Inefficient and Locally Stable Trade Equilibria Under Scale Economies: Comparative Advantage Revisited

William J. Baumol and Ralph E. Gomory*

This paper explores and analyzes the inefficient trade equilibria that can be introduced by scale economies despite the workings of the market mechanism. It thereby seeks to enrich the standard analysis of efficiency of equilibrium, upon which so much of welfare theory has focused. Research in recent decades on trade in commodities whose production entails scale economies—goods that are often the focus of public concern and public policy—have revealed circumstances that are in marked contrast to those in the classical model. It has long been known that trade models with scale economies are characterized by multiple equilibria (see, e.g., Marshall, Matthews 1949–1950, Kemp 1964 and Fother 1979). Indeed, as Krugman (1991, pp. 651–652) remarks,

‘In the emerging literature on increasing returns and externalities, multiple equilibria are not a nuisance but a central part of the story’.

On this point the present authors have shown (1992, 1994) that multiple equilibria do not merely exist but are normally vast in number, increasing exponentially with the number of traded commodities. This paper demonstrates that these many equilibria display some surprising attributes, and describes some previously unrecognized ways in which the market mechanism’s performance can fall short of ideal where scale economies are present.

* Director of the C.V. Starr Center for Applied Economics at New York University, 269 Mercer Street, N.Y., N.Y. 10013, USA, and President, the Alfred P. Sloan Foundation. We are very grateful to the Alfred P. Sloan Foundation and the C.V. Starr Center for support of this work. Some of the material in this article is based on Baumol (1993).
We will show that in a world of scale economies (Section I) market forces cannot always drive the economy to an equilibrium that entails efficient use of the economy’s resources or that follows the dictates of comparative advantage; (Section II) under scale economies many equilibria can be inefficient in the standard sense of the term; (Section III) in contrast, equilibria can be efficient even if they violate the normal comparative advantage conditions.

Usually, an equilibrium is either shown to be efficient or, if it does not meet this standard, it is simply said to be inefficient. In this paper we extend these notions by introducing a measure of the degree to which any particular equilibrium falls short of perfect efficiency. We then apply this new measure to provide suggestive results about the equilibria of some particular trade models. We also formulate a more-flexible definition of comparative advantage, one that is useful in understanding efficiency in this more complex world. We accompany that by another new concept: local efficiency, entailing a comparison of the efficiency of a particular equilibrium point with that of other points in the neighborhood. We think these concepts bring some order to the issue of efficiency in the presence of economies of scale.

1. Efficient and Inefficient Equilibria and Comparative Advantage

Although our general orientation is toward large problems with large numbers of equilibria, we will begin our analysis with some small examples. These simple geometric examples illustrate the results that offer the greatest contrast with the classical case: that equilibria are not always efficient, and that equilibria can be efficient even if they violate comparative advantage.

Two basic definitions. In this paper we will use the term perfectly specialized assignment to denote any division of the task of producing the world’s outputs among producer countries in such a way that no commodity is produced in more than one country. An assignment will be said to satisfy the classical comparative advantage requirements if when Country J produces good 1 and Country J produces good J then the ratio of the average product of Country J labor (assumed in our model to be the only input) in the production of 1 to its average product in J is higher than the corresponding ratio in Country J. We leave for later discussion the appropriate range of input quantity over which the average products should be measured. This issue of range will in fact be critical for the results of Section III.

In Gomory and Baumol (1994) we proved that in a theoretical world of universal scale economies each and every perfectly specialized assignment satisfies the requirements of equilibrium, and that, in addition, every such equilibrium is (locally) stable. Moreover, the number of such specialized assignments is equal to the number of possible country-commodity combinations, and so it grows very rapidly when the number of goods or the number of countries in the model increases.

We will start our discussion with a small model which shows some of the properties of efficiency in the economies of scale case. The discussion is framed in terms of Ricardo’s two countries, England and Portugal, but substitutes for his cloth and wine two more “high-tech” products, computers and Walkman radios, in whose production we may expect scale economies. In such a two-good, two-country model it must be remembered that there are always exactly two perfectly specialized assignments. Portugal can produce all the Walkmans and England all the computers, or the reverse can be true.
Figure 1
One Specialized Equilibrium Dominates the Other

In Figure 1, \( PC^P \) and \( EDE' \) are the production frontiers for Portugal and England. Here, the convexity (downward) of the two frontiers represents the presence of scale economies, because towards the center of the frontier, where the country’s production is unspecialized and the output of each good is relatively small, the output vector is held down (the output point is closer to the origin than it would be in the linear case)\(^3\). The two specialized solutions are \( S' = (E', P) \) and \( S = (P', E) \). In the first of these England produces all the world’s Walkmans and Portugal produces all the computers, while in the second of these the production assignment is reversed. Both of these solutions are obviously locally stable equilibria at prices that clear the markets because if either country tries to produce a small quantity of the other’s product, it will fail because of its high costs. Yet, as shown in Figure 1, where the two frontiers intersect the one specialized point \( S \) clearly dominates the other, \( S' \), so that the latter must be inefficient. This must always be so where the two countries’

\(^3\) In an unpublished paper Avinash Dixit has shown that convexity is sufficient but not necessary for scale economies.
Having just shown that in small-scale models of the sort under discussion an equilibrium that satisfies comparative advantage is necessarily efficient, we can use similar reasoning to show that the converse relationship does not hold—an equilibrium can be efficient even if it violates comparative advantage. In Figure 2, where one of the frontiers, PP', lies entirely above the other, EE', neither equilibrium point dominates the other. We have just seen that in the linearized model derived from the chords of the individual country frontiers solution point $S'$ is efficient because it lies on the world (linearized) production frontier while equilibrium point $S$ is not, because in the linear model $S$ lies below the world frontier. However, because as a result of their (downward) convexity in the original problem the individual country frontiers lie below their chords almost throughout the world frontier, ASS'B does not lie above $S$ (it should be obvious that numerical examples of this sort are easy to construct). So both $S$ and $S'$ really are efficient in the case shown in the graph. Since the production pattern in each of the two equilibria is the reverse of the other, the production patterns of only one of the two equilibria ($S'$) can be in accord with the traditional criterion of comparative advantage, so the other ($S$) has a production pattern that violates comparative advantage (compare the slopes of $EE'$ and $PP'$), yet is efficient. This example shows:

**Theorem 1.3**

Under scale economies an equilibrium can be efficient even if it violates the traditional comparative advantage requirement.

