Chapter 1

Industrial Location: The Location of the Firm in Theory

1.1 Introduction to Classical and Neoclassical Models of Location

The level of output and activity of an area depends on the total quantities of factor inputs employed in the area, and the wealth of an area depends on the total payments received by those factors. Observation suggests that some regions exhibit dense concentrations of factors, with large numbers of people and investment located in the same area, whereas other regions exhibit sparse populations and low levels of investment. At the same time, observation also suggests that people are paid different wages in different areas, while land prices vary significantly between locations. Therefore, in order to understand the economic performance of a region it is necessary to understand why particular quantities of factors are employed in that area, and why the factors there earn the particular rewards that they do.

Production factor inputs are usually defined in terms of three broad types, namely capital, labour, and land, and the factor payments earned by these factors in the production process are profits, wages, and rents, respectively. In some analyses of the production process, additional factor inputs are also identified such as entrepreneurship and technology. However, in our initial discussion of the causes and reasons for particular types of industrial location behaviour, we will not initially distinguish these additional factors from the broad factor groups. We include entrepreneurship in our description of labour, and technology in our description of capital. Later in our discussion of the causes and reasons for particular types of industrial location behaviour, we will also investigate the additional issues associated with entrepreneurship and technology. In this chapter we will concentrate on the determinants of spatial variations in capital investment, and in later sections of the book we will focus on spatial variations in labour stocks, and variations in land prices.

We start our analysis by asking the question—what determines the level and type of capital invested in a particular region? When talking about capital, our most basic unit of microeconomic analysis is the capital embodied in the firm. In order to understand the level of capital investment in an area it is necessary to ask why particular firms are located there and why the particular levels and types of investment in the area are as they are. These are the questions addressed by industrial location theory. We begin by discussing three classical and neoclassical models of industrial location behaviour, namely the Weber model, the Moses model, and the Hotelling model. Each of these models provides us with different insights into the fundamental reasons for, and the consequences of, industrial location behaviour. After analysing each of these models in detail, we will discuss two alternative approaches to analysing industrial location behaviour, namely the behavioural approach and the evolutionary approach. A broad understanding of these various approaches to industrial location behaviour will then allow us to discuss the concept of agglomeration economies.

1.2 The Weber Location-Production Model

triângulo locacional

proporção firm

Our starting point is to adopt the approach to industrial locational analysis originally derived from the nineteenth-century German mathematician Laundhart (1885), but which was formalized and publicized beyond Germany by Alfred Weber (1909). For our analysis to proceed we assume that the firm is defined at a point in space; the firm is therefore viewed as a single establishment. We also adopt the standard microeconomic assumption that the firm aims to maximize its profits. Assuming the profit-maximizing rationale for the firm, the question of where a firm will locate therefore becomes the question of at which location a firm will maximize its profits. In order to answer this question we will begin by using the simplest two-dimensional spatial figure, namely a triangle. This very simple type of two-dimensional approach will subsequently be extended to more general spatial forms.

The model described by Figure 1.1 is often described as a Weber location-production triangle, in which case the firm consumes two inputs in order to produce a single output.

Notation for use with Figures 1.1 to 1.12:

m_1, m_2	weight (tonnes) of material of input goods 1 and 2 consumed by the firm
m_3	weight of output good 3 produced by the firm
p_1, p_2	prices per tonne of the input goods 1 and 2 at their points of production
p_3	price per tonne of the output good 3 at the market location
M_1, M_2	production locations of input goods 1 and 2
M_3	market location for the output good 3
t_1, t_2	transport rates per tonne-mile (or per ton-kilometre) for hauling input goods 1 and 2
t_3	transport rates per tonne-mile (or per tonne-kilometre) for hauling output goods 3
K	the location of the firm.

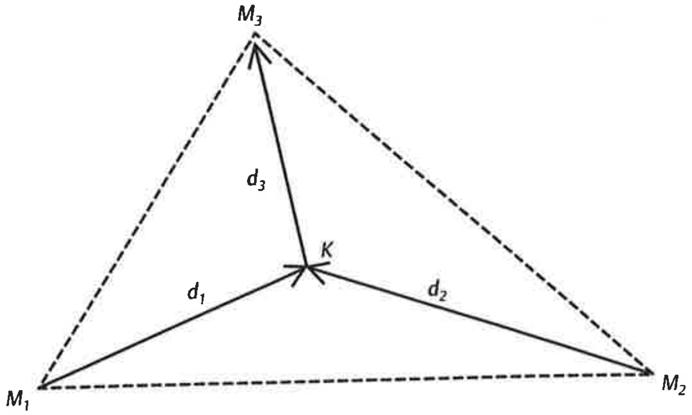

FIG. 1.1 Weber location-production triangle

We assume that the firm consumes material inputs 1 and 2, which are then combined by the firm in order to produce an output commodity 3. In the Weber location-production model, we assume that the coefficients of production are fixed, in that there is a fixed relationship between the quantities of each input required in order to produce a single unit of the output. Our production function therefore takes the general form:

$$m_3 = f(k_1\, m_1,\, k_2\, m_2). \tag{1.1}$$

In the very simplest case $k_1 = k_2 = 1$, in which case our production function becomes

$$m_3 = f(m_1, m_2). \tag{1.2}$$

This represents a situation where the quantity of the output good 3 produced is equal to the combined weight of the inputs 1 and 2. In other words for the purposes of our analysis here, we can rewrite (1.2) as

$$m_3 = m_1 + m_2. \tag{1.3}$$

The production locations of the input sources of 1 and 2, defined as M_1 and M_2, are given, as is the location of the output market M_3, at which output good 3 is sold. The prices per ton of the inputs 1 and 2 are given as p_1 and p_2, at the points of production M_1 and M_2, respectively. The price per tonne of the output good 3 at the market location M_3, is given as p_3. As such, the firm is a price taker. Moreover, we assume that the firm is able to sell unlimited quantities of output 3 at the given price p_3, as in perfect competition. The transport rates are given as t_1, t_2, and t_3, and these transport rates represent the costs of transporting 1 tonne of each commodity 1, 2, and 3, respectively, over 1 mile or 1 kilometre. Finally, the distances d_1, d_2, and d_3, represent the distances over which each of the goods 1, 2, and 3 are shipped.

We also assume that the input production factors of labour and capital are freely available everywhere at factor prices and qualities that do not change with location, and that land is homogeneous. In other words, the price and quality of labour is assumed to be equal everywhere, as is the cost and quality of capital, and the quality and rental price of land. However, there is no reason to suppose that the prices of labour, capital, and land are equal to each other. We simply assume that all locations exhibit the same attributes in terms of their production factor availability. Space is therefore assumed to be homogeneous.

If the firm is able to locate anywhere, then assuming the firm is rational, the firm will locate at whichever location it can earn maximum profits. Given that the prices of all the input and output goods are exogenously set, and the prices of production factors are invariant with respect to space, the only issue which will alter the relative profitability of different locations is the distance of any particular location from the input source and output market points. The reason for this is that different locations will incur different costs of transporting inputs from their production points to the location of the firm, and outputs from the location of the firm to the market point.

If the price per unit of output p_3 is fixed, the location that ensures maximum profits are earned by the firm is the location at which the total input plus output transport costs are minimized, *ceteris paribus*. This is known as the *Weber optimum location*. Finding the Weber optimum location involves comparing the relative total input plus output transport costs at each location. The Weber optimum location will be the particular location at which the sum *(TC)* of these costs is minimized. The cost condition that determines the Weber optimum location can be described as

$$TC = Min \sum_{i=1}^{3} m_i t_i d_i, \tag{1.4}$$

where the subscript i refers to the particular weights, transport rates, and distances over which goods are shipped to and from each location point K. With actual values corresponding to each of the spatial and non-spatial parameters, it is possible to calculate the total production plus transportation costs incurred by the firm associated with being at any arbitrary location K. Given our assumptions that the firm will behave so as to maximize its profits, the minimum cost location will be the actual chosen location of the firm.

In his original analysis Weber characterized the problem of the optimum location in terms of a mechanical analogy. He described a two-dimensional triangular system of pulleys with weights called a Varignon Frame. In this system, the locations of the pulleys reflect the locations of input source and output market points, and the weights attached to each string passing over each of the pulleys corresponds to the transport costs associated with each shipment. The point at which the strings are all knotted together represents the location of the firm. In some cases, the knot will settle at a location inside the triangle, whereas in other cases the knot will settle at one of the corners. This suggests that the optimum location will sometimes be inside the Weber triangle, whereas in other cases the optimum location will be at one of the corners. Nowadays, rather than using such mechanical devices, the optimum location can be calculated using computers. However, although it is always possible to calculate the optimum location of the firm in each particular case, of interest to us here is to understand how the location of the Weber optimum will itself be affected by the levels of, and changes in, any of the parameters described above. In order to explain this, we adopt a hypothetical example.

1.2.1 The location effect of input transport costs

Let us imagine that Figure 1.1 represents a firm that produces automobiles from inputs of steel and plastic. The output good 3 is defined as automobiles and these are sold at the market point M_3. We can assume that input 1 is steel and input 2 is plastic, and these are produced at locations M_1 and M_2, respectively. If the firm produces a car weighing 2 tonnes from 1 tonne of steel and 1 tonne of plastic, and the fixed transport rate for steel t_1 is half that for plastic t_2 (given that plastic is much less dense than steel, and transport rates are normally charged with respect to product bulk), the firm will locate relatively close to the source of the plastic production. In other words, the firm will locate close to M_2. The reason is that the firm will wish to reduce the higher total transport costs associated with shipping plastic inputs relative to steel inputs, *ceteris paribus*. The firm can do this by reducing the value of d_2 relative to d_1. On the other hand, if the firm had a different production function, such that it produces a car weighing 2 tons from 1.5 tonnes of steel and 0.5 tonnes of plastic, then even with the same values for the fixed transport rates t_1 and t_2 as in the previous case, the firm will now be incurring higher total transport costs associated with steel shipments, *ceteris paribus*. The reason for this is that although plastic is twice as expensive to ship per kilometre as steel, the total quantity of steel being shipped is three times that of plastic. The result is that the firm can reduce its total input transport costs by reducing the value of d_1 relative to d_2. The optimum location of the firm will now tend towards the location of production for the steel input M_1.

Within this Weber framework, we can compare the effects of different production function relationships on the location behaviour of the firm. For example, we can imagine that the two types of production function relationships described above—one which is relatively plastic intensive, and one which is relatively steel intensive—actually refer to the different production functions exhibited by two different competing automobile producers. Firm *A* exhibits the plastic-intensive production function, and firm *B* exhibits the steel-intensive production function. As we see in Figure 1.2, from the argument above we know that firm *A* will locate relatively close to M_2, the source of plastic,

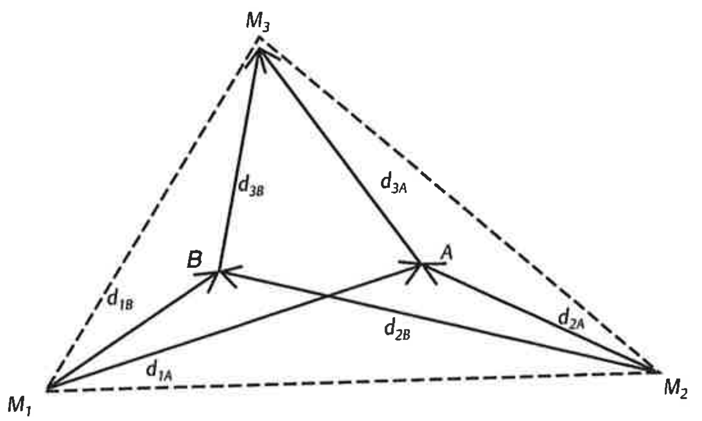

FIG. 1.2 Relative input transport costs and location

while firm B will locate relatively close to M_1, the source of steel. This is because, if we were to consider the case where steel and plastic inputs were shipped over identical distances, i.e. $d_{1A} = d_{2A}$, for firm A the total transport costs associated with plastic transportation would be greater than those associated with steel transportation. It therefore has an incentive to reduce the higher costs associated with plastic shipments by reducing d_{2A} and increasing d_{1B}. Alternatively, for firm B, for identical input shipment distances, i.e. $d_{1B} = d_{2B}$, the total transport costs associated with steel transportation would be greater than those associated with plastic transportation. It therefore has an incentive to reduce the higher transport costs associated with steel by reducing d_{1B} and increasing d_{2B}.

1.2.2 The location effect of output transport costs

Until now we have only considered the transport cost pull of the input sources on the location decision of the firm. However, the market itself will display a pull effect on the location behaviour of the firm. We can imagine the case of a power-generating plant which burns coal and coke, produced at M_1 and M_2, respectively, in order to produce electricity. We can regard the output of the plant as having zero weight or bulk. The output transportation costs of shipping electricity can be regarded as effectively zero, given that the only costs associated with distance will be the negligible costs of booster stations. In this case, the market point of the plant, whether it is a city or a region, will play no role in the decision of where to locate the plant. As such, the optimal location of the plant will be somewhere along the line joining M_1 and M_2. The optimal location problem therefore becomes a one-dimensional location problem. A discussion of this type of problem is given in Appendix 1.1.

In most situations, however, the output of the firm is costly to transport due to the weight and bulk of the output product. Different output weight and bulk will affect the optimum location of the firm relative to the location of the market and the inputs. Once again, we can illustrate this point by using our hypothetical example above of two automobile firms, A and B, each consuming inputs of steel and plastic. However, in this case we can imagine a situation where the input production functions of both firms were the same. In other words the relative input combinations for each firm, given as m_1/m_2, are the same. If both firms pay the same respective transport rates t_1 and t_2 for each input shipped, the relative locational pull of each input will be identical for each firm. However, in this situation we also assume that the firms differ in terms of their technical efficiency, in that firm A discards 70 per cent of the inputs during the production process, whereas firm B discards only 40 per cent of the inputs during the production process. Consequently, the total output weight m_3 of firm B is twice as great as that of firm A, for any total weight of inputs consumed. This greater output weight will encourage firm B to move closer towards the market point and further away from the input points than firm A. As see in Figure 1.3, firm B will therefore be more market-oriented than firm A in its location behaviour.

A more common situation in which similar firms exhibit different location behaviour with respect to the market is where the density of the product changes through the production process at different rates for each of the producers. For example, we can imagine our two automobile firms A and B, producing identical weights of output from

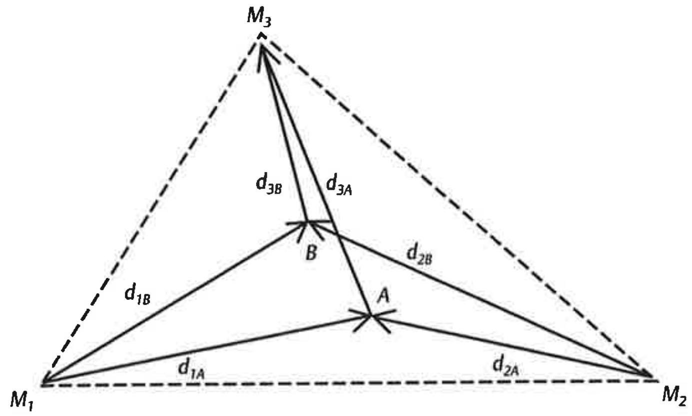

FIG. 1.3 Relative output transport costs and location

identical total weights of inputs. Here, the production functions of both firms are therefore the same. However, we can also assume that firm *A* specializes in the production of small vehicles suited to urban traffic, while firm *B* produces large four-wheel-drive vehicles suitable for rough terrain. As we have already seen, transport rates also depend on the bulk of the product, and products which have a high density will exhibit lower unit transport costs than products with a low density. In this situation firm *B* produces goods which are very bulky, whereas firm *A* produces goods which are relatively dense. Therefore, the output of firm *B* will be more expensive to transport than that of firm *A*, and this will encourage firm *B* to move closer to the market than firm *A*. Once again, as seen in Figure 1.3, firm *B* will be more market-oriented than firm *A*.

1.2.3 The location effect of varying factor prices

Our analysis so far has proceeded on the assumption that labour and land prices are identical across all locations, although in reality we know that factor prices vary significantly over space. The Weber approach also allows us to consider how factor price variations across space will affect the location behaviour of the firm. In order to understand this, it is necessary for us to identify the factor price conditions under which a firm will look for alternative locations.

We assume that the firm is still consuming inputs from M_1 and M_2 and producing an output for the market at M_3. Under these conditions, we know that the Weber optimum K^* is the minimum transport cost location of the firm, and that if all factor prices are equal across space this will be the location of the firm. Our starting point is therefore to consider the factor price variations relative to the Weber optimum K^* which will encourage a firm to move elsewhere. In order to do this, it is first necessary for us to construct a contour map on our Weber triangle, as described by Figure 1.3. These contours are known as *isodapanes*.

