CHAPTER 8

Urban Land Use

Gilles Duranton[*,†], **Diego Puga**[†,‡]
[*]Wharton School, University of Pennsylvania, Philadelphia, PA, USA
[†]CEPR, London, UK
[‡]Centro de Estudios Monetarios y Financieros (CEMFI), Madrid, Spain

Contents

Abstract

We provide an integrated treatment of the theoretical literature on urban land use inspired by the monocentric model, including extensions that deal with multiple endogenous business centers, various dimensions of heterogeneity, and durable housing. After presenting the theory and distilling its key empirical implications, we critically review the empirical literature on differences in prices and development across urban locations, patterns of location choices of heterogeneous households in cities, sprawl and residential decentralization, and employment decentralization.

Keywords

Land use, Urban structure

JEL Classification Code

R14

Handbook of Regional and Urban Economics, Volume 5A
ISSN 1574-0080, http://dx.doi.org/10.1016/B978-0-444-59517-1.00008-8

8.1. INTRODUCTION

In this chapter, we provide an integrated treatment of the theoretical literature on urban land use inspired by the monocentric model, including extensions that deal with multiple endogenous business centers, various dimensions of heterogeneity, and durable housing. After presenting the theory and distilling its key empirical implications, we critically review the empirical literature on differences in prices and development across urban locations, patterns of location choices of heterogeneous households in cities, sprawl and residential decentralization, and employment decentralization.

Urban land use is of fundamental importance. Most obviously, it is at the heart of extremely large allocation decisions made by firms and households. On the residential side, American households devote about a quarter of their consumption expenditure to housing, and the value of the residential housing stock may represent up to 2 years of gross national product. Where development occurs and at what intensity is arguably a first-order determinant of the efficiency of these large allocations. Households also engage in a variety of activities that take place in different locations: they work, they sleep, they play, they go to school, they shop, they visit friends, they go to the dentist, etc.[1] To conduct these activities in different locations, people must travel between them. As a result, land use and transport are intimately connected. American households spend between 5% and 10% of their time awake traveling, and the median household devotes 18% of its budget to transport, most of which goes to road transport.[2]

Beyond this, urban land use is a fundamental determinant of the physical world that surrounds urban dwellers, the majority of the world population. Urban land use determines how the various locations urban dwellers go to or would like to go to are organized and connected with each other. Hence, not only does land use affect the immense resources devoted to housing, commercial property, open space, and transport, it also potentially affects the labor market and the markets for the products we purchase. Land use also arguably affects the ability of firms to produce. In turn, these broader effects of land use may have serious implications for prosperity and equity.

Figure 8.1 depicts the distribution of land across various uses in Paris. The map at the top of the figure classifies land across five uses on a disk with radius 30 km centered on Notre Dame, the conventional center of Paris. We can immediately see that the patterns of land use are quite complex. The next two plots summarize the information by classifying land use by distance from Notre Dame, with the northern half of the map

[1] These travel categories correspond closely to the categories in the US National Household Transportation Survey.

[2] See Combes et al. (2014) for the sources for the housing figures and Couture et al. (2012) for transport. The figures reported here are for the United States, but the magnitudes for other developed and many developing countries are similar.

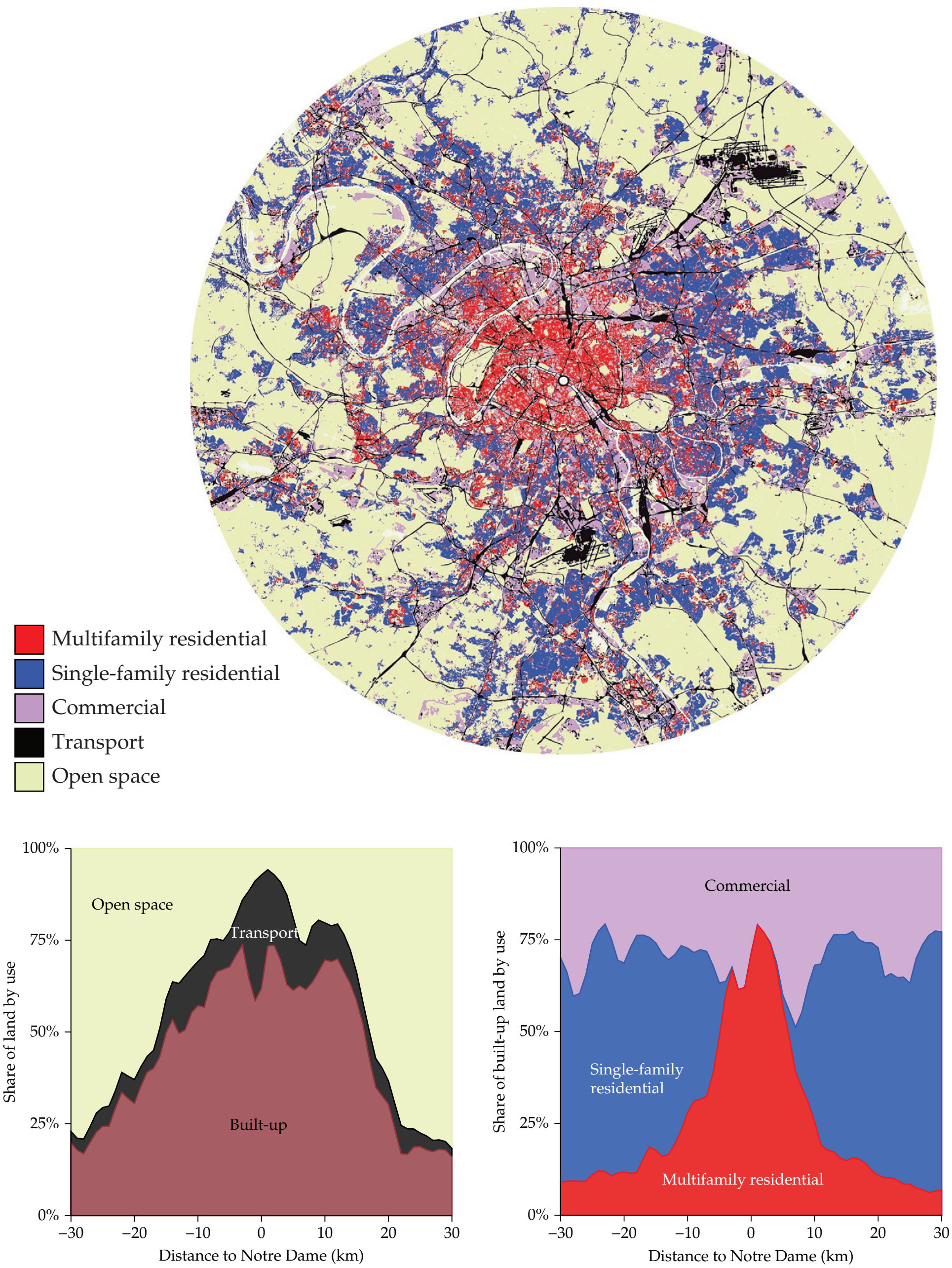

Figure 8.1 Distribution of land across uses in Paris.

plotted on the positive side of the horizontal axis and the southern half plotted on the negative side.[3] The first of these two plots splits all land between open space, land used for transport infrastructure, and built-up land. The last plot further divides the built-up land category between multifamily residential, single-family residential, and commercial uses. Both plots show some very clear gradients. In particular, as we look further away from the center (Notre Dame), the percentage of land that is built up falls, with more land being open space instead. The intensity of residential development also falls very clearly with distance to the center, with multifamily buildings giving way very quickly to single-family homes. The distribution of built-up land between residential and commercial uses does not show much variation by comparison, but we do see two peaks of commercial land (pointing downward, since commercial is plotted at the top) at the sides of a central area with more mixed use. It is also worth noting how much space is taken up by transport infrastructure, particularly close to the center, a very graphic illustration of how closely tied transport is to land use issues in cities. The rest of the chapter will help the reader to understand both the complexity and the order that appear in Figure 8.1.

Before proceeding any further, we will draw some intellectual boundaries for this chapter and justify its organization further. Since everything is located somewhere, land use potentially touches on a large number of topics. At minimum, it could certainly overlap greatly with all the other chapters in this volume. To retain a finite agenda, we think of urban land use as covering mainly the following issues: (a) the differences in land and property prices across locations, (b) the patterns of location choices by types and subgroups of users, (c) the patterns of land conversion across uses, and (d) the patterns of residential and business location changes within cities.

To explore these four sets of issues, we first present an integrated summary of theoretical developments on urban land use before turning to the empirical work on the aforementioned issues. A first reason for using this structure instead of providing a different model for each empirical question is that the theory that underlies the issues listed above is unified. There is no point repeating it several times. Furthermore, the economic analysis of land use first saw some important theoretical developments with empirical

[3] Since they transform spatial data from two dimensions (latitude and longitude) to just one dimension (distance to the center), these plots are more directly comparable with economic models of land use. As we shall see below, typically these models either represent cities as a segment on the real line (e.g., Ogawa and Fujita, 1980) or represent them as a disk that is circularly symmetric by assumption (e.g., Lucas and Rossi-Hansberg, 2002) so that all points at the same distance from the center of the disk have the same land use. It is worth emphasizing that, in addition to the market forces that are the focus of economic models of land use, the allocation of land in Paris to the different uses depicted in Figure 8.1 also reflects public sector land use decisions and important restrictions and regulations. Figure 8.1 is constructed based on the dataset Mode d'occupation du sol 2012 from l'Institut d'aménagement et d'urbanisme de la Région d'Île-de-France.

work lagging behind or developing independently.[4] We endeavor to reconnect empirical work more strongly with theory both by making sure that we highlight the empirical content of the models as we describe them and by trying to tie empirical work to land use models as strongly as possible (or by highlighting the weakness of those links in some cases). Another reason for presenting the theory in a self-contained manner is that it is relevant not only to the issues explored here, but also to many issues explored in other chapters in this volume such as regulation, neighborhood dynamics, and transport, to list just a few.

Following a long and well-established economic literature, our starting point is that accessibility determines land and housing prices at different locations. However, the patterns of accessibility are also affected by the location choices of firms and workers, which are determined by prices. Hence, the land use problem is in essence a hard equilibrium problem with many feedbacks. The literature first solved it by restricting accessibility to be solely about access to jobs and by treating the location of these jobs as exogenous within a simple geography and with frictionless markets. This is the monocentric model that we explore in Section 8.2.

While the simplest version of the monocentric model may be viewed as a reasonable first-order description of many cities and delivers a number of plausible predictions, it remains extremely crude. Even if we are willing to restrict production in cities to take place in a centralized area, the model does not include a number of fundamental urban features. In particular, city dwellers are highly heterogeneous in their incomes, demographics, and preferences. The study of the heterogeneity of urban residents is interesting in itself since, beyond making predictions about prices and the intensity of development, we also expect good models of land use to offer insights into who lives where. The heterogeneity of residents, coupled with that of the housing stock, also implies that land and property markets may not be as frictionless as assumed in the simplest land use models. In addition, the basic model is static in nature, but properties are long-lived and we cannot expect land use in cities to adjust immediately to any shock. This creates further frictions. We explore all these issues in Section 8.3.

But perhaps the most obvious criticism of the monocentric model is that cities have become less and less monocentric. The main problem with the standard approach is not that it cannot accommodate more realistic employment distributions. It can. What the standard approach cannot do easily is allow the distribution of jobs to be endogenous, interacting with the distribution of residents. Much modeling effort has been devoted to this problem since the late 1970s. Residents face a trade-off between accessibility and land and property prices. Businesses benefit from proximity to other businesses

[4] We discuss below the empirical work of Clark (1951), an exception that appeared before the main framework was established. See also the introduction in Quigley (1998) for a brief intellectual history of the early work on urban land use.

because of agglomeration economies but, if they cluster, they must pay higher land prices and also compensate their workers for longer commutes with higher wages. Section 8.4 provides a tractable model of land use in cities under endogenous business location dealing with these complex issues and summarizes other efforts at modeling secondary centers and job decentralization.

No work on urban land use would be complete without a discussion of government intervention. Land and the properties erected on it are usually highly regulated. We explore and discuss the possible reasons for these regulations and their possible effects in Section 8.5.

Our treatment of the theoretical literature gives a dominant role to the accessibility of jobs. While clearly important, job accessibility is not the sole determinant of how land is used and how properties are valued. First, commuting is only one aspect of urban travel. Thus, accessibility should be broadly understood to include proximity to shops, school, amenities, etc. Second, other aspects of location, such as heterogeneity and neighborhood interactions, matter greatly. This said, we believe focusing on accessibility is warranted because it seems uniquely important in shaping cities at a broader scale.

To be useful and become more than a speculation, a conceptualization must confront the empirical reality. This is what the last four sections of our chapter aim to do. Section 8.6 examines the empirical literature that assesses the gradient predictions of the simplest models of urban land use. Section 8.7 then turns to the empirical location patterns of heterogeneous city residents. Section 8.8 looks at recent patterns of residential land development. Finally, Section 8.9 examines changes in business location within cities.

8.2. MODELING URBAN LAND USE: THE MONOCENTRIC MODEL

Land use is one of the oldest topics of economic analysis. Ricardo (1817) and von Thünen (1826), two of the pioneers of the discipline, offered early insights. Ricardo observed that the rent differential between two parcels of land should be equal to the difference in the revenue derived from these two parcels, all else being equal. von Thünen discovered the same idea independently and embedded it in a simple model of farming to show that, on a flat featureless plain, crops that are more costly to transport should be located closer to the center of the village while crops that are cheaper to transport should be cultivated in outer rings.[5] As will become clear below, the insights of Ricardo and Thünen still form the basis of much of our understanding of land use issues today.

The modern approach to urban land use has its origins in the work of Alonso (1964), Mills (1967), and Muth (1969), who first encapsulated city transport, land use, and

[5] von Thünen's work may have been the first ever fully specified formal economic model. He is also credited as being one of the fathers of marginalism (Samuelson, 1983; Fujita, 2012).

population issues into what is known as the monocentric city model. In this section, we first offer a formal exposition of this model. Then, we turn to a number of extensions in Sections 8.3 and 8.4. Our review of these extensions is highly selective because this model has been thoroughly investigated. A full review would take us well beyond the scope of this chapter.

In this section, we expand on the exposition of the monocentric city model we introduced in earlier work (Duranton and Puga, 2014).[6] Consider a linear monocentric city. Land covered by the city is endogenously determined and can be represented by a segment on the positive real line. Production and consumption of a numéraire good take place at a single point $x = 0$, the central business district (CBD). In addition to the numéraire good, individuals living in the city consume housing. The production of housing from land in the city is described below. For now, it is nonetheless important to keep in mind that land and the housing that sits on it are allocated competitively to the highest bidder at each location. Preferences can be represented by a utility function $u(h, z)$ derived from individual consumption of housing, h, and of the numéraire, z. This function is increasing in both its arguments and is strictly quasi concave.

