DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

COMPETITIVE EQUILIBRIUM WITH SEPARABLE EXTERNALITIES

W. David Montgomery

SOCIAL SCIENCE WORKING PAPER 85

May 1975

COMPETITIVE EQUILIBRIUM WITH SEPARABLE EXTERNALITIES

W. David Montgomery
California Institute of Technology

1. INTRODUCTION

The characterization of external effects as "separable" has played an important role in the development of the theory of externalities. The separable case appears particularly well behaved when procedures for achieving an optimum allocation of resources in the presence of externalities are examined. For example, Davis and Whinston [1962] find that separability assures the existence of a certain kind of equilibrium in bargaining between firms which create externalities, and that equilibrium does not exist without separability. Kneese and Bower [1968] argue that with separability the computation of Pigovian taxes to remedy externalities is particularly simple. Marchand and Russell [1974] demonstrate that certain liability rules regarding external effects lead to Pareto optimal outcomes if and only if externalities are separable.

We will argue in this paper that whenever an externality affecting a firm is separable, the production set of that firm is not convex in a neighborhood of zero output.[1] The proposition is established by redefining separability in a manner which allows for the fact that in the long run a firm will shut down rather than accept negative profits. These definitions yield the theorem that separability implies a non-convexity of the production function, which may result in a discontinuous supply correspondence.

Since the discontinuity is a consequence of market structure, i.e. the actions of other firms, rather than a characteristic of technology alone, it is possible that equilibrium will exist despite the discontinuity. We prove that with "complete separability," a case in which the production function is linear in productive inputs, the discontinuity disappears in equilibrium. When the production function exhibits decreasing returns in some range, we argue that the aggregate supply correspondence will be continuous if the number of firms which can earn non-negative profits in equilibrium is independent of equilibrium prices. We conclude with examples of conditions under which this independence will and will not obtain. In passing we establish an intimate connection between externalities and the number of firms in an economy.

II. SEPARABILITY AND NONCONVEXITY

Standard proofs of the existence of equilibrium of a competitive economy employ assumptions regarding the convexity of production sets which are violated when separable externalities are present. Arrow and Hahn [1972] for example, assume in proving existence with externalities that the production set of firm i is convex in the variables controlled by firm i for every activity chosen by other firms.

It appears, therefore, that a case of some importance in applied welfare economics, the case of separable externality, is in an anomalous theoretical position: the fundamental question of whether a competitive equilibrium can exist in this case is unresolved. In this section we will demonstrate the existence of a nonconvexity. In the final three sections we will explore two subjects: the considerations which determine the existence of equilibrium and the implications of nonconvexity for market structure with separable externalities.

Although much of the literature deals with cost functions, it is possible to define a separable externality in two ways.

Let $C(Y_1, Y_2, \ldots, Y_m)$ be the cost function of a firm which produces Y_1, and suffers an external diseconomy which is a function of the output of other firms Y_j, $j = 2, \ldots, m$. The cost function is dual to the production function $F(X_1, \ldots, X_n, Y_2, \ldots, Y_m)$ where X's are inputs.

Conventional definitions of separability become inappropriate in cases in which the firm has the option of going out of business. Hence we define separability on the strictly positive real numbers and extend the separable functions to the non-negative orthant. Let R^n_{++} be the n-dimensional space of strictly positive real vectors which contains $(X_1, \ldots, X_n)$, and let R^{m-1}_+ be the m-1 dimensional space of non-negative vectors which contains $(Y_2, \ldots, Y_m)$.

If we define separability of the production function with respect to externalities in the conventional fashion by stating that a production function is separable iff it can be written as $g(X_1, \ldots, X_n) + h(Y_2, \ldots, Y_m)$ for all $(X_1, \ldots, X_n) \in R^n_+$ and $(Y_2, \ldots, Y_m) \in R^{m-1}_+$, the nonsensical possibility exists that $Y_1 < 0$ for some values of X_i and Y_j. Let $\mathfrak{S}(F) = \{(X_1, \ldots, X_n, Y_2, \ldots, Y_m) \in R^n_+ \times R^{m-1}_+ : F(X_1, \ldots, X_n, Y_2, \ldots, Y_m) > 0\}$. The function $F : R^n_+ \times R^{m-1}_+ \longrightarrow R^1_+$ is defined to be equal to zero on $(R^n_+ \times R^{m-1}_+) \setminus \mathfrak{S}(F)$. Separability is a property which we need only require on $\mathfrak{S}(F)$.

<u>Definition 1</u>: A production function $F : R^n_+ \times R^{m-1}_+ \longrightarrow R^1_+$ is separable if and only if it can be written as $g(X_1, \ldots, X_n) + h(Y_2, \ldots, Y_m)$ for all $(X_1, \ldots, X_n, Y_2, \ldots, Y_m) \in \mathfrak{S}(F)$.

