Working Paper 305

Economies of scale and the existence of supply-side equilibria in a production-constrained spatial interaction model

B. Harris*, J-M. Choukroun**
and A.G. Wilson†

*School of Public and Urban Policy, University of Pennsylvania.
**Wharton School, University of Pennsylvania.
†School of Geography, University of Leeds.

School of Geography
University of Leeds
LEEDS LS2 9JT
ENGLAND.

August 1981

CONTENTS

1. Introduction

In an earlier paper on the supply side of the production-constrained spatial interaction model, two of us showed how equilibrium patterns of the supply centres could be computed and we offered some mechanisms which gave some insights into the patterns which arose in different circumstances (Harris and Wilson, 1978). In this paper, we consider the problem of scale economies in such models - more precisely of *consumer* scale economies - through the influence of the α parameter in the W_j^{α} attractiveness term of the models. It has sometimes been argued that the $\alpha = 1$ case is basically the 'neutral' case and that for $\alpha < 1$ there are scale diseconomies and for $\alpha > 1$ scale economies. We show that the situation is more complicated than that.

It also turns out that the mechanisms used to tackle this issue can be used to throw new light on the nature of the supply-side equilibria which exist. In the Harris and Wilson (1978) paper, we concentrated mainly on the influence of parameters on the equilibrium pattern. Here, we explore the influence of the exogenous demand variables - e_iP_i in the original paper, Q_i here.

It is useful to explain at the outset how this paper originated and the contributions of the three authors. The original motivation came from one author (Harris) from the feeling that zonal attractiveness played a somewhat ambiguous role in the retail trade model. Two of the authors (Harris and Choukroun) accordingly started to investigate the 'economies of scale' problem and decided on an in-depth examination of a two-zone case. Harris worked out the original model for $\alpha = 1$ with a generalised measure of zonal separation and Choukroun generalised this for all α values. The third author (Wilson) showed how to generalise the analysis further to the multi-zonal case and added an exploration to his own relatively undeveloped ideas on the effect of Q_i on the equilibrium pattern (Wilson, 1981, pp. 132-5). It was possible, using the Harris-Choukroun analysis, to take these ideas much further.

In section 2, we present the analysis for the two-zone case in detail. This is extended to the general case in section 3 and there are brief concluding comments in section 4.

2. The analysis of the two-zone case

Take the standard retail trade model (Harris, 1965, Huff, 1964, Lakshmanan and Hansen, 1965) in the form

$$T_{ij} = A_i Q_i W_j^{\alpha} F_{ij} \tag{1}$$

where

$$A_i = 1/\sum_j W_j^{\alpha} F_{ij} \tag{2}$$

to ensure that

$$\sum_j T_{ij} = Q_i \tag{3}$$

T_{ij} is the flow of cash from residents in i to shops in j; Q_i is the total expenditure by residents of i; W_j is the size of the shopping centre at j and W_j^{α}, where α is a parameter, is taken as a measure of attractiveness; F_{ij} is a distance factor which is assumed to decline for increasing separation of i and j.

Total revenue at j, D_j, can be calculated from the model as

$$D_j = \sum_i T_{ij} \tag{4}$$

Thus, using (1)

$$D_j = W_j^{\alpha} \sum_i A_i Q_i F_{ij} \tag{5}$$

Harris and Wilson (1978) explored the consequences of imposing the equilibrium condition that, for each zone, revenue equals cost of supply. If, for simplicity, k is taken as the unit cost of supply, then this condition is

$$D_j = kW_j \tag{6}$$

A most complete analysis can be done for the two-zone case and so we proceed with this before tackling the general case.

Suppose, then, we have two zones only. Since proportionate (or scale) changes in W_j's or F_{ij}'s have no effect (as can be seen from equations (1) and (2)), without loss of generality, we can take

$$W_1 = W \tag{7}$$

$$W_2 = 1 - W \tag{8}$$

We also assume, with some loss of generality, that

$$F_{11} = F_{22} = 1 \tag{9}$$

and that we can take F_{12} such that

$$F_{21} = F_{12} = R < F_{11} \tag{10}$$

We can also take $k = 1$ without loss of generality, and this implies that

$$\sum_i Q_i = \sum_j D_j = \sum_j W_j \tag{11}$$

at equilibrium. Thus, we can take

$$Q_1 = Q \tag{12}$$

$$Q_2 = 1 - Q \tag{13}$$

The equilibrium conditions, (6), can now be written

$$D_1 = W \tag{14}$$

$$D_2 = 1 - W \tag{15}$$

Because of the symmetry in the two-zone case, we only need to work with one equation. Since

$$D_1 = T_{11} + T_{21} \tag{16}$$

we can see by using equation (1) (with the appropriate substitutions from equations (6)-(9)) that equation (13) becomes

$$W = A_1 Q W^{\alpha} + A_2 (1 - Q) W^{\alpha} R \tag{17}$$

Substituting for A_1 and A_2 gives

$$W = \frac{QW^{\alpha}}{W^{\alpha} + (1 - W)^{\alpha} R} + \frac{(1 - Q) W^{\alpha} R}{RW^{\alpha} + (1 - W)^{\alpha}} \tag{18}$$

Given Q, we could solve this for W to find the equilibrium position in the usual way. However, as explained earlier, our main objective here is to explore the functional relationship between Q and W. We can carry this argument further by re-arranging equation (18) so that it is 'solved' for Q. With some manipulation (and assuming $W \neq 0$ since $W = 0$ is always a solution, except for $\alpha = 1$), this generates

$$Q = \frac{[W^{\alpha} + R(1 - W)^{\alpha}][W^{1-\alpha} - R(1 - W)^{1-\alpha}]}{(1 - R^2)} \qquad (19)$$

It can now be seen that Q is a single-valued function of W; but not vice versa. In some cases, the Q-values are outside the range $0 < Q < 1$ and are therefore impermissible in real terms.

For curve sketching, and some later purposes, it is useful to multiply out the right hand side of (19);

$$Q = \frac{1}{(1 - R^2)} \{W - R^2(1 - W) + R(1 - W)^{\alpha}W^{1-\alpha} - RW^{\alpha}(1 - W)^{1-\alpha}\} \qquad (20)$$

which is

$$Q = \left[W \frac{1 + R^2}{1 - R^2} + \frac{R}{1 - R^2}(1 - W)^{\alpha}W^{1-\alpha} - \frac{R}{1 - R^2}W^{\alpha}(1 - W)^{1-\alpha} - \frac{R^2}{1 - R^2} \right] \qquad (21)$$

Note that $\frac{\partial Q}{\partial W}$ can be calculated as

$$\frac{\partial Q}{\partial W} = \frac{1 + R^2}{1 - R^2} + \frac{(1 - \alpha)R}{1 - R^2}[(1 - W^{\alpha})W^{-\alpha} + W^{\alpha}(1 - W)^{-\alpha}]$$
$$- \frac{\alpha R}{1 - R^2}[(1 - W)^{\alpha-1}W^{1-\alpha} + (1 - W)^{1-\alpha}W^{\alpha-1}] \qquad (22)$$

Armed with this basic information, we can now use (19) or (20) to construct plots of Q - W curves for different α-values. We proceed in a number of stages.

(i) It can be seen by inspection in (20) that all curves pass through the point $(\frac{1}{2},\frac{1}{2})$.