Commuting costs increase linearly with distance to the CBD, so a resident living at distance x from the CBD incurs a commuting cost τx. This leaves $w - \tau x$ for expenditure on housing and the numéraire. If we denote by $P(x)$ the rental price of housing at a distance x from the CBD, this resident's budget constraint is thus $w - \tau x = P(x)\ h + z$.

All residents in the city are identical in income and preferences and are freely mobile within the city, and hence must derive a common level of utility $\underline{u}$ from the heterogeneous combinations of housing and the numéraire they consume.

Relative to the standard consumer problem studied in introductory economics, there are two main differences. First, residents must choose the location of their residence as well as allocate their disposable income optimally between housing and the numéraire. The price of housing, and thus the budget constraint they face, varies with their location choice. Without loss of generality, we can think of each resident solving for a standard budget allocation between housing and the numéraire at each location and then choosing the location that offers the highest utility. Second, the price of housing at each location is unknown to the analyst and needs to be computed as part of the equilibrium.

This first version of the monocentric model makes six simplifying assumptions. The first three concern the linearity of cities, the restriction of urban travel to commutes, and a particular specification for commuting costs paid in the numéraire and increasing proportionately with distance. These assumptions are for expositional purposes. Most of the

[6] For further reading, Brueckner (1987) offers a remarkably clear and intuitive graphical presentation of the monocentric model. A more detailed presentation can be found in the classic work of Fujita (1989). See also the appendix in Zenou (2009), where the Marshallian and bid-rent approaches are further developed and compared.

results we derive below readily generalize to two-dimensional cities, other reasons for travel, and other specifications for commuting costs, including commuting costs paid in time or affecting utility. We highlight the few cases where the results depend on these simplifying assumptions. The other three assumptions made here are that of a static model, an exogenous geography of jobs, and homogeneous residents. The next two sections explore extensions that attempt to relax these assumptions. Finally, we also consider that accessibility is the only factor that determines land use. We discuss below how other determinants of land use can be considered together with accessibility.

8.2.1 The Marshallian approach

There are several ways to solve this model. Since they all shed a different light on its mechanics, we describe them in turn. The first path is the Marshallian approach (also sometimes referred to as the indirect approach). This approach solves the individual budget allocation between housing and the numéraire at each location and then obtains house prices by ensuring that, with each consumer allocating optimally his or her disposable income, utility is equalized across locations in the city.

Maximizing utility $u(h, z)$ with respect to h and z subject to $w - \tau x = P(x)\, h + z$ is equivalent to maximizing $u(h, w - \tau x - P(x)\, h)$ with respect to h. The first-order condition of this problem yields a unique Marshallian demand for housing at each location $h(x)$ defined implicitly by $\frac{\partial u(.)}{\partial h} - \frac{\partial u(.)}{\partial z} P(x) = 0$ or, equivalently,

$$P(x) = \frac{\dfrac{\partial u(.)}{\partial h}}{\dfrac{\partial u(.)}{\partial z}} \ . \tag{8.1}$$

This expression is of course the standard first-order condition for utility maximization by residents stating that the marginal utility of more housing per amount spent should be equal to the marginal utility of the numéraire. Using the budget constraint again, we can recover the Marshallian demand for the numéraire $z(x)$ as a function of the Marshallian demand for housing $z(x) = w - \tau x - P(x)\, h(x)$.

In equilibrium, given that all individuals have the same income and are freely mobile, they must obtain the same level of utility $\underline{u}$ defined above (we return later to how it is determined). Thus,

$$u(h(x),\, w - \tau x - P(x)\, h(x)) = \underline{u} \ . \tag{8.2}$$

Totally differentiating Equation (8.2) with respect to x yields

$$\frac{\partial u(h,z)}{\partial h}\frac{\partial h(x)}{\partial x} - \frac{\partial u(h,z)}{\partial z}P(x)\frac{\partial h(x)}{\partial x} - \frac{\partial u(h,z)}{\partial z}\left(\tau + h(x)\frac{\mathrm{d}P(x)}{\mathrm{d}x}\right) = 0 \ . \tag{8.3}$$

By Equation (8.1), the first two terms in Equation (8.3) cancel out (this is just the envelope theorem at work), which implies

$$\frac{dP(x)}{dx} = -\frac{\tau}{h(x)} < 0 \,. \tag{8.4}$$

Equation (8.4) is known as the Alonso–Muth condition.[7] It states that, at the residential equilibrium, if a resident moves marginally away from the CBD, the cost of his or her current housing consumption falls just as much as his or her commuting costs increase. Thus, the price of housing decreases with distance to the CBD. This housing price gradient is the first of a series of gradients that occur in the monocentric model.

The main drawback of the Marshallian approach is that it gets to the solution in a roundabout way. It solves first for the consumer program in a location before recovering the price of housing at this location through the residential equilibrium condition. Then, knowing the price of housing, it returns to the choice of consumption before solving for the optimal location. The main advantage of the Marshallian approach is to make clear that the price of housing at each location is endogenous and emerges within the model.

8.2.2 The bid-rent approach

The Alonso–Muth condition can be derived more directly using the so-called bid-rent approach (also known as the direct approach). We define the bid-rent function for housing $\Psi(x, \underline{u})$ as the maximum price a resident is willing to pay for housing at distance x from the CBD while enjoying utility $\underline{u}$ and satisfying the budget constraint:

$$\Psi(x, \underline{u}) \equiv \max_{h(x),\ z(x)} \left\{ P(x) | u(h,\ z) = \underline{u},\ w - \tau x = P(x)\, h(x) + z(x) \right\} . \tag{8.5}$$

To reduce the problem to a single constraint instead of two constraints, one can solve the budget constraint for $P(x)$ and substitute this into the original program to obtain

$$\Psi(x, \underline{u}) = \max_{h(x), z(x)} \left\{ \frac{w - \tau x - z(x)}{h(x)} \mid u(h,\ z) = \underline{u} \right\} . \tag{8.6}$$

We can then turn the program into an unconstrained maximization problem by replacing $z(x)$ with the restricted Hicksian demand for the numéraire $z(h(x), \underline{u})$:

[7] Although it is also sometimes referred to as the spatial equilibrium condition, we prefer the Alonso–Muth label and reserve spatial equilibrium for the equalization condition (8.2). As made clear below, the Alonso–Muth condition will occur in models where there is no spatial condition involving a full equalization like (8.2). We also note that spatial equilibrium conditions are also present in many other models, including settings that involve amenity or wage differences across locations.

$$\Psi(x, \underline{u}) = \max_{h(x)} \left\{ \frac{w - \tau x - z(h(x), \underline{u})}{h(x)} \right\}. \tag{8.7}$$

Now the Alonso–Muth condition can be obtained by direct application of the envelope theorem to Equation (8.7):

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x} = \frac{\mathrm{d}\Psi(x, \underline{u})}{\mathrm{d}x}\bigg|_{h(x)=h(\Psi(x,\ \underline{u}),\underline{u})} = -\frac{\tau}{h(x)} < 0\,. \tag{8.8}$$

The consumption of housing when the individual pays the bid-rent price can be obtained from the first-order condition for the program of Equation (8.7):

$$\frac{\partial z(h(x), \underline{u})}{\partial h(x)} h(x) + w - \tau x - z(h(x), \underline{u}) = 0\,. \tag{8.9}$$

Note that this can be rewritten as

$$\frac{\partial z(h(x), \underline{u})}{\partial h(x)} = -\frac{w - \tau x - z(h(x), \underline{u})}{h(x)}\,, \tag{8.10}$$

where the left-hand side is the slope of the indifference curve $u(h, z) = \underline{u}$ and the right-hand side is the slope of the budget constraint. Panel (a) in Figure 8.2 illustrates the approach. The vertical axis represents the consumption of numéraire $z(x)$ and the horizontal axis that of housing $h(x)$. The line $z(x) = w - \tau x - P(x)h(x)$ is the budget constraint, which has vertical intercept $w - \tau x$ and slope $-P(x)$. The bid-rent or maximum price that the consumer can pay for housing while staying within the budget constraint and achieving utility $\underline{u}$ is found by pivoting the budget constraint around its vertical intercept until it is tangent to the indifference curve $u(h, z) = \underline{u}$.[8]

While less standard, the bid-rent approach is a more direct way for solving the monocentric model. It also highlights the bidding nature that prevails for the allocation of housing (and land) at each location.

To make the bid-rent approach more concrete, assume, for instance, that $u(h, z) = h^{\alpha} z^{1-\alpha}$, with $0 < \alpha < 1$. Then, $z(h(x), \underline{u}) = h(x)^{-\alpha/(1-\alpha)}\, \underline{u}^{1/(1-\alpha)}$. In this particular case, Equation (8.7) can be written as

$$\Psi(x, \underline{u}) = \max_{h(x)} \left\{ \frac{w - \tau x - h(x)^{\frac{-\alpha}{1-\alpha}}\, \underline{u}^{\frac{1}{1-\alpha}}}{h(x)} \right\}. \tag{8.11}$$

Solving for $h(x)$ in Equation (8.11) and reinserting the solution into it finally yields

[8] Note that this is different from the standard consumer problem solved when deriving the expenditure function, which keeps the slope of the budget constraint fixed and moves it in parallel until reaching the tangency point with $u(h, z) = \underline{u}$. Here, we instead keep the intercept fixed and rotate the budget constraint to reach the tangency point.

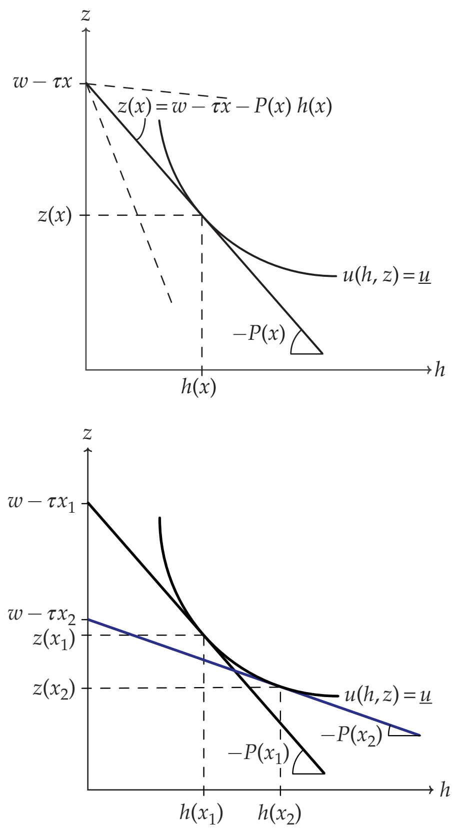

Figure 8.2 A graphical representation of the monocentric model. Panel (a): Deriving housing prices in *x*. Panel (b): Comparative statics.

$$h(x) = \left(\frac{\underline{u}}{(1-\alpha)^{1-\alpha}(w-\tau x)^{1-\alpha}}\right)^{\frac{1}{\alpha}} \text{ and } P(x) = \Psi(x, \underline{u}) = \frac{\alpha(w-\tau x)}{h(x)} = \alpha(1-\alpha)^{\frac{1-\alpha}{\alpha}}\left(\frac{w-\tau x}{\underline{u}}\right)^{\frac{1}{\alpha}}. \tag{8.12}$$

Note that the expenditure on housing $P(x)h(x)$ is a constant share α of the wage net of the commuting cost in the second expression. This is unsurprising given the Cobb–Douglas nature of the utility function assumed in this example. Note also that the housing price gradient is convex.

8.2.3 The dual approach

For the third path to the solution, we use a dual representation of the utility function $v(P(x), w - \tau x)$, where $\frac{\partial v}{\partial P(x)} < 0$, and $\frac{\partial v}{\partial (w - \tau x)} > 0$. The residential equilibrium can be restated as

$$v(P(x),\, w-\tau x)=\underline{u}\;. \tag{8.13}$$

By the definition of the expenditure function $e(P(x),u)$,

$$e(P(x), v(P(x),\, w-\tau x))=w-\tau x\;. \tag{8.14}$$

Substituting Equation (8.13) into (8.14) and totally differentiating with respect to x yields

$$\frac{\partial e(P(x),\underline{u})}{\partial P(x)}\frac{\mathrm{d}P(x)}{\mathrm{d}x}=-\tau\;, \tag{8.15}$$

which implies the Alonso–Muth condition immediately after using Shephard's lemma:

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x}=-\frac{\tau}{\dfrac{\partial e(P(x),\underline{u})}{\partial P(x)}}=-\frac{\tau}{h(P(x),\underline{u})}<0\;. \tag{8.16}$$

While perhaps less intuitive, the dual approach offers the most direct way to derive the Alonso–Muth condition. Because it makes subsequent derivations easier, we mainly retain the dual approach from now on.[9]

Residents react to the lower price of housing by consuming more of it (i.e., living in larger residences) the farther they live from the CBD. To see this, simply differentiate the Hicksian demand for housing with respect to x:

$$\frac{\partial h(P(x),\,\underline{u})}{\partial x}=\frac{\partial h(P(x),\,\underline{u})}{\partial P(x)}\frac{\mathrm{d}P(x)}{\mathrm{d}x}\geq 0\;. \tag{8.17}$$

Equation (8.17) introduces a second gradient: the consumption of housing increases with the distance to the CBD. Note this is a pure substitution effect, since utility is being held constant at $\underline{u}$. Equation (8.17) together with Equation (8.16) immediately implies that the price of housing is convex in distance to the CBD, $\frac{\mathrm{d}^2P(x)}{\mathrm{d}x^2}>0$: house prices do not need to fall as fast as commuting costs increase with distance to the CBD to keep city residents indifferent, since they enjoy having a larger house. This convexity of the housing price gradient already appeared in the example above. This is a generic property of the model, not an artifact of functional forms used to specify preferences.[10]

[9] The use of the dual approach in urban economics was originally introduced by Solow (1973) and was subsequently used by Polinsky and Shavell (1975, 1976), Henderson (1977), and Kanemoto (1980). Duranton and Puga (2014) also use the dual approach in their derivation of the monocentric model but totally differentiate Equation (8.13) and invoke Roy's identity to derive the Alonso–Muth condition (8.16) instead of using the expenditure function and Shephard's lemma.

[10] The housing price gradient may be concave only if commuting costs are sufficiently convex in distance instead of being linear as assumed here. With commuting costs being $\tau(x)$ instead of $\tau\, x$, the Alonso–Muth condition has the marginal cost of commuting $\frac{\mathrm{d}\tau(x)}{\mathrm{d}x}$ instead of τ in the numerator. Hence, only when the commuting function is convex enough can the housing price gradient be concave. Commuting costs highly convex in distance seem unlikely.

