* he effectively minimized its importance. That he may have intended to do so is suggested by his failure to include the diagram in his other writings. At any rate it is not to be found in later editions of the *Principi*. When it was finally restored to the 1936 edition it hardly seemed original. By then, other analysts had independently rediscovered the diagram and had developed it beyond Barone. Only in recent years, with the rediscovery of the extremely scarce 1908 edition of the *Principi*, have scholars been able to confirm the originality of Barone's contribution.

Allyn A. Young

After Fisher and Barone, work on the diagram languished. During the next 20 years (1909-1929) only one new version appeared in print and it was inferior to the earlier ones. In apparent ignorance of the contributions of Fisher and Barone, Allyn A. Young in the appendix to his famous 1928 *Economic Journal* article on "Increasing Returns and Economic Progress" presented a closed-economy version of the diagram that according to him derived straight from Pareto (see Figure 4).

Young did not use his diagram to illustrate comparative advantage or the gains from trade. Still he merits recognition for at least three reasons. He anticipated Gottfried Haberler by two years in defining the slope of the production frontier (his "curve of equal costs") as the opportunity cost of producing a unit increase in either good in terms of the amount of the other good sacrificed. Also he explained better than his predecessors that a concave curve reflects increasing opportunity cost, a linear curve constant cost, and a convex curve decreasing cost. Finally, he indicated how increasing returns in one industry might introduce a convex segment into a otherwise concave curve. In this connection

Figure 4
YOUNG'S DIAGRAM



Curves dd and cc are the production frontiers showing increasing costs and constant costs, respectively. Curves dPi and cPi represent cases in which decreasing costs prevail over part of the production frontier. Tangency with indifference curves II, etc., yields equilibrium at P in the first set of cases, P_1 in the second.

Source: Young [18, p. 540].

he discussed stability of closed-economy equilibrium under increasing, constant, and decreasing costs. He correctly noted that stability is ensured in all cases provided collective indifference curves possess greater convexity than the production frontier.

As for collective indifference curves, he noted that their location on the diagram assumes a fixed distribution of income when in fact that distribution and thus the indifference map itself changes with movements along the production frontier. In other words, a reallocation of production from good X to good Y redistributes income from X producers to Y producers and thus shifts the indifference map. For this reason he thought such curves should be treated as an expository device and not as a rigorous conception. His discussion anticipated Lerner and Tinbergen, both of whom analyzed decreasing costs, and Viner, who criticized the concept of community indifference maps.

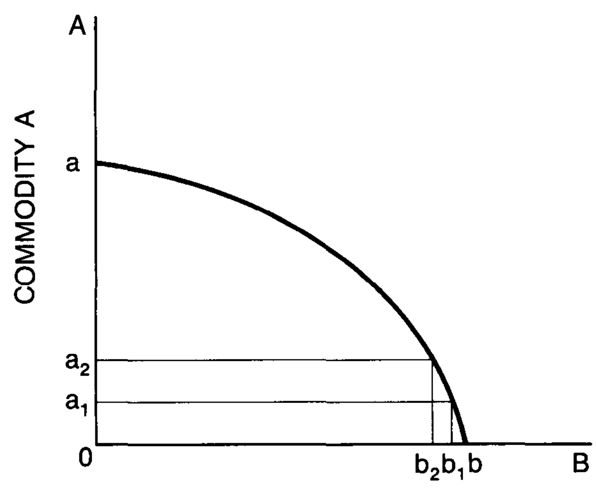
Gottfried Haberler

We have seen how Fisher in 1907 invented the diagram, how Barone in 1908 extended it to international trade, and how Young in 1928 applied it to the closed economy. In 1930, however, Gottfried Haberler in his seminal paper on comparative cost, did what none of his predecessors had done.[6] He introduced into international trade theory a strictly concave production possibility curve showing diminishing returns and increasing costs in the production of both goods (see Figure 5). Fisher and Young, of course, had worked with such concave transformation curves, but not within the context of international trade theory. Barone, on the other hand, had used transformation curves to analyze foreign trade. But the curves he used were not strictly concave.

---

[6] See Haberler [6] for an English translation of his 1930 paper from the original German. Haberler's diagram and its underlying analysis also appears in Chapter 10 of his *The Theory of International Trade, with Its Applications to Commercial Policy* (1936).

Figure 5
HABERLER'S STRICTLY CONCAVE PRODUCTION FRONTIER



The concavity of curve ab shows that successive unit increases in one good require progressively larger decreases in the other. The opportunity cost of each good increases as more is produced.

Source: Haberler [6, p. 10].

Nor had Haberler's predecessors adequately explained the reasons for the curve's concave shape. Such concavity they attributed to diminishing returns and increasing costs without specifing the forces causing these phenomena. Haberler, however, explained the causes of the curve's concavity by invoking the notion of specific and nonspecific factors of production. Specific factors he defined as those tied to a particular industry and suitable to the production of no other good. Nonspecific factors on the other hand are those freely transferable between industries and equally suited to the production of both goods.

Using a two-good, three-factor model, he assumed that each good requires for its production one specific factor which it uses exclusively and a nonspecific factor shared in common with the other industry. Combining increasing amounts of the nonspecific factor with fixed amounts of a specific one to produce more of either good yields decreasing increments of output, i.e., diminishing returns. Thus the amount of one good sacrificed to free enough nonspecific resources to produce a unit increase in the other good must rise as output of the latter good increases. The same thing would happen, Haberler noted, if all resources, though mobile, were not equally well-suited for different employments. For example, suppose that of the nation's fixed stock of resources all initially employed in producing A, part is better suited to producing B. One might think of mountainous land better suited to skiing or mining than to wheat production. Transferring such resources to B at first results in a large rise in the output of that good at the cost of little sacrifice of A. Beyond some point, however, continued expansion of B necessitates the transfer of resources less and less suited to B production and more and more suited to A production. At that point the opportunity costs of B in terms of A sacrificed rises. Either case, Haberler said, yields a smooth concave curve with the marginal opportunity cost of transforming one good into the other rising continuously over the whole range of the curve.

Finally, Haberler better than any of his predecessors explained the place of the transformation curve in the theory of comparative advantage. According to him, the curve together with demand conditions (indifference curves) determines an economy's production point and thus relative commodity costs in the absence of trade. On the assumption that prices equal costs, those curves also determine relative commodity prices. Differences in these autarky relative costs and prices across nations reflect comparative advantages that make trade mutually advantageous. When trade takes place at the equilibrium world price ratio each nation tends to specialize in the production of the commodity of its comparative advantage. As it does so, however, it incurs increasing opportunity costs. Specialization continues up to the point at which marginal opportunity costs equal world prices, i.e., up to the point at which the transformation curve just touches the world price line. Each nation then trades along that line, exporting its comparative advantage commodity in exchange for the other commodity, until it reaches its point of maximum satisfaction.

Haberler's analysis had a galvanizing effect on his contemporaries. In quick succession Jacob Viner, Abba Lerner, and Wassily Leontief combined his concave transformation curve with collective indifference curves to obtain the basic diagram of the trade theorist. Each of these writers, however, put the diagram to somewhat different uses described below.

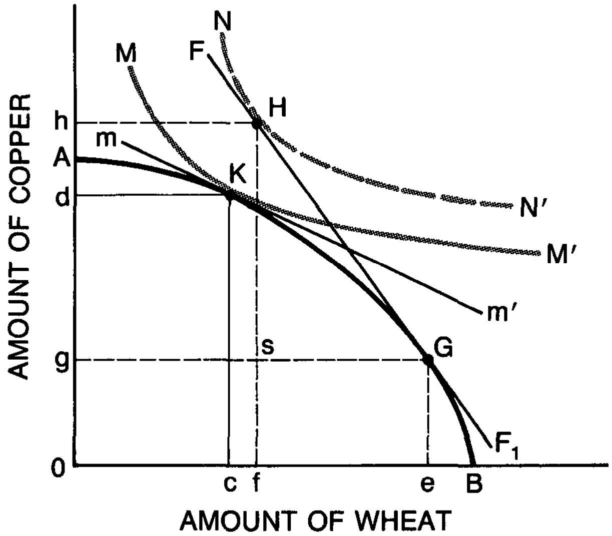
Jacob Viner

Viner's version of the diagram, presented in a lecture at the London School of Economics in January 1931 but not published until the 1937 appearance of his *Studies in the Theory of International Trade*, shows before- and after-trade equilibria for a single country (see Figure 6). Before trade, the country produces and consumes at point K on the highest attainable indifference curve tangent to the production frontier. When presented with the opportunity to trade at a world price ratio different from the autarky one—this difference indicated by the different slopes of the price lines FF_1 and mm'—the country shifts production to point G and then trades along the world price line, exporting Gs units of wheat in exchange for imports of sH units of copper. In so doing, it ends up consuming commodity bundle H lying on a higher indifference curve than the autarky bundle K consumed before trade.

Except for the concavity of the production possibility curve, Viner's diagram is virtually the same as Barone's. But Viner did one thing that neither Barone nor anyone else had done up to that time. He pointed to certain logical flaws in the diagram's construction and questioned its usefulness in showing the gains from trade.

In particular, he focused on the shortcomings of community indifference maps and production possibility curves. Community indifference maps were suspect because they embodied the assumption of a fixed distribution of income when in fact trade would change that distribution and thus the indifference map itself. Likewise the production

Figure 6
VINER'S DIAGRAM



Given the opportunity to trade at world prices shown by the slope of line FF_1, the economy shifts production from autarky bundle K to bundle G, which it then trades for preferred bundle H by exporting Gs wheat for sH copper.

Source: Viner [17, p. 521].

possibility curve was flawed because it assumed perfectly inelastic (fixed) factor supplies when in fact those supplies vary with changes in their prices. Trade, by changing factor prices, would change the quantities of factors supplied and thus the production possibility curve itself. Nor was this the only problem. The curve, Viner noted, also embodied the assumption of factor indifference between alternative uses when in reality factors may prefer one employment to another. Assuming factors employed in the industry of their preference are paid the value of their marginal product there, they must receive a premium over that to induce them to work in the other industry. In that case, factor costs to one industry will not equal sacrificed factor product in the other, and the cost of securing a unit increase in either good is not accurately measured by the quantity of the other good given up.[7] Viner's conclusion was straightforward. Job preferences and the resulting compensating pay differentials drive a wedge between commodity relative prices and the ratio of factor marginal products reflected in the slope of the transformation schedule. In other words, prices would not reflect opportunity costs as Haberler supposed.

Viner's trenchant criticisms proved less than devastating. For the production possibility curve was simply too useful a tool to abandon. Despite its restrictive assumptions, it captured the essence of a country's commodity supply conditions. For that reason, trade theorists chose the diagram and its underlying opportunity cost interpretation over Viner's real cost interpretation.