(ii) For three values of α, $Q(W)$ is straight line. We take each in turn.

$$\text{(a)} \quad \alpha = 0: \quad Q = \left(\frac{1 + R}{1 - R}\right) W - \frac{R}{1 - R} \qquad (23)$$

$$\text{(b)} \quad \alpha = \tfrac{1}{2}: \quad Q = \left(\frac{1 + R^2}{1 - R^2}\right) W - \frac{R^2}{1 - R^2} \qquad (24)$$

$$\text{(c)} \quad \alpha = 1: \quad Q = \left(\frac{1 - R}{1 + R}\right) W + \frac{R}{1 + R} \qquad (25)$$

Recall that R is F_{12} (or F_{21}) and lies between 0 and 1. These lines are then shown plotted on figure 1 for a typical R value (which is, in fact, $R = 0.548812$).

Note that the lines (and curves) are going to be symmetrical about lines parallel to the axes through the point $(\frac{1}{2},\frac{1}{2})$. The various intercepts are given in Table 1.

Table 1.

Label on figure	Q	W	α
A	0	$\frac{R}{1 + R}$	0
B	0	$\frac{R^2}{1 + R^2}$	$\frac{1}{2}$
C	$\frac{R}{1 + R}$	0	1
D	1	$\frac{1}{1 + R}$	0
E	1	$\frac{1}{1 + R^2}$	$\frac{1}{2}$
F	$\frac{1}{1 + R}$	1	1

It is a useful preliminary to note that (20) can be solved for $Q = 0$ so that these intercepts are known for any α (except $\alpha = 1$). They are:

$$Q = 0: \quad W = W_0 = (1 + R^{\frac{1}{\alpha-1}})^{-1} \qquad (26)$$

$$Q = 1: \quad W = W_1 = 1 - (1 + R^{\frac{1}{\alpha-1}})^{-1} \qquad (27)$$

(iii) <u>$Q = Q(W)$ curves for $0 < \alpha < \frac{1}{2}$</u>

In this range, $Q(W)$ is nearly linear and lies in the feasibility triangle PAB (cf. figure 1). All curves are concave to the origin. They are shown in the appropriate section of figure 2.

It is straightforward to see what happens to the curves when R changes. As R decreases and tends to zero (in the case of widely separated zones relative to intrazonal distances), the feasibility triangle PAB tends to collapse towards the line $Q = W$. That is, demand and supply tend to match in each zone.

The same effect would be produced by increasing the β parameter if F_{ij} took one of the usual forms, like $e^{-\beta c_{ij}}$ or $c_{ij}^{-\beta}$.

As R increases and tends to 1 (in the case of zones very close to each other, or a relatively low value of β - $\beta \rightarrow 0$ in the limit), the feasibility triangle tends to collapse into the vertical line $W = \frac{1}{2}$: the supply tends to be equally split between the two zones regardless of their respective levels of demand.

For all values of α in this range, the $Q(W)$ curve is below the line $Q = W$. This means that W is always larger than Q in $0 < Q < \frac{1}{2}$. Thus the zone with the smaller share of demand receives more than its 'fair share' of supply. If Q increases through $Q = \frac{1}{2}$, then a switch to the reverse situation arises at the point of inflection in the curve at P.

(iv) $\underline{\frac{1}{2} < \alpha < 1}$

For α in this range, the Q(W) curves are convex to the origin and lie in the triangle BCP. These are two subcases divided by

$$\alpha_1 = \frac{1 + R}{2} \tag{28}$$

The calculation of α_1 and the proof of corresponding results discussed below are given in the appendix. We consider each in turn.

(a) $\underline{\frac{1}{2} < \alpha < \alpha_1}$

Q(W) lies in the triangle BOP, and so $Q < W$ for all $W < \frac{1}{2}$ and the 'smaller-demand' zones continue to receive more than their fair share.

(b) $\underline{\alpha_1 < \alpha < 1}$

The Q(W) curve now intersects the $Q = W$ line at a point between (0,0) and $(\frac{1}{2},\frac{1}{2})$. Thus, for a portion of the Q-range, the 'small-demand' zone will receive more than its fair share of supply. As α gets close to 1, the Q(W) curve hugs the Q axis and the $\alpha = 1$ line (CP). The intercept (W_0) tends towards the origin as α increases and Q becomes greater than Q for all except very small values of W. Various examples are shown in figure 2.

(v) $\underline{\alpha = 1}$

We return to the straight line case here because it divides two very different families of curves. This provides a clear example of 'economies of scale' in operation. The zone with the larger share of demand always gets more than its fair share of supply. For example, if $F_{ij} = e^{-\beta c_{ij}}$ and $\beta = 0.1$ while $c_{12} - c_{11} = 2$, then $R = 0.83$, and if one zone falls to only 45% of the total demand, it will lose all its retail 'supply' to the other zone.

The provides the first example of a discontinuity and it is worth exploring the interpretation of this case in more detail. We are assuming that if Q changes, a new equilibrium will be worked out in the model and in the two-zone case, for $\alpha = 1$, this is wholly provided by the straight line CPF in figure 1. If for the small zone, Q is declining, it will be in the section CP of the curve.

When Q reaches C, W is zero. In this case the transition is a smooth one. In some of the $\alpha > 1$ cases, it is not, as we shall see.

(vi) $\underline{\alpha > 1}$

There are again two subcases, divided by

$$\alpha_2 = \frac{(1 + R)^2}{4R} \qquad (29)$$

The proof of this result and the computation of α_2 is given in the appendix. Again we consider each in turn.

(a) When $1 < \alpha < \alpha_2$

Equation (19) shows that $Q(W) \rightarrow \infty$ as $W \rightarrow 0$. As W increases, Q(W) decreases to a value $Q_{min}(W_{min})$. It then increases to $(\frac{1}{2},\frac{1}{2})$. Due to the symmetry, there is a coresponding maximum (Q_{max},W_{max}). A range of these curves are shown on figure 2 and a particular example shown in figure 3. For any Q in the range

$$Q_{min} < Q < Q_{max} \qquad (30)$$

there are three feasible values for W. In introducing the notation above, the min and max labels on the W variables simply denote W-values which correspond to the Q-values. However, it turns out that they are in fact the minimum and maximum values of W. If we move along the curve to the left of (Q_{min},W_{min}), then an increase in Q corresponds to a decrease in W; if we move to the right of W_{max}, a decrease in Q corresponds to an increase in W. This behaviour is obviously nonsensical and is shown by a dashed curve in figure 3.

This analysis is based on the equilibrium condition $D_1 = W$, and tests of stability depend on the sign of $D_1 - W$. Since we have re-arranged the equation to solve for Q, it is difficult to proceed in the usual way. It is only necessary, however, to observe that W and Q change monotonically, and this implies $\frac{\partial Q}{\partial W} > 0$. Thus only the positive-gradient part of a Q(W) curve can represent stable states, and these are indicated by the solid curve in figure 3.

This throws new light on the Harris and Wilson (1978) analysis. We obviously require Q to satisfy (30) for a non-zero W to exist. If Q moves outside that range, there is a jump to the other stable state which is always available: W = 0. This is an example of a fold catastrophe. More precisely this means that if

$$Q < Q_{min} \tag{31}$$

W jumps to zero and if

$$Q > Q_{max} \tag{32}$$

1 - W (ie. the other zone), jumps to zero.

The argument can be put in reverse: if all the supply is in one zone (say the '1 - W' zone), so W = 0 in the other, then this is an equilibrium no matter what Q is. If a developer now builds a shopping centre in the 'W' zone, it will only be viable if $Q > Q_{min}$.