Panel (b) in Figure 8.2 illustrates these results. It considers two locations x_1 and $x_2 > x_1$. The price of housing in x_1, $P(x_1)$, is given by the slope (in absolute value) of the budget constraint with intercept $w - \tau x_1$ that is tangent to the indifference curve $u(h, z) = \underline{u}$. The point where the indifference curve and the budget constraint are tangent allows us to read the consumption of numéraire $z(x_1)$ on the vertical axis and the consumption of housing $h(x_1)$ on the horizontal axis. For the resident in x_2, the price of housing, $P(x_2)$, and the consumption of numéraire, $z(x_2)$, and housing, $h(x_2)$, can be read in the same way using this time a budget constraint with intercept $w - \tau x_2$. This lower budget constraint must be flatter—that is, involve a lower price of housing—for it to be tangent to the indifference curve. Then it is also immediate that the consumption of housing is higher in x_2 relative to x_1, whereas the consumption of the numéraire is lower.

We now turn from housing to the land it is built on. To supply housing, a perfectly competitive construction industry uses land and capital under constant returns to scale to produce an amount $f(x)$ of housing floor space per unit of land at a distance x from the CBD. The production function for housing is increasing in both its arguments and is strictly quasi concave. The rental price of land, denoted $R(x)$, varies across the city. The rental price of capital is constant and endogenously given, so we omit it as an argument of the unit cost function in construction $c(R(x))$. The zero-profit condition for the construction sector can then be written as $P(x) = c(R(x))$. Totally differentiating this expression with respect to x yields

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x} = \frac{\partial c(R(x))}{\partial R(x)} \frac{\mathrm{d}R(x)}{\mathrm{d}x}, \tag{8.18}$$

which implies

$$\frac{\mathrm{d}R(x)}{\mathrm{d}x} = \frac{\mathrm{d}P(x)}{\mathrm{d}x} \frac{1}{\dfrac{\partial c(R(x))}{\partial R(x)}} = \frac{\mathrm{d}P(x)}{\mathrm{d}x} f(x) < 0, \tag{8.19}$$

where the simplification follows from Shephard's lemma.[11] Thus, the reduction in the price of housing as one moves away from the CBD is reflected in a reduction in the price of land. The construction industry then reacts to lower land prices by building with a lower capital-to-land ratio further away from the CBD. Put differently, there are two other gradients here as one moves away from the CBD: declining land prices and

[11] Note that the relationship between the land gradient price and the housing price gradient can alternatively be written as $\frac{\mathrm{d}\log R(x)/\mathrm{d}x}{\mathrm{d}\log P(x)/\mathrm{d}x} = \frac{f(x)P(x)}{R(x)}$. Thus, the land price decline is many times the house price decline. In particular, the ratio of the percentage decline in the land price to the percentage decline in the housing price equals the ratio of the price level of housing and the price level of the land the housing is built on.

declining capital intensity in housing (i.e., both larger gardens and properties with fewer stories).

Land is built if the rent $R(x)$ it can fetch in residential use is at least as high as the rent $\underline{R}$ it can fetch in the best alternative use (e.g., agriculture). The edge of the city is thus located at a distance $\overline{x}$ from the CBD such that

$$R(\overline{x}) = \underline{R}\ . \tag{8.20}$$

The physical extent of the city must also be sufficient to hold its population N:

$$N = \int_0^{\overline{x}} n(x)\mathrm{d}x\ , \tag{8.21}$$

where $n(x)$ denotes population density at a distance x from the CBD. Using Equations (8.16) and (8.19), we can express population density as

$$n(x) = \frac{f(x)}{h(x)} = \frac{\dfrac{\mathrm{d}R(x)}{\mathrm{d}x} \Big/ \dfrac{\mathrm{d}P(x)}{\mathrm{d}x}}{-\tau \Big/ \dfrac{\mathrm{d}P(x)}{\mathrm{d}x}} = -\frac{1}{\tau}\frac{\mathrm{d}R(x)}{\mathrm{d}x}\ . \tag{8.22}$$

By Equation (8.19) and the fact that capital intensity declines with distance to the CBD, $\frac{\mathrm{d}f(x)}{\mathrm{d}x} < 0$, density also declines with distance to the CBD, $\frac{\mathrm{d}n(x)}{\mathrm{d}x} < 0$. This gradient of declining density of residents as one moves farther from the CBD is the fifth gradient predicted by this model. As made clear by Equation (8.22), it is a direct implication of two other gradients already discussed: the increase in housing consumption and the decline in the capital intensity of development as one moves farther from the CBD.

Substituting expression (8.22) for $n(x)$ into Equation (8.21), solving the integral, and using Equation (8.20) yields $N = \frac{R(0) - \underline{R}}{\tau}$. This implies a very simple expression for land rent at the CBD ($x = 0$):

$$R(0) = \underline{R} + \tau N\ . \tag{8.23}$$

Valuing the residential equilibrium condition $v(P(x), w - \tau x) = \underline{u}$ at $x = 0$ and using Equation (8.23), we can write the price of housing at the CBD as $P(0) = c(\underline{R} + \tau N)$. The residential equilibrium condition holds for any location in the city, so valuing it at an arbitrary x and at $x = 0$, and using the previous expression for $P(0)$ yields

$$\begin{aligned} v(P(x), w - \tau x) &= \underline{u} = v(P(0), w) \\ &= v(c(\underline{R} + \tau N), w)\ . \end{aligned} \tag{8.24}$$

This can be inverted to solve for house prices $P(x)$ as a function of x, N, w, τ, and $\underline{R}$. That is the "closed city" version of the monocentric city model, which treats population N as a parameter. The "open city" version allows N to be endogenously determined by migration across cities to attain a common utility level. We can write the condition of utility equalization across cities as

$$v(c(\underline{R} + \tau N), w) = \underline{u} \,. \tag{8.25}$$

This spatial equilibrium condition can be inverted to solve for N as a function of $\underline{u}$, w, τ, and $\underline{R}$.

The gradients presented so far compare locations within the city for given parameters. While a full analysis of the comparative statics of the model is beyond our scope here, it is also of interest to describe how the city is affected by changes in these parameters. We focus on the open city case, where we can think of the comparative statics as the consequences of changes that affect a city that is small relative to the entire economy. Then the level of utility $\underline{u}$ is pinned down by what consumers can achieve elsewhere. An increase in the wage in a city increases house prices everywhere in the city: Equation (8.7) immediately implies $\frac{\mathrm{d}P(x)}{\mathrm{d}w} > 0$. Housing needs to become more expensive to offset higher wages as residents need to retain the same level of utility as elsewhere in the economy, and this is attained through a population increase in the city.

A reduction in transport costs—for instance, from road or transit improvements—also increases house prices everywhere in the city: $\frac{\mathrm{d}P(x)}{\mathrm{d}\tau} < 0$. This increase in house prices is again the result of individuals moving into the city in response to the utility gain from lower commuting costs. To accommodate this larger population, cities physically expand outward and also experience rising densities. Of these two channels, the monocentric model predicts that the physical expansion of the city is more important. To see this, consider a segment of the city between the CBD and a point x_C, and think of this segment as the historical central city. We can then think of the segment between x_C and the city edge $\overline{x}$ as the suburbs. Let $N_C = \int_0^{x_C} d(x)\mathrm{d}x$ denote the (endogenous) population of the central city. Then, using Equations (8.22) and (8.23), we can calculate the share of the population in the central city as

$$\frac{N_C}{N} = \frac{R(0) - R(x_C)}{R(0) - \underline{R}} \,. \tag{8.26}$$

A reduction in τ increases land rent at any given point beyond the CBD including x_C, but it does not affect land rent $R(0)$ at the CBD (where there is no need to commute and migration keeps utility unchanged) nor land rent at the city edge, which is fixed at $\underline{R}$. Then Equation (8.26) implies that the share of the population in the central city falls when commuting costs are reduced. This implies that improvements in local transport foster the suburbanization of the population.

To sum up, the monocentric model makes a number of important predictions. It is best known for predicting gradients. We saw five of them above. As one moves away from the CBD, housing prices decline, housing consumption increases, land prices decline, the density of construction declines (i.e., capital intensity in housing declines), and population density declines. These five gradients are potentially measurable and appear intuitively plausible.

8.2.4 Quantitative predictions

Beyond these qualitative predictions, the monocentric model also makes a number of sharp quantitative predictions. First, the Alonso–Muth condition in Equation (8.4) is more precise than simply giving a direction of change. It states that the slope of the housing price gradient is equal to the ratio of the marginal cost of commuting, τ, to the consumption of housing, $h(x)$. More intuitively, as one moves away from the CBD, the marginal decline in house prices should be exactly offset by the marginal increase in the cost of travel. This is observable in principle.

This is not the only quantitative prediction that can be derived from the monocentric model. The second one is apparent in Equation (8.19) and results from optimal investment in housing and zero profit in construction: the ratio of the land price gradient to the housing price gradient, $\frac{\mathrm{d}R(x)/\mathrm{d}x}{\mathrm{d}P(x)/\mathrm{d}x}$, should be equal to the amount of housing $f(x)$ (in terms of floor space per unit of land). Put differently, if for the same housing price gradient the land price gradient is twice as steep, there should be twice as much housing.

The third quantitative prediction appears in Equation (8.22). It states that the population density in x, $n(x)$, is equal to minus the ratio of the land price gradient, $\frac{\mathrm{d}R(x)}{\mathrm{d}x}$, to the marginal cost of commuting, τ. Just like with housing price, land prices adjust to reflect the greater cost of commuting to the CBD but, for land, this relationship needs to be weighted by the number of local residents. This condition is, of course, the land price equivalent of the Alonso–Muth condition (8.4).

While these three quantitative predictions of the monocentric model are "local" in the sense that they apply at each location, the monocentric model also generates two aggregate predictions. The first of these is made clear by Equation (8.23): the differential land rent between the CBD and the edge of the city, $R(0) - \underline{R}$, should be proportional to the city population, N, and to the unit commuting cost, τ. This is a strong quantitative prediction, which can also be potentially applied to the data. Because it concerns city aggregates, this quantitative prediction is more sensitive to our modeling assumptions and, in particular, to the geography that is imposed on cities. For instance, the proportionality of the population to the differential land rent at the center does not carry through to two-dimensional cities.[12] This is because Equation (8.21) becomes $N = \int_0^{\overline{x}} l(x)n(x)\mathrm{d}x$, where $l(x)$ is the amount of land at distance x from the center. If the city sits on a homogeneous plane, then $l(x) = 2\pi x$. However, in reality geographical

[12] The model proposed by Combes et al. (2012) implies the proportionality of land rent at the CBD to population in two-dimensional cities. The three key elements that generate this result are an isoelastic demand for housing, constant returns in construction, and multiplicatively separable commuting costs that directly enter the utility function.

constraints and the fact that parcels of land more suitable for construction tend to be built first imply that cities probably lie somewhere between the one-dimensional and the two-dimensional cases (and in some more extreme cases may be below the one-dimensional case).

The second aggregate quantitative prediction states that total commuting costs are equal to the total differential land rent:

$$\text{TCC} \equiv \int_0^{\bar{x}} \tau x n(x) \mathrm{d}x = \int_0^{\bar{x}} -\frac{x}{\tau} \mathrm{d}R(x) = \int_0^{\bar{x}} (R(x) - \underline{R}) \mathrm{d}x \equiv \text{TDLR}, \qquad (8.27)$$

where the first equality is obtained using (8.22) and the second requires integrating by parts. This result for the proportionality of city aggregates is less sensitive to the modeling of cities as shown by Arnott and Stiglitz (1981). Different geographies imply a different, but still constant, ratio of total differential land rent to total commuting costs. For instance, with two-dimensional circular cities, the total commuting cost is twice the total differential land rent. The linearity of commuting costs remains nonetheless crucial for this aggregate result.

Finally, it remains to be said that the equilibrium of the monocentric model presented here is efficient. This is an expected outcome given the competitive behavior of residents and housing builders and the absence of externalities. It has, however, been well known since Mirrlees (1972) and Dixit (1973) that a non-Rawlsian planner may want to treat ex ante equal residents unequally ex post. As shown by Wildasin (1986), this is because the marginal utility of income differs across locations at the symmetric equilibrium. Hence, a utilitarian planner will allocate workers differently from the competitive equilibrium case explored here, where *ex ante* equal residents end up with the same level of utility ex post (despite living in different locations). A Rawlsian planner will mimic the market outcome. This can be proven by showing that the market allocation minimizes the total social cost (the sum of the commuting cost, the opportunity cost of land, and cost of numéraire consumption) required to achieve a given level of utility $\underline{u}$. We do not develop the proof here for the sake of brevity, but the interested reader can consult Fujita and Thisse (2013, p. 85).

8.3. EXTENDING THE MONOCENTRIC MODEL

The monocentric model presented above is in many respects a remarkable achievement. It brings together housing, construction, transport, and the choice of location and consumption made by households in a tractable way. It derives many predictions. Many of these predictions regard the existence of several gradients, and casual observation is certainly supportive of these predictions. The monocentric model is also a very elegant construct whose apparent simplicity actually reveals a lot of subtlety. Unsurprisingly, the work of Alonso (1964), Mills (1967), and Muth (1969) hailed the creation of a new subfield of economics: urban economics. Although, as the rest of this volume hopefully makes clear, this field has broadened its objects of study to a variety of other issues, most urban economists still recognize the monocentric model as being core to the urban field.

This said, the exposition in Section 8.2 relies on strong simplifying assumptions. Our objective in this section is threefold. First, we aim to review the main extensions of the monocentric models to take stock of the state of the literature. Second, we want to distinguish between the extensions that add to the richness of the monocentric model and work that questions its main results. Our review here puts more emphasis on the latter. Finally, we also emphasize the extensions that are the most relevant for the empirical work which is reviewed from Section 8.6 onward.

8.3.1 Other elements of travel costs and looking beyond accessibility

Our review of extensions to the monocentric model begins with generalizations that do not constitute major theoretical challenges for the baseline model presented in Section 8.2. Rather, they allow us to clarify what the monocentric model is really about and also help highlight a number of issues which are of first-order empirical importance.