Abba Lerner

Unlike Viner, Lerner accepted the trade diagram uncritically. He used it to depict trade equilibrium for the aggregate world economy in a two-country model.[8] His demonstration, as presented in his celebrated 1932 *Economica* article on "The Diagrammatical Representation of Cost Conditions in International Trade," required three steps.

First, he derived the world transformation curve by optimally adding national production possibilities at equal marginal cost ratios. He did so by sliding one country's production possibility block along the other's with the slopes or marginal opportunity cost ratios always kept equal (see Figure 7). In this way he traced out an efficient world production possibility frontier, something nobody had done before.

Second, he confronted this world production frontier with a global community indifference curve which he implicitly derived by aggregating over the underlying country curves (not shown by him). The resulting common point of tangency of the two curves determines the world production and consumption points as well as the equilibrium terms of trade.

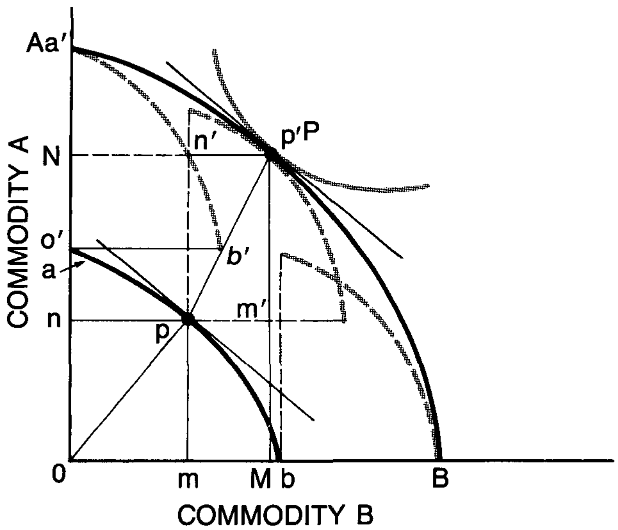
Finally, he located each country's post-trade production point by moving the world terms-of-trade line parallel to itself until it just touched the individual production possibility curves. He did not identify the consumption point or the exports and imports of each

[7] An example will suffice. Industry A pays each unit of labor a real wage w_A equal to its marginal product there. But that same labor unit costs industry B the amount w_A+d, where d is the wage differential or pay premium that compensates for the nonpecuniary disadvantages (subjective disutility) of working in B. Thus labor's cost to B exceeds its foregone product in A by the factor d. Similarly, labor's marginal product in B equals its wage rate there, w_A+d. But that same unit of labor costs A only w_A. Thus labor's cost to A understates its sacrificed alternative product by the factor d. True costs deviate from opportunity cost.

[8] On Lerner see Mundell [13, pp. 147-48] and Samuelson [15, p. 645].

Figure 7

LERNER'S DERIVATION OF THE WORLD TRANSFORMATION CURVE



Moving one country's production block along the other's traces out the world transformation curve AB. The diagram shows three successive positions of the second country's block o'a'b' as it slides along the first country's production frontier ab. Tangency of transformation curve and indifference curve yields world equilibrium at P with country post-trade production points being p' and p, respectively.

Source: Lerner [11, p. 90].

nation. But he did remark that both nations would benefit from trade even if they possessed identical concave transformation curves provided their indifference maps differed. His remark anticipated Wassily Leontief's geometrical demonstration of this case.

He also showed what the world production possibility curve looks like when at least one of the countries produces under conditions of increasing returns such that its production frontier is convex. Richard E. Caves neatly summarizes his analysis.

> He proved that increasing returns necessitate complete specialization by at least one country. This can occur not only when both countries' transformation curves are convex to the origin, but also if one (national) transformation curve is convex while the other shows a constant rate of transformation, or even concavity to the origin, so long as the convexity of the one exceeds the concavity of the other. There will normally be points on the world transformation curve where more than one pattern of international specialization is efficient. No matter which of the two countries specializies completely, the same commodity totals will be produced. Another trait of such a point is if a change in world tastes is moving the world production combination past one, the optimal pattern of specialization may shift markedly [3, pp. 162-63].

Wassily Leontief

In the year after Lerner's article appeared, Leontief in his paper on "The Use of Indifference Curves in the Analysis of Foreign Trade" completed Lerner's demonstration of world trade equilibrium. He did so by depicting for both countries the post-trade consumption points and trade triangles that connect those points with their corresponding production points, something Lerner had failed to do. Unlike Lerner, however, he did not work with world production possibility and taste indifference curves. Instead, he focused on the curves of each country, combining them together in a single chart. In this way he was able to use the diagram to show how trade affects both countries simultaneously.

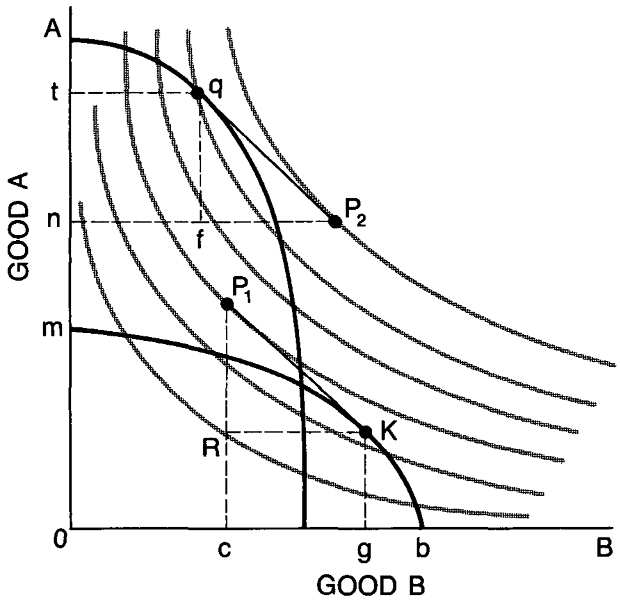
He showed how gains from trade arise when (1) production conditions alone and (2) demand conditions alone differ across countries. In the first case, countries have different production possibility curves but identical indifference maps (see Figure 8A). In the second case (anticipated by Lerner), production possibility curves are the same and only indifference maps differ across countries (see Figure 8B).

Figure 8A depicts the first case. Here the country possessing the vertically elongated transformation curve produces at q where its output valued at world prices is maximized. Then it trades along the relative price line qP_2, exporting qf of good A against imports of fP_2 of good B, and consumes at P_2, a point it could not reach before trade when it was constrained to consume on its production possibility curve. Likewise the other country gains by producing its highest valued output at K, trading along the price line KP_1, and consuming at P_1 beyond its production possibility frontier.

As for equilibrium conditions, Leontief specified that the price lines connecting the production and consumption points must be of the same slope and length for both countries. The first condition ensures that both countries face the same price ratio or terms of trade. The second ensures that exports of one country equal imports of the other. In other words, it ensures that the trade triangles P_1RK and qfP_2 are the same, as required for international equilibrium.

Figure 8a

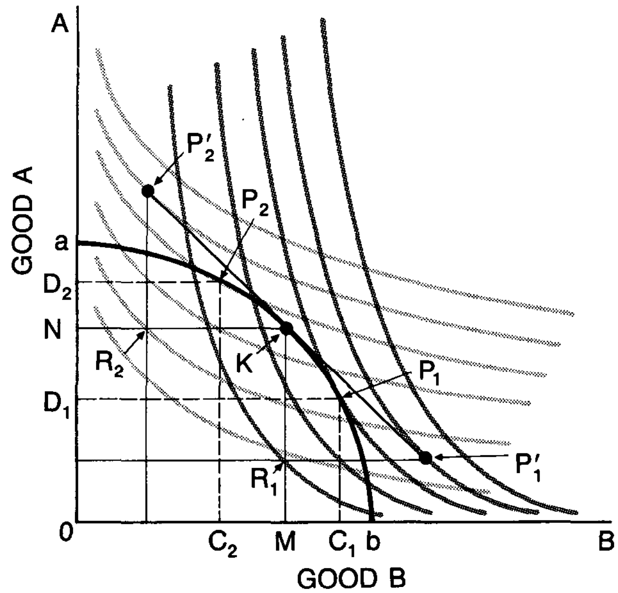
Different Transformation Curves, Identical Indifference Maps



One country produces at q and exports qf of A for fP_2 of B. The other produces at K and exports KR of B for RP_1 of A. The equilibrium world price ratio shown by the common slope of lines qP_2 and P_1K must be such as to make the trade triangles identical.

Figure 8b

Identical Transformation Curves, Different Indifference Maps



Both countries produce at K, one exporting KR_1 of A for $R_1P'_1$ of B, the other exporting KR_2 of B for $R_2P'_2$ of A. The equilibrium world price ratio must be such as to make the trade triangles identical.

Source: Leontief [10, pp. 25, 27].

Trade also enables countries to consume beyond their production possibility curves when only demand conditions (indifference maps) differ. Leontief's second diagram shows why: different demand conditions result in different pre-trade equilibrium points on the production possibility curve. At these different points, comparative costs differ making trade advantageous.

Thus before trade the country with the steeper indifference curves initially consumes and produces at P_1 on its production possibility curve while the other country does the same at P_2. The different slopes of the production possibility curve at those two autarky points show that comparative costs differ across countries making trade profitable. When trade takes place at the equilibrium price ratio given by the slope of line $P'_2P'_1$, each country produces at K and exports the good in which it has a (pre-trade) cost advantage. The first country exports KR_1 of good A for imports of $R_1P'_1$ of good B, reaching consumption point P'_1 in the process. Similarily, the other country exports R_2K of good B in exchange for imports of $R_2P'_2$ of good A, and consumes at P'_2 beyond its production possibility curve. Both countries gain from trade despite having identical production frontiers. Here in Leontief's 1933 diagram is everything and more found in the earlier constructions of his predecessors.

In short, Leontief brought the diagram to its highest stage of development up to the mid-1940s and established it as the standard geometrical tool of the international trade textbooks. It was his version, showing as it does in one Cartesian plane the mutual gains from trade and the international equilibrium conditions for both countries simultaneously, that entered such influential early texts as D.B. Marsh's *World Trade and Investment* (1951) and Charles Kindleberger's *International Economics* (1953). Even today one finds it in such leading texts as Caves' and Jones' *World Trade and Payments* and W. Ethier's *Modern International Economics*.