On a standard geographical map each contour links all of the locations with the same altitude. On the other hand, each isodapane contour here in a Weber map links all the locations which exhibit the same increase in total input plus output transport costs, per unit of output m_3 produced, relative to the Weber optimum location K^*. Increasing isodapanes therefore reflect increased total input plus output transport costs per unit of output m_3 produced, relative to the Weber optimum K^*. As the location of the firm moves away from the Weber optimum in any direction, the firm incurs increasing transport costs relative to the Weber optimum. In other words, the locations become less and less efficient, and the firm exhibits successively lower profits, *ceteris paribus*. We can also say that the firm incurs successively greater opportunity costs as it moves further away from the Weber optimum. If factor prices are equal across space, locations further away from the Weber optimum will become successively less desirable locations for investment. Therefore, we need to ask by how much do *local* factor prices need to fall relative to the Weber optimum location K^* in order for the firm to move there?

If we take the case of location R, we can ask by how much do factor prices at R need to fall relative to the Weber optimum K^*, in order for the firm to move from K^* to R? As we

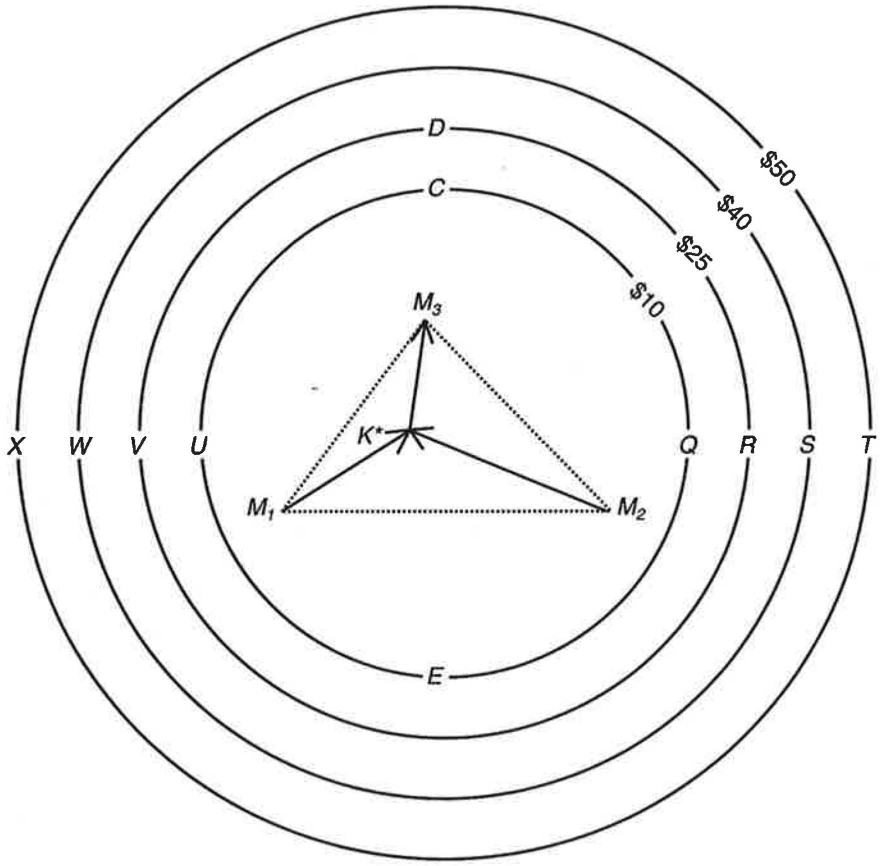

FIG. 1.4 Isodapane analysis

see from Figure 1.4, R is on the \$25 isodapanes. If the costs of the labour and land factor inputs required to produce one unit of output m_3 at R are \$20 less than at K^*, it will not be in the interests of the firm to move from K^* to R. The reason is that the fall in local factor input prices associated with a move from K^* to R will not be sufficient to compensate for the increased total transport costs as we move away from the Weber optimum. If the firm were to move from K^* to R in these circumstances, it would experience profits which were \$5 unit of output m_3 less than at K^*. On the other hand if the local labour and land prices per unit of output as R were \$30 less than at K^*, it would be in the interest of the firm to move. This is because the reduction in the local input factor costs associated with a move from K^* to R will now more than compensate for the increase in total transportation costs incurred by the move. If the firm were to move from K^* to R in these circumstances, it would experience profits which were \$5 per unit of output m_3 greater than at K^*. This type of analysis can be applied to any alternative locations, such as Q, R, S, and T, in order to determine whether a firm should move and to which location.

For example, location Q is on the \$10 isodapane, R is on the \$25 isodapane, S is on the \$40 isodapane, and T is on the \$50 isodapane. Let us assume that the costs of the labour and land factor inputs required to produce one unit of output m_3, at Q, R, S, and T are less than the factor costs at K^* by amounts of \$12, \$20, \$35, and \$55, respectively. We can determine that the alternative locations Q and T are superior locations to K^*, in that both will provide greater profits than K^*, whereas R and S are inferior locations in that they exhibit reduced profits relative to K^*. However, of these superior alternatives, T is the better location because profits here are \$5 per unit of output greater than at K^* whereas those at Q are only \$2 greater. With this particular spatial distribution of local labour and land prices, location T is the optimum location of the firm. T is a superior location to the Weber optimum location at which total transport costs were minimized, because the lower local factor input prices more than compensate for the increased total transport costs associated with the location of T.

This type of approach also allows us to ask and answer a very important question: how will local wages and land prices have to vary over space in order for the firm's profits to be the same for all locations? This can be analysed by modifying Figure 1.4. We can construct Figure 1.5, by employing Figure 1.4, but then altering it by drawing a line from K^* eastwards which passes through Q, R, S, T, as in Figure 1.5. This line is defined in terms of geographical distance. We can then observe how the isodapanes intercept this line.

From the above example, we know that location Q is on the \$10 isodapane, R is on the \$25 isodapane, S is on the \$40 isodapane, and T is on the \$50 isodapane. The firm's profits will be the same in all locations if the local labour and land factor input prices at each location exactly compensate for the increased total transport costs associated with each location. Therefore in Figure 1.5 this allows us to plot the labour and land price gradient with respect to distance which ensures equal profits are made at all locations east of K^*, assuming the wage at K^* is w^*. We can repeat the exercise by drawing a line from K^* which passes west through U, V, W, and X, and plotting the local factor prices which will ensure the firm makes profits equal to those at K^* at all locations west of K^*. Combining this information allows us to construct the interregional factor price curve for our particular firm which ensures that it makes equal profits at all locations in the east–west direction. This is shown in Figure 1.6.

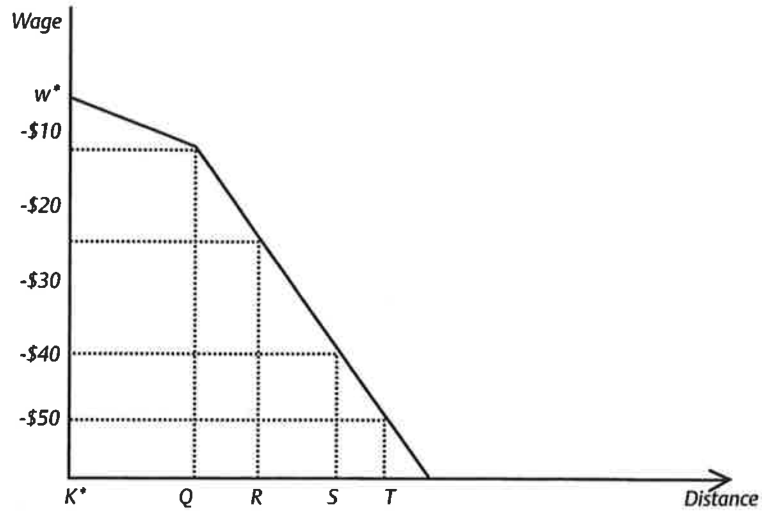

FIG. 1.5 Distance-isodapane equilibrium labour prices

This slope of the line is the interregional *equilibrium* factor price gradient for this particular firm along this particular axis. This equilibrium factor price gradient describes the variation in local factor prices, which ensures that the firm will be *indifferent* between locations. The firm is indifferent between locations along the east–west line, because the profits it can earn are the same everywhere along this line. As such, from the point of view of this firm, all locations along the east–west line are *perfect substitutes* for each other.

In principle, we can also construct similar factor price gradients for movements in any other direction away from K^*, such as movements passing through locations C, D, or E, in order to generate a two-dimensional equilibrium factor price map of the whole spatial economy.

The idea that locations can be perfect substitutes for each other, from the point of view of a firm's profitability, is important in terms of understanding the spatial patterns of industrial investment. For example, if a multinational manufacturing firm is looking for a new production site in order to develop its business in a new area, the likelihood of it going to any particular location will depend on the firm's estimate of the profits it can earn at that location. From the isodapane analysis of our Weber location-production model here, we know that the locations of key input sources such as M_1 and M_2 and market points such as M_3, will automatically mean that some locations are more profitable than others, with the Weber optimum being the most profitable location, *ceteris paribus*. Therefore, in order to make other locations attractive for investment, local factor prices have to fall relative to the Weber optimum. The attractiveness of any particular location as a new investment location for the firm will depend on the extent to which the local factor price falls can compensate for the increased transport (opportunity) costs associated with any suboptimal geographical location. If all local factor prices are interregional equilibrium prices, as described by Figure 1.6, the firm will be indifferent

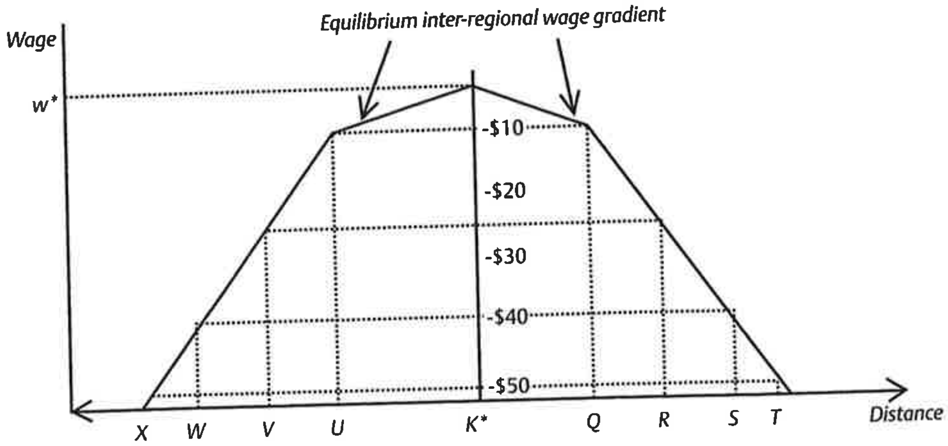

FIG. 1.6 Inter-regional equilibrium wage gradient

between locations. Under these circumstances, the firm will be equally likely to build its new production facility at any location. In other words, the probability of investment will be equal for all locations. Over large numbers of firms with similar input requirements and similar output markets to this particular firm, the level of investment in any location should be the same as in all locations. On the other hand, if wages are not in equilibrium over space, certain areas will automatically appear more attractive as locations for investment, thereby increasing the probability of investment there.

Geography confers different competitive advantages on different locations, which can only be compensated for by variations in local factor prices. However, in the above example, the equilibrium relationship between local factor prices and distance was only applicable to the particular firm in question here. This is because the interregional factor price gradient was calculated with respect to the Weber optimum of this particular firm. As we have seen, different firms will exhibit different Weber optimum locations, and this implies that different equilibrium interregional factor price gradients will exist for different types of firms exhibiting different transport costs, different production functions, and finally different input and output locations.

1.2.4 The locational effect of new input sources and new markets

Our analysis has so far discussed the locational effect of different transport costs, different production functions, and the resulting conditions under which a firm will be willing to move to alternative locations. We will now discuss the question of different input and output locations and the conditions under which a firm will search for alternatives. In the examples above, it was possible to use isodapane analysis to identify the factor price conditions under which a firm will move from one location to another. However, this process of movement itself may engender changes in the input sources employed and the output markets served.

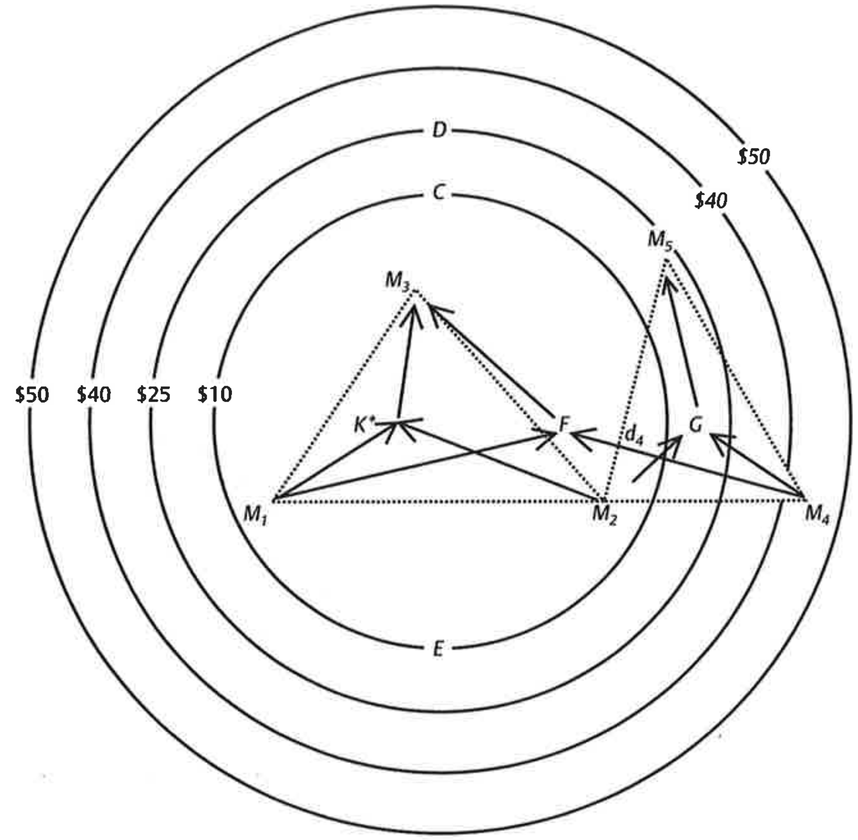

FIG. 1.7 New suppliers and new markets

In Figure 1.7 we can consider the situation where the firm relocates from K^* to F in response to the lower factor prices at F, which more than compensate for the additional input and output transport costs involving in consuming steel and plastic inputs from M_1 and M_2, and serving a market at M_3. Location F has therefore been determined as the new optimum location with respect to M_1, M_2, and M_3. However, in moving from K^* to F, it may be that alternative suppliers of identical inputs now become available. For example, the input supply location M_4 may be able to supply exactly the same steel input as M_1 but from a location whose distance from M_4 to F, which we denote as d_4, is now much closer to the firm than the distance d_1 from M_1 to F. Assuming the delivered price at location F of the steel input produced at M_4, is less than that of the input produced at M_1, i.e. $(p_4 + t_4d_4) < (p_1 + t_1d_1)$, the firm will substitute input supplier M_4 for M_1. This will produce a new Weber location-production problem, with the points M_4, M_2, and M_3 as the spatial reference points. This change in input suppliers will also imply that a new Weber optimum can be found, and that a new series of equilibrium local factor input prices could be calculated with respect to the new Weber optimum location.

With the points M_4, M_2, and M_3 as the spatial reference points, the new Weber optimum is G. At point G, it becomes advantageous for the firm to serve market point M_5, rather than M_3. This is because M_5 is nearer to G than M_3, and $(p_5 - t_5d_5) > (p_3 - t_3d_3)$. Therefore, the firm makes a greater profit from selling automobiles to market M_5 than to market M_3. The firm could switch markets completely from M_3 to M_5. Alternatively, it could decide to supply both markets M_3 to M_5. Under these conditions, it may be that a new optimum location of H arises, in which the firm at H buys from two supplier locations M_4 and M_2, and sells at two market locations, M_3 and M_5. More complex arrangements are possible. For example, in order to guarantee sufficient supplies of steel inputs for the newly expanded automobile market of $(M_3 + M_5)$, the firm may decide to continue to purchase steel from both M_1 and M_4, as well as purchasing plastic from M_2. Now we have a Weber location-production problem with M_1, M_2, M_3, M_4, and M_5 as spatial reference points. Once again, this will move the Weber optimum away from point H, and will also alter the inter-regional equilibrium wage gradient.