The first of these issues regards the specification for commuting costs. Travel is costly, not only in pecuniary terms, but also in time. Transport economists routinely value time spent in privately owned vehicles at half the wage of the traveler (Small and Verhoef, 2007). To consider the time cost of commuting, we assume that travel costs t units of time per unit of distance in addition to out-of-pocket expenses. We also assume that each resident chooses how much labor ℓ to supply and how much leisure s to consume. Leisure enters the utility function as a positive argument, $u(h, z, s)$. Each resident located at distance x from the CBD now faces both a pecuniary budget constraint, $w\,\ell(x) - \tau\, x = P(x)\, h(x) + z(x)$, and a time budget constraint, $\ell(x) + s(x) + t\,x = 1$, following the normalization of the total endowment of time of each resident to unity. We can insert the latter budget constraint into the former and obtain

$$w = P(x)\,h(x) + z(x) + w\,s(x) + w\,t\,x + \tau\,x = P(x)\,h(x) + z(x) + w\,s(x) + T(x), \tag{8.28}$$

where $T(x) \equiv w\,t\,x + \tau\,x$ is the total cost of commuting, and the wage w is also the shadow cost of time. A resident in x maximizes his or her utility $u(h(x), z(x), s(x))$ with respect to his or her consumption of housing, leisure, and other goods subject to the budget constraint (8.28). We can insert the budget constraint into the utility and have this resident maximize $u(h(x), w(1 - s(x)) - P(x)\,h(x) - T(x), s(x))$ with respect to housing h and leisure s. From there, the Marshallian demands can be easily recovered. In equilibrium, utility must also be equalized across locations:

$$u(h(x),\, w(1 - s(x)) - P(x)\,h(x) - T(x), s(x)) = \underline{u}\ . \tag{8.29}$$

This spatial equilibrium condition is the analogue of Equation (8.2) after introduction of a time cost of commuting as part of an endogenous allocation of time. After simplifications using the first-order conditions with respect to housing and leisure, totally differentiating Equation (8.29) with respect to x to derive the optimal choice of residence yields

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x} = -\frac{1}{h(x)}\frac{\mathrm{d}T(x)}{\mathrm{d}x}, \tag{8.30}$$

which is the Alonso–Muth condition that corresponds to Equation (8.4) above with leisure and a time cost of commuting. It is then easy to show that all the other gradients exhibited above are also present here. As argued above, the aggregate properties of the simpler monocentric models are more reliant on the linearity of commuting costs. Note also that similar results with the same five gradients can be obtained by assuming that commuting enters the utility function directly as a negative argument $u(h, z, x)$. Again, solving for the first-order conditions in h, z, and x and totally differentiating the spatial indifference condition analogue to Equation (8.29) yields another version of the Alonso–Muth condition and a negative housing price gradient under mild conditions regarding preferences.[13]

While a more realistic modeling of commuting is more intricate than what the simple monocentric model of Section 8.2 assumes, this does not affect the main properties and predictions of this model. A similar conclusion holds for "accessibility." The simple monocentric model in Section 8.2 equates accessibility with distance to the CBD. This is restrictive for two reasons. First, because in reality employment is far from being concentrated around a single center, so travel to work patterns are more complex than trips to a common center. We deal with generalizations of this aspect in Section 8.4. Second, much of observed travel takes place for reasons other than commuting. Couture, Duranton and Turner (2012) report that in the largest 100 US metropolitan areas, commutes represented less than a quarter of all trips by privately owned vehicles in 2008. Shopping trips, recreational trips, and trips for other personal/family business are about equally important. Even in a city with a single employment center, the location of retail, entertainment, and family and friends is thus likely to matter a lot as well.

While these considerations are empirically of first-order importance, richer notions of accessibility can be readily incorporated into the monocentric model. Assume, for instance, that there is still a single CBD where all employment is located but there are also many equally spaced retail centers. Residents need to travel to work on a daily basis and go shopping, say, every other day. It would be easy to extend the model in Section 8.2 to account for this. Taken alone, the location of jobs still leads to a negative housing price gradient, while the location of retail implies a sawtooth pattern with a small peak at every retail location. Bringing these two elements together, one can easily see that

[13] We are grateful to Frédéric Robert-Nicoud for pointing this out to us.

the job accessibility gradient dominates since travel to retail centers is half as frequent as travel to employment and since there are many retail centers compared with just one employment center. The overall gradient is thus negative but flatter before each retail center and steeper after as one moves away from the CBD. Increasing the frequency of shopping trips or the average distance to a retail center sufficiently can lead to a nonmonotonic overall gradient.

These more complex notions of accessibility can be incorporated in a variety of other ways. For instance, Anas and Moses (1979) and, more recently, Baum-Snow (2007b) consider radial commuting highways in two-dimensional cities. This introduces some heterogeneity in the unit price of housing for properties that are located at the same physical distance to the CBD since they will differ in their distance to the radial highway. Should access to the radial highway be limited by a finite number of exits, the gradient could become nonmonotonic again because the travel distance to the CBD for a location close to a radial highway exit may be less than that of another location that is closer to the CBD but farther from this exit.

Nonmonotonic gradients also occur naturally if one considers several modes of travel that may be combined with one another. For instance, residents may walk to a railway station before their ride to the CBD. By the same type of argument as with radial roads, this generates nonmonotonic gradients with local peaks of housing prices at the railway stations.

As discussed below, it is also possible to embed considerations other than accessibility into a monocentric framework. Large metropolitan areas are often divided into several municipalities or districts, which in some countries have considerable latitude regarding taxation and the local public goods that they offer, including education in the United States. The provision of local public goods will differ across districts and affect housing prices and, in turn, patterns of land use (see, e.g., De Bartolome and Ross, 2003). Local amenities may also have a spatial dimension and will naturally affect housing prices and thus land use and the various gradients (see, e.g., Brueckner et al., 1999).

Although none of these extensions generate theoretically surprising results, they are useful to make two important points that are empirically highly relevant: accessibility cannot be reduced to the distance to the CBD and accessibility is not the only determinant of housing prices and thus of patterns of land use.

8.3.2 Heterogeneous residents, properties, and timing

The model described in Section 8.2 assumes the existence of a representative resident. Even though residents all end up in different locations, they are ex ante symmetric with similar preferences and the same labor income. From the theoretical standpoint, generating asymmetric outcomes from ex ante symmetric agents is a strength. Having residents in different locations and with different consumption levels of housing and other goods is the equilibrium result of a market allocation. It is not the direct and somewhat trivial implication of location-specific tastes, for instance, nor a choice made by an urban planner. While exploring a case with representative residents is theoretically important, it remains nonetheless crucial to understand how household heterogeneity affects residential location patterns.

A first way to model household heterogeneity is to consider different groups with, for instance, heterogeneous income levels.[14] Consider, for instance, N_0 poor residents with a wage w_0 and a utility level $\underline{u}_0$ and N_1 rich residents with a wage $w_1 > w_0$ and a utility level $\underline{u}_1 > \underline{u}_0$. If housing is essential, residents from both groups must consume a positive amount of housing. This implies that in some locations poor residents must outbid rich residents. There must also be other locations where the opposite is true. Consider points that separate rich and poor and denote $\tilde{x}$ one such "boundary" point.[15] At $\tilde{x}$, the bid rents of rich and poor must be the same: $P_0(\tilde{x}, \underline{u}_0) = P_1(\tilde{x}, \underline{u}_1)$. Otherwise, if one group were willing to pay strictly more, there would be only one group of residents, rich or poor, living on both sides of this point, which contradicts the definition of $\tilde{x}$ as a boundary point. This equality in bid rent across groups implies that rich residents must then consume more housing than poor resident at $\tilde{x}$ as long as housing is a normal good. That is, we must have $h(P_1(\tilde{x}, \underline{u}_1)) > h(P_0(\tilde{x}, \underline{u}_0))$. If we use the Alonso–Muth condition (8.16), this implies $\frac{\mathrm{d}P_1(\tilde{x})}{\mathrm{d}x} = -\frac{\tau}{h(P_1(\tilde{x}), \underline{u}_1)} > -\frac{\tau}{h(P_0(\tilde{x}), \underline{u}_0)} = \frac{\mathrm{d}P_0(\tilde{x})}{\mathrm{d}x}$. Thus, the rent gradient should be steeper for poor residents than for rich residents.

If poor residents have a steeper housing price gradient, in equilibrium they will live closer to the CBD, whereas rich residents should live further away. Hence, when commuting costs are the same for both groups and housing is a normal good, poor residents are predicted to occupy small dwellings close to the CBD, where housing is more expensive. The key driver of this somewhat counterintuitive result is that rich residents are more willing to pay greater commuting costs and live further from the CBD because their higher wage allows them to consume more land.

As discussed in Section 8.3.1, in practice the cost of commuting involves a time cost, and the opportunity cost of time is likely larger for rich residents. Working in the opposite direction is the fact that poor residents are more likely to rely on public transport. With different commuting costs for poor and rich τ_0 and τ_1, the condition for poor residents to occupy small central dwellings and rich residents to live in larger residences out in the suburbs is

[14] For the sake of brevity, most of our focus here is on income heterogeneity. There are other forms of heterogeneity that matter. Race is of particular salience in the United States.

[15] The reasoning can be extended to intervals where rich and poor coexist (although they do not occur in equilibrium here).

$$\frac{\mathrm{d}P_1(\tilde{x})}{\mathrm{d}x} = -\frac{\tau_1}{h(P_1(\tilde{x}),\underline{u}_1)} > -\frac{\tau_0}{h(P_0(\tilde{x}),\underline{u}_0)} = \frac{\mathrm{d}P_0(\tilde{x})}{\mathrm{d}x}. \tag{8.31}$$

The literature sometimes uses this condition directly, implying that rich residents live further out if the ratio of commuting costs per unit distance to housing consumption is lower for them (note the minus sign in front of both sides of the inequality in Equation (8.31)). However, it is also quite frequent to see the condition expressed in terms of elasticities. In this case, the condition for rich residents to live further out is that the income elasticity of commuting costs is smaller than the income elasticity of the demand for housing.[16] Finally, some articles prefer to express the condition in terms of the income elasticity of the demand for land (although residents do not value land per se, and consume it indirectly as an input used in the production of housing). Using Equation (8.19), we can rewrite Equation (8.31) as

$$\frac{\mathrm{d}R_1(\tilde{x})}{\mathrm{d}x} = \frac{\mathrm{d}P_1(\tilde{x})}{\mathrm{d}x}f(\tilde{x}) = -\frac{\tau_1 f(\tilde{x})}{h(P_1(\tilde{x}),\underline{u}_1)} > -\frac{\tau_0 f(\tilde{x})}{h(P_0(\tilde{x}),\underline{u}_0)} = \frac{\mathrm{d}P_0(\tilde{x})}{\mathrm{d}x}f(\tilde{x}) = \frac{\mathrm{d}R_0(\tilde{x})}{\mathrm{d}x}. \tag{8.32}$$

Note that $\frac{h(P(x),\underline{u})}{f(x)}$ is the demand for land (embedded in housing consumption), calculated as the product of housing floor space, $h(P(x),\underline{u})$, and land input per unit of floor space produced, $\frac{1}{f(x)}$. Hence, rich residents live further out if the ratio of commuting costs per unit distance to land consumption is lower for them. Alternatively, expressed in terms of elasticities, the condition is that the income elasticity of commuting costs must be smaller than the income elasticity of the demand for land.

While we discuss the empirical relevance of this result below, several properties of this extension to multiple demographic groups must be discussed. First, within each demographic group the monocentric model remains exactly as in Section 8.2 and the results are pinned down by utility equalization across locations. Of course, utility must be equalized within groups but not across groups. Second, at any location, residents from different groups will want to consume different amounts of housing. In turn, this implies bid-rent curves of different steepness. Hence, in equilibrium, the group with the steepest bid-rent curve will have the highest willingness to pay close to the CBD, whereas the group with the flattest bid-rent curve will occupy the locations closest to the urban fringe. The overall bid-rent curve will be formed by the upper envelope of the bid-rent curves from the different groups of residents. This should result in further convexity of the city bid-rent curve. The equilibrium with different groups of residents also implies perfect segmentation of these groups across city segments.

[16] In the specific case where commuting costs are proportional to the wage, this condition implies that rich residents remain on the outskirts of the city provided the income elasticity of the demand for housing is above 1 (Becker, 1965).

Another way to model household heterogeneity is to assume a continuous distribution for the characteristic by which households differ. Let us retain income as the key dimension by which households differ. A continuous distribution presents the monocentric model with a much greater challenge since, under a continuum of incomes, we expect a continuum of utilities and we can no longer rely on the residential equalization condition used above: $u(h,z)=\underline{u}$.

For the problem to remain tractable and for the key intuitions to be conveyed easily, we return to the example used above where residents maximize a Cobb–Douglas utility function $u(h, z) = h^{\alpha} z^{1-\alpha}$ subject to the budget constraint $P(x)\, h + z = w(x) - \tau\, x$. The main difference is that wages are now distributed according to the probability distribution function $g(w)$, whereas before they took a discrete number of levels. For simplicity, we disregard construction and assume that there is one unit of housing available per unit of land.[17]

Solving the model involves characterizing the functions $P(x)$, $w(x)$, and $h(x)$ that describe the housing price, the wage, and the consumption of housing at each location. Formally, we are solving an assignment problem. Unlike the standard assignment problems, where, for instance, a fixed number of workers are assigned to a fixed number of machines, residents choose how much housing to consume.[18] To keep the exposition straightforward, we follow the heuristic originally proposed by Beckmann (1969).[19] This heuristic should be viewed as the counterpart to the Marshallian approach derived above with heterogeneous residents. In essence, we derive the demand for housing at each location and equate it with the supply of housing. Behrens et al. (2015) propose a more formal derivation of the equilibrium where the assignment function is explicitly modeled.[20]

With a Cobb–Douglas utility function, solving for the allocation of disposable income between housing and the numéraire readily yields

[17] We could also solve for housing development, but that would introduce yet another equation into the system below.

[18] The standard assignment problem can be traced to Koopmans and Beckmann (1957). A nice presentation can be found in Sattinger (1993).

[19] Beckmann's choice of functional forms differs slightly from ours. His proposed solution contained minor mistakes later corrected by Montesano (1972).

[20] Brueckner et al. (2002) propose another heuristic that uses a bid-rent approach. More specifically, they rely on the notion that the bid rent of a given type of resident should be maximized at the equilibrium location. That is, they consider that land is allocated through a first-price auction where landowners are able to extract all the surplus. The rest of the literature explicitly or implicitly views housing and land allocation as a second-price (English) auction or a sealed-bid first-price auction, where the seller is not able to extract all the surplus. See section 13.6 in Chapter 13 in this handbook for more on real-estate auctions.

$$h(x) = \frac{\alpha(w(x) - \tau x)}{P(x)}, \tag{8.33}$$

which was already part of expression (8.12). Solving for the location choice of residents implies another version of the Alonso–Muth condition:

$$\frac{dP(x)}{dx} = -\frac{\tau}{h(x)}. \tag{8.34}$$

Using the same type of argument as above, we expect residents with higher wages to reside further from the CBD. That is, and with a slight abuse of language, we expect positive assortative matching between residents (ordered by increasing income) and endogenously defined parcels (ordered by increasing distance to the CBD). More formally, between locations x and $x + dx$, we will find residents with income between w and $w + \frac{dw(x)}{dx}dx$, where $\frac{dw(x)}{dx}$ describes how wages change across locations.