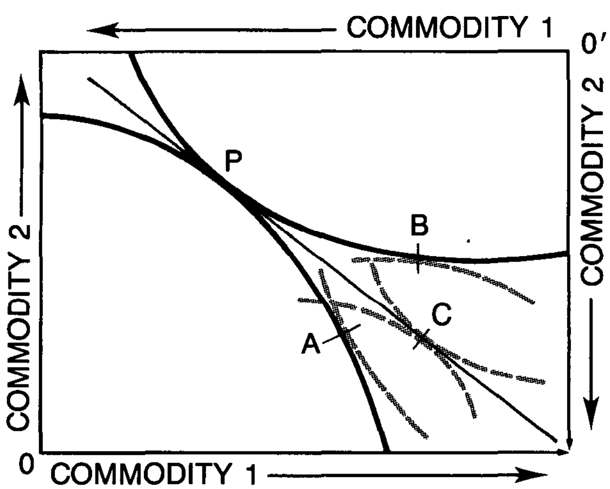
Jan Tinbergen

That Leontief's diagram had by the 1940s already become the standard way to depict international equilbrium under conditions of increasing costs and competitive markets is evident from a glance at Tinbergen's 1945 contribution. His treatment of this case differs in no essential way from Leontief's. Like Leontief he shows the individual open economy producing at the point of tangency of the production frontier and world price line and then trading along that line to reach the consumption point of maximum satisfaction (see Figure 1). And like Leontief he shows that a similar outcome holds for the other country whose exports must also equal the imports of the first and vice versa.

Tinbergen extends Leontief's analysis in two minor respects. He lets production possibility curves *and* indifference maps differ across countries. And he depicts the two-country equilibrium in a box diagram showing the second country's system of coordinates lying diagonally opposite those of the first (see Figure 9). But these are merely trivial differences in mode of presentation. The results he obtains are exactly the same as those shown in Leontief's diagrams.

Figure 9

TINBERGEN'S TWO-COUNTRY DIAGRAM OF WORLD TRADE EQUILIBRIUM



Country A's coordinates are plotted from 0, country B's from 0′. Global equilibrium requires both countries produce and consume at common points of tangency P and C on the world price line PC.

Source: Tinbergen [16, p. 137].

Only when he drops Leontief's assumptions of competitive behavior and increasing costs does he develop some novel results. He considers three cases, two of which yield the perverse outcome that trade may worsen rather than improve a country's welfare. He takes first the case of decreasing costs prevailing in both industries such that the transformation curve becomes convex rather than concave to the origin. He shows here that with trade the only stable equilibria in production are the terminal points on the curve representing complete specialization in one good or the other. Which good the country chooses to produce upon the opening of trade depends on the slope of the world price line and the shape of the indifference curves. Either choice will yield gains from trade.

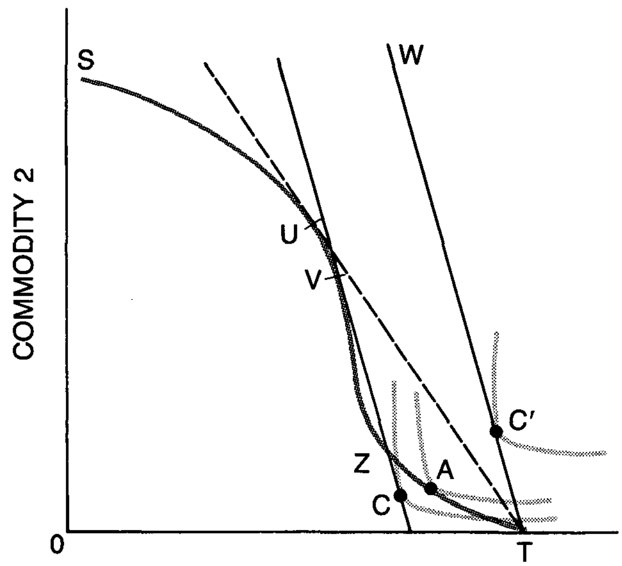
Next he considers the case (anticipated by Young and Lerner) in which decreasing costs prevail in one industry and increasing costs in the other such that the production frontier contains convex and concave segments. He argues that in this case an open trading economy may myopically choose production and consumption points that worsen its welfare compared to the no-trade situation (see Figure 10). That is, given the world price ratio shown by the slope of line VC, the economy chooses production point V and consumption point C which is inferior to autarky point A. But he then notes that this construction assumes that producers and consumers lack perfect knowledge of their opportunities. Otherwise they would produce at T and trade along the price line TW (same slope as VC) to reach consumption point C′ which is superior to the autarky point.

Last he presents a case in which monopoly pricing in the industry possessing a comparative cost advantage distorts relative commodity prices and causes the country to produce and export the wrong good, namely the good in which it has a comparative disadvantage. Depending on the shape of the indifference map, the economy may be better off or worse off than before trade (see Figure 11).

These results of course differ from Leontief's. But Tinbergen reached them with the same geometrical tools. He changed the shape and location of the diagram's curves, to be sure. But he put those curves to their traditional use to depict international equilibrium and the gains (or losses) from trade, albeit for anomalous cases. In this respect, his work continued the tradition stretching from Barone to Leontief.

Figure 10

TRADE EQUILIBRIUM WITH A MIXED TRANSFORMATION CURVE

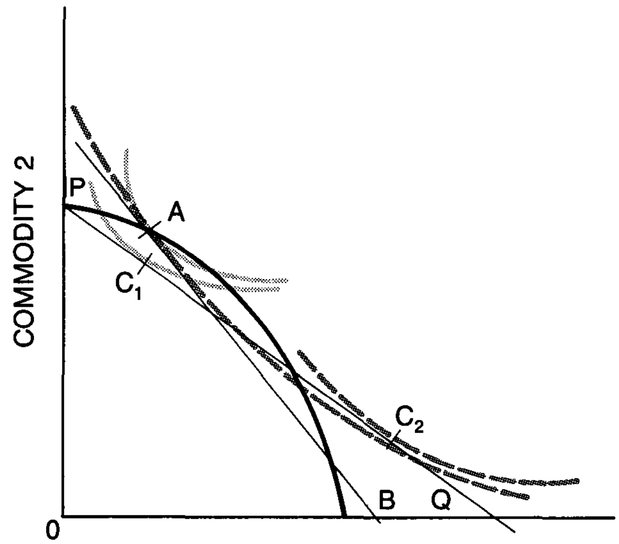


Imperfect knowledge and a mixed (concave-convex) transformation curve can make the country worse off with than without trade. At the world price ratio given by the slope of line VC, the economy produces at V and consumes at point C which lies on a lower indifference curve than the autarky point A. Conversely, with perfect knowledge the economy produces at T and consumes at C′, reaping a clear gain.

Source: Tinbergen [16, p. 133].

Figure 11

MONOPOLY PRICING IN THE COMPARATIVE ADVANTAGE INDUSTRY



Monopoly pricing raises the relative price of good 1 (slope of line AB) above its relative marginal cost (slope of the production frontier at autarky point A) and makes it appear that comparative advantage lies in good 2 when in fact it lies in good 1. Consequently, when trade opens up at the world price ratio given by the slope of line PQ, the economy specializes in the wrong good, producing at P and trading along line PQ to reach point C_1 or C_2 depending on the location of the indifference map. Trade yields losses in the first case, gains in the second.

Source: Tinbergen [16, p. 136].

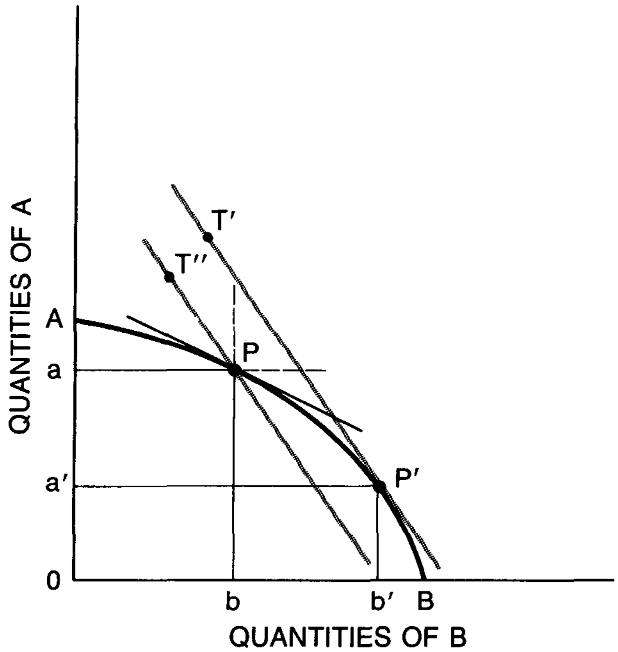
The Diagram Since Tinbergen

After Tinbergen, Haberler in 1950 used the diagram to distinguish between the consumption (exchange) and production (specialization) components of the total gain from trade. The total gain of course is the jump from the autarky consumption point to the preferred point on the (highest attainable) world price line just touching the production possibility curve. Of this total, the consumption gain stems from the opportunity to exchange the pre-trade bundle of goods at world prices. Haberler shows this gain as the movement from P to T″ along a world price line passing through the pre-trade consumption point (see Figure 12). Added to this is the production gain stemming from the opportunity to produce the highest valued bundle of commodities measured at world prices. Haberler shows this gain as the movement from T″ to T′ that results when the economy produces the output mix whose marginal opportunity cost just equals the world terms of trade.

The point of Haberler's demonstration is this: of the two sources of gain, exchange and specialization, the first is fundamental. For, as the diagram shows, exchange yields gains even in the absence of specialization (that is, in the absence of a change of production). The economy simply trades its given autarky bundle for a preferred one at world prices. By contrast, specialization without exchange yields no gains. For it never pays to produce the output mix valued highest at world prices when one cannot trade at those prices: in such cases the autarky mix

Figure 12

HABERLER ON THE GAINS FROM TRADE



Consumption (exchange) gains are shown by the jump from P to T″ as the economy swaps its autarky bundle for a preferred one at world prices. Production (specialization) gains are shown by the further jump to T′ that results when the economy produces and trades the output bundle P′ having the highest value at world prices. Trade yields gains even in the absence of specialization.

Source: Haberler [8, p. 38].

is preferred. On the contrary, specialization without trade yields losses since a closed economy must be self-sufficient (diversified) in all goods. In short, exchange rather than specialization is the necessary and sufficient condition for trade gains.

Haberler's demonstration did not exhaust the diagram's potential: new uses were found for it. Haberler himself employed it again in 1950 to illustrate the infant industry argument for protection. In 1952 James Meade employed it to derive trade indifference curves used in advanced trade geometry. Harry Johnson in 1964 used it to depict noneconomic objectives of tariffs. Jagdish Bhagwati in 1957 used it to show the effects of technological progress on the terms of trade and national welfare. Robert Mundell in 1957 used the diagram to show how international factor mobility negates the protective effects of tariffs. Haberler in 1950, Bhagwati and Ramaswami in 1963, and Johnson in 1965 employed the diagram to analyze domestic market distortions (divergences between private and social marginal costs) arising from external economies or diseconomies and rigid factor prices. The best corrective, they showed, is not a tariff but rather taxes and subsidies in the sector in which the distortions arise.