This type of geometrical arrangement, in which a firm has multiple input sources and multiple output market locations, is the norm for firms in reality. Although our analysis here has been developed primarily with only two input source locations and one output market location, the Weber location-production arguments and the associated isodapane analysis are perfectly applicable to the case of firms with multiple input and output locations. The reason for employing the triangular case of the two input locations and one output market location is that this particular spatial structure is simply the easiest two-dimensional model to explain. The model is designed to help us understand the advantages which geography confers on particular locations as sites for investment. A first key feature of the Weber model is therefore that it allows us to understand the factor price conditions under which other areas will become more attractive as locations for investment. Secondly, the model allows us to see location as an evolutionary process, in which changes in factor prices can engender changes in location behaviour, which themselves can change the supply linkages between suppliers, firms, and markets. Industrial location problems are inherently evolutionary in their nature as firms respond to new markets and products by changing their locations, and by changing the people they buy from and the people they sell to. All of these are spatial issues.

There is one final issue relating to the Weber model which needs to be addressed. In reality, firms are constantly changing their input suppliers and output markets in response to changes in input and output market prices. From our Weber analysis, these changes will also imply that the optimum location of the firm is continuously changing, and that in order to ensure the profitability of any particular location the equilibrium inter-regional factor price gradient must also be continuously changing. However, observation tells us that firms in reality do not move very frequently, and this raises the question of the extent to which the Weber model is a useful analytical tool to describe industrial location behaviour.

The reason why firms are not continuously moving is that the relocation process itself usually incurs very significant costs, such as the dismantling of equipment, the moving of people, and the hiring of new staff. Part of the transactions costs associated with relocation are also related to information and uncertainty, which are topics we will deal with later in the chapter. However, within the above framework we can easily incorporate these relocation costs, by including the annualized cost of these one-off relocation costs into our isodapane model. The existence of these additional costs simply implies that firms will move only when the factor cost advantages of alternative locations also compensate for these additional relocation costs as well as the increased transport costs. In other words, the equilibrium inter-regional wage gradient will be even steeper than under the situation where such costs are negligible. The Weber model therefore still allows us to identify the optimum location, and consequently the profit-maximizing behaviour of the firm in space, even in situations where relocation costs are significant. The observation that firms do not move frequently does not limit the applicability of the Weber model to real-world phenomena.

The one major location issue which the Weber model does not address, is that of the relationship between input substitution and location behaviour. In order to understand this relationship, we now turn to a discussion of the Moses location-production model.

1.3 The Moses Location-Production Model

including substitution behavior

The Weber model assumes that the quantities of each input consumed, m_1 and m_2, are fixed per unit of output m_3 produced. However, we know from standard microeconomic analysis that substitution is a characteristic feature of firm behaviour, and that efficiency conditions mean that firms will substitute in favour of relatively cheaper inputs, *ceteris paribus*. Substitution behaviour was first incorporated coherently into the Weber analysis by Moses (1958), and in order to see how substitution behaviour affects the location behaviour of the firm, we discuss here the main features and conclusions of the Moses approach.

In Figure 1.8, we construct an arc IJ in our triangle M_1, M_2, M_3, which is at a constant distance d_3 from the market point M_3. If we constrain our firm to locate along this arc, the distance from the location of the firm K to the market M_3 will no longer be a variable. Therefore, we can analyse the locational pull on the firm of changes only in the delivered prices of the inputs produced at M_1 and M_2.

For example, if the firm was located at I, the delivered price of input 1, given as $(p_1 + t_1d_1)$, will be a minimum, because the distance d_1 from M_1 to I will be a minimum. Similarly, the delivered price of input 2, given as $(p_2 + t_2d_2)$, will be a maximum, because the distance d_2 from M_2 to I will be a maximum. The delivered price ratio, given as $(p_1 + t_1d_1) / (p_2 + t_2d_2)$, will therefore be a minimum at location I. On the other hand, if the firm now moves to J, the delivered price of input 1 will be a maximum, because the distance d_1 from M_1 to J will be a maximum. At the same time, the delivered price of input 2 will be a minimum, because the distance d_2 from M_2 to I will be a minimum. Therefore, the delivered price ratio, $(p_1 + t_1d_1) / (p_2 + t_2d_2)$, will be a maximum at location J.

In standard microeconomic approaches to firm efficiency, the optimal input combination is determined by finding the point at which the highest isoquant attainable is tangent to the budget constraint. In this standard approach, the slope of the budget constraint is determined by the relative prices of the goods. From the above argument, we

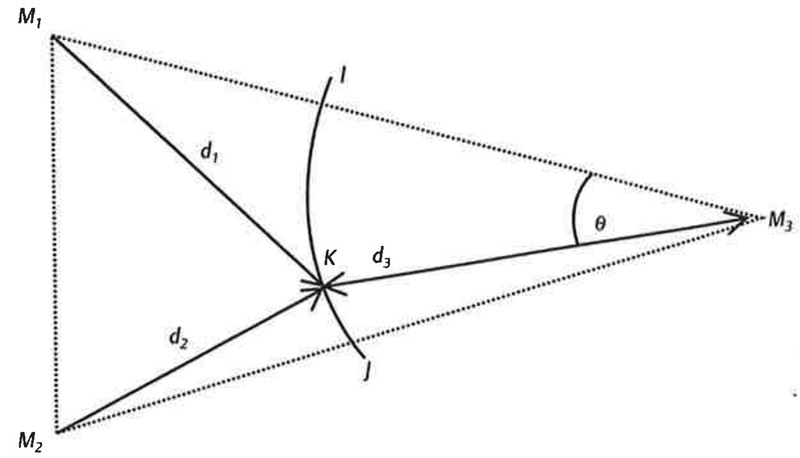

FIG. 1.8 Weber–Moses triangle

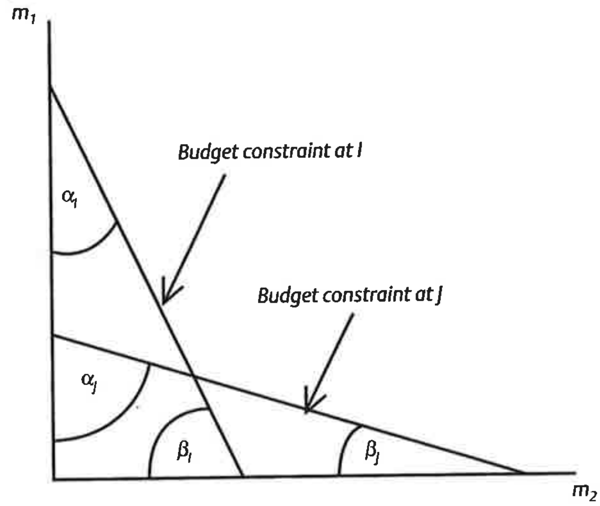

FIG. 1.9 Budget constraints at the end points I and J

can draw the budget constraints at locations I and J as shown in Figure 1.9, which represent equal total expenditure on inputs at each location. The delivered price ratios at locations I and J are given by the ratio of the tangents of the angles a_I/β_I and a_J/β_J, respectively.

Yet, this argument is also applicable to all locations along the arc IJ. If there are different delivered price ratios for different locations, this implies that for given source prices of the inputs p_1 and p_2, the slope of the budget constraints at each location along IJ must be different. As we move along the arc IJ from I to J, the delivered price ratio increases, and for every location along the arc *IJ* there is a unique delivered price ratio. This means that the usual approach to analysing microeconomic efficiency is not applicable to the firm in space, and must be adapted to incorporate the effects of location on the slope of the budget constraint. In order to do this we must construct the *envelope* budget constraint, which just contains all of the budget constraints associated with each of the locations along the arc *IJ*. This is done by drawing each of the budget constraints for each of the location points on the arc *IJ*, as in Figure 1.10, and the outer limits of this set of individual budget constraints will define the envelope budget constraint.

The Moses argument is that we can now apply standard efficiency conditions to this model, by finding the point at which the envelope budget constraint is tangent to the highest isoquant attainable. This is shown in Figure 1.11, where the point of maximum efficiency is at E^*.

At E^*, the optimum input combinations are given as $m_1{}^*$ and $m_2{}^*$. However, E^* also represents an *optimum location* K^*. The reason is that the optimum input combination is found at a particular point on the envelope budget constraint. Yet, every point on the budget constraint also represents a unique location. Therefore, the optimum input mix and the optimum location of the firm are always jointly determined. One is never without the other. This is a profound insight. Where input substitution is possible, all location problems become production problems and all production problems become location problems.

We can illustrate the argument with an example. In our Weber–Moses triangle, we can imagine that a road-building programme takes place in the area around location M_1, the effect of which is generally to reduce the value of t_1 for all shipments of goods from this location, relative to all other locations. If all the other parameters remain constant, this will imply that the delivered price ratio $(p_1 + t_1d_1)\ /(p_2 + t_2d_2)$, at all locations along *IJ* will

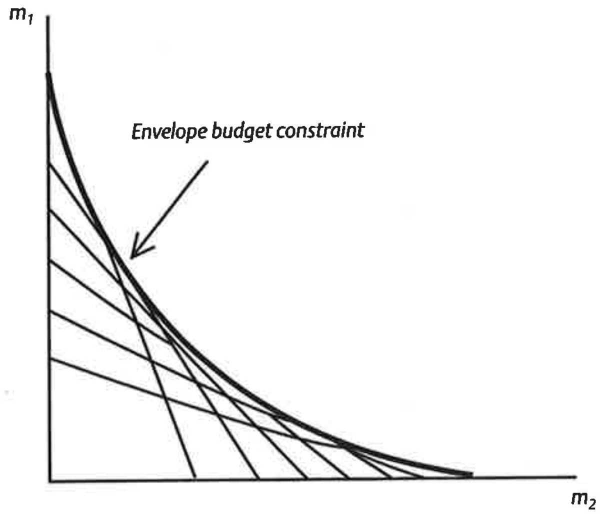

FIG. 1.10 The envelope budget constraint

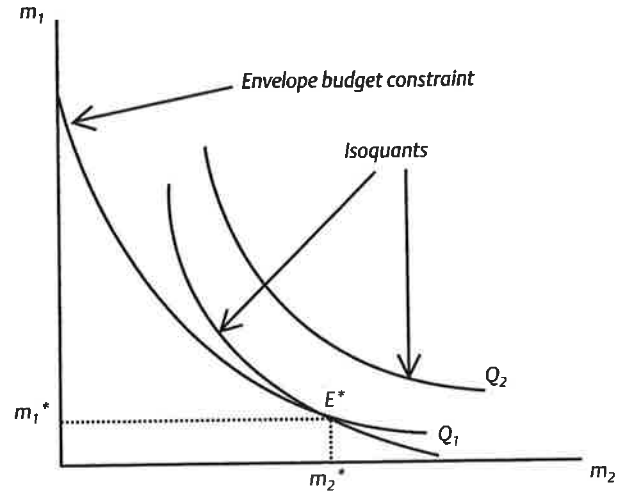

FIG. 1.11 Location-production optimum

fall. In other words, the slope of each budget constraint becomes steeper, *ceteris paribus*, and the envelope budget constraint also becomes steeper and shifts upwards to the left. Strictly speaking, in accordance with the income effect, the envelope will also shift outwards to the right, because the price of the input has fallen. However, in this discussion we focus only on the substitution effect of the change in slope of the envelope. For a given set of production isoquants, the optimum production combination will change from that represented by E^*.

As we see in Figure 1.12, at the new optimum E', the optimum input mix is now m_1' and m_2'. The reason is that the firm substitutes in favour of input 1, which is now relatively cheaper than before, and away from input 2, which is now relatively more expensive than before. In doing so, the firm increases the relative quantities of input 1 it consumes and reduces the relative quantities of input 2 it consumes. However, this also implies that at the original location K^*, the firm now incurs increasing total transport costs ($m_1t_1d_1$) for input 1 relative to the total transport costs ($m_2t_2d_2$) for input 2. Therefore, the firm will move towards M_1, the source of input 1, in order to reduce these costs. The new optimum location of the firm K' is closer to M_1 than E^*, and so the firm moves towards M_1.

The area around M_1 benefits in two different ways. First, the relative quantity of goods produced by the area around M_1 which are bought by the firm increases. This increases regional output for the area. Secondly, the firm itself locates in the vicinity of M_1, thereby increasing the levels of industrial investment in the area.

Exactly the same result would have arisen in the case where, instead of a road-building programme, there was a fall in the local wages at M_1, which reduced the source price p_1, relative to all other locations. Once again, the fall in the delivered price ratio at all locations leads to substitution in favour of the cheaper good and also relocation towards M_1.

We can contrast this Moses result with that of the Weber model. In the simple Weber

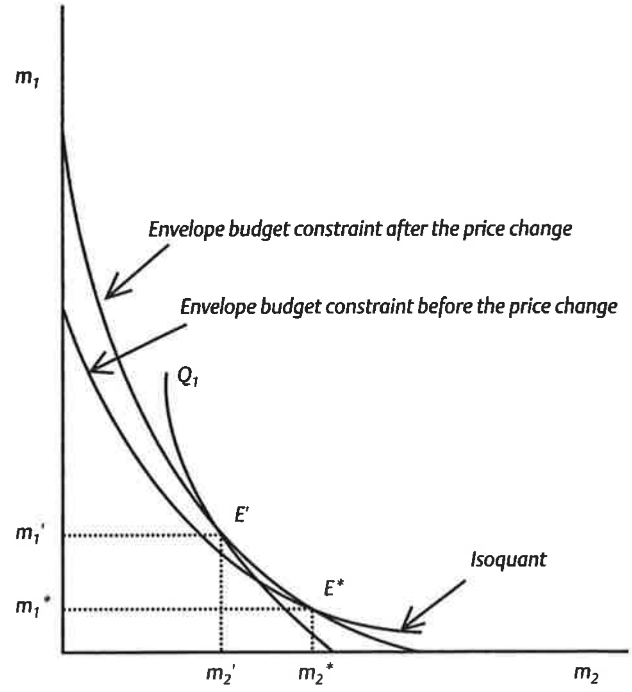

FIG. 1.12 A change in the location-production optimum

model, if the transport rate t_1 falls, *ceteris paribus*, the effect on the location of the firm is to move the locational optimum *away* from M_1. The reason is that input 2 now becomes relatively more expensive to transport, and because the coefficients of production are fixed, such that the relative quantities of m_1 and m_2 consumed remain the same, the firm will move towards the source of input 2 in order to reduce the total transport costs. The difference between the location-production results of the two models is that in the Weber model the fixed coefficients mean that no input substitution is possible, whereas in the Moses model of variable coefficients, input substitution is possible. In the latter case, the input substitution behaviour alters the relative total transport costs and consequently the optimum location behaviour of the firm. In reality, there is a continuum of possible location effects, dependent on the technical substitution possibilities. In situations where the elasticities of substitution are zero or very low, the results will tend to mimic those of the Weber model, whereas in situations where the elasticities of substitution are high, the results will tend towards those of the conclusions of the Moses model.

comparando Weber / Moses.

A second feature of the Moses model is that it allows us to examine the effect of returns to scale on the location-production behaviour of the firm. In particular, we can ask the question, how will the optimum location of the firm be affected by changes in the level of output of the firm? In order to answer this in Figure 1.13 we construct a series of envelope budget constraints, represented by the dotted lines, which correspond to different levels of total expenditure on inputs. Envelope budget constraints further to the right imply

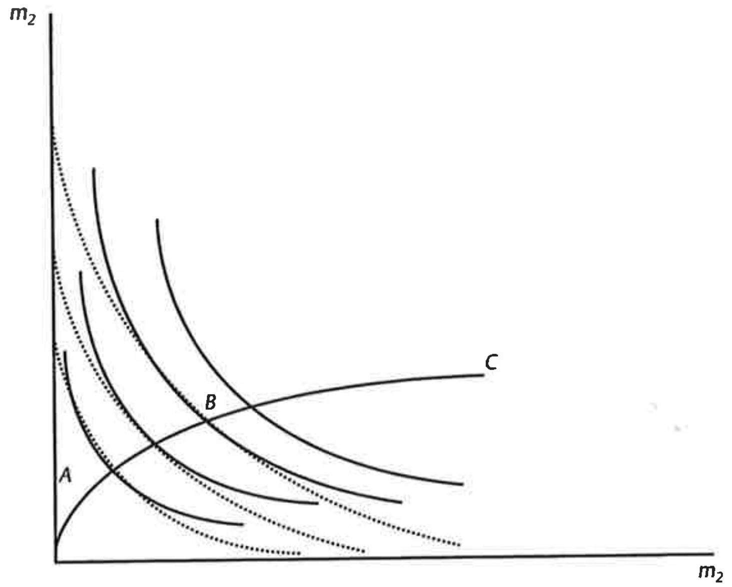

FIG. 1.13 Output changes and location-production behaviour

greater total expenditure levels on inputs. An isoquant map, represented by the solid curves, can now be combined with the envelope map. We can also apply the Moses argument, which states that the optimum point for each level of output and input expenditure is where each particular envelope is tangent to the highest isoquant, to the case of different output levels. By joining all the points of tangency we construct a line *ABC*, which is an output expansion path. Yet, this output expansion path is different from the usual form of an expansion path. Each point on the expansion path defines a particular optimum input combination. However, each point on the expansion path also defines an optimum location.