The preceding discussion also suggests that there is a general tradeoff between (a) the magnitude of the loss incurred as a result of violation of comparative advantage in an efficient specialized equilibrium and (b) the strength of the scale economies needed to render that equilibrium efficient. In terms of our diagrams, the greater the difference between the slopes of the chords of the production frontiers of the two countries, the greater the degree of convexity of those production frontiers that is necessary for an equilibrium that violates comparative advantage to be efficient. This is easily demonstrated geometrically in the two-country two-good case and we conjecture that an effect of this sort is true generally. The intuitive reason for this relationship indicates how scale economies weaken the influence of comparative advantage. For a country specializes in the production of a good in which it has a comparative advantage the resulting loss of efficiency can be made up for as a result of its large output of that good if and only if the economies of scale are sufficiently strong.

**Inefficient Equilibria: Can the Free Market Eliminate Them?** But can't market forces be counted upon to destroy the inefficient equilibria that have just been shown to exist in a world of scale economies? This, surely, is what economists have long been taught about perfectly competitive models with universally diminishing returns. Is there not something similar to be expected here? After all, any inefficient equilibrium is necessarily an unrealized oppor-
tunity for mutual gain. Should not arbitrages or other businessespersons be counted upon to recognize such an opportunity and find ways to take advantage of it?

There are two fundamental reasons why this is not true in a world of scale economies. First, there is the fact that market signals are all local—indicating such things as marginal costs and marginal revenues that correspond to the partial derivatives of the profit function. They tell us in what direction to move in order to go upward on the profit hill on which the decision maker is currently located. But where there are millions of equilibria, each of them the peak of efficiency in its own neighborhood (see Section IV, below) going uphill from an initial position fortuitously selected by history may merely lead us toward the highest point on a nearby little hill. It can easily lead us away from the true global maximum, the top of a much-higher hill that may be far away.

Second, practical reality gives us another key reason why the unsolved market will not generally be up to the task. Moving from one perfectly specialized assignment to another requires firms and countries to embark on the production of goods they are not currently turning out, and about whose production they have little knowledge or experience. Moreover, scale economies mean that to have any chance of success in such an endeavor, one must enter on a large scale. This requires the investors to undertake a very great risk, betting on what constitutes a leap into the unknown. Anyone with much business experience will confirm that such opportunities are hardly always recognized, and when recognized they are hardly always pursued. In short, there is no reason to expect the market mechanism to take automatic advantage of the unrealized opportunities provided by inefficient equilibria. Under scale economies such equilibria can and will be locally stable.

Having completed these preliminaries, let us proceed toward the construction of a more general theory of economic efficiency in the presence of scale economies.

II. ORIENTATION AND THE BASIC MODEL

Orientation. Our orientation in this paper is toward large problems with very large numbers of equilibria. Although we discuss both specialized and non-specialized equilibria, our analysis will emphasize specialized equilibria. There is considerable plausibility to this emphasis since it has been known for over three decades that, from a purely theoretical point of view, specialized equilibria are likely to play a dominant role in models with economies of scale because of the instability of most non-specialized equilibria. Furthermore (Gomory 1994)

showed that in large problems specialized equilibria play the dominant role in shaping the region of equilibria and determining the location of all equilibria, specialized and non-specialized alike. In this paper we will see once again the important role of specialized equilibria since we will show in Section IV that it is only the specialized and almost-specialized equilibria that are likely to be efficient in the usual sense.

However, in addition to specialized equilibria, there is one important class of non-specialized equilibria with properties that complement those of the specialized equilibria. These are equilibria with non-specialized (i.e., shared) production in some industries. These industries have production functions, that, after an initial range of rapidly decreasing average cost with increasing scale of production, eventually reach a minimum average cost, and thereafter become linear (the case of "flat-bottomed average cost"). The instability that characterizes many shared equilibria is weak or non-existent for this type of industry. Equilibria that have shared production in some of these industries and specialized production in all the rest we will call extended-specialized equilibria.

The Basic Model. Our basic model consists of two countries (or two facilities) which we call Country 1 and Country 2. We will have production functions $h_i(t)$ that use the single input $t$ to produce good $i$ in Country $j$. If an amount of labor $h_i$ is used to produce the $i$th good in Country $j$, whose labor supply is $l_j$, the $h_i$ must satisfy the labor-availability inequalities

$$\sum_i h_{i1} \leq l_1 \quad \text{and} \quad \sum_i h_{i2} \leq l_2.$$  

(2.1)

Although we will sometimes refer to equilibria in our discussion or our examples, the only property of an economic equilibrium that we will use is that its labor quantities satisfy (2.1).

We define a feasible production plan $P = [h_i]$ for the quantities $Q_i$ as a set of $h_i$ that satisfy (2.1) and has $h_i(t_i) + f_i(t_3) = Q_i$. We will say that the quantities $Q_i$ are efficient if there is no set of $h_i$ that satisfy (2.1) and make more than the given output quantities, $Q_i$. We will assume that we have a feasible production plan $P$, possibly coming from an economic equilibrium. We would like to know if its resulting outputs, the $Q_i$, are an efficient set of goods.