The next step in the argument is to look at the effect of R. It can be shown that as R increases, Q_{min} and W_{min} both increase towards ½ (and recall that R increasing is equivalent to β decreasing so this makes sense in those terms too). When R decreases, W_{min} tends to zero, so that widely separated zones (or a high β system) have a broad range of feasible non-zero values for W and a Q(W) curve which is close to the line Q = W.

We can also interpret 'economies of scale' phenomena for this case. If W < ½, the Q(W) curve is always above the Q = W line which means, as we would expect, that demand exceeds supply in the smaller zone and economies of scale are in full effect in the larger zone.

(b) <u>$1 < \alpha_2 < \alpha$</u>

In this case, as shown by the final (upper) set of curves in figure 2, Q(W) is a monotonically decreasing function of W over the whole range, with the curve becoming increasingly steep as α increases in value. As in the previous subcase, when the curve has a negative gradient, the equilibrium states are unstable. This is to be interpreted by noting that W = 0, 1 - W = 1 (or vice versa) is the stable solution in this case. In other words, when $\alpha > \alpha_2$, there can only be one centre.

This concludes the analysis for the two-zone case. We can summarise the two kinds of conclusions we have obtained.

(A) Economies of scale

It is sometimes thought that $\alpha = 1$ is the 'neutral' case. It is easy to show that this is a mistaken idea.

If equation (5) is written out in full, substituting for A_i from (2), then it is easy to see that D_j, as a function of W_j, is *not* homogeneous of degree α in W_j. With W_j^α in the numerator, it is not surprising that it turns out that there is 'odd' behaviour with respect to scale economies for $\alpha < 1$. This can be seen particularly clearly in the two-zone case.

What we get from this argument is some clearer insight into how 'economies of scale' do operate in this model. We can summarise these conclusions as follows:

(i) When $\alpha \geqslant 1$, there are always economies of scale, which declines as $R \to 0$.

(ii) For $\alpha < \alpha_1 = (1 + R)/2$, there are always diseconomies of scale.

(iii) For $\alpha_1 < \alpha < 1$, there are economies of scale for some W values.

(iv) The nearest there is to a 'neutral' case is $\frac{1}{2} < \alpha < 1$, and more particularly $\alpha_1 < \alpha < 1$ with α nearer to α_1. There are variations with R, however, which we expect to become complicated in the multi-zone case.

In summary, the closer α approaches 1 (or in all cases when it exceeds it) or the higher the value of R (or the lower the value of β), the more economies of scale will dominate.

(B) The existence of equilibria and the fold catastrophe

Our results are in accord with those of Harris and Wilson (1978) but offer deeper insight into the cases where equilibria disappear. We can summarise our conclusions as follows.

(i) For $\alpha < 1$, there is always a single equilibrium value, W, for any Q.

(ii) For $1 < \alpha < \alpha_2$, and for $Q_{min} < Q < Q_{max}$, there is a stable equilibrium in which both zones have non-zero supply. For $\alpha > \alpha_2$, all the supply is concentrated in one zone. Since α_2, Q_{min} and Q_{max} are functions of α and R (and hence β), these conditions are more detailed than those offered previously. We can also see explicitly the equilibrium as a function of Q (cf. the argument on e_iP_i variation in Wilson, 1981, p. 132).

3. The extension to the general case

We now turn to the general case. Recall that the main Harris and Wilson (1978) task was to solve the equilibrium conditions (5). Using (5), these can be written

$$kW_j = W_j^\alpha \sum_i A_i Q_i F_{ij} \quad (33)$$

Except for $\alpha = 1$, $W_j = 0$ is always a possible solution. For $W_j \neq 0$, (33) can be written

$$kW_j^{1-\alpha} = \sum_i A_i Q_i F_{ij} \quad (34)$$

The $\{W_j\}$ set which constitutes the equilibrium pattern are calculated by solving equations (1)-(4) and (6) iteratively. This is equivalent to solving equations (34) iteratively with the substitution for A_i from (2). Algorithms for this calculation are offered in Harris and Wilson (1978). The W_j's which result can be seen as functions of all parameters and independent variables as well as the W_k for $k \neq j$. To make this explicit, we can write

$$W_j = W_j(Q_1, Q_2, \ldots, Q_N, \alpha, F_{11}, F_{12}, \ldots W_1, W_2, \ldots, W_{j-1}, W_{j+1}, \ldots W_N) \quad (35)$$

In the two zone example, we were able to explore the relationship between Q and W. In this section, we concentrate on studying the more general relationship between W_j and one of the Q_i's, say Q_i.

This extends our work with the two-zone model: first, it offers light on the nature of scale economies in the general case; and secondly, it gives deeper insight into the nature of the equilibrium W_j in a particular zone. This is especially useful when this is near to a critical condition. We can see the effect of a small change of a Q_i (with i possibly near to, possibly distant from j) on the equilibrium value of W_j near to criticality.

Although we want to examine W_j as a function of Q_i, it turns out that Q_i is, formally, as in the two-zone case, a single-valued function of W_j, but not vice versa; so we 'solve' (34) for Q_i, and then we can explore the resulting relationship geometrically and exploit the results in the reverse direction. From (34), we obtain

$$kW_j^{1-\alpha} = A_i Q_i F_{ij} + \sum_{m \neq i} A_m Q_m F_{mj} \tag{36}$$

Thus

$$Q_i = \frac{1}{A_i F_{ij}} \left[kW_j^{1-\alpha} - \sum_{m \neq i} A_m Q_m F_{mj}\right] \tag{37}$$

If we now substitute for A_i (and the A_m's) using (2), we get

$$Q_i = \frac{(W_j^{\alpha} F_{ij} + \sum_{n \neq j} W_n^{\alpha} F_{in})}{F_{ij}} \left(kW_j^{1-\alpha} - \sum_{m \neq i} \frac{Q_m F_{mj}}{\sum_n W_n F_{mn}}\right) \tag{38}$$

where we distinguish the W_j term in A_i but not in the A_m's.

Put

$$K_{ij}(\alpha) = \sum_{m \neq i} \frac{Q_m F_{mj}}{\sum_n W_n^{\alpha} F_{mn}} \tag{39}$$

where, for later use, we have shown K_{ij} as a function of α.

We can also put

$$C_i(\alpha) = \sum_{n \neq j} W_n^{\alpha} F_{in}$$

which is again a function of α, but this time not of W_j. Then (38) can be written

$$Q_i = (W_j^\alpha + \frac{C_i(\alpha)}{F_{ij}})(kW_j^{1-\alpha} - K_{ij}(\alpha)) \quad (41)$$

For later convenience, this can also be multiplied out as

$$Q_i = kW_j + \frac{kC_i(\alpha)}{F_{ij}} W_j^{1-\alpha} - K_{ij}(\alpha)W_j^\alpha - \frac{C_i(\alpha)K_{ij}(\alpha)}{F_{ij}} \quad (42)$$

In the general case, we wish to explore the geometrical relationship between Q_i and W_j for a range of values of α.

At the outset, we note a general difficulty with this kind of analysis first identified in another context by Wilson and Clarke (1979). That is, when curves of this kind are plotted, an assumption has to be made about the values of $\{W_k\}$, $k \neq j$ (and in this example of $\{Q_k\}$, $k \neq i$). In particular, it is necessary to decide whether it is most meaningful in a particular case to keep $\{W_k\}$, $k \neq j$, fixed (which means that $W = \sum_\ell W_\ell$ *varies*, because $\sum_{\ell \neq j} W_\ell$ is fixed, but W_j varies) or whether to vary the backcloth in some specific way but such that W is fixed. We refer to these as the W-fixed and W-varying cases respectively. Correspondingly, there are Q-fixed and Q-varying cases (which makes four possible cases in all). In the two-zone case, we focussed on the W-fixed, Q-fixed case, partly because it has an obvious intuitive interpretation, and partly because it could be analysed very neatly. (We have studied the alternative three cases numerically and fortunately the overall character of the two-zone analysis above is not altered.) In the multi-zone case, it makes more intuitive sense to adopt the W-varying, Q-varying assumption and so we use this for our main results.