If the expansion path is curved downward, such as in the case of *ABC* in Figure 1.13, it implies that as the output of the firm increases, and the total quantity of inputs consumed increases, the optimum input mix changes relatively in favour of input 2. The optimum ratio of m_1/m_2 falls and the optimum location of the firm moves towards M_2. Alternatively, if the expansion path were to curve upwards, this would mean that as the output of the firm increases, the optimum input combination would change in favour of input 1. As the optimum ratio of m_1/m_2 increases, the optimum location of the firm would move towards the market.

This argument immediately leads to the conclusion that if the expansion path is a straight line from the origin, such as *FGH* in Figure 1.14, both the optimum input mix and also the optimum location of the firm will remain constant as output expands, *ceteris paribus*. The actual slope of the expansion path is not important, other than it implies a different optimum location. All that is required to ensure that once the firm has found its optimum location it will always remain at this optimum location as output changes, is

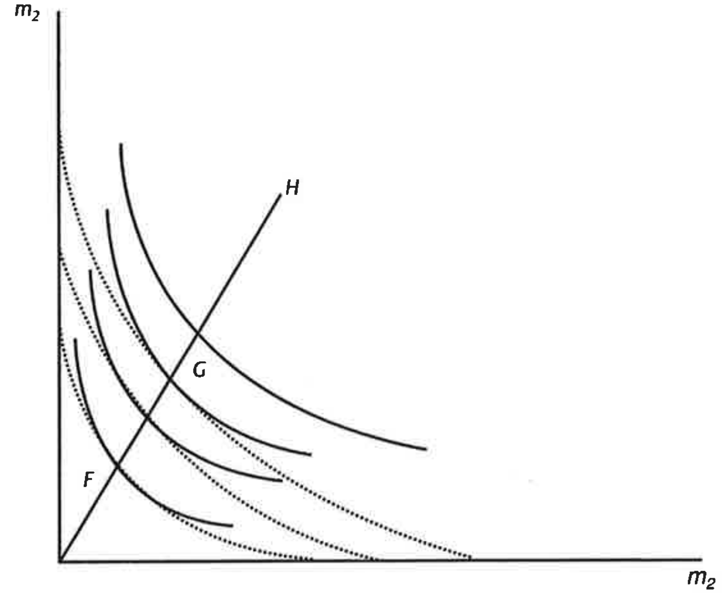

FIG. 1.14 The independent of output optimum location solution

that the production function of the firm exhibits a straight line expansion path from the origin. This is the basic Moses result.

This basic Moses result holds in the case where the firm is constrained to locate on the arc *IJ* at a fixed distance from the market. However, in the more general case where the distance from the market is also part of the location problem, the optimum location of the firm will be independent of the level of output, as long as both the production function of the firm and the transportation technology of the firm exhibit constant returns to scale. The Weber fixed-coefficients production function will satisfy the Moses requirement. However, there are other more general types of production functions allowing for input substitution, which also satisfy this requirement. These results are detailed in Appendix 1.2.

The Moses result can be viewed somewhat as the spatial equivalent of the firm in perfect competition. The firm is a price taker, and once it has determined its optimum production technique and optimum location, the firm will not change its behaviour, *ceteris paribus*. In other words, unless there are external changes in technology which alter the production function relationships, or changes in transportation technology which alter relative transport costs, or externally determined changes in the location of input goods sources and output market points, the firm will always remain at the same location employing the same input–output production techniques. It would be wrong, however, to view these spatial results as implying that the spatial economy is essentially static. From our discussion in section 1.2.4, we saw that the spatial economy exhibits evolutionary characteristics, with firms searching for new optimum locations in response to factor price changes, and subsequently searching for new input supplier and market output locations, in response to their relocation behaviour. The key insights, however, of the Weber and Moses models are that production behaviour and location behaviour are completely intertwined issues. Often this point is overlooked in textbook discussions of industrial economics and the theory of the firm. This is largely because location adds an extra dimension to the optimization problems, making the analysis somewhat more complex.

1.3.1 The logistics-costs model

There are a couple of possible limitations to the applicability of the Weber–Moses framework to real-world phenomena which need to be considered at this point. The first limitation is that the market price or revenue of the output good plays no role in the determination of the optimum location of the firm in either model. In the Weber model, the optimum location is determined solely by the transportation costs associated with the input and output goods, whereas in the Moses model, the input prices do play a role in the optimum location. In neither model does the market price have any effect on the determination of the optimum location. The second limitation of this framework is the emphasis on transport costs as a locational issue. In reality, transport costs tend to be only a very small percentage of total costs for most firms. However, both of these model weaknesses can be largely reconciled within a Weber–Moses framework by employing a broader description of distance-transport costs defined as 'total logistics costs', which includes all of the inventory purchasing and carrying costs associated with transportation (McCann 1993, 1997, 1998). Employing this logistics costs approach, it can be demonstrated both that the market price and market sales revenue do play a crucial role in determining the optimum location, and also that distance costs are very significant. In particular, as we see in Appendix 1.3, the higher value-adding activities will tend to be more market-oriented than lower value-adding activities, and will also tend to be less sensitive to inter-regional labour price changes. As such, market areas will tend to be surrounded by higher-value activities or activities further up the value-chain, whereas supply sources will tend to be surrounded more by lower value-adding activities or firms lower down the supply chain. At the same time, total logistics costs can also be shown to be very much more significant than transport costs alone, because each of the inventory purchasing and carrying cost components can be shown to be functions of distance. A final point here is that the total logistics costs approach can also be employed to account for the economies of distance and scale generally observed in transport pricing (McCann 2001) and discussed in Appendix 1.1.

1.4 Market Area Analysis: Spatial Monopoly Power

In our analysis so far we have assumed that the market location is simply a point in space. However, taking geography and space seriously in our models of firm behaviour also requires that we investigate the explicitly spatial nature of market areas. Market areas frequently differ over space, due to differences in spatial population densities, differences in income distributions across space, and differences in consumer demand across space according to regional variations in consumer tastes. However, even if there were no spatial variations in population densities, income distributions, and consumer demand patterns, space would still be an important competitive issue. The reason is that geography and space can confer *monopoly power* on firms, which encourages firms to engage in spatial competition in order to try to acquire monopoly power through location behaviour. In order to see this we can adopt the approach first used by Palander (1935).

In Figure 1.15, we have two firms *A* and *B* located at points *A* and *B* along a one-dimensional market area defined by *OL*. We assume that both firms are producing an identical product. The production costs p_a of firm *A* at location *A* can be represented by the vertical distance *a*, and the production costs p_b of firm *B* at location *B* can be represented by the vertical distance *b*. As we see, firm *A* is more efficient than firm *B*. The transport costs faced by each firm as we move away in any direction from the location of the firm are represented by the slopes of the transport rate functions. As we see here the transport rates for the two firms in this case are identical, i.e. $t_a = t_b$. For any location at a distance d_a away from firm *A*, the delivered price of the good is given as $(p_a + t_a d_a)$, and for any location at a distance d_b away from firm *B*, the delivered price of the good is given as $(p_b + t_b d_b)$.

If we assume that consumers are evenly distributed along the line *OL*, and we also assume that consumers, being rational, will buy from the firm which is able to supply at that particular location at the lowest delivered price, the total market area will be divided into two sectors *OX* and *XL*. The reason for this is that between *O* and *X*, the delivered price of firm *A*, given as $(p_a + t_a d_a)$, is always lower than that of firm *B*. On the other hand, at all locations between *X* and *L* the delivered price of firm *B*, given as $(p_b + t_b d_b)$, is always lower than that of firm *A*. Although firm *A* is more efficient than firm *B*, and although both firms produce an identical product, firm *A* does not gain all of the market. The reason is that location gives each firm some monopoly power over the area around itself. Firm *A* cannot capture all of firm *B*'s market, even though it is more efficient than firm *B*, because the transport costs associated with shipping goods to market locations close to firm *B* increase the delivered price $(p_a + t_a d_a)$ to an uncompetitive level in market locations close to firm *B*. In terms of selling to consumers in the vicinity of firm *B*, firm *A* is unsuccessful simply because it is too far away. On the other hand, for sales in this area, firm *B* is successful simply because it is in the right location, even though it is less efficient in production.

This type of analysis can be extended to allow for differences in transport rates between

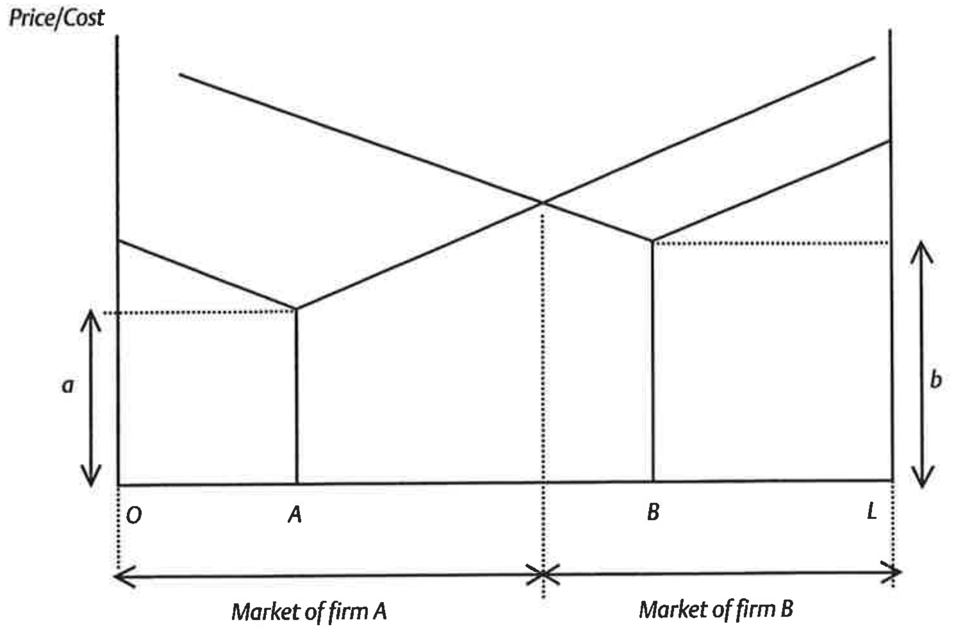

FIG. 1.15 Spatial market areas: a one-dimensional model with equal transport rates

firms as well as differences in production costs. In Figure 1.16a, b, we see that market areas can be divided up in a variety of ways in situations where the production costs and transport rates vary between the firm. Generally, the size of a firm's market area will be larger the lower are the production costs of the firm and the lower are the transport rates faced by the firm. However, only in the case where transport rates are zero is a lower production price sufficient to ensure a firm captures all of the market. The reason is that the existence of transport costs allows less efficient firms such as firm *B* to survive by providing each firm with some monopoly power over particular market areas. In general, the areas over which firms have some monopoly power are the areas in which the firms are located. For example, Figure 1.16b can be regarded as representing a case such as a local bakery, where firm *B* maintains a very small local market area in the face of competition from a national bakery, firm *A*, which produces at much lower unit production costs and transports in large low-cost shipments.

Monopoly power refers to the ability of the firm to increase the production price of the good p_a or p_b, and yet maintain some market share. In general, the greater is the monopoly power of the firm, the steeper is the firm's downward-sloping demand curve. In many textbook descriptions of monopoly or monopolistic power, the slope of the firm's downward-sloping demand curve is viewed as being dependent on brand loyalty, associated with advertising and marketing. However, location is also an important way in which many firms acquire monopoly power. The reason is that transport costs are a form of transactions costs, and from the theory of the firm, we know that the existence of transactions costs such as tariffs and taxes can provide protection for some inefficient

(a)

(b)

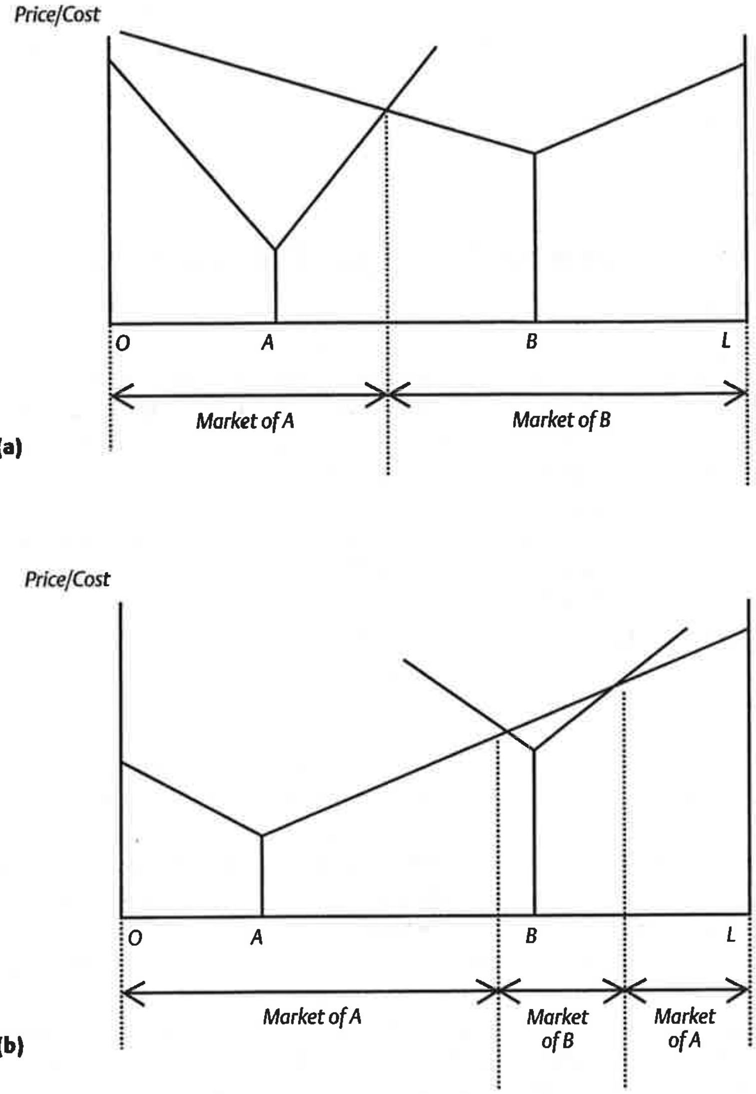

FIG. 1.16a, b Spatial market areas: one-dimensional models with varying transport rates and production costs

firms. Geography acts in a similar manner, because the costs of overcoming space in order to carry out market exchanges incur transport-transactions costs. In the context of Figures 1.15 and 1.16, there are two general rules governing the extent to which distance costs provide a firm with spatial monopoly power:

(i) First, the greater are the values of the transport rates t_a and t_b, the lower will be the fall in the market area of the firm, and the greater will be the monopoly power of the firm, for any marginal increase in the price of either p_a or p_b, *ceteris paribus*.

(ii) Second, the further apart are the firms, the lower will be the fall in the market area of the firm, and the greater will be the monopoly power of the firm, for any marginal increase in the price of either p_a or p_b, *ceteris paribus*.

Therefore, firms which are located at a great distance from each other, and which face significant transport costs, will consequently exhibit significant local spatial monopoly power.

1.4.1 The Hotelling model of spatial competition

The existence of spatial monopoly power provides an incentive for firms to use location as a competitive weapon in order to acquire greater monopoly power. This is particularly important in industries where firms do not compete primarily in terms of price, but instead engage in non-price competition, such a product quality competition. In competitive environments characterized by oligopoly, the interdependence between firms in the determination of output quantities and market share is also a result of locational considerations, as well as interdependence in terms of pricing decisions. The simplest demonstration of this is the Hotelling (1929) model, which describes firms' spatial interdependence within the context of a locational game.