Since we deal entirely with external diseconomy, $h(\cdot) < 0$ and $h' < 0$. The cost function is derived from the production function by finding the minimum cost of producing Y_1 when other firms are producing $(Y_2, \ldots, Y_m)$. When $F > 0$ the cost function is found as the dual of the production function in the conventional manner. We define $\mathcal{C}(0, Y_2, \ldots, Y_m) = 0$ for all $(Y_2, \ldots, Y_m)$.

Definition 2: A cost function $C(Y_1, Y_2, \ldots, Y_m)$ is separable with respect to externalities if and only if it can be written as $C_1(Y_1) + C_2(Y_2, \ldots, Y_m)$ for all $Y_1 > 0$.

Note that a cost function is separable with respect to externalities if and only if $\frac{\partial^2 C}{\partial Y_1 \partial Y_j} = 0$ in $R^1_{++} \times R^{m-1}_{+}$. We assume throughout that F is continuously twice differentiable on $\mathcal{D}(F)$ and that C is continuously twice differentiable on $R^1_{++} \times R^{m-1}_{+}$. A production function is separable with respect to externalities if and only if $\frac{\partial^2 C}{\partial X_i \partial Y_j} = 0$ in $\mathcal{D}(F)$.

To make F and C well behaved in their entire domain we assume that $F \to 0$ as $(X_1, \ldots, X_n, Y_2, \ldots, Y_m)$ approach the boundary of $\mathcal{D}(F)$. That is, let $\dot{\mathcal{D}}(F) = \{(X_1, \ldots, X_n, Y_2, \ldots, Y_m) : g(X_1, \ldots, X_m) + h(Y_2, \ldots, Y_m) = 0\}$. By hypothesis F is continuous on $\mathcal{D}(F)$, and, since $F = 0$ on $(R^n_+ \times R^{m-1}_+) \setminus \mathcal{D}(F)$, on $(R^n_+ \times R^{m-1}_+) \setminus \mathcal{D}(F)$. Then F is continuous everywhere if and only if $F \to 0$ whenever $(X_1, \ldots, X_n, Y_2, \ldots, Y_m) \to \dot{\mathcal{D}}(F)$. We assume also that C is continuous everywhere in its domain, and that F is monotonic increasing in $(X_1, \ldots, X_n)$ on $\mathcal{D}(F)$.

We can now establish the existence of a nonconcavity in the production function.

Theorem 1: a) if the production function is separable, it is not concave in a neighborhood of $\dot{\mathcal{D}}(F)$. b) If the cost function is separable, the production function is not convex in a neighborhood of $Y_1 = 0$.

Proof: a) Let $(\dot{X}, \dot{Y}) = (\dot{X}_1, \ldots, \dot{X}_n, \dot{Y}_2, \ldots, \dot{Y}_m) \in \dot{\mathcal{D}}(F)$. Then $F(\dot{X}_1 + \varepsilon, \dot{X}_2 + \varepsilon, \ldots, \dot{X}_n + \varepsilon, \dot{Y}_2, \ldots, \dot{Y}_m) > 0$, and $F(\dot{X}_1 - \varepsilon, \ldots, \dot{X}_n - \varepsilon, \dot{Y}_2, \ldots, \dot{Y}_m) = 0$ for any $\varepsilon > 0$. If F is concave, $\alpha F(\dot{X} + \varepsilon, \dot{Y}) + (1 - \alpha)F(\dot{X} - \varepsilon, \dot{Y}) < F(\alpha(\dot{X} + \varepsilon) + (1 - \alpha)(\dot{X} - \varepsilon), \dot{Y}) = F((\dot{X} - \varepsilon + 2\alpha\varepsilon), \dot{Y}) = 0$ if $\alpha = 1/2$. But $1/2\, F(\dot{X} + \varepsilon, \dot{Y}) + 1/2\, F(\dot{X} - \varepsilon, \dot{Y}) = 1/2\, F(\dot{X} + \varepsilon, \dot{Y}) > 0$. Hence $F(\dot{X}, \dot{Y})$ is not concave.

b) The argument for $m = 2$ generalizes immediately to general m. Hence let $m = 2$. If the cost function is separable, then $C(Y_1, Y_2) > C_2(Y_2)$ for all $Y_1 > 0$, $Y_2 > 0$. Let $X(Y_1, Y_2)$ be an input vector which is a solution of the problem: minimize $W \cdot X$ subject to $F(X_1, \ldots, X_n, Y_2) = Y_1$, where $Y_1 > 0$. Then with a separable cost function $W \cdot X(Y_1, Y_2) > C_2(Y_2)$ for all $Y_1 > 0$. Since $X(0, Y_2) = (0, \ldots, 0)$ for all $Y_2 > 0$, concavity of F implies that for $X = \overline{X}$, some fixed vector,

$$\alpha F(\overline{X}, Y_2) + (1 - \alpha)F(0, Y_2) < F(\alpha\overline{X}, Y_2)$$

for all $\alpha \in (0,1)$. Hence it must be true that $\alpha F(\overline{X}, Y_2) < F(\alpha\overline{X}, Y_2)$. Now choose α sufficiently small that $W \cdot \alpha\overline{X} < C_2(Y_2)$. This is only possible if $Y_1 = F(\alpha\overline{X}, Y_2) = 0$. But then $\alpha F(\overline{X}, Y_2) > F(\alpha\overline{X}, Y_2) = 0$, and the production function is not concave. ||