There is one case that is completely straightforward. If $P$ does not have equality in both the inequalities of (2.1), the unused labor can be used to make
additional goods. The total output would then strictly dominate the Q; and they would not be efficient. In the remainder of this paper we will assume that the production pattern P that produces the Q; does use up all the labor available in both countries.

We will say that the Q; are efficient, or equivalently that P is efficient, if there is no feasible production plan that uses strictly less than the total labor of both countries. In principle we could test for efficiency by solving a minimization problem: minimize the labor required subject to the non-linear conditions $f_i(i_i,l_i) + f_j(i_j,l_j) = Q_i$ for all i, and to (2.1), and see if that minimal amount is $L+L$. However, because of the non-linearities, this direct approach seems difficult and we will work instead with a linearized model.

III. THE LINEARIZED MODEL

This section of the paper deals with the classical efficiency concept and shows that in the case of scale economies comparative advantage, defined in a somewhat different way, is sufficient (but not necessary) to guarantee the efficiency of an equilibrium. We will also give a strong necessary condition.

Specialized Production. For a given i, one way to produce Q_i is to have it produced solely in Country j. Then Q_i = f_i(f_i) where f_i is the production function for good i in Country j, and f_i is the amount of labor required in Country j to produce Q_i when Country j is the sole producer. If Q_i > f_i(f_i) so that Q_i cannot be produced in Country j, even by all that country's labor, f_i is undefined. If f_i is defined we will say that (i,j) e D, and if it is not defined we will say that (i,j) e D'.

We will proceed by introducing related problems that replace the f_i with linear production functions. These linearized production functions are related to, but are not the same as, those used in the proof of Theorem 1.1. We will refer to the problems obtained in this way as the linearized problems.

Efficiency in the linearized problems will turn out to be closely related to efficiency in the original non-linear problem. The linearized problems are both theoretically and computationally more tractable, and will be the natural basis for the concept of quantified efficiency (λ-efficiency) that we introduce in Section V.

5. One can, instead, use a dual approach that seeks l_i employing all the labor of both countries and produces more than the Q_i.

6. While (i,j) e D' play a prominent role in very small problems they play a very small role in large problems where no single good uses up a large percentage of the labor force.

In defining a linearized problem the critical question is how to linearize. A natural answer is to take as the linearized production function the linear functions that make the Q; with the same amount of labor as the f_i themselves require. This is what we will do whenever possible. The exceptional cases, when the Q; can be produced even with the entire labor force of the country, we will handle slightly differently.

Figure 3 Production frontiers for Linearization

The Linearized Problem. We first define the average productivities $c_{ij}$ by $c_{ij} = f_i(f_i) / f_i = Q_i / f_i$ for (i,j) e D and $c_{ij} = f_i(f_i) / f_i$ for (i,j) e D'. We then define the new linear production functions $f_i(l)$ by $f_i(l) = c_{ij} l$. These linearized production functions clearly depend on the choice of the Q. The linearized problem associated with the output quantities Q_i is obtained by replacing the production functions $f_i(l)$ with the $f_i(l)$ defined from those Q_i. In Figure 3 we see the original production frontiers for a typical economies of scale problem. Here, WSW'W' is the world frontier, obtained from the individual country frontiers EE' and PP'. In Figure 4 the linearized production functions associated with the quantities (Q_1, Q_2) = (2,6) are shown as the dark lines. The individual country linearized production frontiers are obtained by
putting a straight line through the two intersections of the original country production frontier with the rectangle whose vertices are (0,0) and (Q_1, Q_2) = (2,6), because at those intersection points the original production frontiers represent the use of a quantity of labor in the production of the pertinent good is sufficient to produce the given output, Q_1 of good k. In Figure 5 we see the linear problem associated with the quantities (Q_1, Q_2) = (6,3). Clearly, when we use the linearized production function associated with different quantities Q_i, the Q_i are 1-efficient if they are efficient in the associated linearized problem.

Next we define quantities P_{ij} that are analogous to the P_{ij} of the original problem. These P_{ij} are the amounts of labor required to make the Q_i using the linearized production function. For (i,j) ∈ D, L_{ij}(P_{ij}) = c_i P_{ij} = (Q_i / f_i) P_{ij} = Q_i, so P_{ij} = f_i. For (i,j) ∈ D', L_{ij}(P_{ij}) = c_i + l_i = (l_i L_i / L_i L_i) = l_i L_i < Q_i, so P_{ij} > L_i.

\[ 0 \leq y_{ij} \leq 1 \quad y_{ij} = 1 \text{, } \sum_i y_{ij} P_{ij} \leq L_i \text{ and } \sum_j y_{ij} P_{ij} \leq L_i. \quad (3.1) \]

**Figure 4**
Linearization Based on Assignment S = [3,6]

**Figure 5**
Linearization Based on Assignment S' = [6,3]

Since E and F are used repeatedly throughout the following discussion, it is useful to summarize their meaning: E_{ij} is simply the quantity of labor needed to produce the given quantity of good k if it is done in country j alone. F_{ij} is the same, except where the given output of i is so large that it requires more than country j’s entire labor force to produce it. In that case, F_{ij} is the quantity of labor that would be necessary to produce the given quantity of i, using a linearized production function based on the average productivity obtained when using the entire labor force of the country to produce good i.

In contrast with the original problem, we can now easily characterize the labor inputs that will produce the Q_i using the linearized production functions. Consider any y_{ij} that satisfy:

Each labor quantity y_{ij} P_{ij} will, using the linearized production functions L_{ij}(P_{ij}), produce exactly L(y_{ij} P_{ij}) = c_i y_{ij} P_{ij} = y_{ij} Q_i. Therefore, y_{ij} P_{ij} and y_{ij} P_{ij} together will produce (y_{ij} + y_{ij}) Q_i = Q_i. Also the second line of (3.1) shows...
that these \(y_{i1}^*b_{i1}\) will not exceed the quantities of labor available in the two countries. So the \(y_{i1}\) that satisfy (3.1) give the labor inputs \(y_{i1}^*b_{i1}\) that produce the \(Q_i\) using the \(L_i,k_i\). We will often say that the \(y_{i1}\) produce the \(Q_i\) meaning that the labor quantities \(y_{i1}^*b_{i1}\) do.