In Figure 1.17 we adapt Figure 1.15 to the case where both the production costs and transport rates of firm *A* and firm *B* are identical. In other words, $p_a = p_b$ and $t_a = t_b$, and we assume that these prices do not change. As before, we assume that consumers are evenly distributed along *OL* and we also introduce the assumption that the demand of consumers is perfectly inelastic, such that all consumers consume a fixed quantity per time period irrespective of the price. In terms of firm strategy we assume that each firm makes a competitive decision on the basis of the assumption that its competitor firm will not change its behaviour. In the game theory literature this particular set of rules describing the nature of the competitive environment is known as 'Cournot conjectures'. Given that the firms are not competing in terms of their production prices, which are assumed to be fixed, each firm can only adjust its location in order to acquire greater market share. If the firms react to each other in sequential time periods, the location result can be predicted easily.

If we assume that the firms *A* and *B* are initially located at one-quarter and three-quarters of the way along the market, respectively, firm *A* will have monopoly power over *OX* and firm *B* will have monopoly power over *XL*. In this case, both firms will have identical market shares. In time period 1 firm *A* will therefore move from its original location to a location at *C*, just to the left of *B*. In this way firm *A* will increase its market share from *OX* to a new maximum value of *OC*. Firm *B* will still retain market share over *BL*, although its market share is now at a minimum.

Firm *B* will now assume that firm *A* will maintain its location at *C*, and so in time period 2, firm *B* will move just to the left of *C*. In time period 3, firm *A* will respond by moving to the left of firm *B*, and this process will continue until both firms are located at *X*, in the middle of the market. Once both firms are located at *X*, neither firm has any incentive to change its location behaviour, because any location change will involve a reduction in market share relative to their location at *X*. In game theory, any situation in which neither firm has any incentive to change its behaviour is known as a 'Nash equilibrium'.

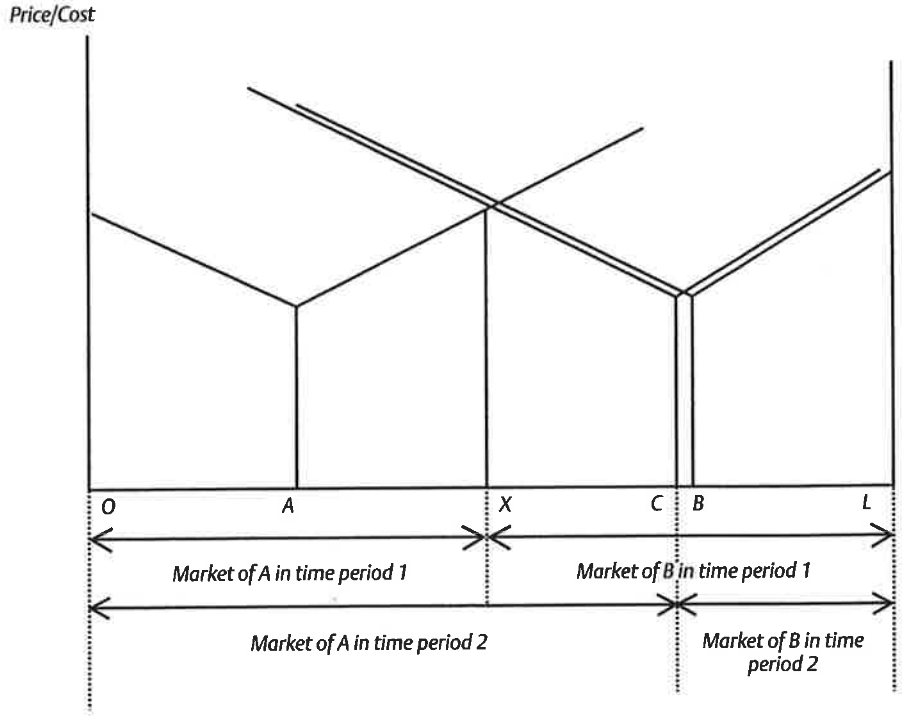

FIG. 1.17 The Hotelling location game

The locational result in which both firms are located at the centre of the market is the Nash equilibrium for this particular locational game. Consequently, once the firms reach this point they no longer continue to move. This is the Hotelling result. The details of this are given in Appendix 1.4.

At the conclusion of the Hotelling game we see that the market share of both firms located at *X* will be half of the market, exactly the same as at the start of the location game. However, from Figure 1.18 we see that the Hotelling result leads to a fall in consumer welfare relative to the original situation. Given that consumers all consume a fixed quantity per time period of the good produced by firms *A* and *B*, there is no substitution effect between the goods produced by firms *A* and *B* and other consumption goods. Therefore, the change in the delivered prices at the each location will accurately reflect the change in welfare of the consumers at each location. The net effect of these welfare gains and losses can be represented by the areas under the delivered price curves, which are arrived at by comparing the delivered prices at the respective locations at the start and the end of the Hotelling location game. The consumers who are located in the centre of the market benefit by generally reduced delivered prices, represented by *egjh* in Figure 1.18, whereas those located at the edges of the market lose by generally higher delivered prices, represented by (*defc*) + (*jklm*) in Figure 1.18. The gain in lower prices for the central consumers is outweighed by the

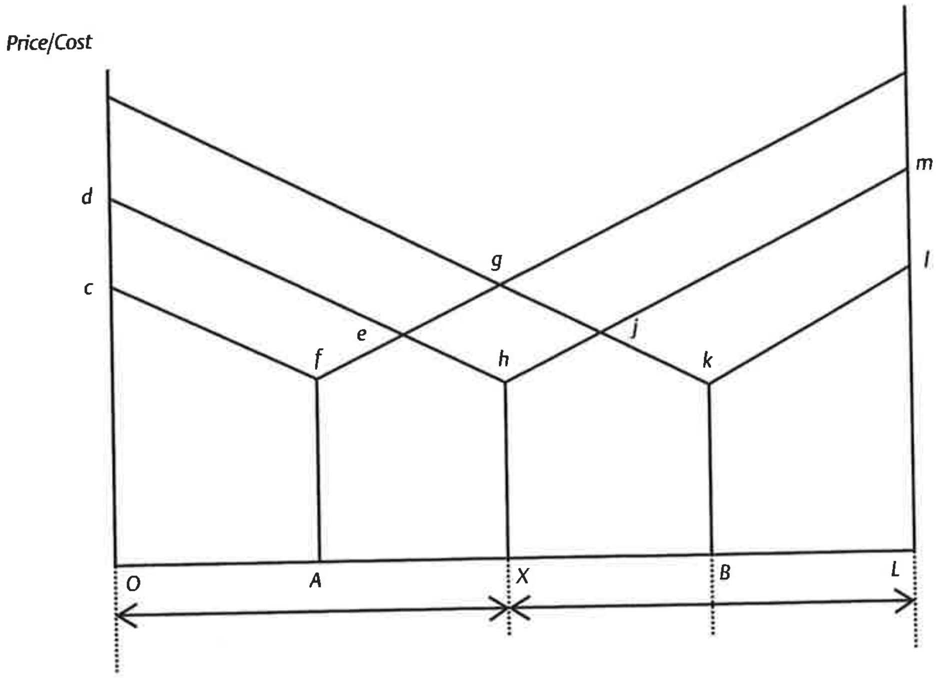

FIG. 1.18 The welfare implications of the Hotelling result

increase in prices for the more peripheral consumers. The net effect is therefore a social welfare loss.

In one-dimensional space discussed here the Hotelling result holds for two firms. Meanwhile, in the two-dimensional case the Hotelling result holds for the case of three firms. However, beyond these numbers, there is no stable equilibrium result as firms keep changing their location. Moreover, even in the one-dimensional case, the Hotelling result only holds as long as the firms do not compete in terms of prices. If price competition is also a possibility, there is no Hotelling result (d'Aspremont *et al.* 1979). In Figure 1.19, we can consider the situation where firm *A* lowers its production price marginally in time period 1 when both firms are located at *X*. In time period 2 firm *A* gains all of the market. From our Cournot conjectures, firm *B* now assumes that firm *A* will maintain both its new lower price and its location at *X*. Therefore in time period 3 firm *B* also lowers its market price below that of firm *A*, and now gains all of the market. This process will continue and the long-run Nash equilibrium of this price war is that both firms will end up selling at zero profit while still being located at point *X*.

If the production costs are not zero, but rather are positive, the long-run result of this co-location competition will be to drive prices down to the marginal costs of production, which is the typical equilibrium result of a competitive market. However, the Hotelling model is implicitly about monopoly power, with firms able to use location as a means of generating monopoly power over a certain portion of their market. As we see in Chapter 2, the greater is their localized monopoly power, the greater will be the possibilities for

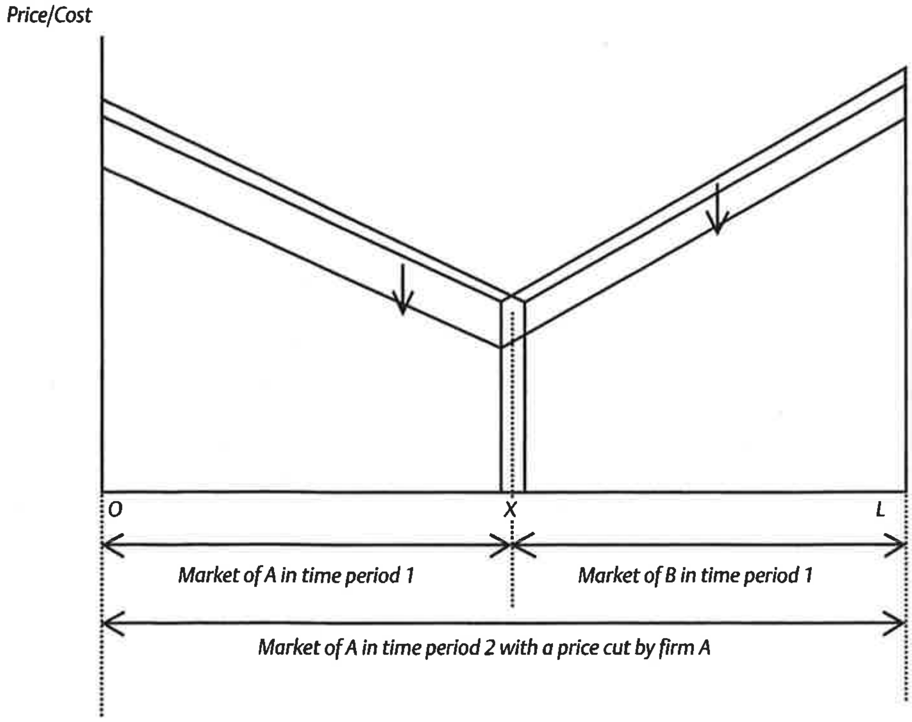

FIG. 1.19 The effect of price competition on the Hotelling result

the firm to raise additional revenues by employing monopoly practices such as price discrimination. Therefore, in order to generate localized monopoly power, as prices spiral downwards due to the Bertrand problem, each firm has some incentive to move away from its competitor in order to maintain monopoly power, and consequently positive profits, over some of the market area. However, neither firm has an incentive to move away first, because in doing so, the other firm will then be able to maintain its current prices at the centre of the market and dominate a larger market area than the firm which moved away from the centre. Therefore, unless there is some way in which the firms can mutually agree to move away from each other, a price war becomes inevitable with disastrous consequences for both firms. The relationship between the co-location of competing activities and the problem of a price war is known as the 'Bertrand problem'.

Competitor firms will consequently only locate next to each other in situations in which price competition is ruled out either by mutual agreement or by other forms of non-price competition. Yet, in these types of non-price competitive situations, the spatial clustering of competitor firms is a natural process. Many types of shops and showrooms, for example, such as those for clothing, electronics goods, automobiles, restaurants, and furniture, compete in industries dominated by non-price competition. In these industries prices are used to indicate product quality, and to indicate the types of consumers for whom the good is intended. As such, prices in these industries tend to be fixed. Firms are unwilling to compete by lowering prices because this suggests that the product quality is falling, and this may actually have an adverse effect on sales. The practice of ascribing prices to products in order to indicate both the product quality and the consumer for whom the product is intended is known as 'price placing', and the problem of lower prices implying lower product quality is related to the famous 'market for lemons' problem described by Akerlof (1970). At the same time, engaging in non-price competition also implies that the products are not identical, and therefore the Hotelling result would appear not to be relevant. However, in many cases of non-price competition, the differences between the products are largely superficial, involving primarily differences in packaging and appearance. The products in essence will essentially still be identical. The fact that firms attempt to make more or less the same products appear very different is known as the 'Hotelling Paradox'. In these situations, firms will tend to cluster together in space. This is exactly how retail parks and central city shopping areas arise.

On the other hand, where firms produce identical products in which non-price competition is extremely difficult, such as the market for gasoline, firms will not cluster together in space. Oil companies which own or franchise out gasoline retail stations will mutually agree not to locate their outlets too close to their competitors, in order to guarantee some market monopoly power for each station in its immediate vicinity. The only time in which gasoline stations will be located close to each other on the same highway is where they are separated from each other by a central reservation, median barrier, or major junction. In these cases, the stations are effectively separated from each other and customers denied the choice between the stations, because drivers are unable easily to switch sides of the road. Therefore, the stations can be considered as not being located together, but rather located away from each other.

The Hotelling result therefore provides us with two important sets of analytical conclusions. First, for competitor firms producing the same type of product and which also engage in non-price competition, the spatial competition for markets may encourage such firms to locate next to each other. In other words, spatial industrial clustering can arise naturally where price competition is not paramount. This is particularly important in many examples of retailing. Moreover, in this situation, the market will be split more or less equally between all of the firms in the spatial cluster. This ensures that no firm will be any worse off than its competitor due to an inferior location, a point we will discuss in section 1.5. On the other hand, for firms which produce more or less identical products for which non-price competition is very difficult to engage in, and in which there are no information problems, spatial competition will encourage such firms to move away from each other. The result of this process is industrial dispersion. Secondly, from a welfare point of view, consumers located close to a spatial cluster of firms will tend to experience a welfare gain relative to those located at a great distance away. The reason for this is that the costs of consuming the goods produced by the firms will tend to be much lower for those who are located close to the firms than for those who are located at a distance away. This is an important observation concerning agglomeration economies, a topic which we will discuss in the next chapter.

When applying these insights of the Hotelling framework to the real world, however, these two observations must be interpreted with caution, because there are some other situations in which price competition and spatial clustering are compatible. This is the case where prices are not predictable and are continually changing, such as in the case of many food markets or gambling activities. In these situations, although price competition is very keen, firms may gain from either short-term first mover advantages, or alternatively customer inertia in the face of rapid and frequent price changes. The co-location of retailing activities in this case is justified, as with the case of non-price retail competition, because this may encourage customers to buy more goods in general than they would otherwise if they were not presented with a broad range of consumption alternatives and the relevant price information about them. As such, all firms in the cluster are expected to gain, and co-location ensures that all firms benefit more or less equally. These arguments are related primarily to the questions of information, clustering, and externalities discussed initially in section 1.5 and at length in Chapter 2.

One final point concerning this Palander and Hotelling type of spatial market analysis is the criticism that in many real-world cases, individual firms charge the same delivered price for a given product at all locations. As such, spatial markets are not divided up according to delivered prices which vary with location. On the other hand, where delivered prices are invariant with respect to distance within a given market area, this implies that the marginal profitability of each delivery will be different according to the location of the customer. This is because the transport costs of outputs must be absorbed by the firm, thereby reducing the net marginal profits from sales as the delivery distance increases. In other words, the profits associated with deliveries to nearby customers will be much higher than those for deliveries to distant customers. As such, for any given spatial distribution of markets, the location of the firm will still determine the overall profitability of the firm. Moreover, as we see in Appendix 1.3 and Appendix 3.4, even for uniform delivered prices, firms are able to employ changes in the quality of service, such as changes in delivery frequencies, in order to mimic the spatial price effects of situations in which customers pay the transport charges in addition to the quoted source prices.

1.5 Behavioural Theories of Firm Location

The models discussed so far rely on the assumption that 'rational' firms will aim to use their location behaviour in order to maximize their profits. We have also assumed that the information available to the firms is sufficient for them to do this. However, in reality the information available to firms is often rather limited. Moreover, different firms will often have different information available to them. For this reason, some commentators have argued that firms cannot and do not make decisions in order to maximize their profits. Rather, they argue that firms make decisions in order to achieve alternative goals, other than simply profit maximization. Therefore, from the perspective of location theory, this critique might suggest that the underlying motivation of our models would need to be reconsidered. The critique has three themes, namely bounded rationality, conflicting goals, and relocation costs. The first two themes can be grouped under the general heading of *Behavioural Theories*, and were not originally directed at location models in particular. The third theme is essentially a spatial question.

