Form Follows Function: Reformulating Population Density Functions

A simple isolated bit of evidence, however striking, is always open to doubt. It is the accumulation of several different lines of evidence that is compelling. (Crick, 1990, p. 37.)

9.1 Cities as Population Density Functions

We could have written this book by beginning with existing theories of city size, shape and distribution and then gradually showing how fractal geometry could be used to reinterpret, generalize and extend the body of theory which geographers, economists and planners have been working with for the last hundred years. But instead we chose a different tack, introducing fractal geometry first and tracing out its implications for how cities might be organized, before we embarked upon the ways in which our theory of the fractal city might link to the mainstream of urban studies. Without our having reviewed the multitude of urban theories in other than cursory terms, it is already clear that fractal geometry has appealing properties with respect to cities in ideas concerning space-filling, self-similarity and density. In fact as Crick (1990) implies above, although these lines of evidence for a fractal theory of cities are highly suggestive, they become compelling when it is realized that much of what has been developed in the mainstream is entirely consistent with fractal geometry.

We are now at the point where we can begin to make these connections, in this chapter to the distribution of population and other activities within the city, in the next to the distribution of population sizes and shapes across cities. In fact as we have implied throughout this book, our foray into the fractal geometry of cities is but a beginning and we anticipate that entire areas of urban theory might be reworked using fractals in the coming years. Here we will confine our efforts to the most basic applications starting with simple gravitational density models which constitute the heart of what has been called social physics. Ever since Newton published his celebrated Laws of Motion in the late 17th century have there been attempts to apply classical physics to social systems in general, city systems in particular.

Social forces between populations in space were described explicitly by Carey in the mid-19th century using gravitational analogies, while the notion that the density of economic activity declines with increasing distance from its market is implicit in the work of von Thunen, circa 1830 (Hall, 1966). In the study of population densities, it is now just over 100 years since Bleicher (1892) wrote: "The rapid decrease in population density with distance from the center is highly characteristic of an old city such as ours" (Edmonston, 1975; Mogridge, 1984).

However another 50 years were to elapse before these observations were associated with specific mathematical functions. There is some evidence that such functions were being used in the 1940s (Ajo, 1944), but it is Colin Clark (1951, 1967) who is accredited with the first use of the negative exponential function as the basic model for population densities. Since then, many variants of this model have been developed. The negative exponential density function is consistent with the theories of strict utility-maximizing associated with urban economic theory (Beckmann, 1969; Muth, 1969), while the development of operational urban models based on entropy-maximizing (Wilson, 1970) and discrete choice theory (Anas, 1982) which we introduced, albeit briefly, in Chapter 4, make widespread use of such functions. Although early analogies with gravitational models based on the inverse square function of distance formed the foundations of social physics in the 1940s and 1950s (Stewart, 1941, 1950), these power functions were quickly replaced with the negative exponential in the 1960s and 1970s due to the analytical convenience of such functions in problems of economic and statistical optimization, as well as to the elegance of the mathematical forms produced.

These developments, however, appear to have lost sight of the fact that the density and flow of economic activity across space must be fundamentally constrained and thus determined by the geometrical properties of their physical systems. It is clear that there is much misunderstanding of these issues, as Stewart (1950) was never slow to point out. Others such as Coleman (1964) who have considered the foundations of social physics have reinforced the point that power functions in their inverse form are the most obvious ones which embody the physical properties of the correct systems of interest (Batty and March, 1976). But such views have never attracted much attention and have remained apart from the mainstream of urban analysis during the last three decades. What a fractal theory of cities offers is a coherent approach to these earlier traditions.

Here we will argue forcibly that the use of the negative exponential function as a model of population densities is fundamentally flawed. We will argue that its use is based on its convenience in problems of optimization, on the elegance of its mathematical properties, not on its appropriateness to empirical data, and certainly not upon our ability to make sensible interpretations of its parameters. Our thesis here is based on the identification of appropriate scaling laws for urban systems based on ideas associated with allometry, although our principal concern will be with linking these laws to the principles of fractal geometry which shows how form follows function. As we demonstrated in Chapter 7, scaling laws based on power functions have been given a new lease of life of late in that their parameters can now be unambiguously associated with the size, shape and

form of the systems they describe. In short, their parameters can be associated with the extent to which their systems exploit the space in which they exist, the fractal dimension being directly associated with these parameters and with the extent to which the form fills the space available. The form or morphology of the system is no longer a prerequisite to models of its functioning, but a consequence of the way the system works in space. In this sense then, we will be dealing with systems in which 'form follows function'.

This chapter will attempt to weave several diverse themes together. Once again, we will note the work on urban allometry in which the size of the urban system is related to the space it occupies (Stewart and Warntz, 1958; Dutton, 1973), we will show how densities based on power functions arise as a natural consequence of such allometry, and we will cast our argument in terms of the new physics of fractional dimension (Davies, 1989) which uses the empirical ideas of Chapter 7 explicitly and the DLA and DBM models implicitly. We will first examine the properties of the negative exponential and inverse power functions, showing that the inverse power function has certain properties for describing urban population densities which have hitherto been overlooked. These properties can be easily exploited through scaling laws appropriate to urban systems, and accordingly, we show how urban form follows quite naturally from such functional descriptions. In this sense, we are also able to show that the density parameters associated with these functions are directly related to the fractional or fractal dimensions of the space within which these city systems exist.

These arguments have profound implications for the way we should measure population density in cities, and thus we then set out to estimate these functions in a variety of forms. The data set we use is for the city of Seoul in South Korea for the year 1982 (Kim, 1985); four variants on the base data set are used in estimating the density parameters of inverse power functions which we also represent in four related forms. These forms are estimated using three methods: the first two – regression of their log–linear forms, and the so-called signature of the underlying fractal growth process – were introduced in Chapters 7 and 8, but a third, conventional method based on entropy-maximizing, is also introduced here. The many variants of data and estimation which we use produce a range of density parameters and fractal dimensions which show remarkable consistency. From the theoretical arguments developed in the last two chapters, we argue that the fractal dimension of a typical city such as Seoul will have a value between 1 and 2, probably between 1.5 and 1.8, and a density parameter between 0.2 and 0.5. We show that this is indeed the case, and this suggests that many previous estimates of population density functions should be reworked with power functions to reveal parameter values which if consistent, could be used to derive more meaningful taxonomies of city size and shape.

9.2 Exponential Functions of Urban Density

Clark's (1951) original model assumes that population density $\rho(r)$ at distance r from the center of the city ($r = 0$) declines monotonically according to the following negative exponential

$$\rho(r) = K \exp(-\lambda r), \tag{9.1}$$

where K is a constant of proportionality which is equal to the central density $\rho(0)$ and λ is a rate at which the effect of distance attenuates. Note that in this chapter we will use λ for the density parameter of the negative exponential and α for the inverse power function. β will be used as previously as a general regression parameter of various scaling relations from which the fractal dimension D can be derived. As previously, we will use the variable r to indicate a variable distance from the CBD, whereas R will be used to define the accumulation of distance which is an integral of r up to the radius R. r should not be confused with its use as a measure of scale in earlier chapters, but its use should be obvious from the context.

Clark's (1951) paper was wide-ranging, idiosyncratic and brilliant; he not only fitted the log transform of equation (9.1) to over 20 cities using linear regression but also speculated on how the parameter λ changed through time, thus charting from the beginning, the direction for all subsequent empirical work on urban density functions. It is still not clear why Clark chose the negative exponential function for he was unaware of Ajo's (1944) doctoral dissertation in which a negative exponential function was used to model traffic flow in the Finnish city of Tampere. It is likely, however, that this function was chosen because it was popular in mathematical economics in models of capital depreciation and the like of which Clark was probably aware.

In the parallel development of social physics, inverse power functions of distance were being widely exploited in gravitational models of traffic flow and in rank-size relations. But there were no attempts to collapse such gravitational models to models of population density until the early 1960s when Smeed (1961, 1963) suggested a suitable such model might be based on

$$\rho(r) = K r^{-\alpha}, \tag{9.2}$$

where K is the constant of proportionality as in equation (9.1) above (but not defined where $r = 0$) and α is the parameter on distance. However, by this time, there was already a major groundswell in urban economics and transport modeling seeking to replace power functions with the negative exponential. The use of log–linear utility functions in urban economics from the work of Alonso (1964) on, and the development of gravity models using entropy-maximizing (Wilson, 1970) both led to the use of the negative exponential. The generalization of spatial interaction models into location as well as traffic distribution models and their collapsing to a single origin and many destinations led quite naturally to Clark's (1951) model (Bussiere and Snickars, 1970). Furthermore, important empirical work by Mills (1970), by Bussiere (1972b) and Bussiere and Stovall (1981), and by Mogridge (1984)

drowned out any further attempts to model population densities by anything other than negative exponential functions.