In all these uses the diagram proved its strength and versatility. So much so that trade theorists will undoubtedly employ it again and again. When they do, they will owe a large debt of gratitude to the pioneers who developed this powerful tool. Even today, if one understands the diagram one understands the logic of comparative advantage and gains from trade.

References

1. Baldwin, R.E. "Gottfried Haberler's Contribution to International Trade Theory and Policy." *Quarterly Journal of Economics* 97 (February 1982): 141-48.

2. Barone, E. *Principi di economia politica*. Roma: Tipografia Nazionale di G. Bertero, 1908.

3. Caves, R.E. *Trade and Economic Structure*. Cambridge, MA.: Harvard University Press, 1960.

4. Fisher, I. *The Rate of Interest*. New York: Macmillian Co., 1907.

5. ———. *The Theory of Interest*. New York: Macmillian Co., 1930.

6. Haberler, G. "The Theory of Comparative Costs and Its Use in the Defense of Free Trade." *Weltwirtschaftliches Archiv* 32 (July 1930): 349-70. As reprinted in *Selected Essays of Gottfried Haberler*, edited by A.Y.C. Koo. Cambridge, MA.: MIT Press, 1985, pp. 1-19.

7. ———. *The Theory of International Trade, with Its Applications to Commercial Policy*. London: William Hodge & Co., 1936.

8. ———. "Some Problems in the Pure Theory of International Trade." *Economic Journal* 60 (June 1950): 223-40. As reprinted in *Selected Essays of Gottfried Haberler*, edited by A.Y.C. Koo. Cambridge, MA.: MIT Press, 1985, pp. 37-54.

9. Hirshleifer, J. "On the Theory of the Optimal Investment Decision." *Journal of Political Economy* 66 (August 1958): 329-52.

10. Leontief, W.W. "The Use of Indifference Curves in the Analysis of Foreign Trade." *Quarterly Journal of Economics* 47 (May 1933): 493-503. As reprinted in *International Trade, Selected Readings*, edited by J. Bhagwati. Harmondsworth, England: Penguin Books, 1969, pp. 21-29.

11. Lerner, A.P. "The Diagrammatical Representation of Cost Conditions in International Trade." *Economica* 34 (August 1932): 346-56. As reprinted in his *Essays in Economic Analysis*. London: Macmillian Co., 1953, pp. 85-100.

12. Maneschi, A., and W.O. Thweatt. "Barone's 1908 Representation of an Economy's Trade Equilibrium and the Gains from Trade." *Journal of International Economics* 22 (May 1987): 375-82.

13. Mundell, R. "Abba Lerner and the Theory of Foreign Trade." In *Theory for Economic Efficiency: Essays in Honor of Abba P. Lerner*, edited by H.I. Greenfield, A.M. Levenson, W. Hamovitch, and E. Rotwein. Cambridge, MA.: MIT Press, 1983.

14. Samuelson, P.A. "A.P. Lerner at Sixty." *Review of Economic Studies* 31 (1964): 169-78.

15. ———. "Irving Fisher and the Theory of Capital." In *Ten Economic Studies in the Tradition of Irving Fisher*. New York: J. Wiley, 1967, pp. 17-37.

16. Tinbergen, J. "Professor Graham's Case for Protection." In Appendix I of his *International Economic Co-operation*. Amsterdam: Elsevier, 1945. Reprinted as "International Trade under Variable Returns in a Very Simple Model" in Appendix II of his *International Economic Intergration*. 2d ed., rev. Amsterdam: Elsevier, 1965, pp. 126-37.

17. Viner, J. *Studies in the Theory of International Trade*. New York: Harper & Brothers, 1937.

18. Young, A.A. "Increasing Returns and Economic Progress." *Economic Journal* 38 (December 1928): 527-42.

THEORETICAL ANALYSIS OF THE DEMAND FOR MONEY

*Bennett T. McCallum and Marvin S. Goodfriend**

In any discussion of the demand for money it is important to be clear about the concept of money that is being utilized; otherwise, misunderstandings can arise because of the various possible meanings that readers could have in mind. Here the term will be taken to refer to an economy's *medium of exchange*; that is, to a tangible asset that is generally accepted in payment for any commodity. Money thus conceived will also serve as a store of value, of course, but may be of minor importance to the economy in that capacity. The monetary asset will usually also serve as the economy's medium of account—i.e., prices will be quoted in terms of money—since additional accounting costs would be incurred if the unit of account were a quantity of some asset other than money. The medium-of-account role is, however, not logically tied to the medium of exchange (Wicksell, 1906; Niehans, 1978).

Throughout much of Western history, most economies have adopted as their principal medium of exchange a commodity that would be valuable even if it were not used as money. Recently, however, fiat money—intrinsically worthless tokens made of paper or some other cheap material—has come to predominate. Under a commodity money arrangement, the exchange value of money will depend upon the demand for the monetary commodity in its nonmonetary, as well as its monetary, uses. But in a discussion of money demand, as distinct from a discussion of the price level, any possible nonmonetary demand for the medium of exchange—which will be absent anyhow in a fiat money system—can legitimately be ignored.

The quantity of money demanded in any economy—indeed, the set of assets that have monetary status—will be dependent upon prevailing institutions, regulations, and technology. Technical progress in the payments industry will, for instance, tend to alter the quantity of money demanded for given values of determinants such as income. This dependence does not, however, imply that the demand for money is a nebulous or unusable concept, any more than the existence of technical progress and regulatory change in the transportation industry does so for the demand for automobiles. In practice, some lack of clarity pertains to the operational measurement of the money stock, as it does to the stock of automobiles or other commodities. But in an economy with a well-established national currency, the principle is relatively clear: assets are part of the money stock if and only if they constitute *claims* to currency, unrestricted legal claims that can be promptly and cheaply exercised (at par). This principle rationalizes the common practice of including demand deposits in the money stock of the United States, while excluding time deposits and various other assets.

The rapid development during the 1960s and 1970s of computer and telecommunications technologies has led some writers (e.g., Fama, 1980) to contemplate economies—anticipated by Wicksell (1906)—in which virtually all purchases are effected not by the transfer of a tangible medium of exchange, but by means of signals to an accounting network—signals that result in appropriate debits and credits to the wealth accounts of buyers and sellers. If there were literally *no* medium of exchange, the wealth accounts being claims to some specified bundle of commodities, the economy in question would be properly regarded and analyzed as a nonmonetary economy—albeit one that avoids the inefficiencies of crude barter. If, by contrast, the accounting network's credits were claims to quantities of a fiat or commodity medium of exchange, then individuals' credit balances would appropriately be included as part of the money stock (McCallum, 1985).

---

* Bennett T. McCallum is H. J. Heinz Professor of Economics, Carnegie-Mellon University, and Research Advisor, Federal Reserve Bank of Richmond. Marvin S. Goodfriend is Economist and Vice President, Federal Reserve Bank of Richmond.

This article is reprinted with permission from *The New Palgrave: A Dictionary of Economics*, edited by John Eatwell, Peter Newman, and Murray Milgate. 4 volumes. London: The Macmillan Press; New York: Stockton Press, 1987.

Basic Principles

An overview of the basic principles of money demand theory can be obtained by considering a hypothetical household that seeks at time t to maximize

$$(1)\quad u(c_t,\ell_t) + \beta u(c_{t+1},\ell_{t+1}) + \beta^2 u(c_{t+2},\ell_{t+2}) + \ldots$$

where c_t and ℓ_t are the household's consumption and leisure during t and where $\beta = 1/(1+\delta)$, with $\delta > 0$ the rate of time preference. The within-period utility function $u(\cdot,\cdot)$ is taken to be well-behaved so that unique positive values will be chosen for c_t and ℓ_t. The household has access to a productive technology described by a production function that is homogeneous of degree one in capital and labor inputs. But for simplicity we assume that labor is supplied inelastically, so this function can be written as $y_t = f(k_{t-1})$, where y_t is production during t and k_{t-1} is the stock of capital held at the end of period $t-1$. The function $f(\cdot)$ is well-behaved, so a unique positive value of k_t will be chosen for the upcoming period. Capital is unconsumed output, so its price is the same as that of the consumption good and its rate of return between t and t+1 is $f'(k_t)$.

Although this setup explicitly recognizes the existence of only one good, it is intended to serve as a simplified representation—one formally justified by the analysis of Lucas (1980)—of an economy in which the household sells its specialized output and makes purchases (at constant relative prices) of a large number of distinct consumption goods. Carrying out these purchases requires *shopping time*, s_t, which subtracts from leisure: $\ell_t = 1 - s_t$, where units are chosen so that there is one unit of time per period available for shopping and leisure together. (If labor were elastically supplied, then labor time would have to be included in the expression.) In a monetary economy, however, the amount of shopping time required for a given amount of consumption will depend negatively upon the quantity of real money balances held by the household (up to some satiation level). For concreteness, we assume that

$$(2)\quad s_t = \psi(c_t,m_t)$$

where $\psi(\cdot,\cdot)$ has partial derivatives $\psi_1 > 0$ and $\psi_2 \leq 0$. In (2), $m_t = M_t/P_t$, where M_t is the nominal stock of money held at the end of t and P_t is the money price of a consumption bundle. (A variant with M_t denoting the start-of-period money stock will be mentioned below.) The transaction variable is here specified as c_t rather than $c_t + \Delta k_t$ to reflect the idea that only a few distinct capital goods will be utilized, so that the transaction-cost to expenditure ratio will be much lower than for consumption goods.

Besides capital and money, there is a third asset available to the household. This asset is a nominal bond, i.e., a one-period security that may be purchased at the price $1/(1+R_t)$ in period t and redeemed for one unit of money in t+1. The symbol B_t will be used to denote the number (possibly negative) of these securities purchased by the household in period t, while $b_t = B_t/P_t$.