The arguments concerning 'bounded rationality' are most closely associated with Simon (1952, 1959). This critique concerns the fact that firms in the real world face limited information, and this limited information itself limits firms' ability to be 'rational' in the sense assumed in microeconomics textbooks. These arguments are a more general critique of rationality within microeconomics as a whole. However, they have been argued to be particularly appropriate to the question of industrial location behaviour. The reason is that information concerning space and location is very limited, due to the inherent heterogeneity of land, real estate, and local economic environments. Therefore, when considering location issues, it would appear that the ability of the firm to be 'rational' is very much 'bounded' by the limited information available to it. In these circumstances, decisions guided by straightforward profit-maximization behaviour appear to be beyond the ability of the firm. Therefore, location models based on this assumption seem to oversimplify the location issue. Location behaviour may be determined primarily by other objectives than simply profit maximization.

Where firms face limited information, Baumol (1959) has argued that firms will focus on sales revenue maximization as the short-run objective of their decision-making. One reason for this is that sales revenue maximization implies the maximum market share for the firm in the short run. Where information is limited, current market share is deemed by many observers to be the best indicator of a firm's long-run performance, because it provides a measure of the monopoly power of the firm. The logic of this approach is that the greater is the market share of the firm, the greater is the current monopoly power of the firm, and the greater will be the firm's long-run ability to deter potential competitors through defensive tactics such as limit-pricing and cross-subsidizing. From the perspective of location models, this may imply that the firm will make location decisions primarily in order to ensure maximum sales revenues rather than maximum profits. In the Hotelling model above these two objectives coincide. However, if the costs of production or transportation faced by the firm were to vary with location along the line *OL*, as they do in the Weber and Moses-type models, the two objectives of sales maximization and profit maximization may not coincide at the same location point in the Hotelling model. The eventual location result will therefore depend on which particular performance measure the firm adopts and chooses to maximize.

The second critique of profit maximization as the decision-making goal of the firm is that of 'conflicting goals'. This critique is most closely associated with the work of Cyert and March (1963). The argument here is that in a world of imperfect information, the separation of ownership from decision-making in most major modern firms means that business objectives are frequently pursued which are different from simply profit maximization. Only shareholders have a desire for maximum profits in the short run. On the other hand, in modern multi-activity, multi-level, multi-plant, and multinational firm organizations, corporate decisions are the result of the many individual decisions made by a complex hierarchy of people, each with particular business objectives, and many of which are different from profit maximization. The reason is that the performance of different employees within a company is measured in different ways. For example, the directors' performance may be evaluated primarily by the firm's market share, whereas the sales manager's performance may be evaluated on sales growth. Similarly, the production manager's performance may be evaluated primarily by inventory throughput efficiency, whereas the personnel officer may be evaluated according to the number of days lost through industrial disputes. Given that each of these different decision-makers is evaluated on different criteria, the success, promotion, and consequently the wages earned by each of these workers will be evaluated differently. Therefore, the objectives pursued by different employees may be quite different from profit maximization. Under these conditions, the 'conflicting goals' critique suggests that firms will aim to 'satisfice'. In other words, the firm will aim to achieve a satisfactory level of performance across a range of measures. In particular, the firm will initially aim to achieve a level of profit sufficient both to avoid shareholder interference in directors' activities and also to avoid the threat of a takeover. Once this objective is achieved, the other various goals of the firm can be satisfied. For example, the firm may aim to achieve market share levels as high as possible without jeopardizing the efficiency cost gains associated with production and logistics operations. Equally, employees' pay may be increased in order to encourage firm loyalty. The point here is that the overall objective of the firm can be specified in various ways.

In Figure 1.20, the firm's total profit function $T\pi$ is constructed as the difference between the total revenue function *TR* and the total cost function *TC*. The firm may choose to produce at output levels Q_1, Q_2, and Q_3, which represent the minimum cost output, the maximum profit output, and the maximum revenue output levels, respectively. All of these levels of output produce a profit level sufficient to maintain a firm's independence π_s, but only one of these output levels Q_2 is the profit-maximization level of output.

From the point of view of location, if we have a set of spatial total cost and revenue curves, such as those described by Figure 1.20, the firm will make different location

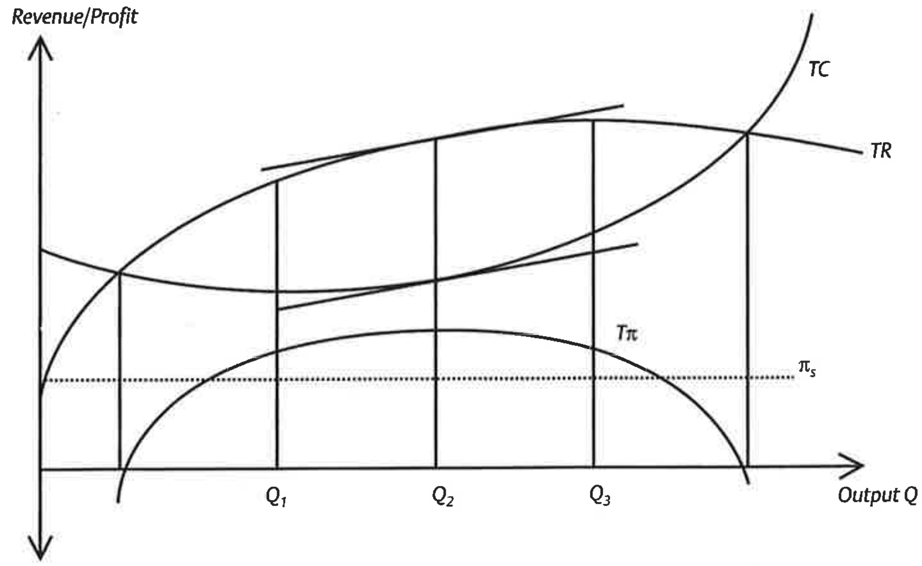

FIG. 1.20 Profit-maximizing, revenue-maximizing, and profit-satisficing

decisions, according to whether the firm is aiming to maximize profits in the short run or whether it is aiming to earn satisfactory profits in the short run along with achieving some other goals. For example, in Figure 1.21, if the firm is aiming to maximize profits in the short run it will locate at point *P*. On the other hand, if it is aiming to maximize sales it will locate at *S*, and if it is aiming to minimize production costs and to maximize production efficiency it will locate at *C*. If the firm had perfect information regarding these different spatial cost and revenue curves, we can argue that the firm will always move to point *P*. However, behavioural theories assume that information is imperfect. Given the limited information available and the conflicting goals within the organization, the actual location behaviour of the firm will depend on which is the particular dominant objective of the firm.

The third critique of the classical and neoclassical location models comes from the question of relocation costs. Relocation costs are the costs incurred every time a firm relocates. The models described above all assume that location is a costless exercise. However, relocation costs can be very significant, comprising the costs of the real-estate site search and acquisition, the dismantling, moving, and reconstruction of existing facilities, the construction of new facilities, and the hiring and training of the new labour employed. These significant transactions costs, along with imperfect information and conflicting goals, will mean that firms are unlikely to move in response to small variations in factor prices or market revenues. In Figure 1.21, the areas in which positive profits are made, i.e. where $TR > TC$, are known as 'spatial margins of profitability' (Rawstron 1958), and are represented by the areas between locations *a* and *b*, *c* and *d*, and *e* and *f*. The relationship between marginal location change and the profitability of the firm in these areas is given by $\delta(\mathrm{TR} - \mathrm{TC})/\delta\delta$, and this is represented by the differences in the slopes of the spatial revenue and spatial cost functions as location changes. In the spatial margins of profitability in which the slopes of the spatial revenue and spatial cost functions are very shallow, the marginal benefit to the firm of relocation will be very

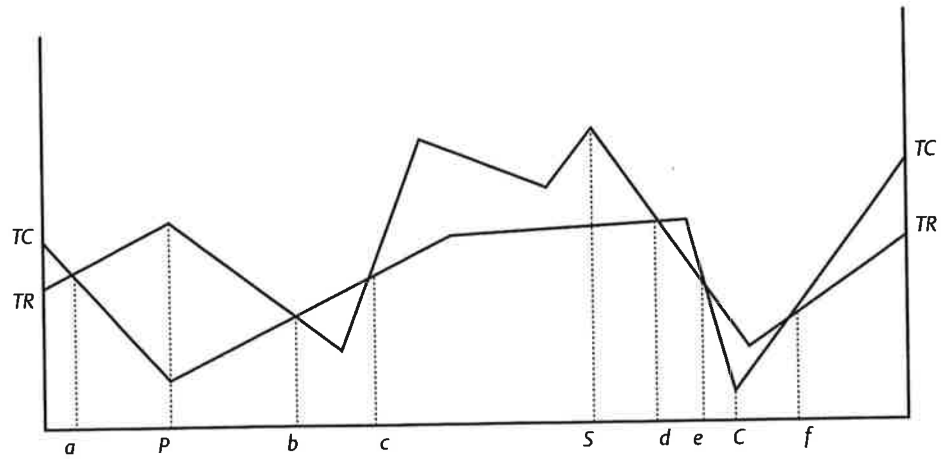

FIG. 1.21 Spatial cost and revenue curves

low. Therefore, in the presence of high relocation costs the firm will not move to a superior location even if the firm knows which alternative is superior. In conditions of imperfect information and bounded rationality, conflicting goals, and significant relocation costs, the behavioural approach would argue that once a firm has chosen a location, the firm will tend to maintain its location as long as profits are positive, and not use relocation as a competitive weapon. Rather, the firm will attempt to reorganize its factor allocations and activities between its current set of existing plants. At the same time, it will focus primarily on other price and non-price issues as competitive weapons, and the relocation of a plant, or the reorganization of multi-plant activities which involves either the closing or opening of a plant, will only be a last-resort strategy. On the other hand, where relocation costs are insignificant, the firm will be take advantage of spatial revenue and spatial cost differences and will be able to move to superior locations as a competitive strategy.

All of the behavioural critiques suffer from the weakness that, unlike the classical and neoclassical location models, the behavioural theories do not of themselves indicate why a firm chooses a particular location in the first place. In this sense the behavioural approach is not prescriptive. However, the classical models do need to be interpreted in the light of the behavioural critique of bounded rationality, imperfect information, conflicting goals, and relocation costs, as these are all features particularly characteristic of the spatial economy. This can be done by considering the evolutionary argument of Alchian (1950).

Alchian's argument is that the behaviour of firms in conditions of uncertainty can be understood by discussing the relationship between a firm and its environment, whereby a firm's environment is understood to encompass all the agents, information, and institutions competing and collaborating in the particular set of markets in which the firm operates. In Alchian's argument, we can characterize the uncertain economy by two broad types of environments. One is an 'adoptive' environment and the other is an 'adaptive' environment. These two classifications are not mutually exclusive, but serve as the two extreme stylized types, between which the real economy will exist.

In the 'adoptive' environment, all firms are more or less identical in that no firm has any particular or systematic information advantage over any other firms. The results of the competitive process will imply *ex post* that some firms will be successful while others will not, although *ex ante* no firms had any a priori knowledge that their products or techniques would be superior to those of their competitors. This characterization of the economy is Darwinian, in that the environment 'adopts' the firms which were better suited to the needs of the economy, even though the firms had no particular knowledge beforehand that this was the case. In statistical terms, in any given time period in the 'adoptive' environment, the probability of a particular single firm making a successful strategic decision is identical to that of all the other individual firms.

On the other hand, in the 'adaptive' environment, some individual firms are able to gather and analyse market information, simply by reason of their size. Large firms in general are able to utilize resources in order to acquire and process information relating to their market environment, and the purpose of these information-gathering activities by the firms is to subsequently use the information to their own advantage, relative to their competitors. In statistical terms, therefore, in any given time period in the 'adaptive' environment, the probability of a particular firm making a profitable strategic decision is increased by reason of its size.

In the real world of heterogeneous firms and imperfect information, smaller firms will tend to perceive themselves to be at an information disadvantage relative to larger firms. Therefore they will tend to make decisions which mimic or dovetail with those of the larger firms, in matters such as styles, protocols, formats, and technology. In part this is because they perceive the market leaders to be the best barometers of market conditions, and also because the behaviour of the market leaders itself often contributes significantly to the overall economic environment simply by reason of their size. By copying the behaviour of the larger firms the small firms therefore perceive that they will maximize the likelihood of their own success. The result is that large firms tend to overcome uncertainty by information-gathering and analysis, and small firms tend to overcome uncertainty by imitation.

This type of leader–follower behaviour is common in models of oligopoly and uncertainty. However, this behaviour is particularly pertinent to questions of location. In environments of uncertainty, larger firms will generally have the information and financial resources to make more considered location decisions than small firms. Major firms will be able to make location decisions more akin to those described by the Weber, Moses, and Hotelling models, given that they will generally have sufficient resources to evaluate the cost and revenue implications of their location choice. These large firms will attempt to make rational and optimal decisions, and the results of their location choices can be analysed by the types of classical and neoclassical models described above. On the other hand, small firms will generally be located where their founders were initially resident. There will have been no explicit initial location decision as such, when the firm began operating. Yet, over time, competition between firms will be partly a result of spatial differences in costs and revenues, and the relationship between profitability and location will eventually become a decision-making issue. In subsequent location decisions, many small firms will tend to choose locations close to the major market leaders for the reasons outlined by Alchian. Imitation therefore also takes place in terms of spatial behaviour. For firms which are risk-averse this is also a particularly good strategy, because as we see from the Hotelling model, locating close to competitor firms ensures that an individual firm's market share is no lower than that of an equivalent firm. Hotelling-type behaviour is therefore common for small firms clustering around major firms.

1.6 Conclusions

The foregoing discussions suggest that we can use classical and neoclassical models of location to consider the spatial behaviour of large firms, or firms in environments with good information and low relocation costs. At the same time, the discussion of the behavioural critique suggests that the leader–follower behaviour typical of many industries will tend to encourage small firms to cluster together in space close to larger firms. This process of industrial clustering, however, will lead to an increased demand for local land; consequently, local real-estate prices will tend to increase, as will local labour prices. These increases in the prices of local factor inputs will reduce profits, *ceteris paribus*, thereby reducing the attractiveness of the area as a location for the firms. This raises the question how long will the cluster of firms continue to exist profitably in the area? This question of industrial clustering is the topic of the next chapter, in which we discuss agglomeration economies, the growth of cities and urban hierarchies, and centre–periphery relationships.

Discussion questions

1. How does the location of input sources and output markets determine the location behaviour of the firm?
2. To what extent are firm-locational changes dependent on the substitution characteristics of the firm's production function?
3. In what ways can space confer monopoly power?
4. What role can location play in the competitive strategy of firms, and how are location and price strategies interrelated?
5. What role do logistics costs play in determining location behaviour?
6. What insights are provided for industrial location analysis by behavioural theories of firm behaviour?

Appendix 1.1 **The One-Dimension Location Problem**

Within the Weber framework, we can summarize the relative strength of the transportation 'pull' towards any particular input source point. If at any particular location, $\Delta(m_1t_1d_1) > -\Delta(m_2t_2d_2)$ as the firm moves away from input source 1 and towards input source 2 (where Δ represents a marginal change), the firm should move towards input source 1. This is because the marginal increase in the total transport costs for the shipment of input 1 is greater than the marginal fall in the total transport costs for the shipment of input 2. Alternatively, if $\Delta(m_1t_1d_1) < -\Delta(m_2t_2d_2)$, the firm will move closer towards input source 2. In the situation where $\Delta(m_1t_1d_1) = -\Delta(m_2t_2d_2)$, the firm can move in either direction, and will be indifferent between adjacent locations.

Within the Weber triangle, we can imagine a situation where the output good is weightless, such as in the case of the electricity generated by a power-station which consumes inputs of coal and coke from M_1 and M_2, respectively. In this case, the plant will be constrained to locate along the line joining M_1 and M_2. Here, the location problem becomes a one-dimensional problem. Initially we can analyse the situation where the transport rates are constant.