We now return to equation (42). We still wish to explore the functional relationship between Q_i and W_j for a range of values of α but because we cannot now make assumptions like (7), (8) and (13), the problem is analytically more difficult. Indeed, it is not really possible to proceed with a completely general analysis.

We noted earlier the problem of specifying $\{Q_k\}$, $k \neq i$ and $\{W_k\}$, $k \neq j$ to be used in (42). If $Q = \sum_i Q_i$ and $Q = \sum_j W_j$ are each considered to be fixed, then this can only be achieved by factoring Q_k (or W_k) downwards as Q_i or (W_j) increases, and vice versa. This leads to problems at the 'end' points, in particular ($W_j = 0$, $W_j = W$) and we concluded that it was best to use the Q-varying, W-varying assumption in which $\{Q_k\}$, $k \neq i$ and $\{W_k\}$, $k \neq j$ are considered to have fixed values throughout the $Q_i - W_j$ calculation. The reader is referred to Wilson and Clarke (1979) for methods of proceeding with alternative assumptions.

In the argument below, we give general indications of the form of the $Q_i - W_j$ relationships by examining the four terms of (42), particularly the last three which are nonlinear in W_j. We give some substance to our conjectures by presenting computations from a four-zone example. The data for this is presented in table 2. Q_1 is plotted against W_2 in all the examples presented below.

<u>Table 2.</u> <u>Data for 4-zone example</u>

$Q_2 = 200$, $Q_3 = 45$, $Q_4 = 100$
$W_1 = 100$, $W_3 = 25$, $W_4 = 500$

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 & 2.5 \\ 3 & 1 & 3 & 2.5 \\ 3 & 3 & 1 & 2.5 \\ 2.5 & 2.5 & 2.5 & 0.5 \end{bmatrix}$$

$F_{ij} = e^{-\beta c_{ij}}$; $\beta = 0.25$; $k = 1$

We can now proceed by examining the separate components of the right hand side of equation (42).

(i) kW_j. This is obviously linear in W_j for all α. It is shown on a scale suitable for 'adding' to the other terms in figure 4.

(ii) $\dfrac{kC_i(\alpha)W_j^{1-\alpha}}{F_{ij}}$. We distinguish two cases:

(a) $0 < \alpha < 1$. In this range, $C_i(\alpha)$ will increase at a modest rate with α. $W_j^{1-\alpha}$ is nearly linear at α near to zero, and nearly constant as α approaches 1. The derivative of the term with respect to W_j is

$\dfrac{kC_i(\alpha)(1 - \alpha)W_j^{-\alpha}}{F_{ij}}$, showing that for $\alpha > 0$ (and $\neq 1$) it is infinite at the origin. This is borne out by the example shown in figure 5(a).

(b) $\alpha > 1$. $C_i(\alpha)$ increases more rapidly with α. $W_j^{1-\alpha}$ decreases with W_j, increasing rapidly so for higher α. The term and the gradient tend to zero for large W_j. An example is shown in figure 5(b).

(iii) $-K_{ij}(\alpha)W_j^{\alpha}$. $K_{ij}(\alpha)$ decreases with increasing α. The function declines with increasing W_j, increasing rapidly so for increasing α. Plots for the example are shown in figures 6(a) and 6(b) for $\alpha < 1$ and $\alpha > 1$ separately.

(iv) $\dfrac{-C_i(\alpha)K_{ij}(\alpha)}{F_{ij}}$. The variation of this term with W_j and α is shown, for the example, in figure 7.

As with the two-zone case, straight lines provide a framework for the Q_i - W_j analysis. $\alpha = 0$ and $\alpha = 1$ each produce straight lines, but in this case $\alpha = \frac{1}{2}$ does not. From (42), we obtain:

<u>$\alpha = 0$</u>

$$Q_i = k\left[1 + \frac{C_i(0)}{F_{ij}}\right]W_j - K_{ij}(0)\left[1 + \frac{C_i(0)}{F_{ij}}\right] \quad (43)$$

$$= \left[kW_j - K_{ij}(0)\right]\left[1 + \frac{C_i(0)}{F_{ij}}\right] \quad (44)$$

<u>$\alpha = 1$</u>

$$Q_i = \left[k - K_{ij}(1) \right] W_j + \frac{C_i(1)}{F_{ij}} \left[k - K_{ij}(1) \right] \quad (45)$$

$$= \left[W_j + \frac{C_i(1)}{F_{ij}} \right] \left[k - K_{ij}(1) \right] \quad (46)$$

It is useful to find the point of intersection by solving equations (43) and (45). This leads to:

$$\left[\frac{kC_i(0)}{F_{ij}} + K_{ij}(1) \right] W_j = \left[\frac{kC_i(1)}{F_{ij}} + K_{ij}(0) + \frac{K_{ij}(0)C_i(0)}{F_{ij}} - \frac{K_{ij}(1)C_i(1)}{F_{ij}} \right] \quad (47)$$

which can be solved for W_j to give, say, W_j^*. (42) can then be solved for the corresponding Q_i, say Q_i^*.

With this framework, we can now explore the $Q_i - W_j$ relationship for other values of α. As usual, we distinguish a number of cases. To fix ideas, the curves for the $Q_1 - W_2$ relationship in the four-zone example are presented in figure 8. These are obtained by summing the curves in figures 4-7.

It is useful to observed immediately that the curves have the same broad form as for the two-zone case, and we can conjecture that this is generally true. (The only oddity is that, for $\alpha < 1$, Q_i is negative for small W_j. However, this can be better stated as W_j being greater than some non-zero value - approximately 100 here for $Q_i = 0$. This will arise from contributions from other Q_k's to this W_j for this particular example.) Now, of course, the curves no longer intersect at a single point, but the overall shapes for different α values, are similar in character to the two-zone case. It is also clear from figure 8(b) that there will be a value of α (> 1), say $\alpha_{2_{ij}}$, such that for $\alpha > \alpha_{2_{ij}}$, there is no 'dip' in the first part of the curve and therefore no part of the curve with a positive gradient. In the example, this occurs at about $\alpha = 1.5$.

Since the shapes of the curves are similar to the two-zone case, this means that the 'economies-of-scale' argument is broadly similar. The results on the existence and stability of equilibria are also similar with one important exception which arises because of the different nature of the multi-zone case. The exception arises because of the difference between figures 2(b) and 8(b) - the two-zone and four-zone examples for $\alpha > 1$. In the two-zone case, when Q passes Q_{max}, the curve turns down again. In the four-zone example, with W_2 varying and W_1, W_3 and W_4 fixed (and hence W varying), for large W_2 and Q_1, there is an almost linear 45^0 relationship between them. This makes sense because with other Q_i's and W_j's fixed, increases in Q_i will nearly all go to j when W_j is so large. This means there is no Q_i^{max}. With this preliminary, we can now proceed with the interpretation. We state the results, as conjectures, for any Q_i in relation to any W_j. The evidence is in figure 8 for the four-zone example.