Theorem 1 can be illustrated heuristically. When the production is separable, the marginal productivity of any factor used by firm 1 is independent of Y_2, the output of firm 2. Therefore at an interior maximum for which the first order conditions are necessary, firm 1's input choice is independent of the acitivty of firm 2. When the cost function is separable, marginal cost of producing Y_1 is independent of Y_2, and firm 1's output choice is independent of firm 2's activity.

If both the production function and the cost function are separable, then input and output of firm 1 are independent of firm 2's activity.[2]

These propositions can be illustrated diagrammatically in a simple fashion for the one-input case. It is possible to find a differential equation which must be satisfied by any production function generating a separable cost function.[3] When there is just one input, denoted X, the differential equation is

$$F_Y F_{YX} - F_Y F_{XX} = 0.$$

The function F which solves this equation is of the form $F = A(X + h(Y))$ where A and h are arbitrary functions restricted to preserve the concavity of F.[4] In each figure the production function satisfies $F(0, Y_2) = 0$; i.e., it goes through the origin. In Figure 1 the production function is characterized by decreasing returns and separability. Since the same input choice must be profit maximizing for all Y_2, changes in Y_2 shift the production function vertically, keeping the slope of $F(X, Y_2)$ the same for constant X.

In Figure 2 a production function of the form $F(X - h(Y_2))$ is drawn. Changing Y_2 in this case shifts the production function horizontally, so that the slope of F is constant for constant Y_1. This production function generates a separable cost function.

Figure 3 illustrates complete separability, i.e. the production function is separable and generates a separable cost function. Such a function must have the property that the slope of $F(X, Y_2)$ equals the slope of $F(X, Y_2')$ for constant X and also for constant Y_1. For this to be true for all X and Y_1, F must be linear in X.[5]

III. EXISTENCE OF EQUILIBRIUM WITH COMPLETE SEPARABILITY

With one input it is only possible to have complete separability when the production function is linear in X_1. In this case the nonconcavity