In the setting described, the household's budget constraint for period t may be written as follows:

$$(3)\quad f(k_{t-1}) + v_t \geq c_t + k_t - k_{t-1} + m_t - (1+\pi_t)^{-1}m_{t-1} + (1+R_t)^{-1}b_t - (1+\pi_t)^{-1}b_{t-1}.$$

Here v_t is the real value of lump-sum transfers (net of taxes) from the government while π_t is the inflation rate, $\pi_t = (P_t - P_{t-1})/P_{t-1}$. Given the objective of maximizing (1), first-order conditions necessary for optimality of the household's choices include the following, in which ϕ_t and λ_t are Lagrangian multipliers associated with the constraints (2) and (3), respectively:

$$(4)\quad u_1(c_t,1-s_t) - \phi_t\psi_1(c_t,m_t) - \lambda_t = 0$$

$$(5)\quad -u_2(c_t,1-s_t) + \phi_t = 0$$

$$(6)\quad -\phi_t\psi_2(c_t,m_t) - \lambda_t + \beta\lambda_{t+1}(1+\pi_{t+1})^{-1} = 0$$

$$(7)\quad -\lambda_t + \beta\lambda_{t+1}[f'(k_t) + 1] = 0$$

$$(8)\quad -\lambda_t(1+R_t)^{-1} + \beta\lambda_{t+1}(1+\pi_{t+1})^{-1} = 0.$$

These conditions, together with the constraints (2) and (3), determine current and planned values of c_t, s_t, m_t, k_t, b_t, ϕ_t, and λ_t for given time paths of v_t, R_t, and π_t (which are exogenous to the household) and the predetermined values of k_{t-1}, m_{t-1}, and b_{t-1}. (There is also a relevant transversality condition, but it can be ignored for the issues at hand.) Also, ℓ_t values can be obtained from $\ell_t = 1 - s_t$ and, with P_{t-1} given, P_t, M_t, and B_t values are implied by the π_t, m_t, and b_t sequences.

The household's optimizing choice of m_t can be described in terms of two distinct concepts of a money demand function. The first of these is a proper demand function, that is, a relationship giving the chosen quantity as a function of variables that are either predetermined or exogenous to the economic unit in question. In the present context, the money demand function of that type will be of the form

$$(9)\quad m_t = \mu(k_{t-1},m_{t-1},b_{t-1},v_t,v_{t+1},\ldots,R_t,R_{t+1},\ldots,\pi_t,\pi_{t+1},\ldots)$$

where the variables dated t+1, t+2,... must be understood as anticipated values. Now, it will be obvious that this relationship does not closely resemble those normally described in the literature as "money demand functions." There is a second type of relationship implied by the model, however, that does have such a resemblance. To obtain this second expression, one can eliminate $\beta\lambda_{t+1}(1+\pi_{t+1})^{-1}$ between equations (6) and (8), then eliminate λ_t and finally ϕ_t from the resultant by using (4) and (5). These steps yield the following:

$$(10)\quad -u_2(c_t,1-s_t)\psi_2(c_t,m_t) = [u_1(c_t,1-s_t) - u_2(c_t,1-s_t)\,\psi_1(c_t,m_t)]\,[1-(1+R_t)^{-1}].$$

Then $\psi(c_t,m_t)$ can be used in place of s_t and the result is a relationship that involves *only* m_t, c_t, and R_t. Consequently, (10) can be expressed in the form

$$(11)\quad h(m_t,c_t,R_t) = 0$$

and if the latter is solvable for m_t one can obtain

$$(12)\quad M_t/P_t = L(c_t,R_t).$$

Thus the model at hand yields a *portfolio-balance* relationship between real money balances demanded, a variable measuring the volume of transactions conducted, and the nominal interest rate (which reflects the cost of holding money rather than bonds). It can be shown, moreover, that for reasonable specifications of the utility and shopping-time functions, $L(\cdot,\cdot)$ will be increasing in its first argument and decreasing in the second.

There are, of course, two problems in moving from a demand function (of either type) for an individual household to one that pertains to the economy as a whole. The first of these involves the usual problem of aggregating over households that may have different tastes and/or levels of wealth. It is well known that the conditions permitting such aggregation are extremely stringent in the context of any sort of behavioral relation; but for many theoretical purposes it is sensible to pretend that they are satisfied. The second problem concerns the existence of economic units other than households—"firms" being the most obvious example. To construct a model analogous to that above for a firm, one would presumably posit maximization of the present value of real net receipts rather than (1), and the constraints would be different. In particular, the shopping-time function (2) would need to be replaced with a more general relationship depicting resources used in conducting transactions as a function of their volume and the real quantity of money held. The transaction measure would not be c_t for firms or, therefore, for the economy as a whole. But the general aspects of the analysis would be similar, so we shall proceed under the presumption that the crucial issues are adequately represented in a setting that recognizes only economic units like the "households" described above.

The distinction between the proper money demand function (9) and the more standard portfolio-balance relation (12) is important in the context of certain issues. As an example, consider the issue of whether wealth or income should appear as a "scale variable" (Meltzer, 1963). From the foregoing, it is clear that wealth is an important determinant of money demand in the sense that k_{t-1}, m_{t-1}, and b_{t-1} are arguments of the demand function (9). Nevertheless, formulation (12) indicates that there is no separate role for wealth in a portfolio-balance relation if appropriate transaction and opportunity-cost variables are included.

An issue that naturally arises concerns the foregoing discussion's neglect of randomness. How would the analysis be affected if it were recognized that future values of variables cannot possibly be known with certainty? In answer, let us suppose that the household knows current values of all relevant variables including P_t, R_t, and v_t when making decisions on m_t and c_t, but that its views concerning variables dated t+1, t+2,... are held in the form of nondegenerate probability distributions. Suppose also that there is uncertainty in production, so that the marginal product of capital in t+1, $f'(k_t)$, is viewed as random. Then the household's problem becomes one of maximizing the expectation of (1), with $u(\cdot,\cdot)$ a Von Neumann-Morgenstern utility function, given information available in period t. Consequently, the first-order conditions (4)-(8) must be replaced with ones that involve conditional expectations. For example, equation (7) would be replaced with

$$(7')\quad -\lambda_t + \beta E_t\{\lambda_{t+1}[f'(k_t) + 1]\} = 0$$

where $E_t(\cdot)$ denotes the expectation of the indicated variable conditional upon known values of P_t, R_t, v_t, and so on. With this modification, the nature of the proper demand function becomes much more complex—indeed, for most specifications no closed form solution analogous to (9) will exist. Nevertheless, the portfolio-balance relation (12) will continue to hold exactly as before, for the steps described in its derivation above remain the same except that it is $E_t[\beta\lambda_{t+1}(1+\pi_{t+1})^{-1}]$ that is eliminated between equations corresponding to (6) and (8). From this result it follows that, according to our model, the relationship of M_t/P_t to the transaction and opportunity-cost variables is invariant to changes in the probability distribution of future variables.

Another specificational variant that should be mentioned reflects the assumption that it is money held at the start of a period, not its end, that facilitates transactions conducted during the period. If that change in specification were made and the foregoing analysis repeated, it would be found that the household's concern in period t would be to have the appropriate level of real money balances at the start of period t+1. The portfolio-balance relation analogous to (12) that would be obtained in the deterministic case would relate m_{t+1} to c_{t+1} and R_t, where $m_{t+1} = M_{t+1}/P_{t+1}$ with M_{t+1} reflecting money holdings at the end of period t. Consequently, M_{t+1}/P_t would be related to R_t, planned c_{t+1}, and P_t/P_{t+1}. Thus the theory does not work out as cleanly as in the case considered above even in the absence of randomness, and is complicated further by the recognition of the latter. The fundamental nature of the relationships are, however, the same as above.

Another point deserving of mention is that if labor is supplied elastically, the portfolio-balance relation analogous to (12) will include the real wage-rate as an additional argument. This has been noted by Karni (1973) and Dutton and Gramm (1973). More generally, the existence of other relevant margins of substitution can bring in other variables. If stocks of commodities held by households affect shopping-time requirements, for example, the inflation rate will appear separately in the counterpart of (12)—see Feige and Parkin (1971).

Finally, it must be recognized that the simplicity of the portfolio-balance relation (12) would be lost if the intertemporal utility function (1) were not time-separable. If, for example, the function $u(c_t,\ell_t)$ in (1) were replaced with $u(c_t,\ell_t,\ell_{t-1})$ or $u(c_t,c_{t-1},\ell_t)$, as has been suggested in the business cycle literature, then the dynamic aspect of the household's choices would be more complex and a relation like (12)—i.e., one that includes only contemporaneous variables—could not be derived.

Historical Development

The approach to money demand analysis outlined above, which features intertemporal optimization choices by individual economic agents whose transactions are facilitated by their holdings of money, has evolved gradually over time. In this section we briefly review that evolution.

While the earlier literature on the quantity theory of money contained many important insights, its emphasis was on the comparison of market equilibria rather than individual choice, i.e., on "market experiments" rather than "individual experiments," in the language of Patinkin (1956). Consequently, there was little explicit consideration of money demand behavior in pre-1900 writings in the quantity theory tradition. Indeed, there was little emphasis on money demand *per se* even in the classic contributions of Mill (1848), Wicksell (1906), and Fisher (1911), despite the clear recognition by those analysts that some particular quantity of real money holdings would be desired by the inhabitants of an economy under any specified set of circumstances. Notable exceptions, discussed by Patinkin (1956, pp. 386-417), were provided by Walras and Schlesinger.

In the English language literature, the notion of money demand came forth more strongly in the "cash balance" approach of Cambridge economists, an approach that featured analysis organized around the concepts of money demand and supply. This organizing principle was present in the early (c. 1871) but unpublished writings of Marshall (see Whitaker, 1975, pp. 165-68) and was laid out with great explicitness by Pigou (1917). The Cambridge approach presumed that the quantity of money demanded would depend primarily on the volume of transactions to be undertaken, but emphasized volition on the part of moneyholders and recognized (sporadically) that the ratio of real balances to transaction volume would be affected by foregone "investment income", i.e., interest earnings. In this regard Cannan (1921), a non-Cambridge economist who was influenced by Marshall, noted that the quantity of money demanded should be negatively related to anticipated inflation—an insight previously expressed by Marshall in his testimony of 1886 for the Royal Commission on the Depression of Trade and Industry (Marshall, 1926). In addition, Cannan developed very clearly the point that the relevant concept is the demand for a *stock* of money.

Although the aforementioned theorists developed several important constituents of a satisfactory money demand theory, none of them unambiguously cast his explanation in terms of marginal analysis. Thus a significant advance was provided by Lavington

(1921, p. 30), in a chapter entitled "The Demand for Money," who attempted a statement of the marginal conditions that must be satisfied for optimality by an individual who consumes, holds money, and holds interest-bearing securities. But despite the merits of his attempt, Lavington confused—as Patinkin (1956, p. 418) points out—the subjective sacrifice of permanently adding a dollar to cash balances with that of adding it for only one period. Thus it was left for Fisher (1930, p. 216) to provide a related but correct statement. The discussions of both Lavington and Fisher are notable for identifying the interest rate as a key determinant of the marginal opportunity cost of holding money.