For $\alpha < 1$, for any Q_i there is one corresponding stable W_j. That is, the intersection of a Q_i = constant line with the $Q_i(W_j)$ curve is unique and is at a positive gradient of the $Q_i(W_j)$ curve. For $\alpha > 1$, there appear to be two cases as before. For $1 < \alpha < \alpha_{2ij}$, for $Q_i > Q_{min\ i}$ (and $W_j > W_{min\ j}$) there is a unique intersection of Q_i = constant with a positive gradient curve. This means there will be a stable non-zero W_j in these circumstances (and the 'circumstances' include the Q_k, $k \neq i$ and W_k, $k \neq j$). If $Q_i < Q_{min\ i}$, then there will be no intersection, and this probably implies that this W_j is zero. If Q_i is increased (and everything else remains fixed) then development becomes possible when Q_i exceeds $Q_{min\ i}$.

For $\alpha > \alpha_{2ij}$, there will always be one intersection of Q_i = constant with the $Q_i(W_j)$ curve but at a negative gradient. This, as in the two-zone case, is interpreted as an unstable state, and so W_j will be zero in these circumstances. This squares with our intuition about high -α cases. There will be relatively few centres. For most zones, when α_{2ij} is calculated, α will exceed it.

In the two-zone case, we were able to examine the influence of R and hence, among other things, the β-parameters. In the multi-zone case, this is more complicated, but could be accomplished through analysis of changing F_{ij} factors which have been retained in explicit form throughout. It would also be possible to analyse the effect of k-variation.

4. Concluding comments

In this paper we have tackled two problems of interest associated with the production-constrained spatial interaction model. First, we have found a way of understanding the representation of consumer scale economies in the model more directly. This turns on the construction of Q_i - W_j curves, particularly the Q - W curve in the two-zone case. This construction is based on an algebraic trick - a re-arrangement of the equilibrium equation which enables the curves to be plotted. This enables us, secondly, to gain new insight into the nature of the W_j-equilibrium. As in the analysis in the Harris and Wilson (1978) paper, the insights arise from the presentation of the geometry.

We emphasised in the paper (in equation (35)), as we have elsewhere, the multi-dimensional nature of the problem. In this paper, as in the Harris and Wilson (1978) paper, we focus on two state variables and investigate their relationship as a function of one parameter. It will be by further expansions of this kind of method - choosing to focus on a small number of variables and parameters and using algebraic and geometrical insights - that we expect further progress to be made.

Acknowledgement

The authors are very grateful to Margaret Spowage of the School of Geography in the University of Leeds for writing and implementing the computer programmes for the various numerical experiments which produced the graphs shown in the various figures.

References

Harris, B. (1965) A model of locational equilibrium for the retail trade, mimeo, Institute for Urban Studies, University of Pennsylvania.

Harris, B. and Wilson, A.G. (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models, *Environment and Planning A, 10,* 371-88.

Huff, D.L. (1964) Defining and estimating a trading area, *Journal of Marketing, 28,* 34-8.

Lakshmanan, T.R. and Hansen, W.G. (1965) A retail market potential model, *Journal of the American Institute of Planners, 31,* 134-43.

Wilson, A.G. (1981) *Catastrophe theory and bifurcation: applications to urban and regional systems,* Croom Helm, London; University of California Press, Berkeley.

Wilson, A.G. and Clarke, M. (1979) Some illustrations of catastrophe theory applied to urban retailing structures, in Breheny, M. (ed.) *Developments in urban and regional analysis,* Pion, London, 5-77.

APPENDIX

Calculation of α_1 and α_2

α_1 is the value of α for which the Q(W) curve is tangent to $Q = W$ at $(\frac{1}{2},\frac{1}{2})$. α_2 is the value of α for which the Q(W) curve is tangent to $Q = \frac{1}{2}$ at $(\frac{1}{2},\frac{1}{2})$. We have:

$$\left.\frac{dQ}{dW}\right|_{\frac{1}{2},\frac{1}{2}} = \frac{1 + R^2 + 2R(1 - 2\alpha)}{1 - R^2} \tag{A1}$$

Hence:

$$\alpha_1 \text{ is the solution for } \left.\frac{dQ}{dW}\right|_{\frac{1}{2},\frac{1}{2}} = 1 \tag{A2}$$

Or:

$$1 + R^2 + 2R(1 - 2\alpha_1) = 1 - R^2 \tag{A3}$$

which implies:

$$\alpha_1 = \frac{1 + R}{2} \tag{A4}$$

α_2 is the solution for

$$\left.\frac{dQ}{dW}\right|_{\frac{1}{2},\frac{1}{2}} = 0 \tag{A5}$$

Or:

$$1 + R^2 + 2R - 4R\alpha_2 = 0 \tag{A6}$$

which gives:

$$\alpha_2 = \frac{(1 + R)^2}{4R} \tag{A7}$$

Figure 1.