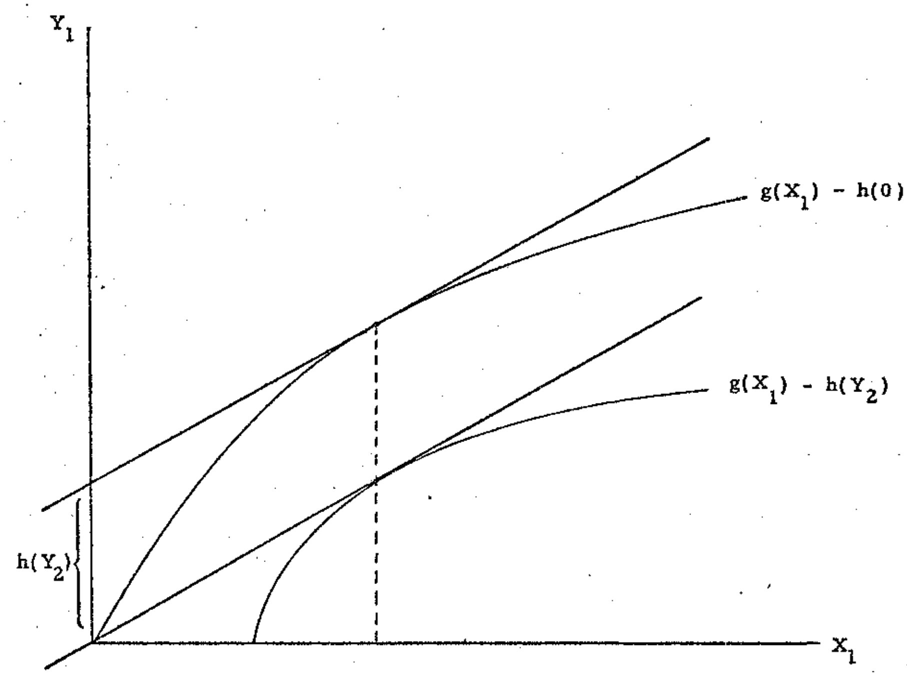

Figure 1: Separable Production Function

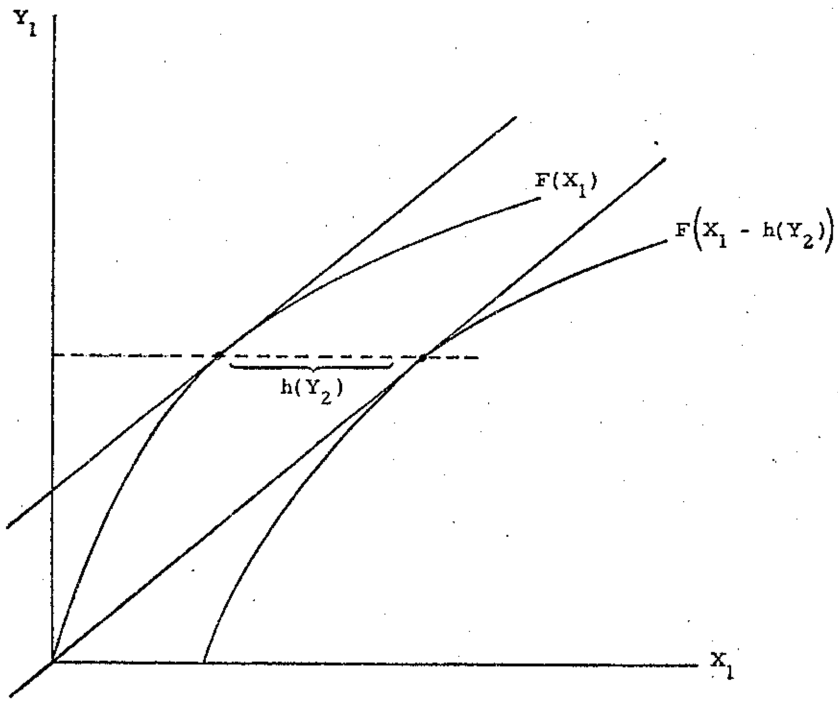

Figure 2: Separable Cost Function

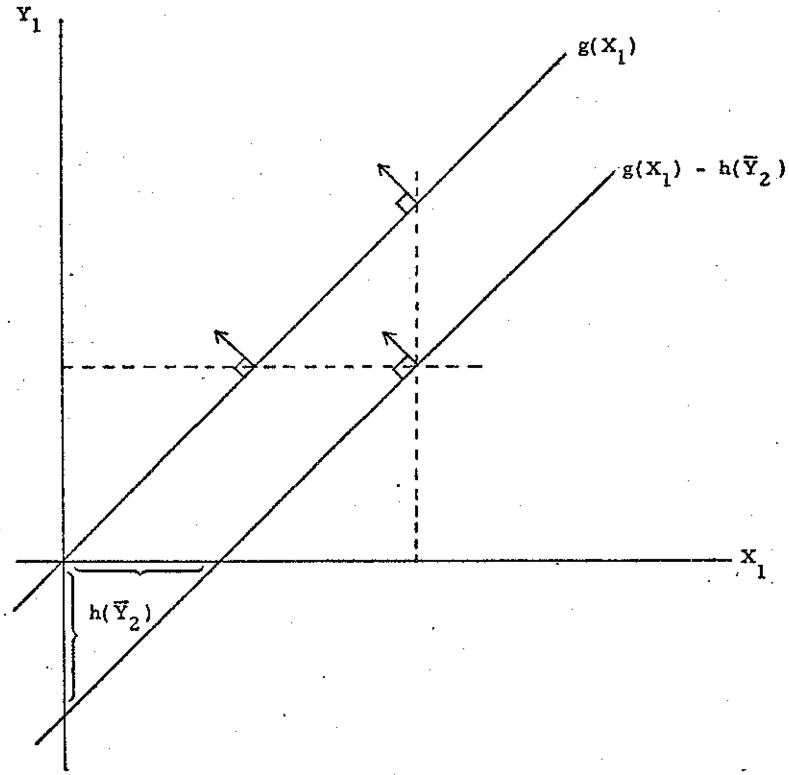

Figure 3: Complete Separability

of the production function disappears in equilibrium. A firm affected by a completely separable externality will have the supply correspondence illustrated in Figure 4. If p_1/w_1 exceeds the (constant) slope of the production function, the firm can earn unbounded profits. If p_1/w_1 is less than the slope of the production function, the firm earns negative profits for all output greater than zero. If $Y_2 > 0$, the same is true when p_1/w_1 equals the slope of the production function. Hence the supply correspondence is not continuous (or even defined) at a point p_1/w_1 equal to that slope.

Despite this discontinuity in the individual supply correspondence, equilibrium will exist. Suppose that firm i has production function $a_i X_i - b_i Y_j$. We ask if there is any price ratio such that when both firms maximize profits $Y_i > 0$ for both. We know that the equilibrium prices must be such that $p_i/w_i \le a_i$, for $i = 1, 2$. Otherwise some firm will produce unbounded output. If $p_1/w_1 < a_1$, then firm 1 produces nothing and firm 2 is unaffected by externality, and hence $p_2/w_2 = a_2$ is an equilibrium price. Now suppose $p_i/w_i = a_i$, $i = 1, 2$. If $Y_1 > 0$ and $Y_2 > 0$, then both firms are earning negative profits and cannot be in equilibrium. If $Y_1 = 0$ and $Y_2 > 0$ then any finite output level gives firm 2 zero profits and is in equilibrium. Moreover, given that $Y_2 > 0$, firm 1 cannot do better than choose $Y_1 = 0$. Hence prices $p_1/w_1 \le a_1$, $p_2/w_2 = a_2$, $Y_1 = 0$, $Y_2 > 0$ is an equilibrium. Permuting the indices gives another equilibrium.