In a justly famous article, Hicks (1935) argued persuasively that progress in the theory of money would require the treatment of money demand as a problem of individual choice at the margin. Building upon some insightful but unclear suggestions in Keynes's *Treatise on Money* (1930), Hicks investigated an agent's decision concerning the relative amounts of money and securities to be held at a point in time. He emphasized the need to explain why individuals willingly hold money when its return is exceeded by those available from other assets and—following Lavington and Fisher—concluded that money provides a service yield not offered by other assets. Hicks also noted that the positive transaction cost of investing in securities makes it unprofitable to undertake such investments for very short periods. Besides identifying the key aspects of marginal analysis of money demand, Hicks (1935) pointed out that an individual's total wealth will influence his demand for money. All of these points were developed further in Chapters 13 and 19 of Hicks's *Value and Capital* (1939). The analysis in the latter is, some misleading statements about the nature of interest notwithstanding, substantively very close to that outlined in the previous section of this article. Hicks did not, however, provide formal conditions relating to money demand in his mathematical appendix.

The period between 1935 and 1939 witnessed, of course, the publication of Keynes's *General Theory* (1936). That work emphasized the importance for macroeconomic analysis of the interest-sensitivity of money demand—"liquidity preference," in Keynes's terminology—and was in that respect, as in many others, enormously influential. Its treatment of money demand *per se* was not highly original, however, in terms of fundamentals. (This statement ignores some peculiarities resulting from a presumably inadvertent attribution of money illusion; on this topic, again see Patinkin, 1956, pp. 173-74).

The importance of several items mentioned above—payments practices, foregone interest, and transaction costs—was explicitly depicted in the formal optimization models developed several years later by Baumol (1952) and Tobin (1956). These models, which were suggested by mathematical inventory theory, assume the presence of two assets (money and an interest-bearing security), a fixed cost of making transfers between money and the security, and a lack of synchronization between (exogenously given) receipt and expenditure streams. In addition, they assume that all payments are made with money. Economic units are depicted as choosing the optimal frequency for money-security transfers so as to maximize interest earnings net of transaction costs.

In Baumol's treatment, which ignores integer constraints on the number of transactions per period, the income and interest-rate elasticities of real money demand are found to be ½ and −½, respectively. Thus the model implies "economies of scale" in making transactions. Tobin's (1956) analysis takes account of integer constraints, by contrast, and thus implies that individuals respond in a discontinuous fashion to alternative values of the interest rate. In his model it appears entirely possible for individual economic units to choose corner solutions in which none of the interest-bearing security is held. A number of extensions of the Baumol-Tobin approach have been made by various authors; for an insightful survey the reader is referred to Barro and Fischer (1976).

Miller and Orr (1966) pioneered the inventory approach to money demand theory in a stochastic context. Specifically, in their analysis a firm's net cash inflow is generated as a random walk, and the firm chooses a policy to minimize the sum of transaction and foregone interest costs. The optimal decision rule is of the (S,s) type: when money balances reach zero or a ceiling, S, the firm makes transactions to return the balance to the level s. In this setting there are again predicted economies of scale, while the interest-rate elasticity is −⅓. For extensions the reader is again referred to Barro and Fischer (1976).

The various inventory models of money demand possess the desirable feature of providing an explicit depiction of the *source* of money's service yield to an individual holder. It has been noted, e.g., by Friedman and Schwartz (1970), that the type of transaction demand described by these models is unable to account for more than a fraction of the transaction balances held in actual economies. Furthermore, their treatment of expenditure and receipt streams as exogenous is unfortunate and they do not generalize easily to fully dynamic settings. These points imply, however, only that the inventory models should not be interpreted too literally. In terms of fundamentals they are closely related to the basic model outlined in the previous section.

A quite different approach was put forth by Tobin (1958), in a paper that views the demand for money as arising from a portfolio allocation decision made under conditions of uncertainty. In the more influential of the paper's models, the individual wealth-holder must allocate his portfolio between a riskless asset, identified as money, and an asset with an uncertain return whose expected value exceeds that of money. Tobin shows how the optimal portfolio mix depends, under the assumption of expected utility maximization, on the individual's degree of risk aversion, his wealth, and the mean-variance characteristics of the risky asset's return distribution. The analysis implies a negative interest sensitivity of money demand, thereby satisfying Tobin's desire to provide an additional rationalization of Keynes's (1936) liquidity preference hypothesis. The approach has, however, two shortcomings. First, in actuality money does not have a yield that is riskless in real terms, which is the relevant concept for rational individuals. Second, and more seriously, in many actual economies there exist assets "that have precisely the same risk characteristics as money and yield higher returns" (Barro and Fischer, 1976, p. 139). Under such conditions, the model implies that no money will be held.

Another influential item from this period was provided by Friedman's well-known "restatement" of the quantity theory (1956). In that paper, as in Tobin's, the principle role of money is as a form of wealth. Friedman's analysis emphasized margins of substitution between money and assets other than bonds—e.g., durable consumption goods and equities. The main contribution of the paper was to help rekindle interest in monetary analysis from a macroeconomic perspective, however, rather than to advance the formal theory of money demand.

A model that may be viewed as a formalization of Hicks's (1935, 1939) approach was outlined by Sidrauski (1967). The main purpose of Sidrauski's paper was to study the interaction of inflation and capital accumulation in a dynamic context, but his analysis gives rise to optimality conditions much like those of equations (4)-(8) of the present article and thus implies money demand functions like (9) and (12). The main difference between Sidrauski's model and ours is merely due to our use of the "shopping-time" specification, which was suggested by Saving (1971). That feature makes real balances an argument of each individual's utility function only indirectly, rather than directly, and indicates the type of phenomenon that advocates of the direct approach presumably have in mind. Thus Sidrauski's implied money demand model is the basis for the one presented above, while a stochastic version of the latter, being fundamentally similar to inventory or direct utility-yield specifications, is broadly representative of current mainstream views.

Ongoing Controversies

Having outlined the current mainstream approach to money demand analysis and its evolution, we now turn to matters that continue to be controversial. The first of these concerns the role of uncertainty. In that regard, one point has already been developed, i.e., that rate-of-return uncertainty on other assets cannot be used to explain why individuals hold money in economies—such as that of the United States—in which there exist very short-term assets that yield positive interest and are essentially riskless in nominal terms. But this does not imply that uncertainty is unimportant for money demand in a more general sense, for there are various ways in which it can affect the analysis. In the basic model outlined above, uncertainty appears explicitly only by way of the assumption that households view asset returns as random. In that case, if money demand and consumption decisions for a period are made simultaneously then the portfolio-balance relation (12) will be—as shown above—invariant to changes in the return distributions. But the same is not true for the proper demand function (9). And the arguments c_t and R_t of (12) will themselves be affected by the extent of uncertainty, for it will affect households' saving—as well as portfolio—decisions. The former, of course, impact not only on c_t but also on the economy's capital stock and thus, via the equilibrium real return on capital, on R_t. In addition, because R_t is set in nominal terms, its level will include a risk premium for inflation uncertainty (Fama and Farber, 1979).

Furthermore, the invariance of (12) to uncertainty breaks down if money must be held at the start of a period to yield its transaction services during that period. In this case, the money demand decision temporally precedes the related consumption decision so the marginal service yield of money is random, with moments that depend on the covariance matrix of forecast errors for consumption and the price level. Thus the extent of uncertainty, as reflected in this covariance matrix, influences the quantity of real balances demanded in relation to R_t and plans for c_{t+1}.

There is, moreover, another type of uncertainty that is even more fundamental than rate-of-return randomness. In particular, the existence of uncertainty regarding exchange opportunities available at an extremely fine level of temporal and spatial disaggregation—uncertainties regarding the "double coincidence of wants" in meetings with potential exchange partners—provides the basic *raison d'être* for a medium of exchange. In addition, the ready verifiability of money enhances the efficiency of the exchange process by permitting individuals to economize on the production of information when there is uncertainty about the reputation of potential trading partners. Thus uncertainty is crucial in explaining why it is that money holdings help to facilitate transactions—to save "shopping time" in our formalization. In this way randomness is critically involved, even when it does not appear explicitly in the analysis. (Alternative treatments of uncertainty in the exchange process have been provided by Patinkin, 1956; Brunner and Meltzer, 1971; and King and Plosser, 1986).

An important concern of macroeconomists in recent years has been to specify models in terms of genuinely structural relationships, i.e., ones that are invariant to policy changes. This desire has led to increased emphasis on explicit analysis of individuals' dynamic optimization problems, with these expressed in terms of basic taste and technology parameters. Analysis of that type is especially problematical in the area of money demand, however, because of the difficulty of specifying rigorously the precise way—at a "deeper" level than (2), for example—in which money facilitates the exchange process. One prominent attempt to surmount this difficulty has featured the application of a class of overlapping-generations models—i.e., dynamic equilibrium models that emphasize the differing perspectives on saving of young and old individuals—to a variety of problems in monetary economics. The particular class of overlapping-generations models in question is one in which, while there is an analytical entity termed "fiat money," the specification deliberately excludes any shopping time or related feature that would represent the transaction-facilitating aspect of money. Thus this approach, promoted most prominently in the work of Wallace (1980), tries to surmount the difficulty of modeling the medium-of-exchange function of money by simply ignoring it, emphasizing instead the asset's function as a store of value.

Models developed under this overlapping-generations approach typically possess highly distinctive implications, of which three particularly striking examples will be mentioned. First, if the monetary authority causes the stock of money to grow at a rate in excess of the economy's rate of output growth, no money will be demanded and the price level will be infinite. Second, steady-state equilibria in which money is valued will be Pareto optimal if and only if the growth rate of the money stock is nonpositive. Third, open-market changes in the money stock will have no effect on the price level. It has been shown, however, that these implications result from the models' neglect of the medium-of-exchange function of money. Specifically, McCallum (1983) demonstrates that all three implications vanish if this neglect is remedied by recognition of shopping-time considerations as above. That conclusion suggests that the class of overlapping-generations models under discussion provides a seriously misleading framework for the analysis of monetary issues. This weakness, it should be added, results not from the generational structure of these models, but from the overly restrictive application of the principle that assets are valued solely on the basis of the returns that they yield; in particular, the models fail to reflect the nonpecuniary return provided by holdings of the medium of exchange. On these points also see Tobin (1980).