So far, the body of work concerned with population density models is almost exclusively based on the negative exponential function or its generalization. As Zielinski (1979) points out, attempts at demonstrating the precedence of one function over another in this field are quite inconclusive and in the absence of further empirical evidence, theoretical considerations suggest the negative exponential function as being superior. There have been some dissenting voices. Smeed's (1961) work has been noted, and Blumenfeld (1972) (summarized in Vaughan, 1987) and Saviranta (1973) have both argued for inverse power functions of density. However, there has been more work at seeking a generalized function such as the gamma whose typical form is

$$\rho(r) = Kr^{-\alpha} \exp(-\lambda r), \tag{9.3}$$

where the parameters K, α and λ fulfill the same roles as in equations (9.1) and (9.2). Tanner (1961), March (1971) and Angel and Hyman (1976) have suggested that functions such as (9.3) are general enough to encompass the various debates about one functional form or the other. Finally, it is worth noting that Parr (1985a, b) amongst others has suggested that the negative exponential function is more appropriate for describing density in the urban area itself, while the inverse power function is more appropriate to the urban fringe and hinterland. In later work, Parr and his colleagues have argued that the most appropriate generalized function of density is the log–normal (Parr, O'Neill and Nairn, 1988: Parr and O'Neill, 1989); this relates to processes which generate such functions, which in turn relate to the use of such functions in fractal geometry.

The most obvious interpretation of the parameter λ in the negative exponential can be made by taking the first derivative of equation (9.1). Then

$$\frac{d\rho(r)}{dr} = -\lambda K \exp(-\lambda r)$$
$$= -\lambda\rho(r), \tag{9.4}$$

and thus

$$\frac{d\rho(r)}{\rho(r)} = -\lambda dr, \tag{9.5}$$

which implies that λ is the percentage change in density for a small change in distance dr. In this sense then, the parameter plays a crucial role in the calculation of elasticities in the fully-fledged urban economic theory of the housing market which gives rise to this function (Muth, 1969).

To find an explicitly spatial interpretation of the parameter λ in equations (9.4) and (9.5), we need to examine how the density function behaves with respect to distance. There are two ways of proceeding. First, we will consider $\rho(r)$ as a one-directional function on a single line of distance from the center $r = 0$, and second we will consider how the same function behaves in the field around its center which involves two directions; this latter possibility involves formulating the density in polar coordinates as $\rho(r, \theta)$ where

r is the line of distance and θ is the angular variation, both about the center. Note that we will use the terms one-directional and two-directional only in those contexts where it is necessary to draw distinctions between direction and dimension. In all other contexts, the term dimension will suffice for both (Batty and Sikdar, 1982).

We will first examine the one-directional function. Note that in the sequel, we will normalize the density over its space to sum to unity. Then

$$\int_0^{R_b} \rho(r) \, dr = \int_0^{R_b} K \exp(-\lambda r) \, dr = 1, \tag{9.6}$$

where R_b marks the distance to the boundary of the spatial system. Evaluating equation (9.6) gives the constant K as

$$K = \frac{\lambda}{1 - \exp(-\lambda R_b)}, \tag{9.7}$$

and therefore the density becomes

$$\rho(r) = \frac{\lambda \exp(-\lambda r)}{1 - \exp(-\lambda R_b)}. \tag{9.8}$$

Clearly when $R_b \rightarrow \infty$, the value of $K = \lambda$. The cumulative distribution of $\rho(r)$, $N(R)$, is given as

$$N(R) = \int_0^R \rho(r) \, dr$$

$$= \frac{1 - \exp(-\lambda R)}{1 - \exp(-\lambda R_b)} \approx 1 - \exp(-\lambda R). \tag{9.9}$$

However, it is the mean density in the city which is of most interest, and this is defined as

$$C(R_b) = \int_0^{R_b} \rho(r) \, r \, dr$$

$$= \frac{1}{\lambda} - \frac{R_b \exp(-\lambda R_b)}{1 - \exp(-\lambda R_b)}. \tag{9.10}$$

In equation (9.10), if $R_b \rightarrow \infty$, then $C(\infty) = 1/\lambda$ and the interpretation of λ is as an inverse measure of the mean density, thus controlling the spread of the function in the linear direction.

The same type of analysis as presented in equations (9.6) to (9.10) can be developed for the two-directional model which we stated earlier in polar coordinates as $\rho(r, \theta)$. The density function is now given as

$$\rho(r, \theta) = K \exp(-\lambda r), \tag{9.11}$$

and its normalization as

$$\int_0^{2\pi} \int_0^{R_b} \rho(r, \theta) \, r \, d\theta \, dr = 2\pi K \int_0^{R_b} \exp(-\lambda r) \, r \, dr = 1. \tag{9.12}$$

The normalizing constant K can be evaluated as

$$K = \frac{\lambda^2}{2\pi \, \{1 - (1 + \lambda R_b) \exp(-\lambda R_b)\}}. \tag{9.13}$$

From equation (9.13) when $R_b \to \infty$, then $K = \lambda^2/2\pi$. The cumulative distribution function $N(R)$ is

$$N(R) = \frac{1 - (1 + \lambda R) \exp(-\lambda R)}{1 - (1 + \lambda R_b) \exp(-\lambda R_b)}$$

$$\approx 1 - (1 + \lambda R) \exp(-\lambda R). \tag{9.14}$$

and the mean density

$$C(R_b) = \int_0^{2\pi} \int_0^{R_b} \rho(r, \theta) \, r^2 \, d\theta \, dr$$

$$= \frac{2}{\lambda} - \frac{\lambda R_b^2 \exp(-\lambda R_b)}{1 - (1 + \lambda R_b) \exp(-\lambda R_b)}. \tag{9.15}$$

As $R_b \to \infty$, $C(\infty) = 2/\lambda$, thus implying that λ is also an inverse measure of mean density controlling the spread of the function over the two directions of space (Batty, 1974).

When the boundary distance $R_b \to \infty$, $\lambda C = 1$ for the one-directional function and $\lambda C = 2$ for the two-directional. In this context, it is possible to associate direction with dimension, the dimension of the spatial system entering the calculation of the parameter λ directly. To anticipate an argument of a later section, the spread of development is unlikely to be over the entire two-dimensional space, but it is likely to be spread over more area than the one-dimensional line. Therefore it is likely that $1 < \lambda C < 2$, and thus the value of λC is a measure of the extent to which the density fills two-dimensional space, if and only if R_b defines a good approximation to the spread across space of the density function. We will now develop the same analysis as contained in equations (9.6) to (9.10) and (9.11) to (9.15) for the inverse power function.

9.3 Power Functions of Urban Density

An immediate interpretation of the parameter α in the scaling function in equation (9.2) is provided by its first derivative

$$\frac{d\rho(r)}{dr} = -\frac{\alpha}{r} \rho(r) = -\alpha K^{-\alpha-1}, \tag{9.16}$$

which can be written as

$$\frac{d\rho(r)}{\rho(r)} \bigg/ \frac{dr}{r} = -\alpha, \tag{9.17}$$

where α is clearly an elasticity, the ratio of the percentage change in density $d\rho(r)/\rho(r)$ to the percentage change in distance dr/r. In equations (9.16) and (9.17), α has a more precise interpretation than the negative exponential parameter λ in equations (9.4) and (9.5), but its role is similar.