We can check for L-efficiency by solving the linear program that minimizes the total labor used in both countries to make the \(Q_i\):

\[
\begin{align*}
\nu = & \text{Min} \sum_{i1} \{y_{i1}^*L_{i1} + y_{i2}^*L_{i2}\}, \\
\text{subject to} & \quad 0 \leq y_{i1} \leq 1, \\
& \quad y_{i1} + y_{i2} = 1, \\
& \quad \sum_{i1} y_{i1}^*L_{i1} \leq L_{i1}, \quad \text{and} \quad \sum_{i1} y_{i1}^*L_{i2} \leq L_{i2}. \\
\end{align*}
\]

(3.2)

From this we can see whether or not the minimizing \(y\) requires the entire labor of both countries. Clearly it is both a necessary and sufficient condition for L-efficiency that the minimizing \(y\) require the entire labor of both countries.

From standard linear programming we know7 that a minimizing \(y\) will have \(y_{i1}\) equal to 0 or 1 for all but at most two 1. Especially for large problems then, this minimizing \(y\) is not very far from being a specialized production pattern. We will use this fact later. Also from standard linear programming we can write down the condition for \(y\) to minimize (3.2). The standard Kuhn-Tucker conditions applied to this specially structured program give: (proof available from the authors).

**Lemma 3.1**

\(y\) is a minimizing solution to (3.2) iff whenever the \(y_{i1}\) and \(y_{i2}\) from (3.2) are both positive we have \(L_{i1}/P_{i1} \leq L_{i2}/P_{i2}\).

This leads directly to:

7. This must be true because if we use \(y_{i1} = y_{i2} = 1\) to eliminate the \(y_{i2}\) we are left with \(y_{i1}\) variables and \(n + 2\) slack variables, call them \(S_i\) and \(S_{i1}\). We cannot have \(y_{i1} = 0\) and \(S_i = 0\) for the same \(i\) since \(S_i = 0\) means \(y_{i1} = 1\). Thus, in a basic solution, in which \(n\) variables \(=0\), we can have at most two cases where \(y_{i1} > 0\) and \(S_i > 0\). Later we will see that there is at most one \(y_{i1}\) that is neither 1 nor 0.
Lemma 3.4: Relation between Efficiency in the Original and the Linearized Plans

If \( P \) is any feasible production plan then \( P^* \) satisfies (3.1) and uses no more labor for any good. Equivalently: if the \( h_i \) are any feasible production pattern then \( y_{i,j} = i(h_i) / Q_i \) satisfies (3.1) and \( y_{i,j} \geq h_j \).

Proof: Consider any production pattern \( h_j \) with \( i(h_j) = i(h_j) = Q_j \) and satisfying (2.1). Now consider \( y_{i,j} = i(h_i) / Q_i \). We will show that \( y_{i,j} \) satisfies the linearized problem (3.1) and uses no more labor, i.e., \( y_{i,j} \geq h_j \).

Clearly, \( 0 \leq y_{i,j} \leq 1 \) and \( y_{i,j} + y_{i,j} = i(h_i) / Q_i + i(h_j) / Q_j = Q_j / Q_j = 1 \). It remains to show \( y_{i,j} P_{i,j} \leq h_j \). This will show both that \( y_{i,j} \) uses no more labor and that it satisfies the inequalities of (3.1).

For \( (i,j) \in D', \quad c_{i,j} = F_{i,j}(L_{i,j}) / L_{i,j} \) Using economies of scale and remembering that in this case \( h_j = c_{i,j} / L_{i,j} \),

\[ L_{i,j} P_{i,j} = y_{i,j} P_{i,j} = y_{i,j} \frac{F_{i,j}(L_{i,j})}{L_{i,j}} P_{i,j} = y_{i,j} F_{i,j} \leq y_{i,j} \frac{F_{i,j}(L_{i,j})}{L_{i,j}} P_{i,j} \]

eliminating the \( \frac{F_{i,j}(L_{i,j})}{L_{i,j}} \) gives \( y_{i,j} P_{i,j} \leq h_j \). (3.3)

For \( (i,j) \in D' \), \( c_{i,j} = F_{i,j}(L_{i,j}) / L_{i,j} \), so

\[ L_{i,j} P_{i,j} = y_{i,j} P_{i,j} = y_{i,j} \frac{F_{i,j}(L_{i,j})}{L_{i,j}} P_{i,j} = y_{i,j} F_{i,j} \leq y_{i,j} \frac{F_{i,j}(L_{i,j})}{L_{i,j}} P_{i,j} \]

eliminating the \( \frac{F_{i,j}(L_{i,j})}{L_{i,j}} \) gives \( y_{i,j} P_{i,j} \leq h_j \). (3.4)

This ends the proof of the lemma.

We have as an immediate consequence:

Theorem 3.5

If a set of \( Q_i \) are \( L \)-efficient then they are efficient.

Proof: We will prove the equivalent statement: if the \( Q_i \) are not efficient in the original problem, they are not efficient in the linearized problem. Let us assume that the \( Q_i \) are not efficient in the original problem so there is a feasible production pattern \( P \) using strictly less than the total labor of the two countries. Then Lemma 3.4 asserts that \( P^* \) is a solution to the linearized problem that uses no more labor, so it too uses less than the total labor of the two countries.

Combining this with Theorem 3.5 gives us a sufficiency condition:

Theorem 3.6

Sufficient conditions for the production plan \( P \) to be efficient are (1) the productivities \( c_{i,j} / c_{i,j} \) satisfy \( c_{i,j} / c_{i,j} \leq c_{i,j} / c_{i,j} \) whenever \( i \) is produced in Country 2 and \( k \) is produced in Country 1, and (2) the linearized plan \( P^* \) uses all the labor of both countries.