2 zone case : $\alpha = 0$, $\alpha = \frac{1}{2}$, $\alpha = 1$ straight lines

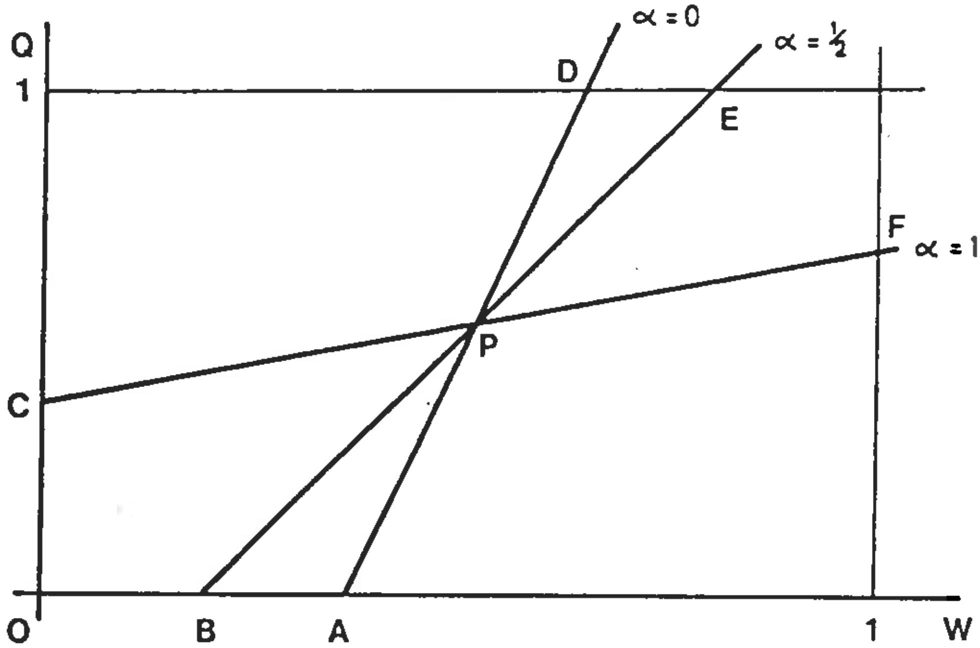


Figure 2. 2-zone case, W fixed; Q fixed

$K = 1$, $\beta = 0.25$

$c_{11} = 0.0$, $c_{12} = 2.4$

$c_{21} = 2.4$, $c_{22} = 0.0$

$F_{ij} = e^{-\beta c_{ij}}$

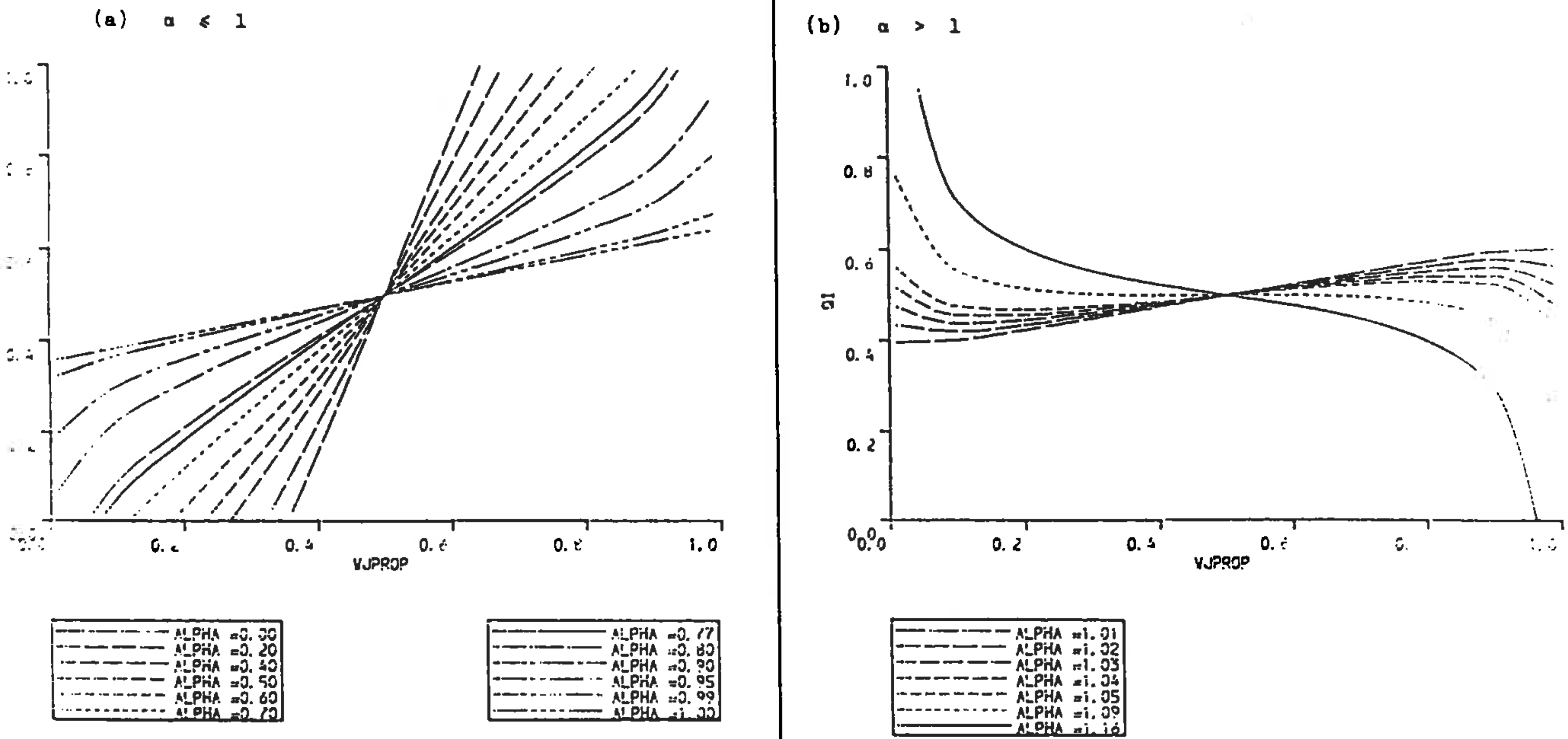


Figure 3.