Recognizing this problem but desiring to avoid specifications like (2), some researchers have been attracted to the use of models incorporating a *cash-in-advance* constraint (e.g., Lucas, 1980; Svensson, 1985). In these models, it is assumed that an individual's purchases in any period cannot exceed the quantity of money brought into that period. Clearly, imposition of this type of constraint gives a medium-of-exchange role to the model's monetary asset and thereby avoids the problems of the Wallace-style overlapping-generations models. Whether it does so in a satisfactory manner is, however, more doubtful. In particular, the cash-in-advance formulation implies that start-of-period money holdings place a *strict* upper limit on purchases during the period. This is a considerably more stringent notion than that implied by (2), which is that such purchases are possible but increasingly expensive in terms of time and/or other resources. Thus the demand for money will tend to be less sensitive to interest-rate changes with the cash-in-advance specification than with one that ties consumption and money holdings together less rigidly. More generally, the cash-in-advance specification can be viewed as an extreme special case of the shopping-time function described in (2), in much the same way as a fixed-coefficient production function is a special case of a more general neoclassical technology. For some issues, use of the special case specification will be convenient and not misleading, but care must be exerted to avoid inappropriate applications. It seems entirely unwarranted, moreover, to opt for the cash-in-advance specification in the hope that it will be more nearly structural and less open to the Lucas critique (1976) than relations such as (2). Both of these specificational devices—and probably any that will be analytically tractable in a macroeconomic context—should be viewed not as literal depictions of technological or social constraints, but as potentially useful metaphors that permit the analyst to recognize in a rough way the benefits of monetary exchange. (On the general topic, see Fischer, 1974).

A final controversy that deserves brief mention pertains to an aspect of money demand theory that has not been formally discussed above, but which is of considerable importance in practical applications. Typically, econometric estimates of money demand functions combine "long-run" specifications such as (12) with a *partial adjustment* process that relates actual money holdings to the implied "long-run" values. Operationally, this approach often results in a regression equation that includes a lagged value of the money stock as an explanatory variable. (Distributed-lag formulations are analytically similar.) Adoption of the partial adjustment mechanism is justified by appeal to portfolio-adjustment costs. Specifically, some authors argue that money balances serve as a "buffer stock" that temporarily accommodates unexpected variations in income, while others attribute sluggish adjustments to search costs.

From the theoretical perspective, however, the foregoing interpretation for the role of lagged money balances (or distributed lags) appears weak. It is difficult to believe that tangible adjustment costs are significant, and in their absence there is no role for lagged money balances, in formulations such as (12), when appropriate transaction and opportunity-cost variables are included. Furthermore, typical estimates suggest adjustment speeds that are too slow to be plausible.

These points have been stressed by Goodfriend (1985), who offers an alternative explanation for the relevant empirical findings. A model in which there is full contemporaneous adjustment of money holdings to transaction and opportunity-cost variables is shown to imply a positive coefficient on lagged money when these determinants are positively autocorrelated and contaminated with measurement error. Under this interpretation, the lagged variable is devoid of behavioral significance; it enters the regression only because it helps to explain the dependent variable in a mongrel equation that mixes together relations pertaining to money demand and other aspects of behavior. (This particular conclusion is shared with the buffer stock approach described by Laidler (1984), which interprets the conventional regression as a confounding of money demand with sluggish price-adjustment behavior.) Furthermore, the measurement error hypothesis can account for positive autocorrelation of residuals in the conventional regression and, if measurement errors are serially correlated, the *magnitude* of the lagged money coefficient typically found in practice.

References

Barro, R. J., and S. Fischer. "Recent Developments in Monetary Theory." *Journal of Monetary Economics* 2 (April 1976): 133-67.

Baumol, W. J. "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Quarterly Journal of Economics* 66 (November 1952): 545-56.

Brunner, K., and A. H. Meltzer. "The Uses of Money: Money in the Theory of an Exchange Economy." *American Economic Review* 61 (December 1971): 784-805.

Cannan, E. "The Application of the Theoretical Apparatus of Supply and Demand to Units of Currency." *Economic Journal* 31 (December 1921): 453-61.

Dutton, D. S., and W. P. Gramm. "Transactions Costs, the Wage Rate, and the Demand for Money." *American Economic Review* 63 (September 1973): 652-65.

Fama, E. F. "Banking in the Theory of Finance." *Journal of Monetary Economics* 6 (January 1980): 39-57.

Fama, E. F., and A. Farber. "Money, Bonds, and Foreign Exchange." *American Economic Review* 69 (September 1979): 639-49.

Feige, E., and M. Parkin. "The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital." *American Economic Review* 61 (June 1971): 335-49.

Fischer, S. "Money and the Production Function." *Economic Inquiry* 12 (December 1974): 517-33.

Fisher, I. *The Purchasing Power of Money*. New York: Macmillan, 1911.

————. *The Theory of Interest*. New York: Macmillan, 1930.

Friedman, M. "The Quantity Theory of Money: A Restatement." In *Studies in the Quantity Theory of Money*, edited by M. Friedman. Chicago: University of Chicago Press, 1956.

Friedman, M., and A. J. Schwartz. *Monetary Statistics of the United States*. New York: National Bureau of Economic Research, 1970.

Goodfriend, M. "Reinterpreting Money Demand Regressions." *Carnegie-Rochester Conference Series on Public Policy* 22 (Spring 1985): 207-42.

Hicks, J. R. "A Suggestion for Simplifying the Theory of Money." *Economica* 2 (February 1935): 1-19.

————. *Value and Capital.* London: Oxford University Press, 1939.

Karni, E. "The Transactions Demand for Cash: Incorporation of the Value of Time into the Inventory Approach." *Journal of Political Economy* 81 (September/October 1973): 1216-25.

Keynes, J. M. *A Treatise on Money*, 2 vols. London: Macmillan, 1930.

————. *The General Theory of Employment, Interest, and Money.* London: Macmillan, 1936.

King, R. G., and C. I. Plosser. "Money as the Mechanism of Exchange." *Journal of Monetary Economics* 17 (January 1986): 93-115.

Laidler, D. "The 'Buffer Stock' Notion in Monetary Economics." *Economic Journal* 94 (Supplement 1984): 17-34.

Lavington, F. *The English Capital Market.* London: Methuen, 1921.

Lucas, R. E., Jr. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1 (Autumn 1976): 19-46.

————. "Equilibrium in a Pure Currency Economy." In *Models of Monetary Economies*, edited by J.H. Kareken and N. Wallace. Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Marshall, A. *Official Papers by Alfred Marshall*, edited by J.M. Keynes. London: Macmillan, 1926.

McCallum, B. T. "The Role of Overlapping-Generations Models in Monetary Economics." *Carnegie-Rochester Conference Series on Public Policy* 18 (Spring 1983): 9-44.

————. "Bank Deregulation, Accounting Systems of Exchange, and the Unit of Account: A Critical Review." *Carnegie-Rochester Conference Series on Public Policy* 23 (Autumn 1985): 13-45.

Meltzer, A. H. "The Demand for Money: The Evidence From the Time Series." *Journal of Political Economy* 71 (June 1963): 219-46.

Mill, J. S. *Principles of Political Economy*, 2 vols. London: John W. Parker, 1848.

Miller, M. H., and D. Orr. "A Model of the Demand for Money by Firms." *Quarterly Journal of Economics* 80 (August 1966): 413-35.

Niehans, J. *The Theory of Money.* Baltimore: Johns Hopkins University Press, 1978.

Patinkin, D. *Money, Interest, and Prices.* New York: Harper and Row, 1956.

Pigou, A. C. "The Value of Money." *Quarterly Journal of Economics* 32 (November 1917): 38-65.

Saving, T. R. "Transactions Costs and the Demand for Money." *American Economic Review* 61 (June 1971): 407-20.

Sidrauski, M. "Rational Choice and Patterns of Growth in a Monetary Economy." *American Economic Review* 57 (May 1967): 534-44.

Svensson, L. E. O. "Money and Asset Prices in a Cash-in-Advance Economy." *Journal of Political Economy* 93 (October 1985): 919-44.

Tobin, J. "The Interest-Elasticity of Transaction Demand for Cash." *Review of Economics and Statistics* 38 (August 1956): 241-47.

————. "Liquidity Preference as Behavior Toward Risk." *Review of Economic Studies* 25 (February 1958): 65-86.

————. "Discussion." In *Models of Monetary Economics*, edited by J.H. Kareken and N. Wallace. Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Wallace, N. "The Overlapping Generations Model of Fiat Money." In *Models of Monetary Economies*, edited by J.H. Kareken and N. Wallace. Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Whitaker, J. K., ed. *The Early Economic Writings of Alfred Marshall, 1867-1890*, 2 vols. New York: Free Press, 1975.

Wicksell, K. *Lectures on Political Economy*, Vol. 2 (1906). Translated by E. Classen. London: Routledge & Kegan Paul, 1935.

AGRICULTURAL SUMMARY AND OUTLOOK

Raymond E. Owens

After deteriorating rapidly in the early 1980s, the financial condition of agriculture leveled off by the middle of the decade. Since then, most measures of the sector's financial condition have strengthened. In 1987, total farm income was higher and production expenses were lower than in the previous year. The improved financial condition of farmers was also apparent in the expanded equity position of their balance sheets, which resulted from a combination of increased asset values and reduced debt burdens.

The turnaround in the agricultural sector was discussed recently at the annual Outlook Conference sponsored by the United States Department of Agriculture (USDA) in Washington, D.C. There, analysts from the public and private sectors assessed the current condition of agriculture and gave their views of its future. Most analysts at the conference believed that the recent financial turnaround was likely to be the start of a longer-term economic improvement for U.S. agriculture. What follows summarizes their consensus view.

World Supply and Demand

Conference analysts cited tightened world supply and demand conditions for agricultural products as a primary reason for an improved farm income picture. Worldwide grain production fell short of consumption of some commodities in 1987, leading to reductions in carryover stocks. Worldwide production was lower due to fewer planted acres and lower yields for some grains. Lower planted acreage resulted from larger acreage set-asides in the United States and the withdrawal of some crop land in other nations. Crop yields held up well in the United States in 1987, but dry weather and storm damage curtailed yields in some other parts of the world.

While production was lower, world coarse grain usage rose in 1987 as the expansion of meat production raised feed usage. Wheat consumption was flat and rice usage was down slightly. Oilseed production matched usage.

World red meat and poultry production rose 1 percent in 1987. Consumption increased by a similar amount so that carryover stocks were unchanged.

The U.S. Agricultural Trade Position

One of the most encouraging developments of 1987 was the rebound in agricultural exports. The total value of agricultural exports rose to $27.5 billion in 1987, an increase of over $1 billion from 1986. The volume of agricultural exports rose by 20 million metric tons to 129.2 million metric tons. Both wheat and corn exports rose sharply, although lower prices dampened increases in export values. Increased crop sales were primarily responsible for the higher export value, but additional meat and horticultural exports accounted for about 40 percent of the value increase. The increase in value for agricultural exports was the first since 1980, when a peak value of $43.8 billion was reached. The value of agricultural imports fell $875 million in 1987 to $20.0 billion. The decrease was the first since 1981/82.