The difficulty with the inverse power function emerges when it is normalized. As previously we will begin with the one-directional form of the

model and then repeat the analysis for two. The normalization of equation (9.2) is given by

$$\int_0^{R_b} \rho(r) \, dr = K r^{-\alpha} \, dr = 1. \tag{9.18}$$

The integral in equation (9.18) cannot be evaluated when $r = 0$, for the function $\rho(0)$ is infinite at this value. This has been one of the main reasons for researchers preferring the negative exponential. However, the problem can easily be dealt with by translating the origin of system to a value of $r > 0$, $r = 1$ being the obvious lower limit. Some have argued that population density is not defined at the center of the city in any case or that population density at a point is meaningless although neither of these assumptions are necessary to the shift of the origin which is arbitrary. Evaluating the integral in equation (9.18) from $r = 1$ to $r = R_b$ gives

$$K = \frac{1 - \alpha}{R_b^{1-\alpha} - 1} \tag{9.19}$$

and

$$\rho(r) = \frac{(1 - \alpha) r^{-\alpha}}{R_b^{1-\alpha} - 1}. \tag{9.20}$$

As Zielinski (1979) points out, the normalization in equations (9.19) and (9.20) is still "analytically awkward", and further indefinite integrals are impossible to evaluate. With a little simplification, however, it is possible to proceed. First if the boundary of the system $r = R_b$ is much greater than $r = 1$, then it is possible to ignore the lower limit; second if it is assumed that α lies between 0 and 1, then equation (9.20) will always act as an inverse power function. Moreover, if $\alpha \geq 1$, then the function breaks down, and is no longer meaningful as a model of population density. Thus the use of equation (9.20) implies that there are tight bounds on the value of α. With these assumptions, the cumulative density function can be written as

$$N(R) = \int_1^R \rho(r) \, dr$$

$$= \frac{R^{(1-\alpha)}}{R_b^{(1-\alpha)}} \approx R^{(1-\alpha)}. \tag{9.21}$$

Finally the mean density $C(R_b)$ is evaluated as

$$C(R_b) = \int_1^{R_b} \rho(r) \, r \, dr$$

$$= \frac{1 - \alpha}{2 - \alpha} R_b. \tag{9.22}$$

and this shows that the parameter α depends directly upon the value of the system boundary R_b. From equation (9.22), the parameter α can be computed as

$$\alpha = \frac{2C(R_b) - R_b}{C(R_b) - R_b}, \tag{9.23}$$

but this is still an approximation whose relevance cannot be judged in advance of any particular application. However, equation (9.23) does provide some bounds on the value of α, namely that as $R_b > C(R_b)$, then $\alpha \leq 1$. The analysis for the two-directional model follows directly. The normalization is given as

$$\int_0^{2\pi} \int_1^{R_b} \rho(r, \theta) \, r \, dr \, d\theta = 2\pi K \int_1^{R_b} r^{-\alpha} \, r \, dr = 1, \tag{9.24}$$

from which K can be evaluated as

$$K = \frac{2 - \alpha}{2\pi \left(R_b^{2-\alpha} - 1 \right)}. \tag{9.25}$$

Assuming that $R_b \gg 1$, then the density $\rho(r, \theta)$ becomes

$$\rho(r, \theta) = \frac{(2 - \alpha) \, r^{-\alpha}}{2\pi R_b^{2-\alpha}} \tag{9.26}$$

and the cumulative density is

$$N(R) = \int_0^{2\pi} \int_1^{R} \rho(r, \theta) \, r \, dr \, d\theta$$

$$= \frac{R^{2-\alpha}}{R_b^{2-\alpha}} \approx R^{2-\alpha}. \tag{9.27}$$

The mean density $C(R_b)$ is now evaluated as

$$C(R_b) = \int_0^{2\pi} \int_1^{R_b} \rho(r, \theta) \, r^2 \, dr \, d\theta$$

$$= \frac{2 - \alpha}{3 - \alpha} \, R_b, \tag{9.28}$$

and this has the same structure as equation (9.22) for the one-directional model. It is possible to compute α from equation (9.28) as

$$\alpha = \frac{3C(R_b) - 2R_b}{C(R_b) - R_b}, \tag{9.29}$$

and from this, it is clear that α is unlikely to be greater than 2. The real significance of equations (9.23) and (9.29) however is that they show that the parameter α of the inverse power model is critically dependent on the boundary of the system R_b, and in certain empirical contexts, it will be possible to estimate α from these equations.

Although the differences between the two functions elaborated through equations (9.4) to (9.15) and (9.16) to (9.29) are substantial in mathematical terms, in practice both have been shown to fit empirical data equally well (or equally badly). In this section we will seek to demonstrate this. In Figure 9.1, one-dimensional negative exponential and inverse power functions from equations (9.8) and (9.20) respectively are plotted up to a regional boundary $R_b = 50$ units of distance. The parameters λ and α have been chosen to fit a mean density $C(R_b) = 20$ units; values of $\lambda = 0.025$ and
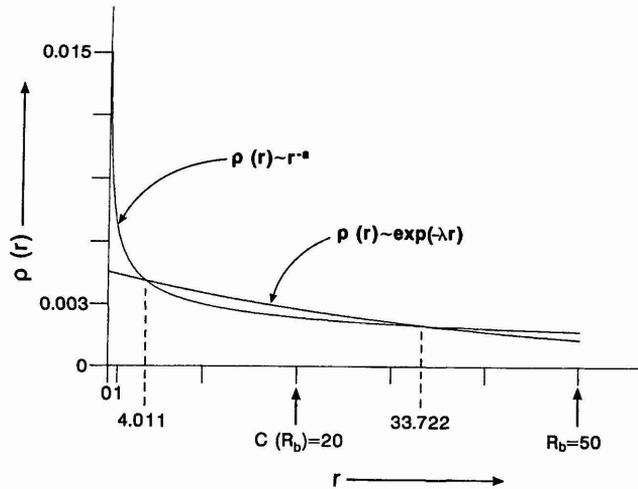
Figure 9.1. A comparison of the negative exponential and inverse power functions.

$\alpha = 0.349$ have been estimated which give simulated mean densities for both functions within 0.01% of the predetermined mean. Before we explore these functions further, it is worth noting for the inverse power function that if equation (9.23) is used to predict α, which is based on the assumption that the lower limit of r is 1, then the value of this parameter is 0.333 which is within 5% of the computed value.

No simple solution for the parameter λ exists for the negative exponential model. A first approximation to λ based on $1/C$ gives 0.05 while the use of the two-dimensional equation $2/C$ gives 0.1. The actual computed value is half the one-dimensional approximation, thus implying that the negative exponential over the given range from $r = 0$ to $r = 50$ provides only a partial approximation to its overall spread. This can be easily seen in Figure 9.1 where the inverse power function is much steeper in the area of its origin compared to the negative exponential which has a more gradual slope over the same distance to which it is applied. A more direct comparison of these functions can be made over their intermediate ranges where, from Figure 9.1, it is clear that the functions are similar. These functions will cross at $r = 4.011$ and $r = 33.722$, which are solutions to the equation of the two model functions $\{\lambda \exp(-\lambda r)/[1 - \exp(-\lambda R_b)]\} = \{(1 - \alpha)r^{1-\alpha}/R_b^{1-\alpha}\}$. It is between these two values that the functions are similar.

From Figure 9.1, it is quite clear that both functions are likely to be poor predictors of the central densities in the neighborhood of the origin $r = 0$. The inverse power function will overpredict while the negative exponential is unlikely to give a good prediction for there is little flexibility in controlling the value it takes when $r = 0$. Figure 9.1 also shows the problem with the negative exponential at the regional boundary. The fall-off in density from this model is likely to be too great. The pattern of population density in the hinterlands of western cities has been shown to be more even and higher than the negative exponential is able to predict, hence the suggestion that the negative exponential be only used to model intraurban

variation in density, the inverse power being used to model peripheral variation (Parr, 1985a).

One of the main reasons for the popularity of the negative exponential function in population density models relates to its elegant properties. In fact, as we have shown above, these properties can only be fully exploited when it is assumed that the regional boundary distance is infinite; in short, this is only the case when the infinite boundary provides a good approximation to the observed density field. In the given hypothetical example as well as in the applications to Seoul which follow, this is clearly not the case and it is the inverse power model which yields more tractable equations which can be exploited in rapid estimation. Furthermore, although the negative exponential emerges 'naturally' from entropy-maximizing, this is due to the constraint on total travel distance adopted. There is in fact considerable evidence to suggest that distance perception is logarithmic, this being the basis of the Weber–Fechner law in psychology. If the logarithmic constraint on travel distance is used, then it is easy to show that the inverse power function is the appropriate derivation using entropy-maximizing. On balance, it is our view that both functions are useful in different contexts. Moreover, we should not ignore Zielinski's (1980) comment: "Since both (functions) can give practically identical fits to data, what criteria should be adopted to prefer one to the other? In this case the pragmatist must take the back seat to the theoretician. And theoreticians must endeavor to prove their assumptions". This we will attempt to do in the following sections.