In this case we can have multiple equilibria, each with just one firm producing nonzero output. It does not matter which firm is out of business. This is characteristic of constant returns, since one firm can produce any level of output using the same total input which would be used if the output were divided among many firms. Normally constant returns imply that the number of firms in an industry is indeterminate. With constant returns and separable externality in an industry, the size of an industry is one firm.

The arguments given for two firms also apply to the case of n firms. If it is impossible to have two firms simultaneously maximizing

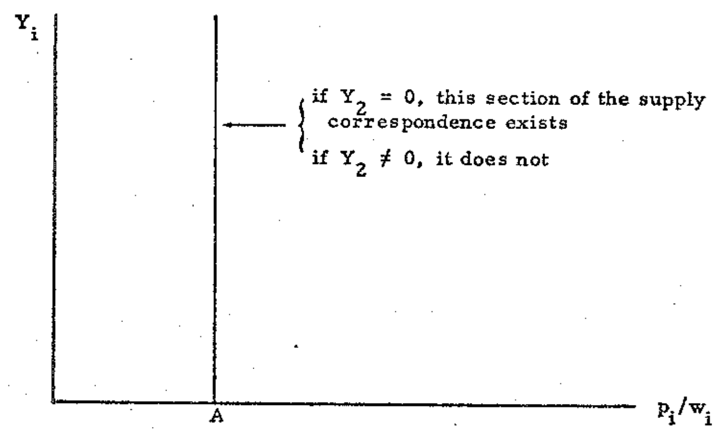

Figure 4

profits and producing positive output, it is impossible to have n, since comparing any two firms, we find only one is in operation. Thus we have proved the following theorem.

Theorem 2: If all externalities are completely separable, and $h_i \neq 0$ in the neighborhood of equilibrium, then in equilibrium only one firm affected by externalities produces nonzero output.

IV. DISCONTINUITY IN AGGREGATE SUPPLY

Problems of existence of equilibrium can only arise when separability is combined with nonconstant returns, as in Figures 1 and 2. Because of the simpler functional forms involved, we concentrate on the case of a separable production function $F = f(X_1, \ldots, X_n) + g(Y_2, \ldots, Y_m)$ where f is strictly concave. We will show that if the number of firms which can earn non-negative profits changes as prices change, the supply correspondence will not be continuous. Consider an example of two firms, only one of which is affected significantly by the externality. That is, assume

$$F_1 = g_1(X_1) - h_1(Y_2)$$

and

$$F_2 = g_2(X_2)$$

Assume for simplicity that both firms produce identical goods, and use identical inputs. Then aggregate supply by these firms is $Y_1 + Y_2$ and aggregate input demand is $X_1 + X_2$. With externality going only one way it is not difficult to determine how $Y_1 + Y_2$ will vary with p/w. At any price ratio firm 2 will set

$$\frac{\partial F_2}{\partial X_2} = \frac{w}{p}$$

The output thus determined is taken as a parameter by firm 1 in choosing its optimal supply. We take a simple example:

$$Y_1 = X_1^{\frac{1}{\alpha}} - hY_2^{\beta}$$

$$Y_2 = X_2^{\frac{1}{\alpha}}$$

where $\alpha > 1$. Then maximizing profits firm 2 will choose

$$X_2 = \left(\frac{p}{\alpha w}\right)^{\frac{\alpha}{\alpha-1}}$$

and

$$Y_2 = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}$$

Firm 1 will choose

$$X_1 = 0$$

or

$$X_1 = \left(\frac{p}{\alpha w}\right)^{\frac{\alpha}{\alpha-1}}$$

depending on whether or not it can earn non-negative profits. We can write firm 1's profits, Π_1, as a function of p/w.

Since

$$Y_1 = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\left(\frac{p}{\alpha w}\right)^{\frac{\beta}{\alpha-1}}$$

$$\Pi_1 = p\left[\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\left(\frac{p}{\alpha w}\right)^{\frac{\beta}{\alpha-1}}\right] - w\left(\frac{p}{\alpha w}\right)^{\frac{\alpha}{\alpha-1}}$$

$$= \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}\left[p - ph\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}} - \frac{p}{\alpha}\right]$$

Since $\alpha > 1$, this expression is positive if

$$\left(1 - \frac{1}{\alpha}\right) > h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$$

Three cases, which differ in terms of the change in marginal damage when Y_2 increases, can be distinguished. Marginal damage, defined as $\frac{\partial F_1}{\partial Y_2}$, is increasing if $\frac{\partial^2 F_1}{\partial Y_2^2} > 0$; constant if $\frac{\partial^2 F_1}{\partial Y_2^2} = 0$, and decreasing if $\frac{\partial^2 F_1}{\partial Y_2^2} < 0$.