The next theorem states that if the feasible production plan \( P \) is efficient, then \( P^* \) uses up the labor of the two countries, and satisfies average comparative advantage, it is efficient.

Theorem 3.7: Sufficiency of Average Competitive Advantage

Sufficient conditions for the specialized or extended-specialized production plan \( P \) to be efficient are (1) it has productivities \( c_{i,j} / c_{i,j} \) that satisfy \( c_{i,j} / c_{i,j} \leq c_{i,j} / c_{i,j} \) whenever \( i \) is produced in Country 2 and \( k \) is produced in Country 1, and (2) it uses all the labor of both countries.

Proof: If \( P \) is extended-specialized (which includes specialized) we have for goods with specialized production, \( i(h_j) = i(h_j) / i(h_j) / h_j \) because \( h_j = h_j \). For goods with shared production we have either \( i(h_j) = i(h_j) / i(h_j) / h_j \) if \( (i,j) \in D \) and \( i(h_j) = i(h_j) / i(h_j) / h_j \) because \( h_j \) is already in the linear range. In either case we obtain \( y_{i,j} P_{i,j} = h_j \) instead of \( y_{i,j} P_{i,j} \leq h_j \) in (3.2) or (3.4). Therefore, if \( P \) uses up all the labor so does \( P^* \) and then Theorem 3.6 applies. Q.E.D.

Necessary Conditions: If we next consider necessary conditions there are certain things that are straightforward. A production plan \( P \) can not be efficient if an interchange of two industries between the two countries produces more of both goods, or equivalently, the same goods with less labor, so clearly not-good
efficiency requires satisfaction of the two-good efficiency condition for every pair of goods as a necessary condition. However, there is also a necessary condition related to the linear problem.

We start with this Lemma:

**Lemma 3.8**

Lemma: If the \( y_{ij} \) are integer (i.e., 0 or 1) and satisfy the labor-availability constraints (3.1), then, \( P^* = \{A_i \} \) where \( A_i = y_{ij}W_{ij}A_i \) is a feasible (specialized) production plan for the original problem.

Proof: Since the \( y_{ij}W_{ij}A_i \) satisfy (3.1) the \( A_i \) clearly satisfy (2.1) - they are no more than the available labor. It only remains to show that \( A_i \) can make the \( Q_i \) i.e., that \( f_1(y_{ij}W_{ij}A_i) + f_2(y_{ij}W_{ij}A_i) = Q_i \) For each \( i \) one of the two \( y_{ij}W_{ij}A_i \) is 1, and the other is 0. Let us suppose \( y_{ij}W_{ij}A_i = 1 \). If \( f_1(1) \in D \), \( f_1 = f_1' \) so \( f_2(y_{ij}W_{ij}A_i) = Q_i \); and \( f_2(0) = 0 \) so the \( A_i \) make the \( Q_i \). Clearly we have the same outcome if \( y_{ij}W_{ij}A_i = 0 \) and \( y_{ij}W_{ij}A_i = 1 \) provided that (2.2) is D. We next consider the possibility \( y_{ij}W_{ij}A_i = 1 \) and \( y_{ij}W_{ij}A_i = 0 \) and \( (i,j) \in D' \). As we remarked earlier when we were defining the linear problem, \( (i,j) \in D' \) implies that \( f_1'(1) \geq f_2(1) \). This in turn implies, for \( y_{ij}W_{ij}A_i = 1 \), that \( y_{ij}W_{ij}A_i > 1 \). This means that the first inequality in (3.1) cannot be satisfied. That, however, contradicts the assumption that the \( y_{ij}W_{ij}A_i \) satisfy (3.1) so this case cannot occur. The same reasoning applies to the remaining case \( y_{ij}W_{ij}A_i = 0 \) and \( y_{ij}W_{ij}A_i = 1 \) and (2.2) \in D'. Q.E.D.

This Lemma enables us to prove:

**Theorem 3.9**

If \( P \) is an efficient production plan, then the corresponding linear plan \( P^* \) uses no more labor in (3.2) than the minimizing integer solution to (3.2).

Proof: If we start with some efficient production plan, \( P \), then by Lemma 3.4 its linearized plan, \( P^* \), satisfies the linearized problem and uses no more labor than \( P \). If there were an integer \( P \) that solved the linearized problem with strictly less labor, it would give us by Lemma 3.8 a new feasible production plan \( P^* \) using no more labor than that, and therefore using less labor than the original \( P \). This contradicts the assumed efficiency of the original \( P \). Q.E.D.

**Inefficient and Locally Stable Trade Equilibria**

If we start with a specialized \( P = \{a_{ij} \} \), it produces a \( P^* \) with \( y_{ij}W_{ij}A_i / Q_i \). However, we can verify that these \( y_{ij}W_{ij}A_i \) are exactly the \( a_{ij} \) of \( P \) itself. For example if \( f_1(a_{ij}) = Q_i, y_{ij}W_{ij}A_i = 1 \) so \( y_{ij}W_{ij}A_i = f_1 = a_{ij} \). This leads immediately to:

**Theorem 3.10: A Necessary Condition for Efficiency**

If a specialized \( P \) is efficient, its \( P^* \) must be integer (perfectly specialized) and must be minimal among integer solutions to (3.2).

In other words \( P^* \) solves the integer programming problem represented by (3.2).

Proof: This is a restatement of Theorem 3.9 for specialized \( P \) using the fact that, for specialized \( P \), \( P^* \) is integer.

This theorem tells us that for a specialized plan, \( P \), to be efficient in the original problem, its linearized counterpart, \( P^* \), must be the minimizing solution of the integer (perfectly specialized) version of the (efficiency) programming problem (3.2).