9.4 Urban Allometry, Density and Dimension

In the growth and evolution of natural systems, there has been considerable research into the ways various features of such systems scale with increasing size, the study of relative sizes being allometry (Gould, 1966). We briefly alluded to this line of research in Chapters 2, 6 and 7, but here we provide a more complete summary in preparation for this and the next chapter. Allometric relationships relate the size of an object to a familiar yardstick such as length. For example, taking a measure of size in terms of the dimension of a system E, in Euclidean geometry we can define points where $E = 0$, lines where $E = 1$, planes where $E = 2$ and so on. If we take the yardstick as a measure of length in one dimension, that is R say, then objects which scale as a point vary as R^0, as length itself R^1, as the plane R^2 and so on, in general the scaling relation being R^E. If the size of the object in question is expected to scale as R^E, we refer to this as isometric scaling. If the object scales as R^D with $D < E$, then this is called negative allometry and if $D > E$, this is positive allometry.

In its most basic form, population density can be considered as the scaling between population $N(R)$ and area $A(R)$, however defined, where R is some measure of the linear dimension of the space. Then the key scaling relation between $N(R)$ and $A(R)$ can be written as

$$N(R) \sim A(R)^\phi. \tag{9.30}$$

From *a priori* considerations, we can argue that the value of the parameter ϕ indicates how population fills the two-dimensional space available to it. If $\phi > 1$, then the population fills more than the two dimensions of the space, while if $\phi < 1$, population is filling less than its available space. There have been several studies of this allometric relation over different sizes of cities based on the general assumption that as cities grow into the third dimension, then equation (9.30) should exhibit positive allometry. This is borne out in the work of Stewart (1947) and Stewart and Warntz (1958) where they show that $\phi = 4/3$ for British and American cities with 1951 data. With Swedish data for 1966, Nordbeck (1971) concludes that $\phi = 1.506$ which is a little larger than $3/2$ which would be the value if cities filled the three dimensions available to their development. More recently, Jones (1975) following work by Best, Jones and Rogers (1974), has derived ϕ as 1.193 for cities in England and Wales from the 1971 Population Census, and Craig and Haskey (1978) who reworked these data concluded that ϕ has remained approximately stable since 1951 with a value between 1.41 and 1.45. However, Woldenberg (1973) using data for American cities has shown that ϕ varies from around 0.8 to 1.2 depending upon the data set used.

All these studies refer to data sets based on wide ranges of city size. In the intra-urban case, where population densities decline with increasing distance from the center, equation (9.30) will show negative allometry. First we will replace the measure of area in equation (9.30) with the yardstick of length R which enables us to link this analysis to population density functions introduced above. Then equation (9.30) can now be written

$$N(R) = \gamma A^{\phi}$$

$$= \gamma(\pi R^2)^{\phi} = \varphi R^D, \qquad (9.31)$$

where we have assumed that the distance associated with the area is given as $R = \sqrt{A}$, and γ and φ are constants of proportionality. From equation (9.31) it is clear that $D = 2\phi$ and if $\phi < 1$, then $D < 2$. Takayasu (1989) refers to D as the 'effective dimension', but to anticipate our argument, this of course is the fractal dimension. As in the two previous chapters, the value of D is thus a measure of the extent to which the city fills its two-dimensional area.

Without anticipating the value of D for any particular city, we can now present the equation for population density in terms of the scaling relation based on the yardstick of length R and explicitly containing the parameter D. From equations (9.30) and (9.31), we will write the density as in equations (2.33) and (7.9)

$$\rho(R) = \frac{N(R)}{A(R)} \sim A(R)^{\phi-1}, \qquad (9.32)$$

and from this it is clear that if population is isometric with area and $\phi = 1$, then $\rho(R)$ is constant. For population density to decline with increasing distance from the center, then the exponent in equation (9.32) must be less than 0, that is $\phi - 1 < 1$, while it is most unlikely that ϕ would be greater than 1 which would imply increasing density with distance from the center. Writing (9.32) explicitly in terms of the yardstick R gives

$$\rho(R) = \frac{\varphi R^D}{\pi R^2} = \xi R^{D-2},\qquad(9.33)$$

and for the density to decline with distance, it is clear that the fractal dimension D should be less than 2, implying that the city will not fill the two-dimensional space available.

The real value of this analysis is in revealing how density should be defined. Casual observation should suggest whether the city is filling two or three dimensions. If the city fills less than the plane, then the appropriate space on which to compute density is the plane; thus the definition of area as $A(R) = \pi R^2$ in equation (9.32) is correct and the value of D will be less than 2. However, if the city is filling three dimensions of space then the appropriate normalization by area would be $A(R) = \pi R^3$, the volume, and then the value of D would probably be greater than 2 but less than 3. In this case, the more general form of density equation in (7.18) would be appropriate. In fact, in all studies of density to date (of which the authors are aware), the implicit assumption is that cities fill the plane, not the volume, thus suggesting that the argument made by Nordbeck (1971) amongst others that cities fill the third dimension is spurious.

The other important conclusion from this analysis is that the measure of area to be used represents the space within which the city grows, its field, not the more restricted area which is associated with its built-up form; in fact most studies of density based on the population and area of Census tracts treat the area of the field, although in some studies which use built-up area, the parameters derived cannot be compared with those here. What is certain, however, is that in most studies of population density to date, researchers have paid very little attention to the definition of area, thus throwing into question the validity of the parameter values estimated, at least in terms of the sorts of theory invoked here. Comparisons between different studies are therefore difficult to make.

9.5 The Basic Scaling Relations Revisited

We are now in a position to relate the density equations stated earlier in terms of the inverse power functions, to those derived from urban allometry. We will use the two-directional function given in equation (9.26). Equating this to the density in equation (9.33), we will integrate over θ. Then

$$\rho(R) = \int_0^{2\pi} \rho(R, \theta)\, d\theta = K' R^{-\alpha} = \xi R^{D-2},\qquad(9.34)$$

and if the equation holds, then $D - 2 = -\alpha$ or $\alpha = 2 - D$. As D is a measure of the extent to which space is filled, then it is expected to be less than 2, and thus α is its complement, a measure of the extent to which the available space is not filled. The cumulative density functions can also be equated, and the same conclusions drawn. Comparing equation (9.27) with (9.31) and using appropriate notation, then

$$N(R) = K''R^{2-\alpha} = \varphi R^D. \tag{9.35}$$

Equations (9.34) and (9.35) suggest a coherent and useful coincidence between the fractal dimension D and the parameter of the associated inverse power function of population density; that is, that the two dimensions of the associated space can be divided into a space-filling component D and a non-space-filling component α which sum to the Euclidean dimension E of the space within which the city exists, in short, that $E = D + \alpha = 2$.

We have already defined three scaling relations, namely the population–area relation in equation (9.30), the population–radius relation in equations (9.31) and (9.35), and the density relation in equations (9.33) and (9.34). We will add to these a fourth relation, based on the first derivative of the cumulative population–radius relation in equation (9.35). This relation, referred to as the incremental population–radius equation, is derived as

$$n(R) = \frac{dN(R)}{dR} = hR^{D-1} = hR^{\eta}. \tag{9.36}$$

h is a constant of proportionality and η is the scaling parameter. We now have four scaling relations; three of these which relate the cumulative population $N(R)$, the incremental population $n(R)$, and the density $\rho(R)$ to the radius R and based on equations (9.31), (9.36) and (9.33), are those used in Chapters 7 and 8 to describe the distribution of particles in the DLA and DBM simulations. These will be used here in the empirical analysis of density in Seoul, but note that the alternate density variable $Q(R)$ which was used in Chapters 7 and 8 and defined previously in equation (7.26) is not used. The fourth relation is the classic allometric equation which relates $N(R)$ to $A(R)$ as in equation (9.30).