Case 1: If marginal damage is constant, $\beta = 1$ and the sign of profit is independent of p/w, depending only on whether $1 - 1/\alpha > h$, in which case profits are always positive, or $1 - 1/\alpha < h$, in which case profits are always negative. Since firm 2 produces non-zero output for all $p/w \neq 0$, $1 - 1/\alpha > h$ implies that firm 1 can always produce some output. The supply function then is

$$Y_1 + Y_2 = 2\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}$$

$$= (2 - h)\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}$$

If $1 - 1/\alpha > h$, the supply function is

$$Y_1 + Y_2 = Y_2 = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}.$$

Both functions are continuous, and no problems can arise.

Case 2: If marginal damage is increasing, then $\beta > 1$ and $h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$ is an increasing function of p/w. It is always possible to choose p/w small enough that

$$\left(1 - \frac{1}{\alpha}\right) > h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$$

Therefore for low values of p/w firm 1 can earn positive profits.

As p/w increases, however, we can choose a large enough value to make

$$\left(1 - \frac{1}{\alpha}\right) < h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$$

so that profits become negative. Let R be that value of p/w for which

$$1 - \frac{1}{\alpha} = h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$$

Then for $p/w \leq R$, total supply is

$$Y_1 + Y_2 = 2\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\left(\frac{p}{\alpha w}\right)^{\frac{\beta}{\alpha-1}} \qquad (1)$$

For $p/w \geq R$,

$$Y_1 + Y_2 = Y_2 = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} \qquad (2)$$

Solving for R in

$$1 - \frac{1}{\alpha} = h\left(\frac{1}{\alpha}R\right)^{\frac{\beta-1}{\alpha-1}}$$

gives

$$\frac{1}{\alpha}R = \left(\frac{\alpha - 1}{\alpha h}\right)^{\frac{\alpha-1}{\beta-1}}$$

Substituting R for p/w in (1') gives

$$Y_1 + Y_2 = 2\left(\frac{\alpha - 1}{\alpha h}\right)^{\frac{1}{\beta-1}} - h\left(\frac{\alpha - 1}{\alpha h}\right)^{\frac{\beta}{\beta-1}}$$

$$= \left(1 + \frac{1}{\alpha}\right)\left(\frac{\alpha - 1}{\alpha h}\right)^{\frac{1}{\beta-1}} \qquad (3)$$

and in (2) gives

$$Y_1 + Y_2 = \left(\frac{\alpha - 1}{\alpha h}\right)^{\frac{1}{\beta-1}}$$

Since (3) > (4), the total supply function will jump at R, since (1), which defines the supply function up to R, is strictly greater than (2) which defines the supply function from R on. This is illustrated in Figure 5.

Case 3: If marginal damage is decreasing, then $\beta < 1$ and $h\left(\frac{p}{\alpha w}\right)^{\frac{\beta-1}{\alpha-1}}$ is a decreasing function of p/w. By a train of reasoning identical to

that used in Case 2 we can establish that for p/w < R firm 1 cannot earn positive profits; for p/w > R it can. Therefore the supply function is (3) for 0 < p/w ≤ R and (2) for p/w > R. We have established that (2) > (3) at R. Therefore the supply function jumps, as in Figure 6. These discontinuities can interfere with the existence of equilibrium. If the demand curve for $Y_1 + Y_2$ goes through the discontinuity, it will never be possible to achieve exact equality between supply and demand. This is more likely in Case 3, of course, since in Case 2 a demand curve must have a positive slope to pass through the discontinuity and not intersect the supply curve.

Such a discontinuity does not necessarily preclude the possibility of proving the existence of equilibrium, although the proof becomes more difficult. The standard argument (Arrow and Hahn [1971], p. 169-171, Rothenberg [1960]) is that as more firms and consumers are introduced into the economy it becomes possible to choose various combinations of firms producing zero output and firms producing positive output to approximate demand at the price for which the discontinuity exists. Such an argument cannot necessarily be made in this case. As Rothenberg [1960] pointed out, there is no guarantee that non-convexities arising from externalities will vanish as the number of agents in the economy increases. The discontinuity caused by one-way, separable externality will change in different ways depending on how the economy is expanded. If we let the number of polluting firms increase while holding constant the number of sufferers, the discontinuity will vanish since the firms affected by the nonconvexity become small relative to the economy. If both classes are increased at the same rate, the size of the discontinuity relative to aggregate supply may remain roughly constant.