Special Cases: We will next discuss some illuminating special cases. These cases illustrate the variety of outcomes that can occur when the classical efficiency concept is used in the scale economies model. In particular, they demonstrate that the proportion of efficient specialized equilibria in each case can be very large, and in some cases very small.

**Identical Production Functions:** Assume \( f_1(l) = f_2(l) \) for all \( l \). Any equilibria or any feasible production plan \( P = \{a_{ij} \} \) point provides a set of \( Q_i \), and satisfies our condition (2.1). Let us suppose that \( P \) is specialized and that Country 1 is the producer in the ith industry. Then \( a_{ij} = a_{i1} \) and \( a_{ij} = 0 \). Clearly, (1.1) \in D. If we also have (2.2) \in D, it takes the same amount of labor \( a_{ij} \) to produce \( Q_i \) in both countries because the production functions are identical, so \( a_{ij} = a_{i1} \). Prior to this, we have no economies of scale over \( a_{i1} = a_{i2} = a_{i3} \). If \( a_{i2} = a_{i3} \) or \( a_{i3} = a_{i4} \), we must be true that \( a_{ij} = a_{i1} \). Then we have a economies of scale over the producing country. Country 1, has an \( a_{i1} + a_{i2} + a_{i3} \) and the non-producer, in this case Country 2, an \( a_{i2} + a_{i3} + a_{i4} \). This argument can, of course, be repeated for all \( i \). We conclude that the competitive advantage conditions of Theorem 3.7 are necessarily satisfied and that we have efficiency. So if \( f_1(l) = f_2(l) \) we have efficiency for all specialized production plans or specialized equilibria.
To explain the next two examples we must allude briefly to some of the results of Gomory (1991, 1992) and Gomory and Baumol (1992). In these papers we showed not only that the n-good model has 2^n-2 specialized equilibria, but also that if we calculate for each equilibrium its relative national income \( z_2 = Y_1 / (Y_1 + Y_2) \) and its (Cobb-Douglas) utility \( U_i \) for Country 1, and then plot the points \((z_2, U_1)\) in the \( z_2-U_1 \) plane, the resulting 2^n-2 points lie in a well-defined region of the plane. This region has a characteristic shape and well defined upper and lower boundaries that can be computed. We also showed that the region between the upper and lower boundaries tends to fill up solidly with these equilibria as n increases.

A similar statement can be made about a \( z_2-U_2 \) plane, if we choose to plot the equilibria with \( z_2 = Y_2 / (Y_1 + Y_2) \) as the horizontal axis and \( U_2 \), the Cobb-Douglas utility for Country 2 as the vertical axis. In Figure 6, and later in Figures 7, 8, and 9, we combine the \((z_2, U_1)\) and \((z_2, U_2)\) plots. In Figure 6 the dark dots are the points \((z_2, U_1)\) and the lighter dots the points \((z_2, U_2)\). The horizontal axis is \( z_2 \) if read from left to right, and \( z_2 = 1 - z_2 \) if read from right to left. Figure 6 represents the equilibria for an 11-industry model. These equilibria give us a large set of specialized production patterns, whose efficiency has considerable natural interest.

With this background we can now discuss the next examples.

Productivity ratios that do not depend on the \( Q \). Efficiency is also easier to understand when the \( c_{il} \) depend on the size of the \( Q \) but the \( c_{il}/c_{ij} \) ratios do not. This occurs when, in each industry, one country has a consistent degree of advantage over the other over all \( Q \), or at least over a wide range of \( Q \). An example of this is the case \( c_{il} = c_{ij} \). In this case we find directly that the \( c_{il}/c_{ij} \) ratios for each i are given by the expression \( (c_{il}/c_{ij})^{1/n} \), which does not involve \( Q \). These ratios are independent of the actual \( Q \) values and therefore are the same for all the different equilibria or specialized production plans. The ratios can be used to rank the industries in order of decreasing comparative advantage, \( c_{il}/c_{ij} \). If we assign production of the first n-1 industries to this ordering in Country 1 and the rest to Country 2, we obtain an equilibrium that certainly satisfies the comparative advantage condition of Theorem 3.7. We can create n-1 efficient equilibria in this way. These equilibria are shown in Figure 7 for a 27-good model.

This example illustrates an important point. There are efficient equilibria that give very low utility for both countries, those that are near \( z_2 = 1 \) or \( z_2 = 0 \). At these points one country makes almost everything and the other makes only one or two products. However, these products are made in such large quantities that the point is efficient, since there is no way to make more of those one or two while maintaining the quantities of the others. The disturbing implication

is that efficiency and welfare are almost completely divorced, except where \( z_1 \) is held fixed, i.e., at a given level of relative national income.

Ratios that depend on the \( Q \). One might think at this point that there is always some string of efficient equilibria to be obtained by finding the equilibria whose ordering corresponds to the comparative advantage ordering of the productivity ratios, \( c_{il}/c_{ij} \). However, what has simplified our work to this point is that the \( c_{il}/c_{ij} \) ratios with which we have dealt have been independent of \( Q \), while in general the \( c_{il}/c_{ij} \) depend also on the size of the \( Q \). This dependence can in fact exclude the possibility of efficient equilibrium points over wide ranges of relative national incomes. We can provide an example in which there are no efficient specialized equilibria in the equilibrium region of the \( z_2-U_1 \) plane for \( z_2 > 0.5 \). While this example is somewhat artificial it does show the sensitivity to the details of the classical efficiency concept in this setting. It therefore seems worthwhile to consider other concepts that may generalize the classical concept and improve its adaptation to a world of scale economies.