Before we broach the question of estimation, we must make clear the fact that we expect the value of the fractal dimension estimated from the Seoul data to be similar to that of the DLA simulations, that is $D \approx 1.7$, as implied in many of the estimates given earlier in Table 7.1. Thus the value of the density parameter would be around –0.3, that is $\alpha = D - 2$. Earlier we speculated that in the case of the negative exponential density function, the dimension of the system would fall between $1 < \lambda C < 2$ from arguments relating to the mean density in one- and two-dimensional systems. We can now speculate that $\lambda C \approx 1.7$, a result which has already been borne out in work by one of the authors almost 20 years ago (Batty, 1976). Similar relations are not obvious for the inverse power function, although D and α are immediate results of their estimation and further insights must await further research. To provide some sense of closure to this theoretical section and to progress the research to applications, methods of estimating equations (9.31), (9.36), (9.33) and (9.30) in that order will now be discussed.

9.6 Methods of Parameter Estimation

There are two main issues to clarify before we embark on fitting the basic scaling relations to the empirical data: these deal with aggregating the data,

and choosing appropriate estimation methods. We will deal with these issues in turn. Researchers fitting population density models have either dealt with the data in their raw form by Census tracts or have aggregated the data into concentric rings. The latter method produces better fits and hence more reliable estimates because local variations in the data are reduced, but the use of one or the other data set ultimately depends upon the purpose for which the analysis is undertaken. Here we will employ both, but we consider that the method of concentric rings is likely to give much 'truer' estimates of the values of the parameters D and α which is of most interest to the analysis performed here. In fact, this was the method which we used to organize the data in Chapters 7 and 8 for comparing the theoretical simulations with Taunton and Cardiff. A related issue involves the way the data are aggregated over distance and whether relationships are fitted in their cumulative or incremental form. Cumulative relations are bound to give better fits and this will be our emphasis although we will also examine some incremental estimates. In this sense then, we will follow the practice of Bussiere (1972) and Bussiere and Stovall (1981), rather than the traditional treatment used by Clark (1951).

The first method of estimation we will use and one that will act as our baseline, involves taking logarithmic transformations of equations (9.31), (9.36), (9.33) and (9.30) and estimating their parameters using linear regression. The four relations can now be summarized in discrete form as

$$\log N_i = \alpha_1 + \beta_1 \log R_i, \tag{9.37}$$

$$\log n_i = \alpha_2 + \beta_2 \log R_i, \tag{9.38}$$

$$\log \rho_i = \alpha_3 + \beta_3 \log R_i, \tag{9.39}$$

$$\log N_i = \alpha_4 + \beta_4 \log A_i. \tag{9.40}$$

Note first that the intercepts or constant terms α_1, α_2, α_3 and α_4 are not directly related to the parameter α of the inverse power function but are so defined to make them consistent with their usage in Chapters 7 and 8 through equations (7.49). Throughout the analysis, distances are ordered so that $R_i < R_{i+1} < R_{i+2}$, the cumulative populations are given as $N_i = \Sigma_j n_j$, $(j = 1, \ldots, i)$, the areas for each cumulative ring as $A_i = \pi R_i^2$, the actual area of each ring as $a_i = A_i - A_{i-1}$, $(i > 0$ and $A_0 = 0)$, and the density $\rho_i = n_i / a_i$. We also need to be clear about the various parameter estimates in these equations and the look-up table reproduced as Table 9.1 shows the conver-

Table 9.1.   Relationships between the parameters

Equation number[1]	Intercept parameter	Slope parameter	Dimension D	Density coefficient α
9.37 (9.31)	$\alpha_1 = \log \varphi$	$\beta_1 = D$	β_1	$2 - \beta_1$
9.38 (9.36)	$\alpha_2 = \log h$	$\beta_2 = \eta$	$1 + \beta_2$	$1 - \beta_2$
9.39 (9.33)	$\alpha_3 = \log \xi$	$\beta_3 = D - 2$	$2 + \beta_3$	β_3
9.40 (9.30)	$\alpha_4 = \log \gamma$	$\beta_4 = \phi$	$2\beta_4$	$2(1 - \beta_4)$

[1] The first equation number in each row is the number of the log transform of the second equation number (which is in brackets).

sion between these parameters, the dimension D and the coefficient α. In the sequel, all estimates are given to three decimal places.

There are several other methods of estimation which we consider less reliable than log–linear regression. Mills (1970), Mills and Tan (1980), and Weiss (1961) all use methods which involve only two values of density or population given at two different points within the city. For example, if the density $\rho(R)$ or the cumulative population $N(R)$ are known at, say, distances R and R_b, then using equations (9.26) or (9.27) in the two-dimensional case, the constant and scaling parameters can be found from the solution of two equations in two unknowns using the Newton–Raphson method (Batty, 1976). There is little work on how reliable these types of estimates are in comparison with those from regression, but we consider that such methods are really only appropriate when no more than two observations are available.

There are, however, two other methods which we will use, the first of which was introduced in Chapters 7 and 8 and which can only be used if population data are available on a very fine lattice. Where each occupied lattice point is the location of a single equal-sized household, then a good approximation to the density in equation (9.33) is $\rho(R) \sim R^{D-2}$ where the constant of proportionality is approximately equal to unity. Then for any distance R from the center, D can be calculated from a log transform of this equation, given earlier in equation (8.13) which is restated as

$$D(R) = 2 + \frac{\log \rho(R)}{\log R}. \tag{8.13}$$

In earlier chapters we called the graph of $D(R)$ against R the 'signature' of the density function. In the vicinity of its origin, $D(R)$ can fluctuate wildly, but it soon settles down to a characteristic value ($D(R) \approx 1.7$ for the case of a DLA cluster). At the edge of the cluster, the value of the dimension is also unreliable because this is the area where the city is likely to be developing most rapidly. Thus the best value of the dimension will be given at the distance value of the mean density, that is at $D\{C(R_b)\}$. We have not yet said anything about the data in the applications which follow, but two of the four data sets we will use are based on a crude approximation to the shape of the city using lattice point data which reveals the morphology of the city in question. This will become clear in the sequel.

The last method is well-known in that it is based on deriving the density and population model by entropy-maximizing. Formulating the model in terms of the location of population n_i in individual zones i, the model can be stated as

$$n_i = N \frac{R_i^\eta}{\sum\limits_i R_i^\eta}, \left(\sum\limits_i n_i = N \right), \tag{9.41}$$

where N is the total population in its absolute value (or normalized form where $N = 1$). The parameter η in equation (9.41) is found by solving the model subject to either of the following two constraint equations

$$\bar{C} = \sum_i \frac{n_i}{N} R_i \qquad\qquad (9.42)$$

or

$$\bar{C}_{\log} = \sum_i \frac{n_i}{N} \log R_i. \qquad\qquad (9.43)$$

Strictly speaking, equation (9.41) is associated with constraint equation (9.43) not (9.42); that is, equation (9.41) is derived by maximizing entropy subject to equation (9.43), thus leading to the derivation of the inverse power function (Wilson, 1970). In fact, we will fit equation (9.41) by separately solving for both equation (9.42) and (9.43), thus giving two estimates of η (and thence of D and α through Table 9.1). We will not concern ourselves any further with entropy-maximizing except to point out that there is a strong relationship between entropy, information and fractal dimension which finds its clearest expression in the derivation of power functions such as those treated here (Batty and Sikdar, 1982; Takayasu, 1989).

9.7 Applications to Large Cities: the Seoul Data Base

Seoul was selected for our empirical work because the negative exponential model has already been fitted to its population density, thus giving us the opportunity of making some casual comparisons to the inverse power density models which we will estimate here (Kim, 1985). We also had access to reasonable population data for the city at three dates 1970, 1977 and 1982 from which we selected the 1982 data for our applications. Moreover, Seoul is a rapidly growing city, its population increasing from around 5.5 million in 1970 to almost nine million in 1982, and we consider that cities such as this one where rapid growth has taken place under few market imperfections to be ideal testing ground for the ideas introduced here. The basic set of zones for the 1982 population data are shown in Figure 9.2(a) which are aggregations of Census tract data. In the analysis, the center of the city was found using centrographic analysis on the 1982 data (Kim, 1985). Distances to all other zones were then defined as crow-fly distances from this center to each individual zone centroid. From this original data set, the zones are ranked according to distance from the center and these are aggregated in the given order for estimating three of the four scaling relations.