A higher value of exports combined with a lower value of imports to widen the 1987 agricultural trade surplus to $7.2 billion, almost $2.0 billion above the 1986 figure. While improved from 1986, the trade surplus nevertheless remains low by historical standards. As recently as 1981, the agricultural trade surplus stood at $26.5 billion.

Farm Income

Analysts from USDA reported that farm cash income totaled about $57 billion and net farm income reached about $45 billion in 1987. As shown in Table I, the income measures were boosted by higher cash receipts from the sale of livestock, record levels of government payments, and lower production expenses (lines 1, 2, and 7). Farm production expenses fell in 1987 primarily due to decreased usage of crop inputs and lower costs for feed used by livestock producers. After adjusting for inflation, net farm income reached a record level in 1987.

Crop sales were buoyed by larger than expected domestic usage and exports. A total of 129.2 million tons of grains moved out of U.S. ports in 1986/87 compared to just 109.5 million tons in 1985/86. The higher tonnage figures translated into a higher world market share for U.S. agricultural products. Lower prices, a weaker dollar, and government export

Table I

FARM INCOME STATISTICS

(Billions of dollars)

Item	1981	1982	1983	1984	1985	1986	1987F	1988F
1. Farm receipts	144.1	147.1	141.1	146.7	149.2	140.2	137	138
Crops (incl. net CCC loans)	72.5	72.3	67.1	69.4	74.4	63.6	58	62
Livestock	69.2	70.3	69.4	72.9	69.8	71.6	74	71
Farm related[1]	2.5	4.5	4.5	4.4	5.0	5.1	5	5
2. Direct Government payments	1.9	3.5	9.3	8.4	7.7	11.8	17	15
Cash payments	1.9	3.5	4.1	4.0	7.6	8.1	9	7
Value of PIK commodities	0.0	0.0	5.2	4.5	0.1	3.7	9	8
3. Total gross farm income (4+5+6)[2]	166.3	163.5	153.1	174.7	166.0	159.5	163	162
4. Gross cash income (1+2)	146.0	150.6	150.4	155.1	156.9	152.0	155	153
5. Nonmoney income[3]	13.8	14.3	13.5	13.4	11.8	10.8	10	8
6. Value of inventory change	6.5	-1.4	-10.9	6.2	-2.7	-3.3	-1	1
7. Cash expenses[4]	113.2	112.5	113.3	116.3	109.6	100.1	97	99
8. Total expenses	139.4	140.0	140.4	142.7	133.7	122.1	118	118
9. Net cash income (4-7)	32.8	38.1	37.1	38.8	47.3	52.0	58	52.5
10. Net farm income (3-8)	26.9	23.5	12.7	32.0	32.3	37.5	45	42.5
Deflated (1982$)	28.6	23.5	12.2	29.7	29.1	32.9	39	36
11. Off-farm income	35.8	36.4	37.0	38.3	42.5	44.7	48	49
Loan changes:[5]								
12. Real estate	9.4	4.0	2.5	-0.8	-5.6	-7.3	-6	-6
13. Non-real estate	6.2	3.4	1.0	-0.8	-9.2	-10.5	-9	-5
14. Rental income plus monetary change	6.4	6.3	5.3	8.9	8.8	7.8	7	8
15. Capital expenditures[5]	16.8	13.3	12.7	12.5	9.6	8.6	7	8
16. Net cash flow (9+12+13+14-15)	37.9	38.4	33.6	33.6	31.6	33.4	44	42.5

[1] Income from machine hire, custom work, sales of forest products, and other misc. cash sources.

[2] Numbers in parentheses indicate the combination of items required to calculate a given item.

[3] Value of home consumption of self-produced food and imputed gross rental value of farm dwellings.

[4] Excludes capital consumption, prerequisites to hired labor, and farm household expenses.

[5] Excludes farm households.

F = midpoint of forecast range.

Note: Totals may not add due to rounding.

Source: U. S. Department of Agriculture, Economic Research Service.

subsidies were, in part, responsible for the increase in export usage. A second factor spurring crop sales was increased domestic usage for livestock. Domestic grain sales were boosted by low feed prices and increased livestock production in 1987.

The net effect of generally lower grain prices dominating higher grain usage was a reduction of crop cash receipts in 1987. Cash receipts totaled just $58 billion, down from the 1986 total of $63.6 billion.

Livestock cash receipts were up moderately in 1987, as higher production and strong prices coincided. Strong demand for meat led to expanded production and generally higher prices for producers. Livestock cash receipts rose to a record $74 billion in 1987, up from $71.6 billion in 1986. Direct government payments totaled $17 billion in 1987 compared to $11.8 billion in 1986. Both figures are large by historical standards. As recently as 1981, direct government payments were only $1.9 billion.

Total farm production expenses fell in 1987. Cash expenses were pressured downward by reduced crop plantings by farmers attempting to qualify for government price support programs. Cash expenses totaled $97 billion in 1987, down $3 billion from 1986.

Balance Sheet

Farmers' total equity, as shown in Table II, rose in 1987 for the first time in the 1980s. Stable to slightly higher farm real estate values (line 1) and lower farm debt burdens (line 10) contributed to the increase.

The value of farm real estate peaked at $785 billion in 1981 but then declined steadily to $510 billion in 1986. According to the Department of Agriculture estimates, this decline was reversed in 1987 when the total value of farm real estate rose to $530 billion. Improved farm income prospects and lower interest

Table II

BALANCE SHEET OF THE U. S. FARMING SECTOR

Item	1981	1982	1983	1984	1985	1986P	1987F
				$ billion			
Assets							
1. Real estate[1]	784.7	748.8	739.6	639.6	558.9	510.1	530
2. Non-real estate	212.0	212.2	205.4	208.9	191.2	181.5	179.5
3. Livestock and poultry	53.5	53.0	49.7	49.6	46.3	47.6	48.5
4. Machinery and motor vehicles	101.4	102.0	100.8	96.9	87.7	80.4	76
5. Crops stored[1]	29.1	27.7	23.7	29.6	23.1	18.4	19
6. Financial assets	28.0	29.5	31.3	32.8	34.2	35.0	36
7. Total farm assets	996.7	961.0	945.0	848.5	750.1	691.6	709.5
Liabilities							
8. Real estate[2]	98.7	102.5	104.8	103.7	97.7	88.1	83
9. Non-real estate[3]	83.6	87.0	87.9	87.1	77.5	66.8	58
10. Total farm liabilities	182.3	189.5	192.7	190.8	175.2	155.0	141
11. Total farm equity	814.4	771.5	752.3	657.7	574.9	536.6	568.5
				Percent			
Selected ratios							
12. Debt-to-assets	18.3	19.7	20.4	22.5	23.4	22.4	20
13. Debt-to-equity	22.4	24.6	25.6	29.0	30.5	28.9	25
14. Debt-to-net cash income	556	497	519	492	370	298	245

[1] Non-CCC crops held on farms plus value above loan rates for crops held under CCC.

[2] *Excludes debt on operator dwellings, but includes CCC storage and dryng facilities loans.*

[3] Excludes debt for nonfarm purposes.

P = preliminary

F = midpoint of forecast range.

Source: U. S. Department of Agriculture, Economic Research Service.

rates may have assisted the rebound. While the 1987 increase in farmland value is not large in percentage terms, it accounts for the rise in total farm asset values from $692 billion in 1986 to around $710 billion in 1987.

Farm liabilities were lower in both the real estate and non-real estate categories in 1987. Lower debt loads resulted from loan paydowns by farmers with available cash and further farm debt writedowns by financial institutions. Real estate liabilities fell an estimated $3 to $7 billion to about $83 billion and non-real estate liabilities declined $8 to $10 billion to about $58 billion.

Total farm equity reached an estimated $568 billion in 1987, up from $537 billion the year before. While far short of the $829 billion equity peak reached in 1980, the 1987 figure is encouraging: stabilization in the farm balance sheet is one of the most important developments of 1987.

OUTLOOK FOR THE FARM SECTOR IN 1988

Trade

The outlook for agriculture in 1988 bears some similarities to the summary for 1987. A primary similarity is that agricultural trade should continue to improve. Agricultural exports are projected to reach $31 billion in 1988, up sharply from last year's $27.9 billion. The higher export value is anticipated to result from a combination of increased volume and higher prices. Export volume is expected to reach 137.0 tons in 1988, up from 129.2 tons in 1987. Volume growth should be centered in larger quantities of grains and cotton. Demand for these commodities will be helped by the Export Enhancement Program that subsidizes U.S. agricultural exports to qualifying purchasers. Higher prices are anticipated in grains, cotton, and soybeans as world supply tightens relative to demand.

Imports of agricultural commodities are expected to remain level or decline slightly. In 1987, $20.6 billion of agricultural commodities were imported into the United States. The Department of Agriculture is forecasting this value to fall to $20.5 billion in 1988.

Income

Farm income prospects are bright in 1988. Net farm income is projected to range between $40 and $45 billion, only a little below the record $45 billion of 1987. Total government payments are likely to contribute significantly to net income in 1988, accounting for about $13 billion. Production expenses should remain about even with 1987.

Balance Sheet

If all goes as predicted, the farm balance sheet should be stronger by the end of 1988. Asset values are projected to rise further provided farmland values increase in response to strong farm income prospects, as most analysts believe they will. Further, debt reductions are anticipated as farmers pay down more of their existing debt and financial institutions continue to charge off some of their uncollectible farm loans.

Farm Policy

The provisions of the 1985 Farm Bill remain in effect in 1988. Under these provisions, farmers will continue to operate under programs which provide crop price supports in exchange for reduced plantings. Smaller crop plantings should reduce production and pressure crop prices upward. Since direct government payments are the difference between the target and actual market prices farmers receive, rising grain prices should reduce the differential and thus also direct government payments.

Livestock producers will also feel the effects of these government programs. Such programs, which push grain prices up, affect livestock producers' profit margins. Rising grain prices, for example, translate into higher feed costs, which reduce the profit margins of livestock producers.

The cost of farm policy has risen dramatically in recent years. While USDA expects these costs to fall somewhat in 1988, they still will remain high by historical standards. The increasing concern over the federal budget deficit has focused attention on farm program outlays and many analysts warn that the current expenditures may be subject to reductions in the next several years.

Food Prices

Conference analysts look for only modest food price increases in 1988. Stronger grain prices will place some upward pressure on cereal-based foods, but expected expansion in meat production and only slight increases in the prices of fresh fruit and vegetables should limit the overall increase in food prices.

Nonfarm factors, including processing, packaging, and distribution costs should show only modest increases in 1988. The net effect of farm and nonfarm influences on food prices should be a 2 to 4 percent increase.