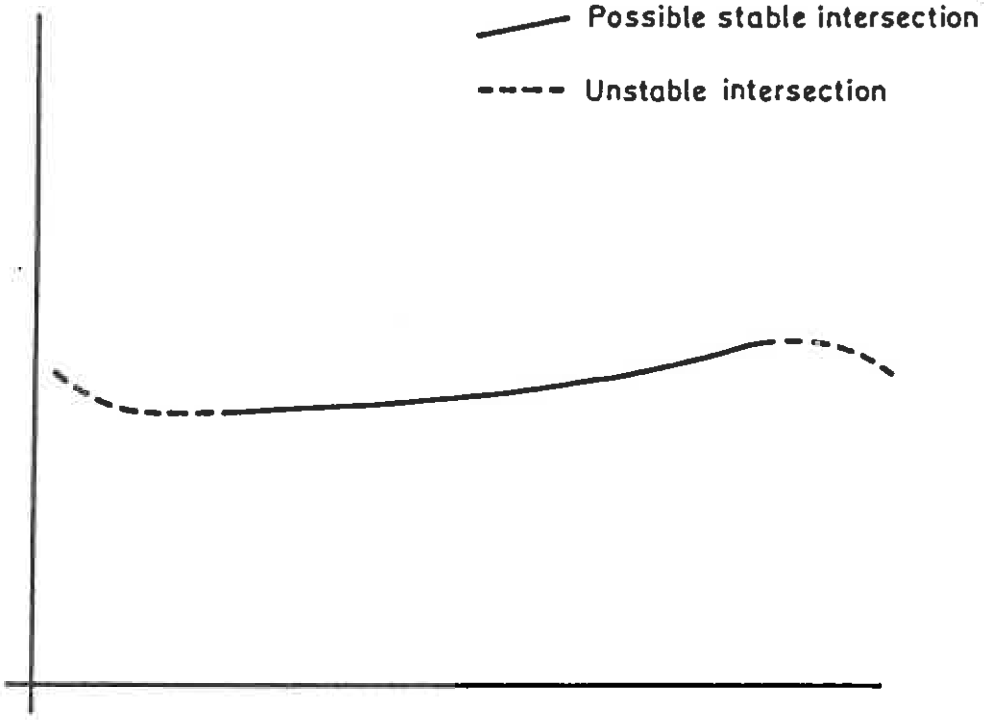


Figure 4. 4-zone case. kW_j term

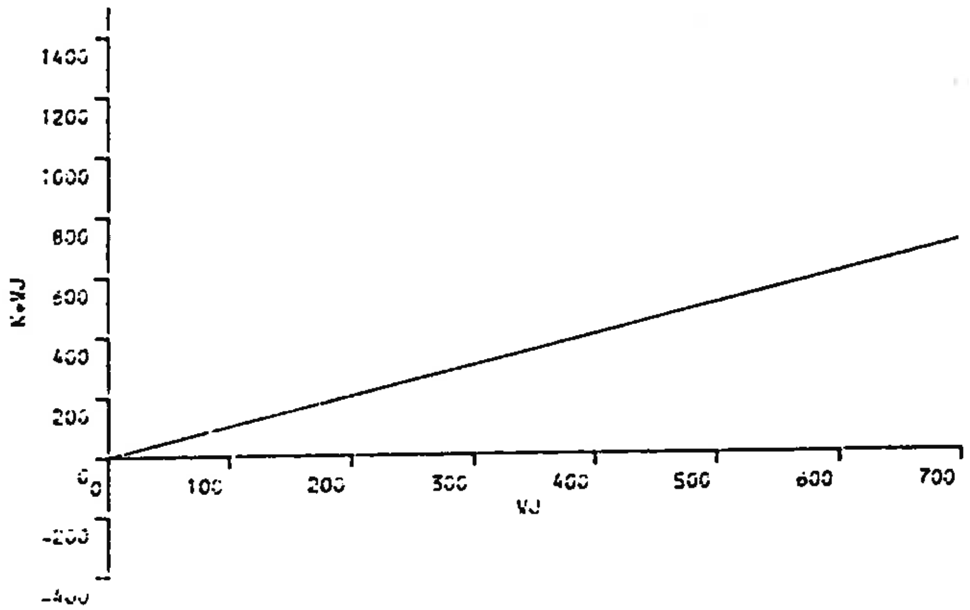


Figure 5. 4-zone case. $kC_i(\alpha)W_j^{1-\alpha}/F_{ij}$

(a) $\alpha \leqslant 1$

(b) $\alpha > 1$

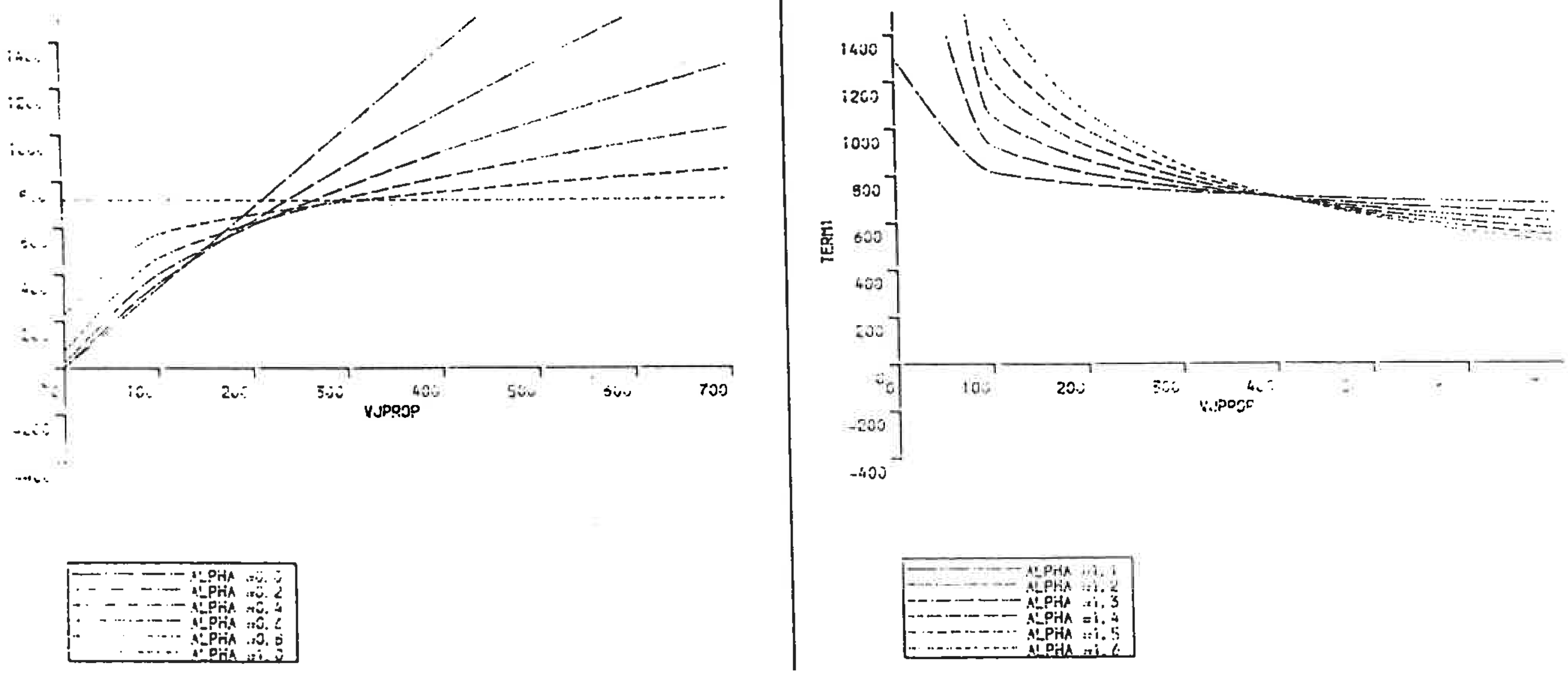
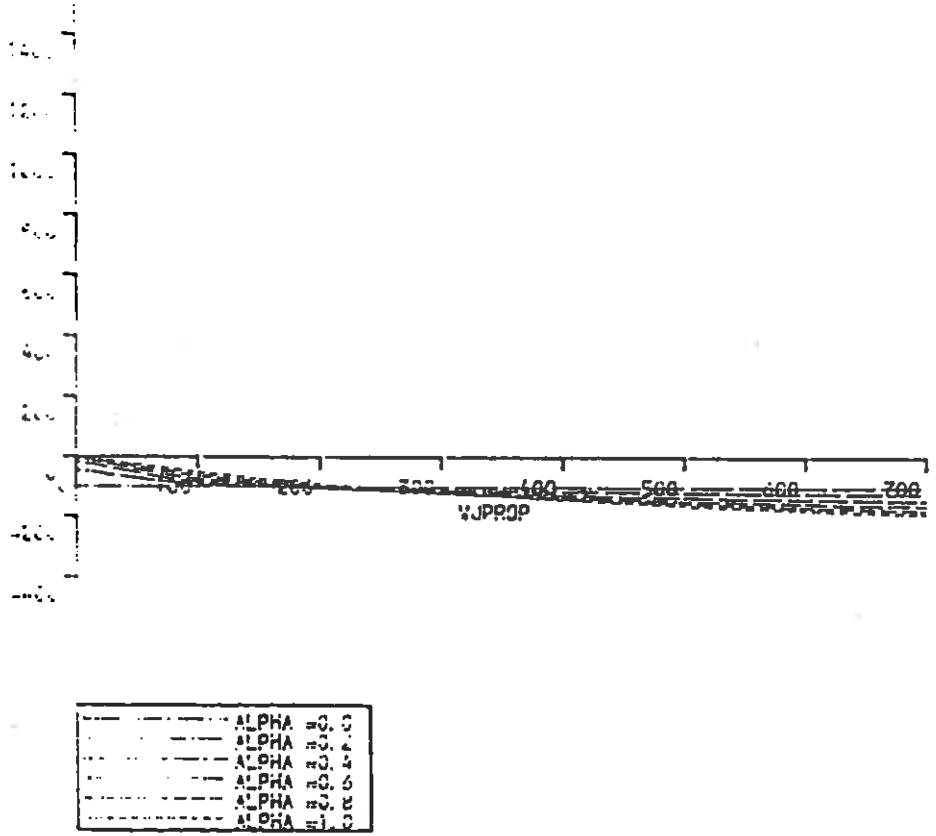