We can now equate the supply and demand of housing between x and $x + dx$. Following the simplifying assumption made above of disregarding the construction sector and the assumption of a linear city, the supply of housing between x and $x + dx$ is dx. There is a density $g(w(x))$ of residents with income between $w(x)$ and $w(x) + \frac{dw(x)}{dx}dx$. Hence, equating demand and supply leads to $h(x)\,g(w(x))\,\frac{dw}{dx}dx = dx$, or after simplifications

$$\frac{dw(x)}{dx} = \frac{1}{h(x)\,g(w)}. \tag{8.35}$$

Substituting (8.33) into (8.34) and (8.35), we obtain a system of two differential equations for $P(x)$ and $w(x)$. In the case where the distribution of wages is a Pareto distribution, $g(w) = (b+1)\frac{\underline{w}^{b+1}}{w^b}$, it is easy to verify that the system formed by Equations (8.33)–(8.35) admits a solution of the following type:

$$w(x) = \underline{w}\,x, \quad P(x) = P(0)\,x^{-b}, \quad \text{and} \quad h(x) = h(0)\,x^{b+1}. \tag{8.36}$$

These expressions imply that the housing price gradient and the housing consumption gradients are both power functions of the distance to the CBD. This may not be very surprising since a constant share of disposable income is spent on housing and the distribution of wages is assumed to follow a power law. This expression should also make it clear why with other distributions of wages a closed-form solution will typically be unavailable. It must also be the case that if the housing consumption and housing price gradients are to be power laws, the wage gradient must be linear in x for Equation (8.33) to factorize and simplify.

Relative to the canonical model with a representative resident derived above, it is easy to see that all its key local properties are retained here with heterogeneous residents despite the absence of utility equalization across locations within the city.[21] That is, the model still predicts an Alonso–Muth condition for the housing price gradient. It also still predicts gradients of housing consumption, density of residents, etc. A development gradient could also be derived in a more complete model where construction is explicitly modeled.

Note that these gradients differ from the situation with homogeneous residents explored above. With Cobb-Douglas utility, the house price gradient with homogeneous residents given in expression (8.12) depends on the income net of commuting costs $w - \tau x$ elevated to the power $1/\alpha$, the inverse share of housing in consumption. With heterogeneous residents, it is equal to distance elevated to minus the slope parameter b of the Pareto distribution of income. Interestingly, this does not depend on the commuting cost parameter. To explain this surprising result, note first that, for a given resident to be at his or her optimal location, he or she should have no incentive to move further away from the CBD. Hence, the Alonso–Muth condition continues to hold for that person. But then, recall that land is no longer allocated through an indifference condition where residents must be indifferent everywhere. Instead, residents with different levels of income compete to occupy land. More specifically, a resident competes for land with his or her poorer neighbor closer to the CBD and with his or her other, richer, neighbor located on the other side. How much this resident will bid in equilibrium will depend on how much richer he or she is relative to his or her poorer neighbor and how much poorer he or she is relative to his or her richer neighbor. In equilibrium, the shape parameter of the distribution of income solely drives the housing price gradient.

Finally, the ordered sorting of residents by increasing income as one moves away from the CBD is extreme. In a richer model with residents that also differ in their commuting costs, Behrens et al. (2015) show that it is easy to relax this result to obtain some social mixing. In this case, the model predicts only broad trends where income rises only on average as one considers locations further from the CBD.

Assignment models have the great advantage of being able to deal more naturally with the inherent heterogeneity of city residents. This advantage comes at the cost of a much greater technical complexity. Setting up the equilibrium conditions in models less rudimentary than the one considered here is often challenging, and closed-form solutions are available only in specific cases. Since in assignment models, a resident in a given location pays only what the second person with the highest willingness to pay is willing to offer, these models naturally link different submarkets without forcing full equalization. Since Braid (1981), these models have been used to understand price changes in different quality segments of the housing market. Recent contributions include those of Määttänen and Terviö (2014) and Landvoigt et al. (2011). We do not dwell on these models further here because they do not model land use explicitly.

[21] The more aggregate properties of the monocentric models with homogeneous residents that relate total land rents and total commuting costs do not hold in general in this type of setting (Behrens et al., 2015).

While much of the literature has been concerned with heterogeneous users of land, land parcels and the properties that sit on them are also highly heterogeneous. Starting with land, Combes et al. (2012) report that for vacant parcels that were developed into single-family homes in France in 2008, the mean parcel area is 1100 m^2, with a large standard deviation of 1200 m^2. Even after the city where parcels are located and their distance to the center have been controlled for, there remains considerable heterogeneity. Furthermore, after their location and area have bene accounted for, parcels also differ in a number of other dimensions, including their shape. The shape of parcels is more difficult to characterize. Combes et al. (2012) suggest an *ad hoc* measure using the ratio of the road frontage of a property to the square root of its area. The mean of this ratio for parcels with newly built single-family homes in France is 0.68. The standard deviation is nearly as large at 0.50, with a first decile at 0.23 and a ninth decile at 1.07. Again this suggests considerable heterogeneity, this time in the shape of parcels.[22]

This heterogeneity of parcels matters. According to data used by Combes et al. (2012), land prices per square meter decline fairly sharply beyond a certain size threshold. Similarly, a standard deviation of the frontage to area ratio is valued at 8% of the sale price. When properties are considered instead of only land, the heterogeneity in land parcels is compounded by the heterogeneity of the structure that sits on them. This heterogeneity of structure concerns their size, quality, level of depreciation, and style.

Despite such heterogeneity being widely mentioned to explain why property markets do not work like other asset markets, there has been very little to no work on the determinants and implications of parcel and property heterogeneity for land use in cities.[23] Several articles by Asami and coauthors (see, e.g., Asami and Ohtaki, 2000) attempt to provide axiomatic metrics for parcel shapes and explore their development implications. There is also a debate about continuous versus discrete models of land use. Berliant (1985) questioned the consistency of continuous land use models as an approximation for large discrete economies since each resident (in continuum) can consume only a zero amount of land in equilibrium. In their responses, Papageorgiou and Pines (1990) and Asami et al. (1991) showed under which conditions appropriately defined urban models with a discrete number of residents approximate classical continuous models. Because even in the discrete models considered by these authors land parcels

[22] There is also considerable heterogeneity in residential and job densities. Anas et al. (1998) provide striking evidence of such heterogeneity in Los Angeles. A strongly smoothed three-dimensional representation of employment density makes the city look strongly monocentric with only a few subcenters. Higher levels of resolution reveal instead a highly "jagged" picture. This feature is also apparent when looking at the details of the map of Paris in Figure 8.1.

[23] Instead, there is a long tradition that treats parcel heterogeneity as a nuisance that needs to be conditioned out. This is usually done through hedonic regressions. However, the models presented here suggest that parcel location, parcel area, and their intensity of construction are all determined simultaneously. This points to some obvious endogeneity problems for hedonic regressions.

are determined endogenously without friction, this debate is about the foundations of the monocentric model, not about the implications of parcel heterogeneity.

To explain the heterogeneity of parcels at a given distance from the CBD, natural geography and roads must play a prominent role. Parcels are heterogeneous because the land itself is heterogeneous. In addition, roads and other man-made obstacles need to cut through the land.

One might object that despite this unevenness of land, we may be able to come reasonably close to the optimal parcel size and optimal development for each parcel suggested by the model above. Even if the land is uneven and divided by the roadway, the blocks of parcels are usually large enough to be able accommodate a fairly close approximation to the optimal number of parcels. The heterogeneity of parcels and properties would then be a reflection of heterogeneous residents choosing to live in the same location as, for instance, in Behrens et al. (2015), where heterogeneity in income and commuting costs implies the presence of different residents occupying parcels of different sizes with different amounts of housing living next to each other.[24] Note that this approach to parcel heterogeneity is in line with the modeling approach taken in Section 8.2 where parcels arise endogenously following choices made by residents.

There are, however, reasons to believe that the heterogeneity of demand is not the entire explanation for parcel and property heterogeneity. There might be some elements of (supply) exogeneity in the sizes and shapes of parcels. Taken literally, the monocentric model presented above implies that parcels and properties should be modified as incomes grow or commuting costs decline. The debate about the foundations of the monocentric model mentioned above has also stressed the indivisible nature of parcels and housing units (Ellickson, 1977; Berliant, 1985). In reality, most established neighborhoods often see little change in many years despite changes in income, progress in transport technology, or wild fluctuations in gasoline prices. Brooks and Lutz (2012) provide more direct evidence from the land assembly process and report that assembled parcels trade at an about 40% premium. Cunningham (2013), using different data, finds a lower but still large premium of about 20%. Whether this large premium is due to a holdout problem or to parcels hosting properties at different stages of their life cycle is an open question.[25]

[24] An alternative is to assume that residents have idiosyncratic preferences for locations within the city as in Anas (1990), where this heterogeneity in tastes is modeled using a logit specification. Greater taste heterogeneity relaxes the competition for land close to the CBD and flattens the bid-rent curve. With income heterogeneity it should also weaken the sorting of residents by income. A difficulty with the logit framework is that the utility valuation of a parcel is independent of the valuation of the parcel next to it.

[25] See Strange (1995) for a consistent model of the holdout problem. In this model, a developer first makes offers for parcels, which are accepted or rejected by landowners, who do not know how much the developer stands to gain. Landowners can make counteroffers which are accepted or rejected by the developer. There are many possible equilibria. The weakly dominant equilibrium has a number of interesting properties. The first-stage offer is not informative. Small landowners ask for relatively more and often hold out on projects. In many cases, socially profitable projects do not get implemented.

This nonetheless suggests the existence of considerable frictions regardless of from where exactly they stem. Empirically, we would like to know how much of the observed heterogeneity of residents in similar locations is caused by the exogenous supply of heterogeneous parcels and how much is caused by the demand from residents that are heterogeneous in two or more dimensions and endogenously make parcels and properties heterogeneous.[26]

Although, to our knowledge, the implications of exogenous parcel heterogeneity have not been formally explored, we can form a number of conjectures about their effects. First, parcel heterogeneity will generate some mixing of heterogeneous residents. Second, the frictions that underlie parcel heterogeneity may also be at the root of significant inefficiencies in land use, particularly in areas that were developed long ago—that is, the central part of cities. Parcels may have been of optimal size when the city was first developed. However, given changes in income, transport, or construction technology, parcels that were of optimal size 100 years ago are unlikely to be of optimal size today. They may be instead be grossly suboptimal. It would be important to know how far existing parcels are from unconstrained optimality. Getting an idea of the implied welfare losses would also be important because large deviations from optimal parcel size may only imply small welfare losses if the effects of suboptimal parcel size can be partially offset through several other margins (changing intensity of development, rising housing quality, preservation of historical buildings, etc.). More generally, the theoretical literature (and the empirical literature) has paid much more attention to trends (i.e., gradients) than to variations around those trends (i.e., the heterogeneity between neighboring properties) even though understanding variations around those trends is potentially very informative about the mechanisms that drive land use and its implications.

That both residents and properties are highly heterogeneous is a challenge to the bidding assumptions made above regarding the determination of land prices. The representative monocentric model of Section 8.2 and the assignment model developed above both assume that residents bid competitively for land (or properties) and that the process of allocating land is decided (implicitly) by a second-price auction. Heterogeneity in both supply and demand will make land and property markets thin, and the assumption of competitive bidding may no longer be warranted. The thinness of land and property markets is compounded by the fact that not everyone tries to buy and sell at the same time. This calls into question the static nature of the models used so far.

[26] Again, with only one dimension of heterogeneity as in the assignment model described above, extreme ordered sorting is expected in equilibrium. It is only when residents differ in two dimensions of heterogeneity that some mixing will occur. See Behrens et al. (2015) or Epple and Platt (1998) for different modeling approaches. See also Davis and Dingel (2013) for a model of income mixing across cities.

At this stage, we need to recognize that many land and property markets may be better characterized as search markets. In the simple case of a resident seeking to buy a starter home, this resident will first search for a number of properties. When he or she sees one that is "good enough," he or she will view it as a potential match and make an offer. Sometimes there will be other potential buyers interested in the same property. More often perhaps, this prospective buyer will negotiate on a one-to-one basis with the seller to reach an agreement on the price before going forward with the transaction. This portrayal corresponds closely to the mechanics of the canonical job search model (Mortensen and Pissarides, 1994; Pissarides, 2000). More specifically, after the tags "properties," "seller," and "resident" have been replaced by "job," "employer," and "job seeker," respectively, this stylized description of the housing market is the one usually applied to a job search and the formation of standard employment relationships. The main difference is that, for the labor markets, supply and demand are independent and employers can create jobs by investing, whereas for the housing market, sellers of houses are often also buyers.[27]

The key model that describes the housing market as a two-sided search process is due to Wheaton (1990). Beyond being consistent with the simple depiction of the housing market in the previous paragraph, this model is useful in replicating a number of interesting stylized facts about property markets, such as the prevalence of bargaining and the existence of persistent vacancies with properties coming in and out of the market.

There has been a steady stream of research on searches in the housing market. This literature is discussed in Chapter 13 in this handbook. Because of the close resemblance of a housing search with many other forms of searches in economic life, the lessons of this broader literature are arguably relevant for housing (see Rogerson et al., 2005, for a review), keeping in mind nonetheless the substantive differences between housing and labor such as the buyer–seller problem mentioned above. There is little empirical work on housing searches that closely relates to theory. A first exception is Carrillo (2012), who calibrates a housing search model to infer some of its unobserved parameters. One can also cite recent work by Genesove and Han (2012a,b) or Merlo et al. (2013). A key reason behind the paucity of empirical work is that teasing out the empirical content of search models is notoriously difficult (Postel-Vinay and Robin, 2002). The last issue with this class of model is that, to our knowledge, no one has uncovered the implications of a housing search for land use.[28]

[27] The joint buyer–seller problem is studied theoretically and empirically by Anenberg and Bayer (2013), who suggest that it may amplify housing market fluctuations both in prices and in volumes.

[28] For instance, there is only a small fraction of properties available on the market at any point in time. This implies that residents may be able to get the most suitable property on the market at the time of their search but not the overall best property for them. Hence, residential land might be misallocated because of search and relocation frictions. A similar misallocation will also occur with commercial properties. This may then affect productivity. In turn, this may prevent house builders from providing the optimal amount of heterogeneity in housing. More specifically, they will refrain from supplying more "extreme houses" since there may be no buyer for them at the time they try to sell them.

Another class of search models has implications for land use: labor search models. The reason is that depending on where an unemployed resident lives, searching for a job may be more or less costly. In a simple monocentric setting, imagine, for instance, that searching for work involves going to the CBD. Then, the cost of searching for a job increases with the distance to the CBD. This has a direct effect on the search effort of unemployed residents. In turn, this will affect residential patterns in cities and land development. It is also perhaps reasonable to assume that the distance to the CBD affects the efficiency of the job search process and not only its costs.