Figure 6
Regions of Equilibria from an 11-Industry Model
IV. LOCAL EFFICIENCY: THE EFFICIENCY COSTS OF NONSPECIALIZATION

We will call the $h_i$ locally efficient at $Q_i$ if, roughly speaking, there is no nearby $h_i$, say, $h_{i^*}$, that provides more than the quantities $Q_i$. More precisely $h_i$ is locally efficient if there is some $\epsilon_i$ such that $N_i^h e_i h_{i} \leq \epsilon_i$ implies that $h_i N_i^{h_{i^*}} e_i h_{i^*} \leq Q_i$ for all $i$.

Local efficiency still rests on the idea that generates interest in the concept of efficiency. It still asks whether or not a better arrangement, i.e., one that generates larger quantities of goods than the $Q_i$, is possible. However, a local efficiency test only compares nearby arrangements, as common sense may suggest. Local efficiency makes behavior of our large numbers of equilibria much more coherent.

To see this we first need another concept—an almost specialized production pattern (aspp). In such a pattern, with the exception of at most one $i$, all production is specialized; $h_i > 0$ implies $h_{i^*} = 0$ and $h_{i^*} > 0$ implies $h_i = 0$. We will also assume as part of the definition that for positive $h_i$, $h_{i^*} h_{i^*} > 0$, i.e., that a positive labor input always yields a positive output at the point in question, since otherwise the point is automatically inefficient. Note that a perfectly specialized production pattern is always an aspp. With this definition we can state the main theorem.

**Theorem 4.1**

A sufficient condition for $h_i$ to be locally efficient is that the $h_i$ be an almost perfectly specialized production pattern.

This has the important Corollary: The production pattern of any perfectly specialized equilibrium point is locally efficient.

Proof of the theorem: The theorem is almost proved if we demonstrate the following lemma:

**Lemma 4.2**

If $h_i$ is an aspp, then for sufficiently small, any changes $\delta h_i$ that strictly decrease the total labor quantity in Country 1, while maintaining the total output
we suppose that it is Country 1, then $b_t$ would give us a set of $n_t$ that
underlies the labor of Country 1, and which makes the $Q_t$ while not increasing
the use of labor in Country 2. This contradicts the lemma and proves the
theorem.

That the app condition of Theorem 4.1 is not too arbitrary can be seen from
the following theorem:

**Theorem 4.3**

If the production functions $f_t$ have increasing derivatives (rising marginal
products of labor), then app is a necessary and sufficient condition for local
efficiency.

The proof of this Theorem is available from the authors.

Since efficiency implies local efficiency Theorem 4.3 has as an immediate
corollary a result about classical efficiency:

**Theorem 4.4**

If the production functions $f_t$ have increasing derivatives there are no efficient
equilibria with more than one shared industry.

This result helps to explain our emphasis on specialized and near specialized
(app) production patterns. In the case of increasing derivatives they are the
only production patterns with even the possibility of being efficient.

Examples of production functions with increasing derivatives are all production
functions of the form $e_t e^{n_t}$, with $n_t > 1$, or any production function of
that form preceded by an interval of zero output. An example of a reasonable
production function not immediately meeting the criterion is a production
function with a flat-bottomed average cost curve or one that starts out at zero,
then increases sharply, and then becomes linear with a positive slope $e_t$ that is
less than the derivative in the preceding steep portion. These are the functions
of the type that underlies the extended-specialized production plans. However,

---

**V. QUANTIFIED EFFICIENCY (Q-EFFICIENCY)**

Although we see that all specialized production patterns are locally efficient,
we are still interested in their efficiency in a more global sense. Now, however,
we will not ask whether a given production pattern is efficient or not efficient
but, rather, how bad or how good is it in terms of efficiency. We now seek a
quantitative rather than a binary answer.

For this purpose we introduce as our quantitative measure, $\lambda$-efficiency.
Consider any set of $Q_t$ and its production pattern $b_t$. We will assume that the $b_t$
use up the total labor available in both countries. We will look for other
production patterns $b_t$ that make the same set of $Q_t$ possibly using less of the
total labor of the two countries, i.e., $\sum b_{t1} < L_t + L_2$. We propose the

**Definition:** $\lambda = (\min \sum b_{t1}) / (L_t + L_2)$ where the minimization extends over
all production patterns that make the $Q_t$ and do not require more than the
available amount of labor in either country. We define $\lambda$ to be the $\lambda$-efficiency
of $Q$. Clearly $\lambda = 1$ coincides with the classical notion of efficiency, but now
we can describe a point as having a $\lambda$-efficiency of, say, 0.77. This measures
the percent of the total workforce required to make the original set of goods
when they are made in the most efficient possible way. Thus, $1-\lambda$ is the
proportion of the available labor wasted by the inefficiency of the equilibrium
under study.

We will next discuss ways of finding the $\lambda$ associated with a production
pattern $P = (L_t L_2)$. We will not be able to determine the magnitude of $\lambda$ precisely,
but will be able to obtain workable upper and lower bounds.

**Underestimating $\lambda$:** If $v$ is the minimizing value of the linear programming
problem (3.2) $v$ is also an underestimate of the labor required to make
the quantities $Q_t$ using the $b_t$. This follows immediately from Lemma 3.4 which
says that any feasible production plan $P$ gives rise to a $P^*$ that satisfies the
linearized plan and uses less labor. This gives us, by the definition of $\lambda$:
Theorem 5.1

\[ \lambda_0 = v_1 f(1,1+L_0) \] is an underestimate of \( \lambda \).

Overestimating \( \lambda \): Let \( v_1 \) be the minimizing value of (3.2) when the minimization is not over all \( y \) but only over all integer \( y \).

Theorem 5.2

\[ \lambda_0 = v_1 f(1,1+L_0) \] is an overestimate of \( \lambda \).

Proof: By Lemma 3.8 any integer solution of (3.2) is a feasible production plan. Any feasible production plan is an overestimate of the minimal amount of labor required. Therefore the minimum integer solution is a production plan and therefore an overestimate.