Figure 6. $4\text{-zone case} - K_{ij}(\alpha)W_j^{\alpha}$

(a) $\alpha \leqslant 1$



(b) $\alpha > 1$

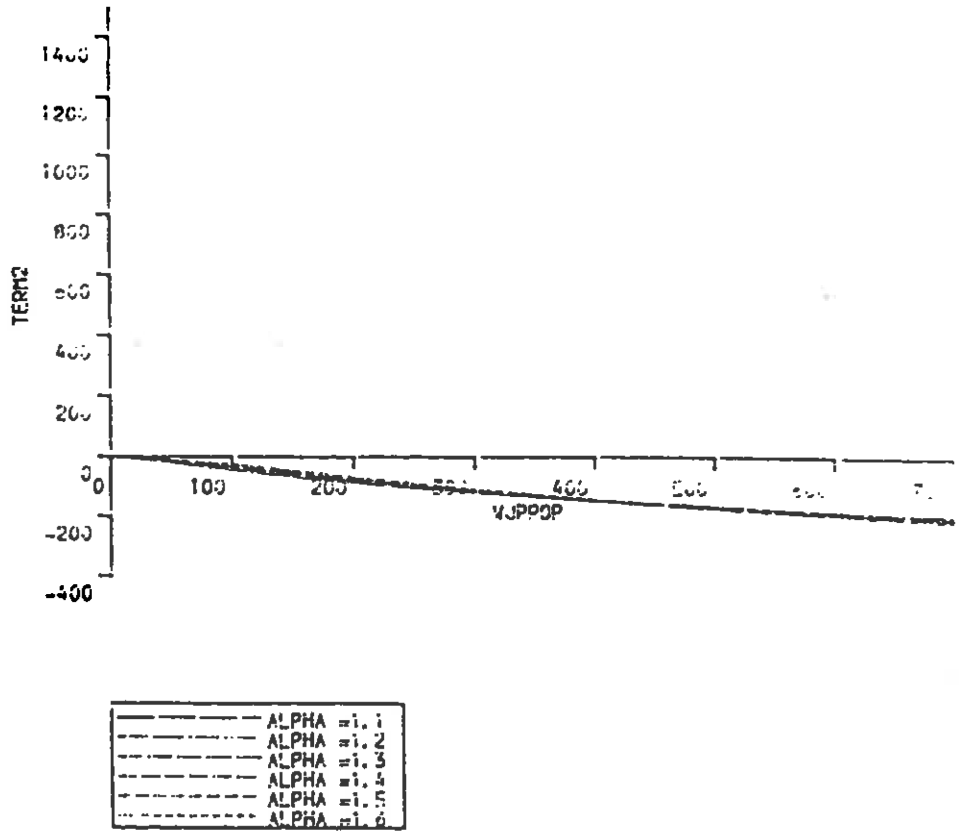
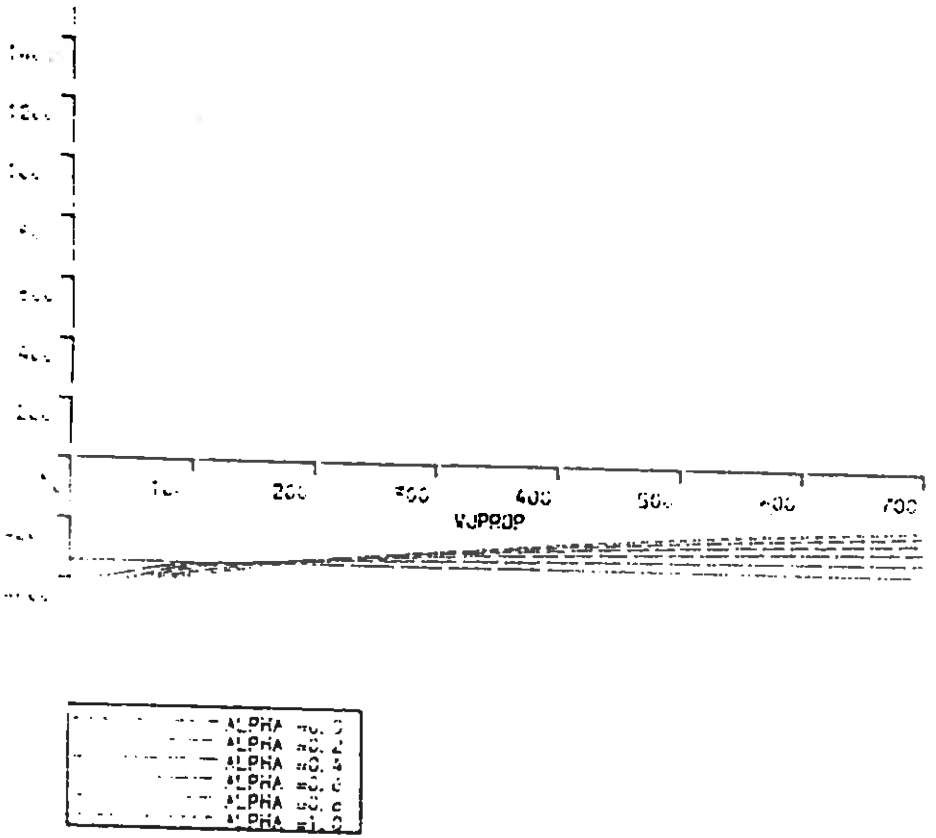


Figure 7. 4-zone case - $C_i(\alpha)K_{ij}(\alpha)/F_{ij}$

(a) $\alpha \leqslant 1$



(b) $\alpha > 1$

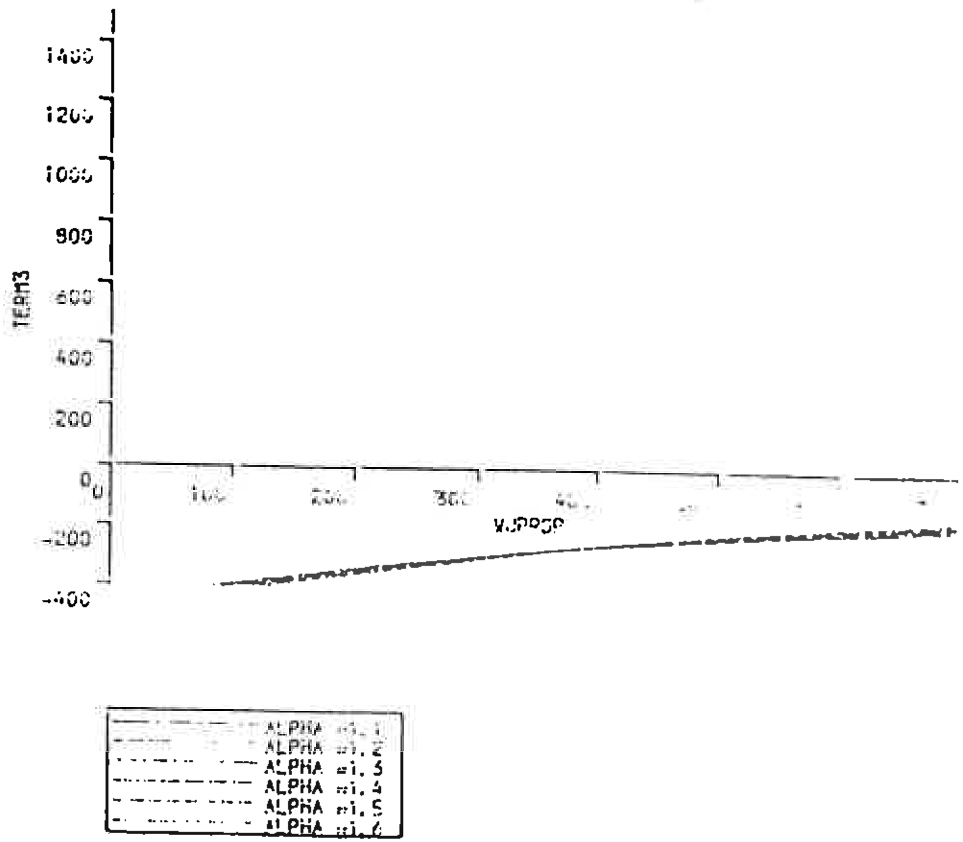


Figure 8. 4-zone case: $Q_i - W_j$ (Sum of figures 4-7)