In this situation, any small change, denoted here by D, in the input shipment distance

d_1, will be associated with an equal and opposite change in the input shipment distance d_2. If $\Delta(m_1t_1d_1) > -\Delta(m_2t_2d_2)$, at any location along the line joining M_1 and M_2, as the firm moves away from input source 1, the firm will locate at M_1. Alternatively, if $\Delta(m_1t_1d_1) < -\Delta(m_2t_2d_2)$, at any location along the line joining M_1 and M_2, as the firm moves away from input source 1, the firm will locate at M_2. The reason for this is that if m_1, t_1, m_2, and t_2 are fixed, the only cause of change in the total transport costs for each input shipment is the change in the relative distances, which are always equal and opposite in this case. Therefore, the inequality which holds at any particular point on the line M_1M_2, will hold at all points on the line. This will encourage the firm to continue to move in the same direction. The optimum location behaviour of the firm is therefore to locate at the particular end-point M_1 or M_2 which has the lowest total transport costs. As we see in Figure A.1.1.1, in a one-dimensional space such as the line joining M_1 and M_2, where transport rates are constant, there is always an end-point optimal location solution, in this case at M_2. In microeconomics this is called a corner solution, because the optimum location will never be between, or interior to, the end points M_1 and M_2.

The situation becomes somewhat more complicated where the transport rates change with the distance of haulage. Transport rates per tonne-kilometre normally fall with increasing haulage distance, implying that the total input transport costs increase less than proportionately with the distance. On the other hand, in some circumstances transport rates per tonne-kilometre increase with distance, implying that total transport costs increase more than proportionately with distance. In these situations, we must also consider the effect of the change in transport rates with changes in distance, as the distance itself changes. As above, an optimal location for the firm will only be at an interior location, i.e. between the end points M_1 and M_2, where the marginal increase in the total transport costs for the shipment of input 1, as we move away from M_1, is equal to the marginal fall in the total transport costs for the shipment of input 2. If we denote the total transport costs associated with the input shipment of good 1 as $TC_1 = (m_1t_1d_1)$,

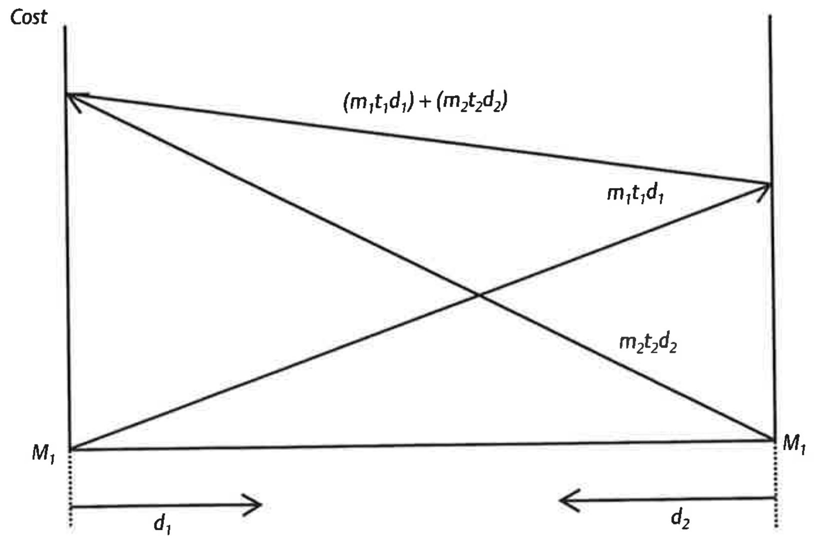

FIG. A.1.1.1 One-dimensional location problem with constant transport rates and those associated with the input shipment of good 2 as $TC_2 = (m_2t_2d_2)$, the condition for an optimum location internal to M_1 and M_2, is where $\Delta TC_1 = -\Delta TC_2$, for a small location change. In order to identify such a situation it is necessary to use calculus to observe the first derivative of each of the relationships between total transport costs and distance in the situation where transport rates vary with haulage distance. By partial differentiation

$$\delta(TC_1)/\delta d_1 = m_1(t_1 + d_1(\delta t_1/\delta d_1))$$

and:

$$\delta(TC_2)/\delta d_2 = m_2(t_2 + d_2(\delta t_2/\delta d_2)).$$

An interior optimum location is possible where

$$m_1(t_1 + d_1(\delta t_1/\delta d_1)) = -m_2(t_2 + d_2(\delta t_2/\delta d_2)).$$

$$m_1t_1 + m_2t_2 = -(d_1(\delta t_1/\delta d_1) + d_2(\delta t_2/\delta d_2)).$$

Given that the left-hand side is positive, the right-hand side must also be positive for an interior optimum location. This implies that at least one of the terms $(\delta t_1/\delta d_1)$ or $(\delta t_2/\delta d_2)$ must be positive. In other words, the marginal change in transport costs for at least one of the input goods must be increasing with distance. The transport costs associated with at least one of the inputs must be increasing more than proportionately with distance, as distance increases, in order for there to be an optimum location between M_1 and M_2. We can see this in Figure A.1.1.2 where the interior optimum is at d^*.

On the other hand, if transport rates exhibit economies of distance, i.e. $(\Delta t_1/\Delta d_1)$ and $(\Delta t_2/\Delta d_2)$ are both negative, or fixed transport rates, i.e. $(\Delta t_1/\Delta d_1)$ and $(\Delta t_2/\Delta d_2)$ are zero, there is no interior solution. As we see in Figures A.1.1.3 and A.1.1.1, in these cases, which are the usual two situations, the optimal location will always be at an end-point such as M_1 and M_2. In A.1.1.3 the optimum location is at M_1 whereas in A.1.1.1 it is at M_2.

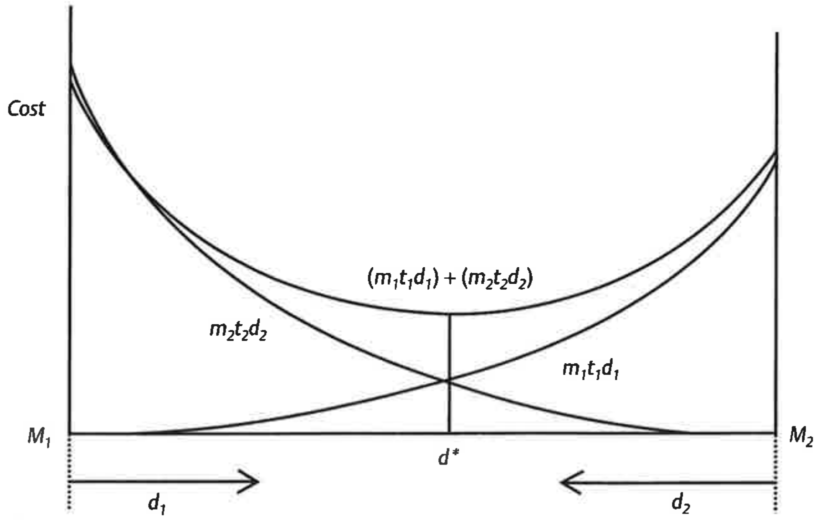

FIG. A.1.1.2 One-dimensional location problem with increasing transport rates

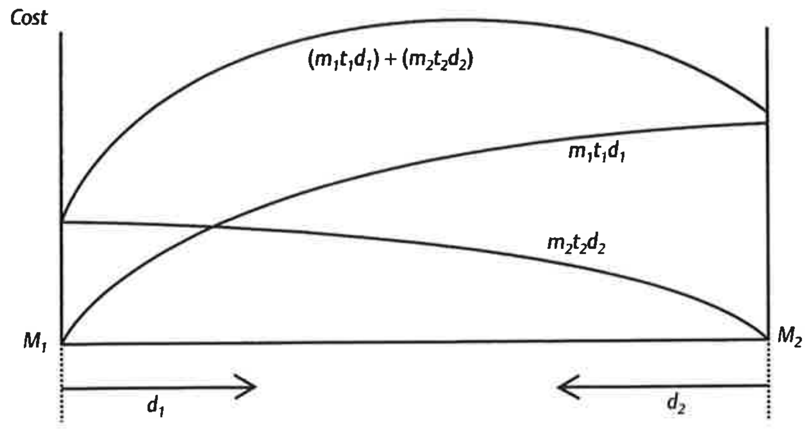

FIG. A.1.1.3 One-dimensional location problem with decreasing transport rates

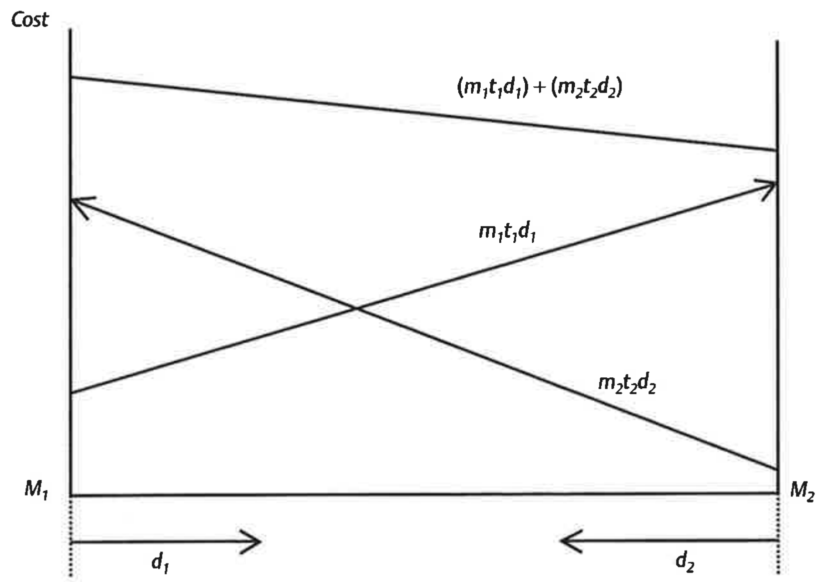

FIG. A.1.1.4 One-dimensional location problems with terminal costs and linear transport rates

The final possibility is where there are trans-shipment costs associated with the loading and unloading of goods at ports or terminals. In these situations, the 'terminal' costs associated with these trans-shipment points may alter the transport rates in a variety of ways. Optimal locations with terminal costs can be either at end-points or at interior locations. As we see in Figures A.1.1.4 to A.1.1.6, the optimal location will depend on the structure of the transport costs.

In Figure A.1.1.4. the transport rates are constant, although not equal to each other, and both transport cost functions exhibit terminal costs. In this case, the optimal location which minimizes total transport costs is at the end point M_2.

In Figure A.1.1.5, the transport rates are falling with distance, such that total transport costs are concave with distance, although they are not equal to each other. Both transport cost functions exhibit terminal costs. In this case, the optimal location which minimizes total transport costs is at the end point M_1.

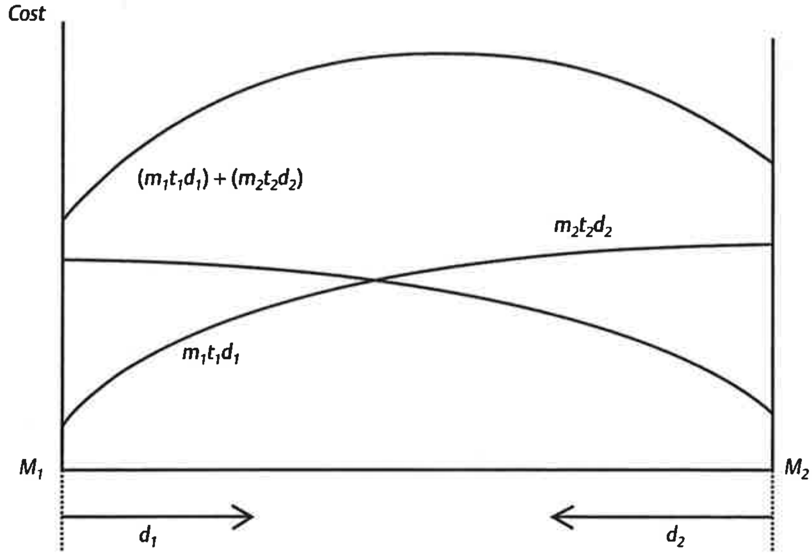

FIG. A.1.1.5 One-dimensional location problems with terminal costs and falling transport rates

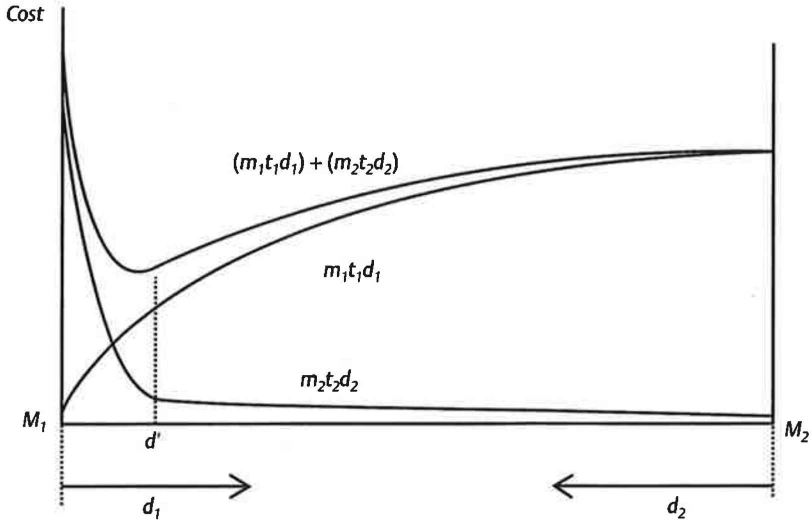

FIG. A.1.1.6 One-dimensional location problems with terminal costs and increasing and decreasing transport rates.

In Figure A.1.1.6, one of the transport rates is falling with distance, whereas the other is increasing with distance. Both transport cost functions exhibit terminal costs. In this case, the optimal location which minimizes total transport costs is at the interior location d'. The classic proof of the one-dimensional location problem is given by Sakashita (1968).

Appendix 1.2. The General Solution to the Weber–Moses Problem

We can write the profit (π) function of the firm as

$$\pi = p_3m_3 - (p_1 + t_1d_1)m_1 - (p_2 + t_2d_2)m_2 - t_3d_3m_3. \tag{A.1.2.1}$$

Any profit maximization production-location point will need to satisfy the optimization conditions both with respect to the input combinations, m_1 and m_2, and also the locational coordinates. In our Weber–Moses triangle (Figure 1.8) we can define the locational coordinates in terms of two variables, namely the angle q and the output shipment distance d_3. Any changes in the input distances d_1 and d_2 can be defined in terms of changes in these two variables. For an optimum location-production result, the partial derivatives of the profit function with respect to the four variables m_1, m_2, θ, and d_3, must be equal to zero. Following Miller and Jensen (1978), by partial differentiation, the first-order conditions for profit maximization are

$$\frac{\partial(\pi)}{\partial m_1} = -(p_1 + t_1d_1) - m_1d_1\left(\frac{\partial t_1}{\partial m_1}\right) - t_3d_3\left(\frac{\partial m_3}{\partial m_1}\right) - m_3d_3\left(\frac{\partial t_3}{\partial m_3}\right)\left(\frac{\partial m_3}{\partial m_1}\right) = 0. \tag{A.1.2.2}$$

$$\frac{\partial(\pi)}{\partial m_2} = -(p_2 + t_2d_2) - m_2d_2\left(\frac{\partial t_2}{\partial m_2}\right) - t_3d_3\left(\frac{\partial m_3}{\partial m_2}\right) - m_3d_3\left(\frac{\partial t_3}{\partial m_3}\right)\left(\frac{\partial m_3}{\partial m_2}\right) = 0. \tag{A.1.2.3}$$

$$\frac{\partial(\pi)}{\partial \theta} = -m_1t_1\left(\frac{\partial d_1}{\partial \theta}\right) - m_1d_1\left(\frac{\partial t_1}{\partial d_1}\right)\left(\frac{\partial d_1}{\partial \theta}\right) - m_2t_2\left(\frac{\partial d_2}{\partial \theta}\right) - m_2d_2\left(\frac{\partial t_3}{\partial d_2}\right)\left(\frac{\partial d_2}{\partial \theta}\right) = 0. \tag{A.1.2.4}$$

$$\begin{aligned}\frac{\partial(\pi)}{\partial d_3} = &-m_1t_1\left(\frac{\partial d_1}{\partial d_3}\right) - m_1d_1\left(\frac{\partial t_1}{\partial d_1}\right)\left(\frac{\partial d_1}{\partial d_3}\right) - m_2t_2\left(\frac{\partial d_2}{\partial d_3}\right) - m_2d_2\left(\frac{\partial t_2}{\partial d_2}\right)\left(\frac{\partial d_2}{\partial d_3}\right)\\ &- m_3t_3 - m_3d_3\left(\frac{\partial t_3}{\partial d_3}\right) = 0.\end{aligned} \tag{A.1.2.5}$$