The second data set used is quite different. The urban form of Seoul taken from a 1982 cadastral map was defined by placing a 72×72 lattice across the built-up area and coding each lattice point according to whether the surrounding cell was developed or undeveloped. The size of the lattice spacing was half a kilometer, and a lattice point was classified as developed if more than half the surrounding cell was built up. The basic logic behind this type of coding is that if the lattice spacing is fine enough, the lattice will detect the location of population in such a manner as to provide a
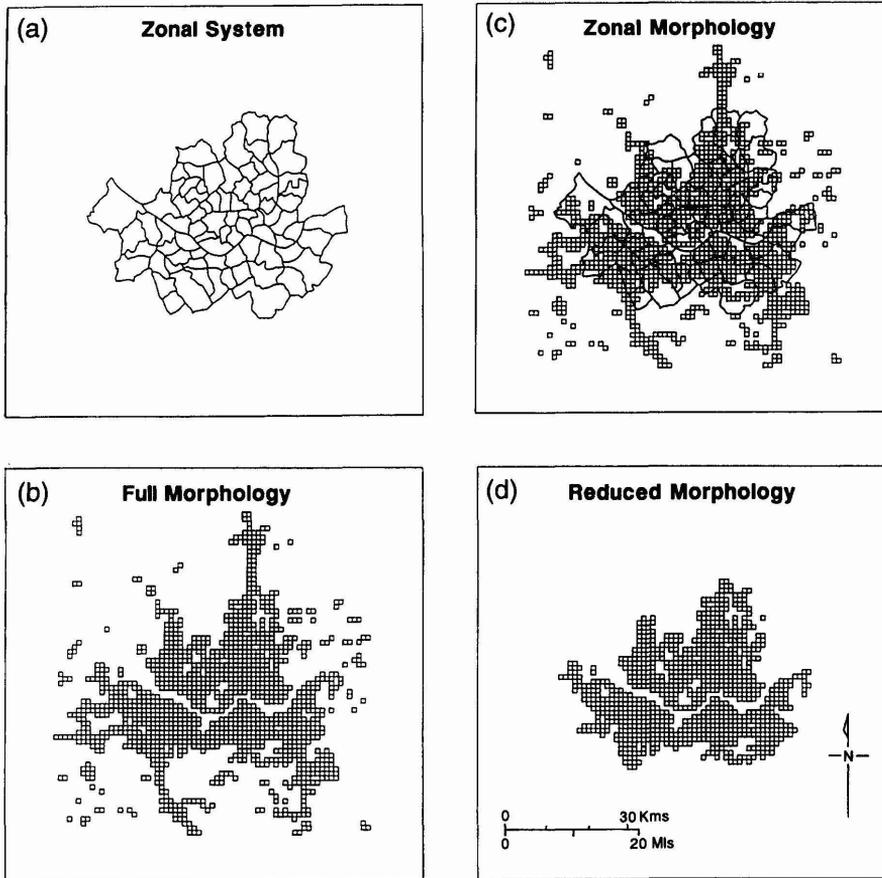
Figure 9.2.   Zonal and lattice systems for the four data sets.

detailed morphology of the city. In short, although each lattice point is assumed to have an identical population density, when aggregated for purposes of analysis into concentric rings say, then the variable density of population will be detected. This is clearly not the case here where the lattice is far too coarse; for it to pick up the kind of variation needed, the spacing should be as fine as $10 \text{ m} \times 10 \text{ m}$. Our gross approximation here is thus illustrative rather than definitive, and although it does pick up the morphology and variable density at a crude scale, the results of using this data set must be interpreted with caution. The coding is shown in Figure 9.2(b).

Two related data sets have also been produced. First, by combining the measured population in the first 80 zone data set with the morphology defined by the second, a third data set in which the actual population is allocated to the points of the lattice can be defined. This is done by overlaying the zonal structure in Figure 9.2(a) on the morphology in 9.2(b), associating each lattice point with one of the original 80 zones, and counting the number of lattice points in each zone. The population in each original zone is then allocated equally to as many lattice points as there are in the zone in question. Figure 9.2(c) shows the overlay of the two zoning systems

which produces the third, noting that the periphery of the city is excluded from the original urban morphology shown in Figure 9.2(b) due to the fact that population data is not available for these outer areas. The fourth and last data set is similar to the second except for the peripheral areas which have been excluded. This is simply formed by overlaying the first data set on the second, excluding the peripheral areas, thus forming a reduced morphology based on lattice spacing. This is shown in Figure 9.2(d).

For the second, third and fourth data sets which are all based in some way on the lattice morphology, concentric rings can be easily defined at increasing units of equal distance around the center. In these cases, we have divided the greatest radial distance to the edge of the city into 25 equal units on which we have arranged 25 concentric rings. The populations and related variables for each ring have then been derived by counting the number of lattice units (and their populations in the case of the third data set) which fall into each ring. In subsequent analysis, we will refer to these four data sets as follows; the first data set based on the zones shown in Figure 9.2(a) is called the Zonal System, the second in Figure 9.2(b) the Full Morphology, the third in Figure 9.2(c) the Zonal Morphology, and the fourth in Figure 9.2(d) the Reduced Morphology.

It is also important to be clear about which estimation method is to be applied to which data set at this stage. All four scaling relations in equations (9.31), (9.36), (9.33) and (9.30) will be fitted to all four data sets using the logarithmic transformations given in equations (9.37) to (9.40). This will represent our set of baseline estimates for we consider this type of estimation as providing us with the most comprehensive set of estimates. This will generate four estimates of the dimension D and the density parameter α for each of the data sets. The estimates based on solving the entropy-maximizing constraint equations in (9.42) and (9.43) will also be applied to each of the four data sets. This will provide us with two sets of parameters D and α derived from η, one for each solution of the appropriate constraint equation. Finally, the method of calculating D (and hence α) from the signature given in equation (8.13) is only valid for the Full and Reduced Morphology data sets, the second and fourth, because this requires that the data be based on equally dense lattice points. This concludes our survey of the data and we will now embark on presenting the baseline estimates.

9.8 The Density Model Estimates

Throughout the process of fitting the baseline equations in (9.37) to (9.40) to the four data sets, the analysis was aided by frequent graphical interpretations of the base data, model predictions and residuals. To provide some sense of the similarities between the data sets, we have plotted the cumulative population $\{N_i\}$ against the radius $\{R_i\}$ as given in equation (9.37). The graphs of these data are plotted in Figure 9.3 and show very characteristic profiles which go some way to indicating that the four data sets are detecting the rudiments of the urban form and density of Seoul, as well as being similar to related profiles in other cities (for example, see the exten-
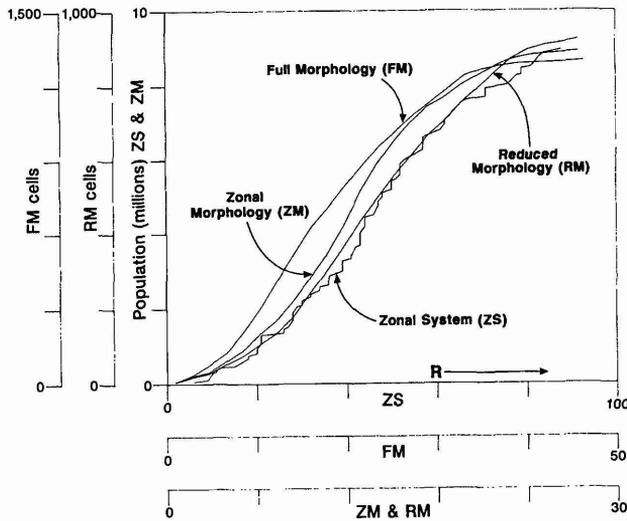
Figure 9.3. Cumulative population profiles for the four data sets.

sive examples in Bussiere and Stovall, 1981). Using Figure 9.3, if a detailed analysis is made of the variation in density and population in the vicinity of the origin and at the edge of the city, it is clear that density is volatile and shows no characteristic trend in these areas. Population density is notoriously difficult to model near the center of the city, while at the edge, the city is growing at its fastest rate, and the density is nowhere close to the density which will ultimately prevail.

The models therefore should not be applied in these areas; at the center whatever function is used is likely to be inappropriate, while at the edge, densities are likely to be too low, and thus the function will underpredict these. We will use these characteristics to exclude certain observations from our data set in the results presented below as we did with the simulation results and the empirical data pertaining to Taunton and Cardiff in Chapters 7 and 8. This is a particularly important issue in modeling population density, for we must assume that wherever the density function applies, the city must be in equilibrium. In previous work, this issue has been entirely ignored for most research has not been cast into an appropriate dynamic framework which indicates that densities in the areas of the origin and periphery of a growing city are likely to be subject to rapid change. The theory of the growing city which we have alluded to here based on DLA simulations provides a strong rationale for systematically excluding these areas from our estimation and accordingly, we will proceed by doing so.