The first model of labor search with an explicit modeling of land is due to Wasmer and Zenou (2002). Further developments can be found in Zenou (2009). In the simplest model proposed by Zenou (2009), the unemployed locate either close to or far from jobs depending on the fundamental parameters that govern the functioning of the labor market. More specifically, if these parameters imply a labor market that is sufficiently tight, unemployed individuals (or workers with a greater propensity of becoming unemployed) will prefer to live close to the CBD. This proximity to the CBD will lower their search costs, facilitating a more intense search, which will be rewarded with a job more often. Instead, if the labor market is not sufficiently tight and unemployment is high in equilibrium, the incentives to search intensively are muted, and unemployed individuals will prefer to live further from the CBD. In turn, this greater distance to jobs will weaken their incentives to search for jobs.[29]

To sum up, there are several take-away points that emerge form this discussion of heterogeneity in land use models. First, the exploration of several dimensions of heterogeneity is still in its infancy. This is true in particular for the heterogeneity of parcels and for the assignment of heterogeneous parcels to heterogeneous residents. Progress will be slow because this type of work is often technically challenging and relies on techniques that are new to the field. Next, a key lesson from existing work is that for many first-order questions such as the location choices of different socioeconomic groups, the results are sometimes sensitive to fine details about the assumptions or to some key parameter values. Allowing for heterogeneous residents often modifies the results or even, sometimes, changes their qualitative nature. Recall, for instance, that all the results for the proportionalities of the urban aggregates no longer hold with heterogeneous residents or that the land price gradients no longer depend on commuting costs in simple models of assignments with a continuous distribution of income. Finally, and most importantly, despite the aforementioned changes to secondary results, all the key insights in Section 8.2 still hold with heterogeneous residents. At the individual level, all the key trade-offs remain qualitatively the same, and an appropriately redefined Alonso–Muth condition still

[29] Although it may sound from this description that multiple equilibria are possible with a high unemployment-peripheral unemployed configuration versus a low unemployment-central unemployed configuration, this is not the case in the model of Wasmer and Zenou (2002), where the equilibrium is unique.

governs location choices and the gradients for land prices, housing prices, the intensity of development, population density, and parcel sizes.

8.3.3 Durable housing

When we do comparative statics in the monocentric model, we are implicitly letting the city be completely rebuilt from scratch to fit the new conditions. This is often seen as a reasonable simplification because the comparative statics fit well with comparisons of actual cities built under different conditions. However, some important details of the standard monocentric model conflict with reality. For instance, the model predicts that cities will be built contiguously and with building heights monotonically decreasing from the center. In practice, however, we see some centrally located parcels left vacant while others further out are developed, and building heights can both increase and decrease as we travel outward from the CBD. Replacing the assumption that housing is completely malleable with the more realistic assumption that housing is durable helps address these shortcomings. In addition, it provides useful additional insights into how cities react to changing conditions.[30]

Once we acknowledge the durability of housing, we must take into account that housing developers will recover their investment over an extended period of time. Then it becomes important to consider how they form their expectations about the future evolution of prices. A simple possibility is that developers have myopic expectations, as in Anas (1978), behaving as if current conditions will last forever. Then, a developer who owns a parcel of land at a distance x from the CBD will develop it at time T if and only if

$$\int_T^\infty R_T(x)e^{-rt}\mathrm{d}t = \int_T^\infty P_T(x)f(x)e^{-rt}\mathrm{d}t - \int_T^\infty r\,k_T(x)e^{-rt}\mathrm{d}t \geq \int_T^\infty \underline{R}_T e^{-rt}\mathrm{d}t\,. \quad (8.37)$$

The return, $R_T(x)$, that the developer expects to obtain from a parcel of land at a distance x from the CBD from time T onward is the difference between the expected present value of the rent from the $f(x)$ units of housing floor space developed on the parcel of land and the expected present value of the cost at a constant interest rate r of the capital $k_T(x)$ used in the development (note that, by Shephard's lemma, $k_T(x) = \frac{\partial c(R_T(x), r)}{\partial r}$, which varies with $R_T(x)$). For land to be developed, this return must be greater than the expected present value of the agricultural rent, $\underline{R}_T$. Owing to myopic expectations, all variables have time subindex T. Integrating Equation (8.37) and simplifying the result implies that, at time T, the edge of the city is still given by the same condition as in the standard static monocentric model:

[30] In this section, we review briefly extensions to the monocentric model featuring durable housing. For additional details, see the survey by Brueckner (2000).

$$R_T(\overline{x}) = P_T(x) - rk_T(x) = \underline{R}_T \ , \tag{8.38}$$

which is the same as Equation (8.20) of the standard model but with a time subindex T. Consequently, a city that grows over time has contiguous development, as in the static monocentric model. The key difference is that the capital intensity of development $k_T(x)$ at each point x reflects the conditions at the time T when the city edge was at this point instead of current conditions. Several situations may occur, with building heights and population density both decreasing (as in the static model), staying constant, or increasing with distance from the CBD. To understand under which conditions these outcomes occur, recall that the optimal intensity of development at the city edge is given by $k_T(\overline{x}) = \frac{\partial c(R_T(\overline{x}), r)}{\partial r} = \frac{\partial c(\underline{R}_T, r)}{\partial r}$ after using Equation (8.38). Hence, unless there is a change in the interest rate or in the agricultural land rent, new development at the urban edge driven by higher wages or lower commuting costs will occur at the same level of capital intensity. A combination of higher wages (for the city to expand) and a lower interest rate will lead to more capital-intensive development at edge of the city. A combination of higher wages (for the city to expand) and higher interest rates will, on the other hand, lead to less capital-intensive new development at the urban edge.

The assumption of myopic foresight, while convenient, is not very satisfactory. For instance, in a city that is growing gradually over time, one would expect developers to take this growth trajectory into account. Developers may be able to predict the future quite well. Taken to the extreme, this implies assuming that developers have perfect foresight. A tractable monocentric model with irreversible development and perfect foresight is that of Capozza and Helsley (1989), who assume for simplicity that dwelling size and floor space per unit of land are both equal to unity: $h_t(x) = f_t(x) = 1$. They also assume that conversion of one unit of land from nonurban use to urban use involves a fixed amount of capital k instead of an endogenously chosen amount $k_t(x)$. Denoting by T the endogenous date at which a parcel located at a distance x from the CBD will be converted to urban use, we can express the expected present value of returns for a developer who owns that parcel as

$$\int_0^T \underline{R}_t \mathrm{e}^{-rt} \mathrm{d}t + \int_T^\infty P_t(x) \mathrm{e}^{-rt} \mathrm{d}t - k\,\mathrm{e}^{-rT} \ . \tag{8.39}$$

The first term in Equation (8.39) is the land rent obtained up until time T while the parcel is still used for agriculture. The second term is the rent obtained from the development at time T onward for the unit of housing built on the land. The third term is the present value of the conversion cost. Note that, unlike in the case of myopic expectations, developers fully anticipate the evolution of rents of agricultural land and housing, $\underline{R}_t$ and $P_t(x)$. The first-order condition for the developer can be found by differentiating Equation (8.39) with respect to T and equating it to zero, implying

$$P_T(x) - rk = \underline{R}_T \ . \tag{8.40}$$

Note this condition is the same as Equation (8.38) for the case of myopic foresight, so land is still developed when the urban land rent minus the cost of conversion from agricultural to residential use equals the agricultural land rent. However, perfect foresight introduces an important difference: the price of land is no longer proportional to the rent. Instead, the price of land is equal to the maximized value of Equation (8.39). As a result, in a growing city, land beyond the city edge is priced above the expected present value of the agricultural rent, reflecting the anticipation of its conversion to urban use. This model also generates a price gradient with higher prices for undeveloped land as we approach the city edge from outside the developed area.

If we let the structural characteristics of housing change endogenously, the dynamic monocentric model with perfect foresight can generate leapfrog development, where parcels are not developed contiguously starting from the CBD (Fujita, 1982; Wheaton, 1982; Turnbull, 1988). Letting the structural characteristics of housing vary introduces a second choice variable in the expected present value of the developer's return relative to Equation (8.39). This leads to an additional first-order condition for optimal structural characteristics of development in addition to (8.40), which regards the optimal timing of development. Since for any given time T these two first-order conditions can intersect for more than one value of x, it follows that parcels of land separate from each other can be developed simultaneously, while some parcels in between may be left vacant.

In this section, we have so far assumed that developers either have no anticipation of the future and expect current conditions to last forever (the myopic foresight case) or can predict the future precisely (the perfect foresight case). A more realistic case sits in between these two, with developers anticipating the future but realizing that there is uncertainty about the evolution of rents. Capozza and Helsley (1990) explore such a monocentric model with irreversible development under uncertainty. As in Capozza and Helsley (1989), they fix the dwelling size and floor space per unit of land ($h_t(x) = f_t(x) = 1$) so that the only choice variable for the developer is when to convert land from agricultural to urban use. Conversion again involves a constant capital cost k. They focus on an open city within a large urban system where the ongoing level of utility is constant at $\underline{u}$. With every resident consuming one unit of housing built on one unit of land, this implies a constant level of consumption of the numéraire $\underline{z}$. Income w for every resident in the city is assumed to vary. More specifically, it follows a Brownian motion with drift $g > 0$ and variance σ^2. When income rises in the city, this makes the city more attractive relative to other cities, bringing in more residents, which raises land rents until utility is restored to level $\underline{u}$. When income falls, the city loses population and land rents fall. From Equations (8.16) and (8.18) with $h_t(x) = 1$ and $\frac{\partial c(R_t(x))}{\partial R_t(x)} = 1$, the bid-rent curve is linear: $\frac{\mathrm{d}P_t(x)}{\mathrm{d}x} = \frac{\mathrm{d}R_t(x)}{\mathrm{d}x} = -\tau$. Hence,

$$P_t(x) = w - \underline{z} - \tau x \,. \tag{8.41}$$

Capozza and Helsley (1990) show that the optimal conversion date T for the developer owning a parcel at a distance x from the CBD satisfies

$$P_T(x) - rk = \underline{R}_T + \frac{r - \psi g}{\psi r}, \tag{8.42}$$

where

$$\psi \equiv \frac{\sqrt{g^2 + 2\sigma^2 r} - g}{\sigma^2} \leq \frac{r}{g}. \tag{8.43}$$

Comparison of Equation (8.42), for irreversible development under uncertainty, and Equation (8.40), for irreversible development under perfect foresight, shows that developers require higher urban rents to develop land under uncertainty. This implies that, in a growing city, uncertainty delays urban development. The reason is that a developer who converts land to urban use may be surprised by lower rents than expected and come to regret the conversion. As in the case of perfect foresight, land beyond the city edge is priced above the expected present value of the agricultural rent. This now reflects not just the anticipation of its conversion to urban use as a result of expected urban growth (as was the case with perfect foresight) but also an additional option value of agricultural land arising from the uncertainty about future urban rents. Holding a parcel of agricultural land implies holding an option to convert it to urban use. Urban development is equivalent to exercising that option.

A key feature of land development is the existence of significant lags between the time when a development project is decided on and the time when new floor space can be put on the market. These lags are caused by construction and the time it takes to obtain a building permit. They imply that a developer will face a potentially very different rent relative to the one that prevailed when the project was decided on. In addition, developers also have the option to cut their losses and stop a project should the circumstances become too unfavorable. Bar-Ilan and Strange (1996) extend the model of Capozza and Helsley (1990) to account for these two features. The main result is that development lags reduce the amount of delay. As in Capozza and Helsley (1990), developers still want to delay their investment for fear of lower rents in the future. However, development lags make the opportunity cost of a delay more expensive when rents are high. The option of stopping a project also puts a floor on the returns from a new development. In turn, this makes returns from development convex in rents and leads the value of land development to increase with uncertainty. Since rents are lower further away from the CBD, the same income uncertainty results in greater rent uncertainty in less central locations. In the presence of long development lags, developers may prefer to convert to urban use land that is further away, leaving more central locations undeveloped. This provides a motivation for patterns of leapfrog development.

The assumption that development is irreversible is justified because buildings are long-lived, and once it has been converted to urban use, land very rarely reverts to nonurban use (Burchfield et al., 2006). Nevertheless, buildings deteriorate over time and require periodic maintenance, and while no development is rare, redevelopment is very common. It is therefore important to consider not just developers' decisions about the initial development but also their decisions about redevelopment. This can again be done under different scenarios regarding developers' expectations. Brueckner (1980) studies the case of myopic foresight. The inconsistency of developers' behavior is enhanced by redevelopment because now myopic developers do not just assume that current rent levels will persist forever, they also decide whether to redevelop now while ignoring that they will again be deciding whether to redevelop in the future. Nevertheless, studying redevelopment with myopic developers is a useful exercise because some key conclusions carry over to more sophisticated treatments of developers' expectations. In a growing city, land is still initially developed as in the case of the myopic irreversible development of Anas (1978) when the value of land in urban use minus the conversion costs equals the value of land in nonurban use. The main difference is that there will be periodic redevelopment. Consider, for instance, a city where the agricultural land rent is constant at $\underline{R}$ but wages keep rising over time. Wage growth causes the city to expand outward. However, since developers are myopic, they expect the current state of the city to be the permanent state, so they keep building at the edge of the city with the same constant capital intensity $k_t(x) = \frac{\partial c(R_t(\overline{x}), r)}{\partial r} = \frac{\partial c(\underline{R}, r)}{\partial r}$. At the same time, they may find it optimal to compensate for their past myopic behavior by redeveloping land closer to the CBD at a higher intensity. Thus, close to the city center there will be some tall new buildings standing next to shorter older buildings. If we isolate structures built around the same time, more central buildings are taller than those in the periphery, as in the standard static monocentric model. However, since many generations of building coexist, the overall pattern is one of sawtooth building heights.

Combining perfect foresight and redevelopment greatly complicates the developer's problem. A possible solution is to focus on a city in a stationary state, as in Brueckner (1981) and Arnott et al. (1983). Alternatively, Braid (2001) studies a city in a nonstationary state by using specific functional forms that make the developer's problem at any distance from the CBD a simple transformation of the problem at unit distance.

Housing redevelopment has two important components: the deterioration of existing structures over time and their eventual replacement with new structures. Even in the absence of deterioration, changing conditions could be sufficient to generate redevelopment. Deterioration strengthens the incentives for redevelopment and also raises additional issues—for instance, enriching the analysis of income sorting we considered in Section 8.3.2. A developer may initially build housing at a high quality level to target high-income residents. However, as the passing of time deteriorates the building and

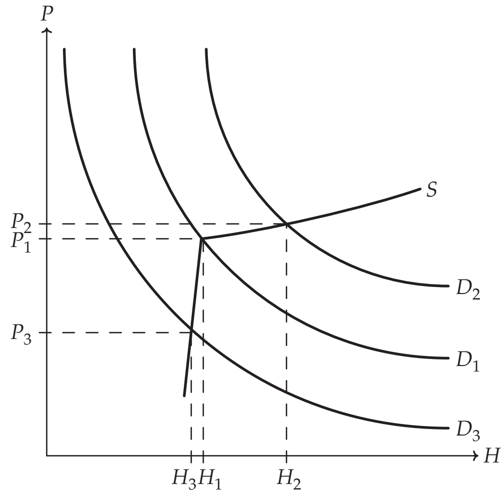

Figure 8.3 The asymmetry between growing and declining cities.

some of its features become obsolete, the effective level of quality it provides falls. This may lead its residents to move to a higher-quality dwelling, while lower-income residents move in to replace them. This filtering process is studied formally by Sweeney (1974a,b) and is thought to be crucial for the provision of housing to low-income groups. Rosenthal (2014) examines the filtering process empirically and shows that it is amplified by dwellings shifting over time from being owner-occupied to being rented, with the filtering process accelerating once units are rented. Combined with the amplifying effect on filtering from a low income elasticity of housing, Rosenthal (2014) concludes that filtering can be quite effective in providing suitable low-income housing. Brueckner and Rosenthal (2009) develop a model where filtering leads to cyclical changes in the location of higher-income households. In their model, dwellings have a fixed time span, during which fixed-size dwellings provide housing services that decline with their age. Thus, residents who wish to consume more housing must do so by residing in newer dwellings. In a growing city, locations close to the CBD are developed first. As they age, high-income households relocate toward the city edge to consume more housing. Eventually, central dwellings reach the end of their life and are replaced by new structures, prompting high-income households to relocate toward the city center away from now ageing suburban dwellings.[31]

One of the main implications of the durability of housing is that there are important asymmetries between growing cities and declining cities. This point, on which Figure 8.3 is based, is made by Glaeser and Gyourko (2005). The figure represents housing supply

[31] See Chapter 16 in this handbook for more on filtering in the housing market.

and demand in a city, with the price of housing P measured on the vertical axis and the quantity of housing H measured on the horizontal axis. Suppose that initially the city has a housing stock H_1 priced at P_1 per unit of floor space. If the city experiences a positive housing demand shock that moves the demand curve upward from D_1 to D_2, then, provided that suitable land is available and new construction is not significantly constrained by regulation and zoning, additional housing will be built, increasing the stock from H_1 to H_2 possibly with a small price increase from P_1 to P_2.