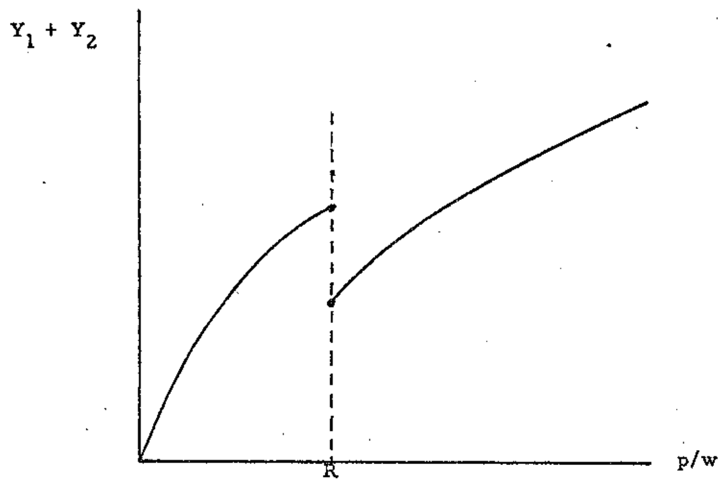

Figure 5

Case 2: Increasing Marginal Damage

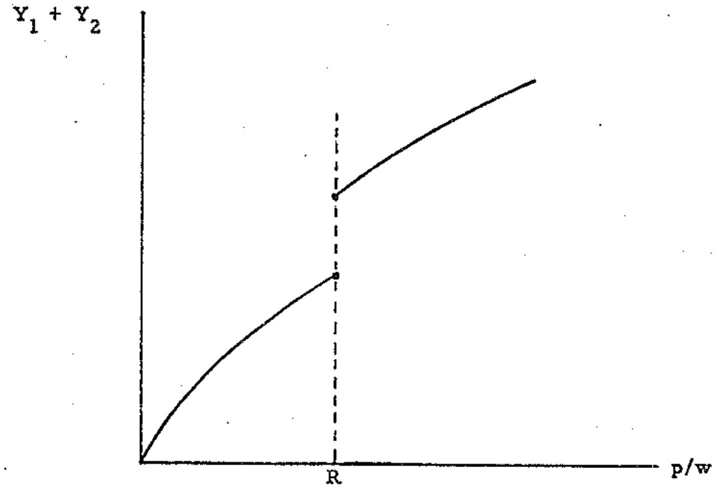

Figure 6

Case 3: Decreasing Marginal Damage

V. THE CASE OF CONSTANT MARGINAL DAMAGE

When marginal damage is constant, it becomes possible to find conditions under which the number of firms earning non-negative profits is independent of prices, so that no discontinuities which cause problems of existence of equilibrium will arise. In such a case the number of firms may be determined by the degree of convexity of the production function. We assume now that all firms are identical and that externalities go in both directions.

The analysis of cases in which there are mutual externalities and decreasing returns is complex, since to define the total supply correspondence we must make sure that the supply response of any firm to given prices is profit maximizing with respect not only to the prices but also with respect to the decision of the other firm (Montgomery [1975b]). When this is done, it appears that for some production functions it is impossible for more than one firm to earn positive profits at any prices; in some cases two firms can earn positive profits at all prices; in others more firms can. Whenever market structure is thus independent of prices, the system is well behaved and the aggregate supply correspondence exhibits no discontinuities.

Assume two identical firms $i = 1, 2$, with production functions

$$Y_i = X_i^{1/\alpha} - hY_j.$$

At any price ratio p/w the input choice of firm i will be either

$$X_i = \left(\frac{p}{\alpha w}\right)^{\alpha/\alpha-1}$$

or $X_i = 0$, depending on whether Y_j is such that firm i can or cannot earn non-negative profits.

We begin by assuming that both firms are operating at interior maxima. We know the input of X_i; but output Y_i depends on the output of firm j, which in turn depends on Y_i. We cut through this chain by solving

$$Y_i = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\left[\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - hY_i\right] \qquad (4)$$

Therefore

$$Y_i = \frac{1}{1+h}\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}}$$

and

$$\Pi_i = p\left(\frac{p}{w}\right)^{\frac{1}{\alpha-1}}\left[\frac{\alpha^{\frac{1}{1-\alpha}}}{1+h} - \alpha^{\frac{\alpha}{1-\alpha}}\right]$$

Profits Π_i will be positive, with both firms in operation, for all prices when h is less than some critical value. We solve

$$\frac{\alpha^{\frac{1}{1-\alpha}}}{1+h} - \alpha^{\frac{\alpha}{1-\alpha}} = 0$$

to find this critical value, which is $h = \alpha - 1$. Since α is a measure of how strongly returns to scale decrease, it is a relation between the magnitude of the externality and returns to scale which determines how many firms can operate in equilibrium. Since the possibility of earning non-negative profits is independent of p/w, the economy will always have two firms operating if $h < \alpha - 1$, and one firm if $h > \alpha - 1$.