Simplifying the Calculations: Both the linear programming solution to (3.2) which gives \( \lambda_0 \) and the integer programming solution to (3.2) which gives \( \lambda_0 \) can be greatly simplified. The first problem reduces to the well-known knapsack problem and the integer problem can be reduced to a dynamic programming problem with one state variable. A description of the method can be supplied to interested readers.

Convergence of \( \lambda_0 \) and \( \lambda_0 \) for large problems: \( \lambda_0 \) and \( \lambda_0 \) will tend toward each other, and therefore toward \( \lambda \) for large problems in which no one or two industries use up a large proportion of the total labor force. Intuitively, this reflects the fact, mentioned in Section III, that the linear programming problem will have at most two non-integer components, and so its solution will not be far from being an integer solution itself. Therefore, if the two components don't matter much, the linear programming underestimate is not far from an integer programming overestimate. While this is plausible there is considerable work in getting these thoughts to be precise. A formal description is available from the authors.

Efficiency in the Equilibria of a Region: We can now use the overestimates and underestimates to examine the efficiencies of the various equilibria populating the regions of equilibrium. In Figure 8 we show the 93 equilibria in an 11-good model whose underestimate of \( \lambda \), \( \lambda_0 \), is > 0.98. In Figure 9 we show the 449 equilibria whose overestimate of \( \lambda \), \( \lambda_0 \), is > 0.98. In these figures, and in other similar figures that we have examined, there is no particular tendency for the equilibria near the middle to be more efficient than the rest. There is, however, some tendency for the more efficient points to be near the upper boundary. This tendency is much more pronounced in the middle region than at either the right or left ends.
However, taking the region as a whole we may ask how efficient these equilibria are on average. In Figure 10 (a cumulative distribution) we have considered all the equilibria of the 11-good model. The height y of the upper curve gives the percent of the 2048 equilibria whose $\lambda_e$ is less than or equal to $x$. The lower curve is a similar plot for $\lambda_e$. A plot of actual efficiency, if we could obtain it, would be in between. The two vertical bars are the average $\lambda_e$ on the left and the average $\lambda_e$ on the right. The average efficiency lies in between, i.e., between 0.91 and 0.93. Considering the range of production efficiencies in the model parameters, the high average efficiency is little surprising.

9. In the 11-country model the production function for the ith industry in the jth country was $\alpha_j^{\alpha_j}$. The $\alpha_j$s were between 1 and 2. The 11 values for the $\alpha_i$ were 1.00, 1.05, 1.10, 1.15, 1.20, 0.95, 0.90, 0.85, 0.80, 0.75, and 0.70, and the 11 values for the $\alpha_j$ were 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75. The 27-country model was similar.
In Figure 11, we have the same plot as in Figure 10. However, this time we have taken a random sample of 2,000 points from a larger (27-good) model. The convergence of the two curves and the improved estimates of the average efficiency are exactly what one expects from the convergence discussed above.

VI. CONCLUSION

As promised in the introduction, we have demonstrated that there is a wide range of states of efficiency in the classical sense, that are possible for the many equilibria that exist under economies of scale and for similar production relationships. We have shown the profound differences in efficiency theory from that in the classical case of diminishing or constant returns, proving that equilibria need not be efficient and that even efficient equilibria need not satisfy comparative advantage. We have discussed examples in which all 29, 2-specialized equilibria are efficient, and examples in which there are very few efficient equilibria. We have shown how the concept of comparative advantage can be modified to adapt it to the n-good case and especially to the specialized or extended-specialized solutions that are important in a world of scale economies, demonstrating that comparative advantage is sufficient but not necessary for efficiency. We have introduced two new efficiency measures, local efficiency and scale efficiency, that seem to function more uniformly than the standard classical concept in the economics of scale setting, while retaining many of the properties that provide interest to the notion of efficiency.

The subject is clearly not of academic interest alone. It relates immediately to the role of laissez-faire and government intervention in international trade. The classical analysis showed good reasons to believe that in a world of diminishing returns unspecialized market forces can be relied upon to do a reasonably good job in promoting efficiency in the trade process. The current paper, along with a number of recent writings dealing with the role of scale economies in international trade show that here matters are less simple. It does not follow that mindless government intervention, much less unfettered protectionism, is the way to go. But the analysis shows that economic welfare may be enhanced in some circumstances if the governmental takes on some role. Exploration of the appropriate role is only beginning and this paper is intended as a contribution to the process.

REFERENCES


SUMMARY

In the presence of scale economies a country that happens to be the exclusive producer of a commodity will be able to retain its monopoly against entry of others to enter on a small scale even if that firm has neither absolute nor comparative advantage in its production. Hence equilibria can violate comparative advantage; be inefficient and yet be stable. Moreover, equilibria that violate comparative advantage can be efficient. In the article, sufficient efficiency conditions for the scale economies case are provided, and two new concepts, local efficiency and (quantitative) degree of efficiency are explained. The analysis confirms that if a world of scale economies market forces cannot be relied upon to yield an efficient equilibrium.

ZUSAMMENFASSUNG

WILLIAM J. BAUMOL AND RALPH E. GOMORY


RÉSUMÉ

Au cas d'économies de l'échelle, un pays qui est le seul producteur d'un bien peut être capable de préserver son monopole malgré les efforts d'autres pays d'entrer dans la production à une échelle plus petite, même si le pays tenant le monopole n'a ni un avantage absolu ni relatif dans la production. Il en suit que ce genre d'équilibre peut contredire le principe de l'avantage comparatif, être non-efficient et être stable. De plus, des équilibres qui violent le principe de l'avantage comparatif peuvent être efficients. Dans l'article les conditions suffisantes pour le cas d'économies de l'échelle sont données, et deux conceptions nouvelles, celle de l'efficacité locale et celle du degré (quantitatif) de l'inefficacité sont expliquées. L'analyse confirme que dans un monde avec des économies de l'échelle les forces du marché ne produisent pas toujours des équilibres efficient.