Consider now that the city experiences a negative housing demand shock. This shock moves the demand curve downward from D_1 to D_3. Then, because housing is durable, existing housing will remain in place. At most, housing may deteriorate owing to lack of maintenance, but this will reduce the housing stock only very slowly. Thus, a negative demand shock will be almost completely reflected in a sharp drop in prices from P_1 to P_3, with almost no change in quantities other than depreciation.

Putting everything together, we find the key is that housing supply will be kinked at the level of the current housing stock, being relatively elastic above this current level and very inelastic below it. This has several implications. First, as shown in Figure 8.3, positive shocks increase the population more than they increase housing prices, while negative shocks decrease housing prices more than they decrease the population. Second, cities grow more quickly than they decline because the durability of existing housing slows the fall in the population. Third, the abundance of cheap housing priced well below construction costs is a sign of large negative demand shocks that, since the population in declining cities falls slowly, can be taken as a sign of future decline. Glaeser and Gyourko (2005) show that these features hold empirically in the United States.

8.4. AGGLOMERATION AND COMMERCIAL LAND USE: MODELING POLYCENTRIC CITIES

The monocentric model is a remarkable achievement and a very useful stylized representation of cities. However, two aspects of the model are in particular tension with important empirical facts about cities. First, the monocentric city model explains patterns of residential land use and commuting within a city. However, it does not explain why individuals wish to be in a city to start with. If we treat the wage w as a parameter independent of a city's population in Equation (8.24), then $\frac{\mathrm{d}v}{\mathrm{d}N} < 0$. This implies that any individual prefers to live alone than to live in a city of any size. To explain why cities exist at all, we must introduce agglomeration economies. This is particularly important given the growing evidence about the importance of such agglomeration economies.[32] A simple way to incorporate agglomeration economies into the monocentric model is to have the wage depend positively on the city's population: $w = w(N)$, with $\frac{\mathrm{d}w}{\mathrm{d}N} > 0$. This leads to a trade-off in Equation (8.24) between urban costs ($\frac{\mathrm{d}v}{\mathrm{d}P(0)}\frac{\mathrm{d}P(0)}{\mathrm{d}N} < 0$) and stronger agglomeration

[32] They are reviewed in Chapter 5 in this volume.

economies ($\frac{\mathrm{d}v}{\mathrm{d}w}\frac{\mathrm{d}w}{\mathrm{d}N} > 0$). For a more detailed exposition of the implications of introducing agglomeration economies in the monocentric city model, including the modeling of systems of cities and the analysis of systematic and stochastic determinants of city growth, we refer the reader to Duranton and Puga (2014) and Chapter 4 in this volume.

A second aspect of the monocentric city model that appears at odds with modern cities is precisely its monocentric structure: in the monocentric model described in Section 8.2, firms do not use any land and locate, by assumption, at a single central point. In reality, firms use land as an input, and the division of land between residential and commercial uses within a city follows complex patterns. In particular, as we discuss in Section 8.9, land use patterns in actual cities are far less extreme than are assumed by the monocentric city model. In 1996, only about 25% of employees in US metropolitan areas worked within 5 km of their CBD (Glaeser and Kahn, 2001). Increasingly, secondary employment centers have emerged in metropolitan areas, and they have absorbed a growing number of jobs (Anas et al., 1998).

Extending the monocentric model so that land is used in production is straightforward and simply involves modeling the CBD as a segment or a disk instead of as a point. Having more than one employment center or area also does not change the basic mechanics of the model either, provided the location of secondary centers is exogenously given (see White, 1976, for an early example). A simple way to endogenize the location of secondary centers is to take a central city and its capacity as given and examine the problem of a developer who sets up an edge city, choosing its distance with respect to the central city and its capacity. Henderson and Mitra (1996) examine this problem and highlight the trade-off faced by the edge-city developer: locating the secondary center further away from the central city alleviates competition for land and lowers costs, but it also weakens productivity spillovers between the central city and the edge city. The weakening of spillovers reduces the productivity of the edge city but, by also reducing the productivity of the central city, strengthens the developer's monopsony power.

A more difficult problem is to endogenize the location of both firms and workers throughout the city, with areas emerging endogenously with commercial, residential, or mixed land use as a result of the interactions of the location decisions of all agents directly with each other and through land markets. This problem was first tackled independently by Ogawa and Fujita (1980) and Imai (1982) in a framework where firms benefit from proximity to each other due to communication externalities that decay linearly with distance. Fujita and Ogawa (1982) studied a case with exponential decay instead of linear decay in externalities. This was revived and generalized by Lucas and Rossi-Hansberg (2002). We now describe a simple version of the Ogawa and Fujita (1980) and Imai (1982) framework. For a neat exposition of the Fujita and Ogawa (1982) framework, see Fujita and Thisse (2013).

The city occupies a segment of endogenous length on the real line with one unit of land available at each location x. Denote by $m(x)$ the endogenous density of firms and by $n(x)$ the endogenous density of residents at location x. In equilibrium, there can be areas with mixed land use ($m(x) > 0$ and $n(x) > 0$), areas with only commercial development ($m(x) > 0$ and $n(x) = 0$), and areas with only residential development ($m(x) = 0$ and $n(x) > 0$).

Agglomeration economies arise owing to spillovers that raise a firm's productivity when its workers are able to interact more closely with other workers in the city. In particular, suppose that, using one unit of labor and λ units of land, each firm produces one unit of output for every unit of communication spillovers involving its workers. Hence its cost function is $(w(x) + \lambda\, P(x))/A(x)$ where $A(x)$ denotes communication spillovers. In turn, communication spillovers between workers depends on how far their jobs are located, starting at β units of communication spillovers for workers employed at the same location and decreasing at a rate γ per unit of distance between their job locations. The output of a firm choosing to locate at x then depends on the location of all other firms as follows:[33]

$$A(x) = \int_{-\infty}^{\infty} (\beta - \gamma|x - y|)m(y)\mathrm{d}y\ . \tag{8.44}$$

Differentiation of Equation (8.44) yields $\frac{\mathrm{d}A(x)}{\mathrm{d}x} = -\gamma\left(\int_{-\infty}^{x} m(y)\mathrm{d}y - \int_{x}^{\infty} m(y)\mathrm{d}y\right)$ and $\frac{\mathrm{d}^2A(x)}{\mathrm{d}x^2} = -2\gamma\, m(x)$. This implies that $A(x)$ reaches a global maximum at the point in the city where half the firms are located to its left and half the firms are located to its right. Without loss of generality, assign coordinate $x = 0$ to this point, so that $A(x)$ increases with x for $x < 0$ and decreases with x for $x > 0$. Furthermore, $A(x)$ is a concave function of x in areas wherever there is commercial development ($m(x) > 0$) and a linear function of x wherever there is no commercial development.

Free entry of firms exhausts their profits. The bid-rent function for commercial land $\Phi(x)$ is the maximum price a firm can pay for land at each location x while making zero profit:

$$\Phi(x) = \frac{1}{\lambda}[A(x) - w(x)]\ . \tag{8.45}$$

In the standard monocentric model each worker commutes from his or her location x to the exogenous CBD located at 0. Now, instead, a worker residing at x chooses the work location that best suits him or her. We maintain the assumption of commuting costs increasing linearly with distance at a rate τ. Let $T(x)$ denote the utility-maximizing job location of a worker as a function of his or her residential location x:

$$T(x) \equiv \arg\max_{y}\ \{w(y) - \tau|x - y|\}\ . \tag{8.46}$$

[33] β is assumed to be large enough that $A(x)$ does not end up being negative.

Thus, for a given residential location, workers choose their job location by trading off wages against commuting costs.

Let us simplify the residential location problem by assuming that all residences have the same unit size. Then, maximizing utility $u(1, z)$ subject to the budget constraint $w(x) - \tau|x - T(x)| = P(x) \times 1 + z(x)$ is equivalent to maximizing consumption of the numéraire $z(x) = w(x) - \tau|x - T(x)| - P(x)$. Suppose all houses are built with one unit of land and a fixed amount of capital. To avoid carrying around additional constants, we set both the constant cost of capital throughout the city and the price of land in agriculture to zero, so that the price of housing and the price of land coincide instead of differing by a constant and so that the price of land at the city edge is zero: $P(x) = R(x)$ and $R(\bar{x}) = 0$. The bid-rent function for housing and for residential land $\Psi(x, \underline{u})$ is the maximum price a resident can pay for housing at each location x while consuming the amount of numéraire $z(\underline{u})$ that allows him or her to enjoy utility $\underline{u}$ and while also satisfying the budget constraint:

$$\Psi(x, \underline{u}) = w(T(x)) - \tau|x - T(x)| - z(\underline{u}) . \tag{8.47}$$

Land will be allocated to the highest bidder. This implies that the rental price of land is given by

$$R(x) = \max(\Phi(x), \Psi(x, \underline{u})) , \tag{8.48}$$

$$R(x) = \Phi(x) \text{ if } m(x) > 0 , \tag{8.49}$$

$$R(x) = \Psi(x, \underline{u}) \text{ if } n(x) > 0 . \tag{8.50}$$

In turn, land use is described by

$$\lambda m(x) + n(x) = 1 \quad \text{if} \quad R(x) \geq 0 , \tag{8.51}$$

$$m(x) = n(x) = 0 \quad \text{if} \quad R(x) < 0 . \tag{8.52}$$

Labor market clearing implies

$$\int_X n(x)\mathrm{d}x = \int_{T(X)} m(x)\mathrm{d}x , \tag{8.53}$$

for every interval X. Finally, we must consider the aggregate population constraint,

$$\int_{-\infty}^{\infty} n(x)\mathrm{d}x = N, \tag{8.54}$$

and the aggregate firm constraint,

$$\int_{-\infty}^{\infty} m(x)\mathrm{d}x = N. \tag{8.55}$$

Equations (8.44)–(8.55) are the equilibrium conditions of this framework.

The general form of the equilibrium is the following. Patterns of land use are symmetric around the point $x=0$, which splits the distribution of firms into halves. There is a central area of mixed land use centered at $x=0$ and extending from $-x_0$ until x_0, where firms and residences coexist continuously. Beyond this central mixed-use area, and at both sides of it, there are areas of pure commercial land use, extending from $-x_1$ to $-x_0$ and from x_0 to x_1. Finally, beyond the commercial areas there are areas of pure residential land use extending from $-\overline{x}$ to $-x_1$ and from x_1 to $\overline{x}$. Workers living in the mixed-use area work where they live, whereas workers living in the purely residential area commute to work in the purely commercial area. These patterns of land use can be expressed more formally as

$$m(x)=\begin{cases}\dfrac{1}{1+\lambda} & x\in[-x_0,x_0]\ ,\\ \dfrac{1}{\lambda} & x\in[-x_1,-x_0]\cup[x_0,x_1]\ ,\\ 0 & x\in[-\overline{x},-x_1]\cup[x_1,\overline{x}]\ ,\end{cases} \tag{8.56}$$

$$n(x)=\begin{cases}\dfrac{1}{1+\lambda} & x\in[-x_0,x_0]\ ,\\ 0 & x\in[-x_1,-x_0]\cup[x_0,x_1]\ ,\\ 1 & x\in[-\overline{x},-x_1]\cup[x_1,\overline{x}]\ ,\end{cases} \tag{8.57}$$

where the densities in each interval follow from the above description and Equations (8.51) and (8.53). Given that each firm uses λ units of land and each resident uses one unit, the shares of commercial and residential land are $\lambda m(x)$ and $n(x)$, respectively. Panel (a) in Figure 8.4 plots the share of commercial land in the equilibrium we have just described. We now show that such an equilibrium exists and derive the equilibrium values of the communication spillovers $A(x)$, wages $w(x)$, and land prices $R(x)$, as well as the values of x_0, x_1, and $\overline{x}$.