Table 9.2 presents the results of fitting equations (9.37) to (9.40) to the Zonal System data whose spatial system is subdivided as in Figure 9.2(a). In the sequel, all parameter estimates from these equations will be converted into their appropriate fractal dimension D using Table 9.1, and the reader can easily note the density parameter α by calculating it as $D - 2$. Table 9.2 shows immediately that the fit of the cumulative populations with respect to area and radius are much better than those of the incremental population and the density. This difference has been widely observed in

Table 9.2. Dimensions associated with the Zonal System data

Number of zones	Population–radius (9.37)	Incremental population (9.38)	Population density (9.39)	Population–area (9.40)
80	1.765	1.096	1.633	1.504
	97.7	00.9	18.9	87.0
77 less 1–3	1.667	1.088	1.490	1.352
	98.0	00.5	25.9	84.2
71 less 72–80	1.850	1.203	1.976	1.976
	98.5	04.8	02.3	99.3
68 less 1–3, 72–80	1.758	1.228	1.932	1.914
	98.9	04.9	23.7	99.4

Note: The statistics shown below each of the estimates of dimension are coefficients of determination $100r^2$ calculated for each set of dependent and independent variables. These are also shown in Tables 9.3, 9.4 and 9.5.

the literature and results from aggregating variables, thus reducing their variance (Muth, 1969). However, our quest is to increase the fit of all four equations by systematically removing those observations which are most suspect. To this end, we have first removed the three central zones, then the nine peripheral zones, and then these central and peripheral zones together. The fit of all the equations is improved by these exclusions, although the estimates of dimension for the incremental population equation (9.38) differ most from the other three whose values are closest to one another. The best fits occur for all four equations when both the central and peripheral zones are excluded with the cumulative population–radius relation giving a dimension value closest to that of the DLA model (1.758 compared to 1.71), with all four of its dimensions for the four variants on this data set, falling between 1.667 and 1.850. What is extremely encouraging for this data set is that all the dimensions lie between 1 and 2, the range which suggests that cities do not fill their entire two-dimensional space. As we will see, our confidence that the 'true' dimension of Seoul lies between 1.5 and 1.8 will be progressively increased as we examine each data set.

The second set of data – the Full Morphology – detects the form of the city, but is based on the assumption that each lattice cell has the same density of population. These 1300 cells have been aggregated into 25 equally spaced concentric rings, and the same type of progressive reduction in the rings used is made in the attempt to exclude the most problematic observations. The rationale for these exclusions only becomes clear when the data are plotted, which has been done in every case but is not illustrated here for lack of space. The results are shown in Table 9.3, and these indicate once again that it is the incremental population relation in equation (9.38) which has the lowest fit, and the cumulative population relation (9.37) which has the best. Of the 24 estimates of dimension given in Table 9.3, only two of these fall outside the range of $1 < D < 2$. The average of all dimensions in this table is 1.551, and once again, the estimates from this data set which detect only the form are similar to those in the first set which

Table 9.3. Dimensions associated with the Full Morphology data

Number of concentric rings	Population–radius (9.37)	Incremental population (9.38)	Population density (9.39)	Population–area (9.40)
25	1.333 97.3	0.668 06.2	1.593 63.2	1.513 95.5
24 less 1	1.290 95.5	0.148 25.1	1.449 73.9	1.389 94.3
15 less 16–25	1.498 99.5	1.597 71.3	1.835 55.6	1.793 99.2
14 less 1, 16–25	1.552 99.2	1.384 41.7	1.753 65.8	1.721 98.8
9 less 10–25	1.529 99.3	1.918 98.1	1.957 60.1	1.943 99.9
8 less 1, 10–25	1.682 99.7	1.876 94.8	1.940 54.6	1.930 99.9

detect only the density. In Table 9.3, the best estimates across all four scaling equations are given by the data which excludes the 16 peripheral zones.

Perhaps the most consistent data set for the models developed here is the third – the Zonal Morphology – which combines both density and form and which is shown in Figure 9.2(c). The results of fitting the four equations to this set are shown in Table 9.4 where it is clear that all dimensions estimated, with the exception of one, fall between the limits of 1 and 2. As observations are excluded, the performance of the models increases significantly with the best fitting range of estimates achieved when the 15 periph-

Table 9.4. Dimensions associated with the Zonal Morphology data

Number of concentric rings	Population–radius (9.37)	Incremental population (9.38)	Population density (9.39)	Population–area (9.40)
25	1.434 98.4	1.048 00.2	1.694 68.3	1.641 98.1
24 less 1	1.477 97.7 .	0.725 04.6	1.632 68.1	1.597 97.3
13 less 14–25	1.486 98.5	1.873 93.7	1.846 78.1	1.817 99.7
12 less 1, 14–25	1.640 98.9	1.941 89.5	1.856 59.3	1.842 99.4
10 less 11–25	1.416 98.6	1.831 92.2	1.822 82.2	1.780 99.6
9 less 1, 11–25	1.566 98.7	1.887 84.6	1.810 67.8	1.788 99.3

eral rings are excluded. Here the dimensions vary from 1.416 to 1.822, and if the central ring is excluded as well, the range narrows to 1.566 to 1.810 with an average of 1.763, again close to our theoretical value of 1.71. The last data set – the Reduced Morphology, shown in Figure 9.2(d), is simply the Full Morphology reduced by the exclusion of the periphery. However, the data are still organized into 25 concentric rings, and as such, this represents a scaled-up version of the Full Morphology. The model estimates for the progressive exclusion of rings from this data set are shown in Table 9.5 where all 24 dimensions computed lie between 1.140 and 1.899. The same sorts of interpretation as in the previous three data sets emerge again here: the incremental population and density relations show the poorest fits with the population–area and population–radius the best. The overall best fit occurs when 11 peripheral rings are excluded, and the average dimension when the central ring is excluded as well is 1.798.

The range of dimension values computed from these four data sets is remarkably narrow, with only three values from the 88 estimated falling outside the range of $1 < D < 2$, these values being less than 1 in each case. This is fairly conclusive evidence that the inverse power density function when fitted to population density data computed with respect to the urban field (and not just the residential built-up area) will yield a fractional dimension between 1 and 2, with the likely value between 1.5 and 1.8. This also implies that the density parameter should lie between 0.2 and 0.5. We would argue that if estimates of the inverse power function yield values of $\alpha > 1$ or $\alpha < 0$, the data or the estimation procedure is likely to be suspect, or the data in question implies an urban morphology which is unusual. However, before we conclude, we will also present the other two methods of estimation, the first based on the signature equation in (8.13), the second on the entropy estimation method relating to equations (9.41) to (9.43).

We are only able to estimate dimensions for the signature of the density

Table 9.5. Dimensions associated with the Reduced Morphology data

Number of concentric rings	Population–radius (9.37)	Incremental population (9.38)	Population density (9.39)	Population–area (9.40)
25	1.472 98.8	1.330 16.3	1.733 76.5	1.689 99.1
24 less 1	1.545 98.7	1.140 02.3	1.699 72.4	1.673 98.6
18 less 19–25	1.504 98.6	1.724 86.3	1.812 83.0	1.781 99.7
17 less 1, 19–25	1.663 99.1	1.707 76.0	1.817 71.1	1.800 99.5
14 less 15–25	1.475 98.3	1.835 92.7	1.823 78.6	1.791 99.6
13 less 1, 15–25	1.630 98.7	1.899 88.2	1.838 59.8	1.824 99.3

for the Full and Reduced Morphologies, the second and fourth data sets as presented in Figure 9.2(b) and (d). From equation (8.13), we have plotted the signature $D(R)$ against R and this is shown in Figure 9.4. This signature relates to the Full Morphology, but as the Reduced Morphology is a subset of this, the figure is relevant to both. The regional boundaries differ in that $R_b = 47.424$ for the Full, $R_b = 28.071$ for the Reduced, and these are also indicated in Figure 9.4. At these values of R, the dimension for the Full system can be calculated as 1.756 in comparison with 1.704 for the Reduced. However, it is clear from Figure 9.4 that a better value of R to take would be the mean value $C(R_b)$, and these yield dimensions of $D\{C(R_b)\}$ of 1.845 for $C(18.590)$, the Full data set, in contrast to 1.856 for $C(14.185)$ for the Reduced set. These estimates are close to and consistent with the regression estimates in Tables 9.2 to 9.5, and some exploration of the signature in Figure 9.4 reveals that in the region of the origin, the estimates of dimension are volatile, although by $R = 9$, these estimates have settled to around $D(9) = 1.85$. Clearly the mean values of $C(R_b)$ are the most appropriate to use. Together with the estimates of the simulations using the DLA and DBM methods in the two previous chapters, these signature functions have proved to be the most robust methods of estimation.