Equations (A.1.2.2) to (A.1.2.5) can be rearranged to give

$$\frac{\partial(\pi)}{\partial m_1} = -\left(\frac{\partial m_3}{\partial m_1}\right)\left[t_3d_3 + m_3d_3\left(\frac{\partial t_3}{\partial m_3}\right)\right] - m_1d_1\left(\frac{\partial t_1}{\partial m_1}\right) - (p_1 + t_1d_1) = 0. \tag{A.1.2.6}$$

$$\frac{\partial(\pi)}{\partial m_2} = -\left(\frac{\partial m_3}{\partial m_2}\right)\left[t_3d_3 + m_3d_3\left(\frac{\partial t_3}{\partial m_3}\right)\right] - m_2d_2\left(\frac{\partial t_2}{\partial m_2}\right) - (p_2 + t_2d_2) = 0. \tag{A.1.2.7}$$

$$\frac{\partial(\pi)}{\partial \theta} = -m_1\left(\frac{\partial d_1}{\partial \theta}\right)\left[t_1d_1\left(\frac{\partial t_1}{\partial d_1}\right)\right] - m_2\left(\frac{\partial d_2}{\partial \theta}\right)\left[t_2 + d_2\left(\frac{\partial t_2}{\partial \theta}\right)\right] = 0. \tag{A.1.2.8}$$

$$\begin{aligned}\frac{\partial(\pi)}{\partial d_3} = &-m_1\left(\frac{\partial d_1}{\partial d_3}\right)\left[t_1 + d_1\left(\frac{\partial t_1}{\partial d_1}\right)\right] - m_2\left(\frac{\partial d_2}{\partial d_3}\right)\left[t_2 + d_2\left(\frac{\partial t_2}{\partial d_1}\right)\right]\\ &- m_3\left[t_3 + d_3\left(\frac{\partial t_3}{\partial d_3}\right)\right] = 0.\end{aligned} \tag{A.1.2.9}$$

Equations (A.1.2.6) and (A.1.2.7) together define the production relationships at the optimum between each of the inputs and the output. Meanwhile equations (A.1.2.8) and (A.1.2.9) together define the location of the firm at the optimum. To understand the production function characteristics which will ensure that the optimum location of the firm is independent of the level of output, we need to observe the conditions under which the marginal rate of substitution between the inputs remains constant for all levels of output. For an optimum location which is independent of the level of output, equations (A.1.2.8) and (A.1.2.9) must be satisfied because the firm will not move. Therefore, we need only observe the production relationships. By rearranging equations (A.1.2.6) and (A.1.2.7) we arrive at

$$\frac{\partial m_3}{\partial m_1} = \frac{-m_1 d_1\left(\frac{\partial t_1}{\partial m_1}\right) - (p_1 + t_1 d_1)}{\left[t_3 d_3 + m_3 d_3\left(\frac{\partial t_3}{\partial m_3}\right)\right]} \tag{A.1.2.10}$$

and

$$\frac{\partial m_3}{\partial m_2} = \frac{-m_2 d_2\left(\frac{\partial t_2}{\partial m_2}\right) - (p_2 + t_2 d_2)}{\left[t_3 d_3 + m_3 d_3\left(\frac{\partial t_3}{\partial m_3}\right)\right]}. \tag{A.1.2.11}$$

The term $(\delta m_3/\delta m_1)$ is the marginal product of input 1 at the optimum, and the term $(\delta m_3/\delta m_2)$ is the marginal product of input 2 at the optimum. Dividing (A.1.2.11) by (A.1.2.10) gives us an expression for the marginal rate of substitution between the two inputs at the optimum, thus

$$\frac{\partial m_3/\partial m_2}{\partial m_3/\partial m_1} = \frac{-m_2 d_2\left(\frac{\partial t_2}{\partial m_2}\right) - (p_2 + t_2 d_2)}{-m_1 d_1\left(\frac{\partial t_1}{\partial m_1}\right) - (p_1 + t_1 d_1)}. \tag{A.1.2.12}$$

The expression (A.1.2.12) is constant, i.e. the marginal rate of substitution between the inputs is constant, in the case where there are no economies of scale in transportation. In this situation, the ratio $(\delta t_1/\delta m_1) = (\delta t_2/\delta m_2) = 0$, and the expression reduces to

$$\frac{\partial m_3/\partial m_1}{(p_1 + t_1 d_1)} = \frac{\partial m_3/\partial m_2}{(p_2 + t_2 d_2)} \tag{A.1.2.13}$$

In other words, the marginal product of input 1 divided by the *delivered* price of input 1 is equal to the marginal product of input 2 divided by the *delivered* price of input 2. These conditions are the spatial equivalent of standard microeconomic efficiency conditions, and this general Weber–Moses result has also been shown to hold in the case of multiple inputs and outputs (Eswaran *et al.* 1981). Note that economies of distance play no role in the result, because once the firm is located at the optimum location, the distance relationships are all unchanging. Production functions which satisfy the conditions here are functions which are linear or which are homogeneous of degree one. On the other hand, in situations where transport rates exhibit economies of scale, the results of the Weber–Moses problem become much more restrictive.

Appendix 1.3 **The Logistics-Costs Location-Production Model**

An alternative broader specification of the distance costs associated with moving goods can be provided by including all of the inventory costs associated with the shipment of goods. The justification for this is that moving goods over space takes time, and between individual shipments, firms must hold inventories of goods to maintain supplies. The holding of these inventories itself incurs costs, so the firm must consider the relationship between the costs of moving goods, and the costs of not moving, i.e. holding inventories. If we include all of the costs associated with the holding of inventories plus the shipment of input good *i* over space, we can define the total logistics costs of the input shipments per time period as:

$$TLC_i = \frac{m_i}{Q_i} S_i + \frac{IQ_i(p_i + t_i d_i)}{2} + m_i t_i d_i, \qquad \text{(A.1.3.1)}$$

where the parameters m, p, t, d are the same as in the above sections, and Qi is the weight of an individual input shipment, Si is procurement costs of inputs, and I holding cost coefficient of input inventories.

The first terms on the right-hand side of (A.1.3.1) represents the ordering and procurement costs which are incurred each time an input shipment is received, but which are independent of the shipment size. In manufacturing firms, these costs will also include machinery set-up costs, and can be shown to be very significant. As these costs are independent of the shipment size but are incurred each time an input shipment is received, these total ordering costs are a multiple of the shipment frequency, i.e. the number of shipments per time period. The second term on the right-hand side represents the inventory capital holding costs, which are a function of the average value of inventories held per time period. These costs are the capital interest plus insurance costs associated with holding inventories. Assuming we consume inventories at a constant rate, and that stocks are replenished in a timely manner such that our inventory levels stay constant, these costs can be seen to be a function of the delivered price of the goods. In other words, inventory costs are a function of transport costs. Finally, the third term on the right-hand side of (A.1.3.1) represents the familiar transport costs term used above.

Using a similar logic we can also define the logistics-costs associated with shipments of output goods, denoted with the subscript '*o*', as

$$TLC_0 = \frac{m_0}{Q_0} S_0 + \frac{IQ_0(p_0 - t_0 d_{0i})}{2} + m_0 t_0 d_0. \qquad \text{(A.1.3.2)}$$

In this case, the capital costs associated with holding inventories are the opportunity costs of output revenue which are incurred by not shipping outputs in a continuous manner and selling them at the market price of p_o.

For both input and output goods, the aim of the firm is to determine the optimum shipment size Q^* which minimizes the sum of the total logistics costs for any given locational arrangement of input supply points and output markets. However, this is not as straightforward as it might initially appear, because it can be shown that while the optimal shipment size Q^* is a function of the transport rates, transport rates are also a function of the optimal shipment size. In order to circumvent these problems, the definition of transport rates must be respecified to allow for discrete shipments. While the details of this problem are beyond the scope of this book (McCann 1993, 1998, 2001), it can be demonstrated under very general conditions that the optimal shipment size (McCann 2001), and therefore the optimal average weight of inventories $Q^*/2$ to be held, is a positive function of the distance of the shipment, and the transport costs associated with the shipment. This also implies that the ordering costs, which are an inverse multiple of the size of Q, are also a function of the transport costs. The combined sum of all the interrelated components of total logistics costs can be shown to be a concave function of distance transport costs (McCann 1998), with distance cost curves similar in shape to those in Figure A.1.1.5.

By employing this broader logistics-costs description of the costs associated with transporting goods over geographical distances, we can now re-evaluate the Weber–Moses problem within a logistics-costs framework. Under very general conditions (McCann 1993, 1998), we can show that there is no solution to the independent of output optimum location problem. Moreover, as we see in Figure A.1.3.1, as the value added by a firm increases, the optimum location of the firm K^* moves towards the market.

As the value added by the firm increases, or the higher up the input–output value-chain is the firm, the steeper is the negatively sloping inter-regional equilibrium wage gradient, which would allow a firm to move away from a central market point (McCann 1997, 1998). We can see this in Figure A.1.3.2, in which the point M represents a location containing both markets and input source points.

The inter-regional equilibrium wage gradient associated with total logistics costs, rather than simply transport costs, becomes steeper, i.e. it changes form R_1 to R_2, as the value added by the firm increases, or as the firm moves up the value-adding chain. In other words, in order to encourage a firm to move away from a Weber optimum location at M where wages are w_M, inter-regional wage differences will need to be greater in order for a high value-adding firm to relocate. On the other hand, a low value-adding firm, or a

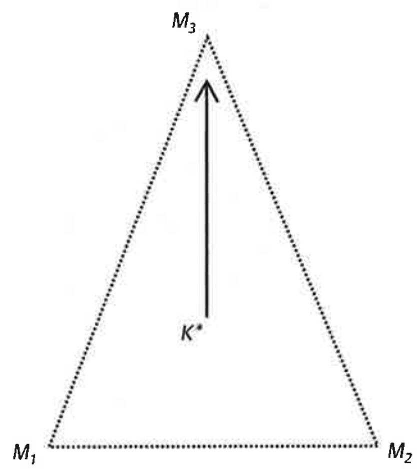

FIG. A.1.3.1 Logistics-costs optimum location and value added by the firm

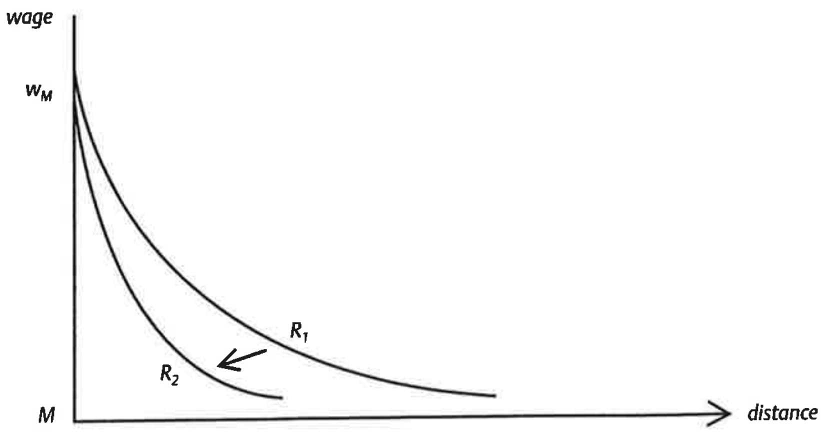

FIG. A.1.3.2 Inter-regional equilibrium wage gradient associated with logistics-costs

firm lower down the value-adding input–output chain, will be able to move in response to relatively minor inter-regional wage differences (McCann 1997, 1998). Moreover, in Appendix 3.4, knowing that the optimum shipment size Q^* is inversely related to the trip frequency f, we are able to show that the inter-regional equilibrium wage gradient must be convex with respect to distance, as in Figure A.1.3.2.

The overall location-production conclusions to come out of the logistics-costs approach to the Weber–Moses framework is therefore that high value-adding firms tend to be both more market-oriented, and much less responsive to regional wage differences, i.e. they are much less footloose, than low value-adding firms.

Appendix 1.4 **The Hotelling Location Game**

Before a Hotelling game of spatial competition takes place along a market OL, as we see in Figure A.1.4.1, we assume that we have two firms A and B, with A located to the left of B. The distance from O to A is denoted as a, and the distance from L to B is denoted as b. The distance from O to the market boundary is denoted as x', and the total distance OL is denoted as d. In order for this duopoly to exist there must be three conditions satisfied. The first condition is that a consumer located at point O must always buy from firm A. In other words, the delivered price of the output of A at O must always be less than the delivered price of the output of B at O. This can be written as

$$P_A + t_A a < P_B + t_B(d - b). \qquad \text{(A.1.4.1)}$$

Secondly, a consumer located at point L must always buy from firm B. In other words, the delivered price of the output of B at L must always be less than the delivered price of the output of A at L. This can be written as

$$P_B + t_B b < P_B + t_A(d - a). \qquad \text{(A.1.4.2)}$$

At the same time, thirdly, there must also be an indifferent consumer at a distance x'

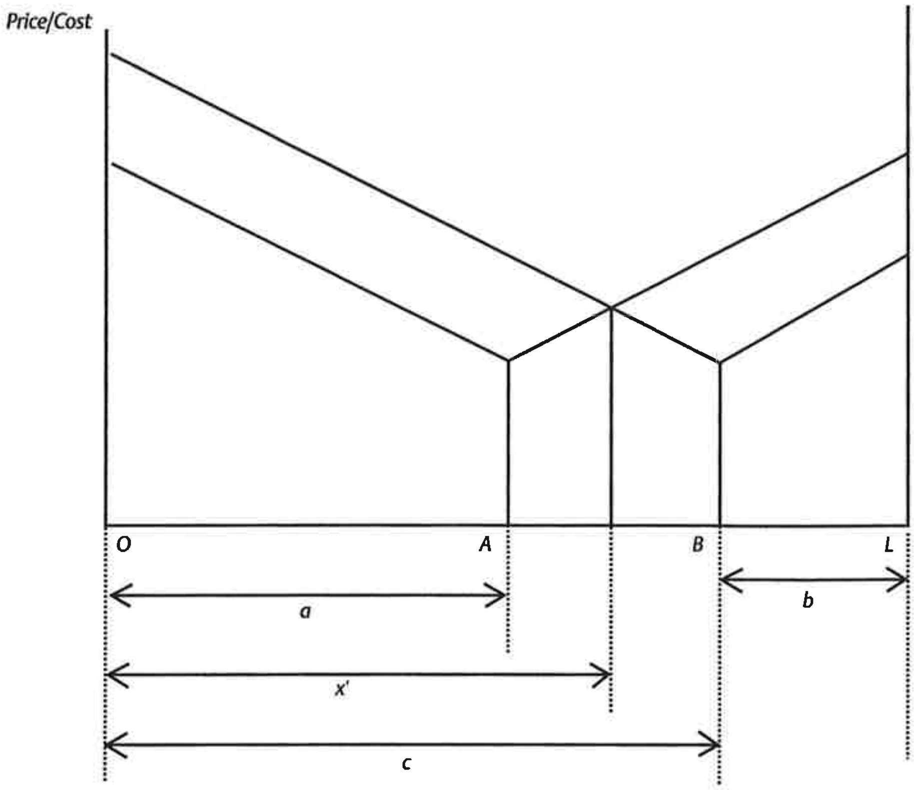

FIG. A.1.4.1 The Hotelling spatial framework

somewhere between A and B. For this indifferent consumer the delivered prices must be the same. In other words

$$P_A + t_A(x' - a) = P_B + t_B(d - b - x'). \tag{A.1.4.3}$$

If we set $t_A = t_B$, then rearranging equation (A.1.4.3) gives

$$P_A - P_B + 2tx' - ta - td + tb = 0. \tag{A.1.4.4}$$

If the transport rates t_A and t_B for the two firms are the same, and the source prices, P_A and p_B, of the two firms are also the same, we have

$$x' = \frac{a + d - b}{2}. \tag{A. 1.4.5}$$

The value of x' given in equation (A.1.4.5) represents the size of the market captured by firm A, and the size of the market captured by firm B can thus be represented as

$$d - x' = d - \left(\frac{a + d - b}{2}\right) = \frac{b + d - a}{2}. \tag{A. 1.4.6}$$

Recalling from Figure A.1.4.1 that $c = (d - \mathrm{b})$, we can rewrite (A.1.4.5) as

$$x' = \frac{a + d - (d - c)}{2} = \frac{a + c}{2}$$

As such, if the transport rates are the same and also the product source prices are the same, the boundary between the two firms is exactly halfway between the two firms, and is independent of the transport rates, as we would expect.

For a given source production price P_A, known as a 'mill' price, the market revenue of firm *A* depends on maximizing the value of x'. From (A.1.4.5) and the arguments in section 1.4.1, we see that this is increased by increasing a and reducing b as much as possible, while still ensuring that firm *A* is to the left of firm *B*. This location change then triggers the leapfrogging behaviour described in section 1.4.1.