This suggests that if α is sufficiently large, it may be possible to have more than two firms in operation. Suppose there are n + 1 identical firms, with production functions

$$Y_i = X_i^{\frac{1}{\alpha}} - h\sum_{j\neq i} Y_j$$

Clearly every firm which is in operation will choose identical X_i. Thus we can again solve

$$Y_i = \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - h\sum_{j\neq i}\left[\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - hnY_1\right]$$

$$= \left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - hn\left[\left(\frac{p}{\alpha w}\right)^{\frac{1}{\alpha-1}} - hnY_1\right] \qquad (5)$$

where the substitution $nY_1 = \sum_{i\neq j} Y_j$ follows from the fact that all firms in operation will produce identical output. Since (5) is identical to (4) except that h is replaced by nh, it follows that

$$\Pi_i > 0 \quad \text{for all } i \text{ if} \quad nh < \alpha - 1$$

For example, if h is just less than 1, and $\alpha = 3$, $n + 1 = 3$, firms can earn positive profits simultaneously.

Two tentative conclusions about conditions under which competitive equilibrium will exist follow.

1) If, relative to the number of firms affected by separable externalities, the number of firms not affected by separable externality is large, and those firms are able to produce the same outputs and use the same inputs as firms affected by externality, equilibrium will exist.

2) If the separable production function is linear in the output of other firms, so that marginal damage is constant, equilibrium will exist.

Acknowledgments

I am indebted to Joel Franklin, David Grether, and James Quirk for substantial mathematical assistance and moral support. Research support was provided by the Environmental Quality Laboratory of the California Institute of Technology.

FOOTNOTES

1. Wellisz, Starrett, and Inada and Kuga have also noticed that externalities imply nonconvexity. Starrett, and Inada and Kuga, however, consider a different type of nonconvexity. They examine a general equilibrium system in which a market exists for each externality, and demonstrate that when a firm ceases operation rather than suffer an externality, a nonconvexity can arise independent of separability. The nonconvexity considered in this paper arises even in a system in which no markets exist in which externalities are traded. We will examine the consequences of this nonconvexity for the existence of competitive equilibrium in an economy in which the trading of externalities is impossible. Wellisz comes closer to recognizing the nonconvexity we consider, since he points out that separable externalities impose fixed costs which may drive some firms out of business.

2. These propositions corect the claim by Wellisz [1964] that separable externalities do not affect resource allocation. It is true that the cost-function separability which he considers implies that the allocation of output is unaltered, but in general the allocation of input is affected by devices which "internalize" external effects.

3. See Montgomery [1975a].

4. There are two obvious ways of generalizing the form $F = A(X + h(Y_2))$ to the n-input case: $F = B(X_1 + h_1(Y_2), \ldots, X_n + h_n(Y_2))$ where B

and $h_1, \ldots, h_n$ are arbitrary functions, or $F = A(g(X_1, \ldots, X_n) + h(Y_2))$ where A and h are arbitrary functions and the Hessian determinant $|g_{ij}| = 0$. Both are solutions of (1). Though they bear some obvious resemblances to each other, it does not appear that either is the most general form of a solution of (1). It is known that any linear combination is also a solution (Ford [1955]).

5. Proved in Montgomery [1975a].

References

Arrow, K. J., and Hahn, F. [1971]. General Competitive Analysis. (San Francisco: Holden-Day, 1971).

Davis, O. A., and Whinston, A. [1962]. "Externalities, Welfare, and the Theory of Games." Journal of Political Economy, vol. 70 (June 1962).

Ford, Lester [1955]. Differential Equations. (New York, 1955).

Inada, K., and Kuga, K. [1973]. "Limitations of the 'Coase Theorem' on Liability Rules." Journal of Economic Theory, vol. 6 (December, 1973).

Kneese, S., and Bower, B. [1968]. Managing Water Quality: Economies, Technology, Institution. (Baltimore, 1968).

Marchand, J. R., and Russell, K. P. [1973]. "Externalities, Liability, Separability, and Resource Allocation." American Economic Review, vol. 63 (September 1973).

Montgomery, W. D. [1975a]. "Separability and Vanishing Externalities," California Institute of Technology Social Science Working Paper No. 74, 1975., forthcoming in American Economic Review, December 1975.

Montgomery, W. D. [1975b]. "Stability of Pure Trade Equilibrium with Externalities," California Institute of Technology Social Science Working Paper No. 76, 1975.

Rothenberg, J. [1960]. "Non-convexity, Aggregation, and Pareto Optimality." Journal of Political Economy, vol. 68, (October 1960).

Starrett, D. [1972]. "Fundamental Non-convexities in the Theory of Externalities." Journal of Economic Theory, vol. 4, (April 1972).

Wellisz, S. [1964]. "On External Diseconomies and the Government-Assisted Invisible Hand." Economica, vol. 31, (November 1964).