The entropy estimation also produces well-fitting models with dimension values which accord to the results so far. These estimations involve solving equations (9.42) or (9.43) for the parameter η by some non-linear method – here we use the Newton–Raphson method – then using Table 9.1 to calculate the dimensions D (= $1 + \eta$). Two estimates are made for each of the four data sets, these estimates being based on solving for the constraint in equation (9.42), then (9.43). All four data sets with their exclusions as given previously in Tables 9.2 to 9.5 are used in the estimation, thus giving 44 estimates of D in total. These results are shown in Table 9.6. In this table, under each dimension value, we first give the coefficient of determination ($100r^2$) for the predictions of the incremental population model which is
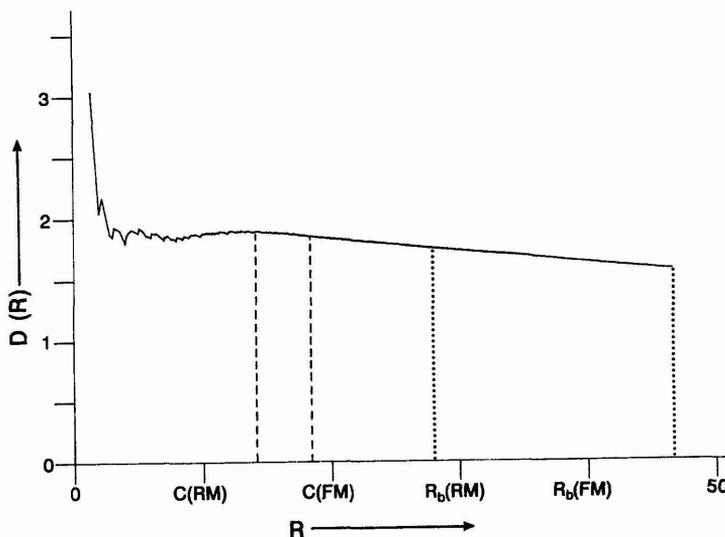
Figure 9.4. The signature of the full and reduced morphologies.

Table 9.6. Entropy estimates of the dimension

Zonal system			Full Morphology			Zonal Morphology			Reduced Morphology		
No. of zones	Equations (9.42)	(9.43)	No. of rings	Equations (9.42)	(9.43)	No. of rings	Equations (9.42)	(9.43)	No. of rings	Equations (9.42)	(9.43)
80	1.132	1.703	25	0.898	0.621	25	1.085	0.822	25	1.245	1.029
	01.8	01.9		00.9	00.1		00.8	00.3		08.1	01.7
	99.8	99.8		94.6	96.5		95.1	96.1		97.6	98.0
77	1.134	1.064	24	0.698	0.422	24	0.940	0.689	24	1.151	0.947
	01.3	01.4		08.4	04.1		00.2	00.1		02.7	04.8
	99.8	99.8		95.9	97.1		95.6	96.4		97.7	98.1
71	1.188	1.134	15	1.405	1.252	13	1.942	1.973	18	1.640	1.544
	03.5	03.6		39.4	44.2		94.3	94.4		66.5	68.2
	99.8	99.8		98.7	99.1		99.9	99.9		99.3	99.4
68	1.199	1.130	14	1.283	1.157	12	1.972	1.990	17	1.614	1.152
	02.8	02.9		21.7	24.5		93.1	93.1		59.2	60.8
	99.8	99.8		98.9	99.2		99.9	99.9		99.3	99.4
			9	1.870	1.829	10	1.882	1.912	14	1.881	1.870
				93.3	93.6		94.3	94.4		90.6	90.6
				99.8	99.8		99.9	99.9		99.8	99.8
			8	1.838	1.803	9	1.910	1.931	13	1.903	1.890
				90.3	90.5		92.6	93.0		88.4	88.4
				99.8	99.8		99.9	99.9		99.8	99.8

Note: the first line of each cell is the fractal dimension, the second line is $100r^2$ for the increment of population, and the third line is $100r^2$ for the cumulative population.

the form in which the model is specified in equation (9.41), and below this, we give the same coefficient for the population in cumulative form.

The results are similar to those in Tables 9.2 to 9.5 in that the Zonal System data performs least well. The Full, Zonal and Reduced Morphology data sets give parameter values and fits which are both good and similar to one another, and there is a consistent increase in the performance of all models as zones or rings are progressively excluded from the estimation. Of the 44 estimates, eight fall below $D = 1$ and none is greater than 2. As in previous estimates using regression, the best fits are those obtained when the most zones or rings are excluded. These results again strengthen our confidence in various hypotheses that suggest that cities never fill their two-dimensional space in which they grow, regardless of any vertical growth into the third dimension which has occurred within the last century, and that their fractional dimension lies between 1.5 and 1.8. Moreover, if the models and data sets to which they are applied are correctly specified, it is easy to give order of magnitude estimates for the density and dimension parameter values before the analysis begins.

9.9 Fractals and City Size

The theoretical analysis with which we introduced this chapter concentrated on the relative advantages and limitations of the two classical func-

tional forms which have dominated social physics, the negative exponential which has become the conventional wisdom for representing density and interaction, and the inverse power which was associated with much early research. Notwithstanding our view that the inverse power function has less direct but important properties which have hitherto hardly been exploited or even recognized, the parameters of both functions reflect the properties of the space within which such functions are defined. However, it is clear that the parameters of the inverse power function are direct measures of the extent to which the phenomena whose density is being measured, fills the space available, and using arguments from urban allometry, the link between such functions and fractal geometry can be made. Thus the function used determines the form of the system being modeled, and it is in this sense that we say 'form follows function'.

The empirical work in this chapter in which the inverse power function has been fitted to population density profiles in the city of Seoul has produced quite startling results. Although all the theory we have developed in the last three chapters suggests that cities have a fractional dimension between 1 and 2, we did not expect our results to be quite so conclusive, over such a large range of scaling relations and estimation methods. Using the four data sets and variants of these based on excluding certain observations, we have provided 136 estimates of the dimension D and density parameter α. Of these, only 11 fall outside the postulated range $1 < D < 2$, and these are all less than 1. If we look at the results based on excluding the most problematic observations, of the 24 values of the dimension produced, 21 fall within the range 1.566 to 1.940, thus suggesting that the 'true' value of the dimension must be nearer 2 than 1.

However, perhaps the most important value of this analysis is not in demonstrating the consistency of results produced by both theoretical and empirical analysis, but in the need to be extremely careful in the way data are collected and density defined. Strictly, we need data bases in which every household and household size is recorded in terms of its location before we can develop any definitive analysis of density which exploits the fractal model most appropriately. This represents an immense task, but with better data and data systems becoming available, it is now within the bounds of feasibility. On both the theoretical and empirical sides of this argument, we also need to explore the link between entropy, information and dimension (Takayasu, 1989); for in doing so, we are likely to generate a clearer picture of the role of the conventional model of population density based on the negative exponential, as well as exploring further the properties of the inverse power function. This will also enable us to link our approach to the mainstream where entropy and utility maximizing are widely used in the derivation and estimation of spatial economic, urban and transportation models.

Finally, we need to explore the relationship between city size, fractal dimension, changing densities and changing form. This implies once again that we broach directly the question of an appropriate model dynamics which encompasses processes of fractal growth (and decline), as well as questions of reversibility. Only in this way will we be able to connect these ideas to the large body of knowledge originally developed by Clark (1951) which concerns the changing shape of population density profiles over

time. If we can make progress here, it may be possible to begin to fashion a theory of city size, shape and form which enables us to classify cities in terms of the efficiency and perhaps economy of their form. This will be the goal of our next and final chapter where we will speculate upon the groundwork which is required for a fractal theory of city size.