Substituting Equation (8.56) into (8.44) yields the equilibrium value of the communication spillovers at each point in the city:

$$A(x)=\begin{cases}\beta N-\gamma\left(\dfrac{1}{\lambda}x_1^2-\dfrac{1}{\lambda(1+\lambda)}x_0^2+\dfrac{1}{1+\lambda}x^2\right), & x\in[-x_0,x_0]\ ,\\ \beta N-\gamma\left(\dfrac{1}{\lambda}x_1^2-\dfrac{2}{\lambda(1+\lambda)}x_0|x|+\dfrac{1}{\lambda}x^2\right), & x\in[-x_1,-x_0]\cup[x_0,x_1]\ ,\\ \beta N-\gamma\left(\dfrac{2}{\lambda}x_1-\dfrac{2}{\lambda(1+\lambda)}x_0\right)|x|, & x\in[-\overline{x},-x_1]\cup[x_1,\overline{x}]\ .\end{cases} \tag{8.58}$$

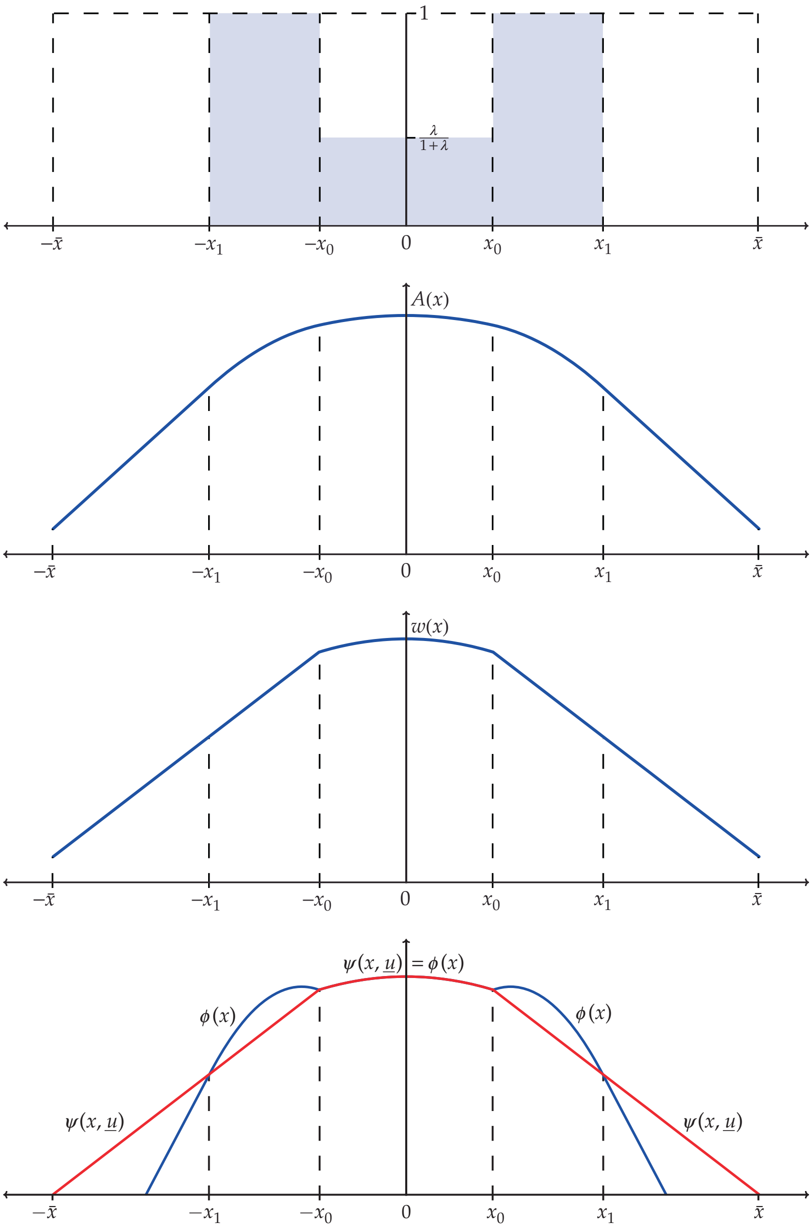

Figure 8.4 Equilibrium land use patterns and gradients in Ogawa and Fujita (1980). Panel (a) shows the share of land in commercial use. Panel (b) shows spillovers. Panel (c) shows wages. Panel (d) shows bid-rent gradients.

Panel (b) in Figure 8.4 plots the equilibrium spillovers $A(x)$. As shown above, $A(x)$ is concave over the interval where firms are located, $x \in [-x_1, -x_1]$, with a maximum at the (endogenous) center of the city and it linearly decreases with the distance to the center over the interval where all land is in residential use, $x \in [-\overline{x}, -x_1] \cup [x_1, \overline{x}]$. It also follows immediately from Equation (8.58) that $A(x)$ is continuous in x and so is its first derivative.

Turning to wages, consider two points x and y with firms to which workers are commuting. Then, by the definition of the commuting cost function in Equation (8.46), $w(x) - \tau|T^{-1}(x) - x| \geq w(y) - \tau|T^{-1}(x) - y|$ and $w(y) - \tau|T^{-1}(y) - y| \geq w(x) - \tau|T^{-1}(y) - x|$. These two inequalities together imply $w(x) - w(y) = -\tau(x - y)$. Equivalently, $\frac{\mathrm{d}w(x)}{x} = -\tau$ for any $x > 0$ such that $m(x) > 0$ and $T^{-1}(x) \neq x$ and $\frac{\mathrm{d}w(x)}{x} = \tau$ for any $x < 0$ such that $m(x) > 0$ and $T^{-1}(x) \neq x$. Thus, in any area with firms to which workers are commuting, the wage is a linear function of x decreasing as one moves away from $x = 0$ with a slope equal in absolute value to the commuting costs parameter τ. Consider now firms whose workers instead do not need to commute because their workers reside at the same location—that is, where $T(x) = x$. By Equations (8.48)–(8.50), (8.56), and (8.57), $\Psi(x) = \Phi(x)$ for $x \in [-x_0, x_0]$. Substituting Equations (8.45) and (8.47) and $T(x) = x$ into this equality turns it into $w(x) - z(\underline{u}) = \frac{1}{\lambda}[A(x) - w(x)]$, which we can solve for $w(x)$. Thus, the wage gradient in the city is given by

$$w(x) = \begin{cases} \dfrac{1}{1+\lambda} A(x) + \dfrac{\lambda}{1+\lambda} z(\underline{u}), & x \in [-x_0, x_0], \\ w(x_0) - \tau(|x| - x_0), & x \in [-\overline{x}, -x_0] \cup [x_0, \overline{x}]. \end{cases} \tag{8.59}$$

Panel (c) in Figure 8.4 plots the equilibrium wage $w(x)$. Note from Equation (8.59) that $w(x)$ continuous in x. Also, given that $A(x)$ is a concave function of x for $x \in [-x_1, x_1]$ with a maximum at $x = 0$, it follows from Equation (8.59) that the wage $w(x)$ is a concave function of x for $x \in [-x_0, x_0]$ with a maximum at $x = 0$. Outside this central area of mixed land use, the wage decreases linearly with x as one moves away from the center with a slope equal in absolute value to the commuting costs parameter τ.

Consider next the bid-rent functions for land. On the residential side, substituting Equation (8.59) into (8.47) and using $T(x) = x$ for $x \in [-x_0, x_0]$ yields

$$\Psi(x, \underline{u}) = \begin{cases} \dfrac{1}{1+\lambda} \left[A(x) - z(\underline{u})\right], & x \in [-x_0, x_0], \\ \Psi(x_0, \underline{u}) - \tau(|x| - x_0), & x \in [-\overline{x}, -x_0] \cup [x_0, \overline{x}]. \end{cases} \tag{8.60}$$

Note that $\Psi(x, \underline{u})$ is a continuous function of x. Also, it follows from Equation (8.60) and the concavity of $A(x)$ for $x \in [-x_1, x_1]$ that, like the wage, the bid-rent function for land in residential use $\Psi(x, \underline{u})$ is also a concave function of x for $x \in [-x_0, x_0]$ with a maximum at $x = 0$. Within the mixed-use area, workers living and working closer to the center obtain a higher wage, but this is exactly offset by a higher price for their residence. Outside this central area of mixed land use, $\Psi(x, \underline{u})$ decreases linearly with x as one moves away from the center with a slope equal in absolute value to the commuting costs parameter τ. Within the purely residential area, workers living closer to the center are able to obtain a higher wage at any given commuting distance from their home, but this higher wage is exactly offset by a higher price for their residence.

Regarding the bid-rent function for land in commercial use, substituting Equation (8.59) into (8.45) results in

$$\Phi(x) = \begin{cases} \frac{1}{1+\lambda}\left[A(x) - z(\underline{u})\right], & x \in [-x_0, x_0] , \\ \Phi(x_0) + \frac{1}{\lambda}[A(x) - A(x_0)] + \frac{\tau}{\lambda}(|x| - x_0), & x \in [-\overline{x}, -x_0] \cup [x_0, \overline{x}] . \end{cases} \tag{8.61}$$

Note that $\Phi(x)$ is also a continuous function of x. From Equations (8.58) and (8.61), $\Phi(x)$ is a concave function of x for $x \in [-x_0, x_0]$ and also for $x \in [-x_1, -x_0] \cup [x_0, x_1]$, although its slope changes discretely at $-x_0$ and x_0. Beyond x_1 and $-x_1$, $\Phi(x)$ becomes a linear function of x, maintaining the same slope it has at x_1 and $-x_1$, respectively.

Panel (d) in Figure 8.4 plots the bid-rent functions for land in residential use and in commercial use, $\Psi(x, \underline{u})$ and $\Phi(x)$. For $x \in [-x_0, x_0]$, they both coincide and land is in mixed use. For $x \in [-x_1, -x_0] \cup [x_0, x_1]$, firms bid for land more than residents and land is in purely commercial use. Finally, for $x \in [-\overline{x}, -x_1] \cup [x_1, \overline{x}]$, residents bid for land more than firms and land is in purely residential use.

The edge of the city can be obtained by integrating both sides of (8.51) between $-\overline{x}$ and $\overline{x}$ and using the aggregate constraints (8.54) and (8.55):

$$\overline{x} = \frac{1+\lambda}{2} N . \tag{8.62}$$

For the land use patterns of Equations (8.56) and (8.57) to be consistent with the equilibrium conditions (8.48)–(8.51), we must have

$$R(x) = \Phi(x) = \Psi(x, \underline{u}), \qquad x \in [-x_0, x_0] , \tag{8.63}$$

$$R(x) = \Phi(x) \geq \Psi(x, \underline{u}), \qquad x \in [-x_1, -x_0] \cup [x_0, x_1] , \tag{8.64}$$

$$R(x) = \Psi(x, \underline{u}) \geq \Phi(x), \qquad x \in [-\overline{x}, -x_1] \cup [x_1, \overline{x}] , \tag{8.65}$$

$$R(\overline{x}) = \Psi(\overline{x}, \underline{u}) = 0 . \tag{8.66}$$

Equations (8.64) and (8.65) and the continuity of $\Psi(x, \underline{u})$ and $\Phi(x)$ imply $\Phi(x_1) = \Psi(x_1, \underline{u}))$. Substituting Equations (8.60) and (8.61) into this equality yields $A(x_0) - A(x_1) = (1 + \lambda)\tau(x_1 - x_0)$. Substituting (8.58) into this gives a first equation linking x_0 and x_1:

$$\frac{\gamma}{\lambda}\left(x_1^2 - x_0^2 - \frac{2}{1+\lambda}x_0(x_1 - x_0)\right) = (1+\lambda)\tau(x_1 - x_0)\ . \tag{8.67}$$

A second equation linking x_0 and x_1 is obtained by substituting Equations (8.56) and (8.57) into the aggregate firm constraint (8.55):

$$2\left(\frac{1}{1+\lambda}x_0 + \frac{1}{\lambda}(x_1 - x_0)\right) = N\ . \tag{8.68}$$

Equations (8.67) and (8.68) have two solutions. The first solution has the entire city under mixed land use: $x_0 = x_1 = \frac{1+\lambda}{2}N = \overline{x}$. For this solution to be an equilibrium, we must make sure that starting from such a configuration, a firm at the edge of the city is not willing to outbid workers for land in order to concentrate its production, thus pushing workers out into a purely residential area. That is, the bid-rent gradient for firms cannot be steeper (have a more negative slope) at $x = x_1$ when $x_0 = x_1 = \overline{x}$. Differentiating Equations (8.60) and (8.61), we can express the condition $\frac{\mathrm{d}\Phi(x)}{\mathrm{d}x}|_{x=x_1} \geq \frac{\mathrm{d}\Psi(x,\underline{u})}{\mathrm{d}x}|_{x=x_1}$ as $\frac{2\gamma}{\lambda}\left(x_1 - \frac{1}{1+\lambda}x_0\right) \leq \tau(1+\lambda)$. Valuing this at $x_0 = x_1 = \overline{x} = \frac{1+\lambda}{2}$, we find the condition for a fully integrated equilibrium becomes $N \leq \frac{\tau(1+\lambda)}{\gamma}$.

The second solution to Equations (8.67) and (8.68) takes the values $x_0 = \frac{\tau}{\gamma}(1+\lambda)^2 - \frac{1+\lambda}{2}N$ and $x_1 = \frac{\tau}{\gamma}(1+\lambda) - \frac{1-\lambda}{2}N$. We must check that this second solution involves values in the admissible range—that is, such that $0 \leq x_0 \leq x_1 \leq \overline{x}$. From the solution itself, we see that $0 \leq x_0$ is equivalent to $\frac{N}{2(1+\lambda)} \leq \frac{\tau}{\gamma}$, and $x_0 \leq x_1$ is equivalent to $\frac{\tau}{\gamma} \leq \frac{N}{1+\lambda}$. Using Equation (8.62), we find $x_1 \leq \overline{x}$ is also equivalent to $\frac{\tau}{\gamma} \leq \frac{N}{1+\lambda}$. In addition, we must again check that land is allocated to the highest bidder, as reflected in conditions (8.63)–(8.66). By inspection of Equations (8.60) and (8.61), we can see that condition (8.63) is satisfied. Since $\Phi(x_0) = \Psi(x_0, \underline{u})$) and $\Phi(x_1) = \Psi(x_1, \underline{u})$), conditions (8.64) and (8.65) are equivalent to $\frac{\mathrm{d}\Phi(x)}{\mathrm{d}x}|_{x=x_0^+} \geq \frac{\mathrm{d}\Psi(x,\underline{u})}{\mathrm{d}x}|_{x=x_0^+}$ and $\frac{\mathrm{d}\Phi(x)}{\mathrm{d}x}|_{x=x_1} \leq \frac{\mathrm{d}\Psi(x,\underline{u})}{\mathrm{d}x}|_{x=x_1}$, respectively. Using Equations (8.58), (8.60), (8.61), we can see that they are both satisfied provided that $\frac{\tau(1+\lambda)}{\gamma} \leq N$. And population has been determined by ensuring that Equation (8.66) is satisfied.

Pulling all of the above together, we can summarize it as follows:

$$x_0 = \begin{cases} 0 & \text{if } \dfrac{\tau(1+\lambda)}{\gamma} \leq \dfrac{N}{2}\ , \\ \dfrac{\tau}{\gamma}(1+\lambda)^2 - \dfrac{1+\lambda}{2}N & \text{if } \dfrac{N}{2} < \dfrac{\tau(1+\lambda)}{\gamma} < N\ , \\ \dfrac{1+\lambda}{2}N = \overline{x} & \text{if } N \leq \dfrac{\tau(1+\lambda)}{\gamma}\ . \end{cases} \tag{8.69}$$

$$x_1 = \begin{cases} \dfrac{\lambda}{2}N & \text{if } \dfrac{\tau(1+\lambda)}{\gamma} \leq \dfrac{N}{2}, \\ \dfrac{\tau}{\gamma}(1+\lambda) - \dfrac{1-\lambda}{2}N & \text{if } \dfrac{N}{2} < \dfrac{\tau(1+\lambda)}{\gamma} < N, \\ \dfrac{1+\lambda}{2}N = \overline{x} & \text{if } N \leq \dfrac{\tau(1+\lambda)}{\gamma}. \end{cases} \tag{8.70}$$

The thresholds $\overline{x}$, x_0, and x_1, in Equations (8.62), (8.69), and (8.70), respectively, depend on the population. By substituting these three equations, as well as Equations (8.58) and (8.60) into (8.66), we can obtain the equilibrium population N in the city as a function of the level of utility $\underline{u}$ that can be achieved elsewhere.

Before we discuss these equilibria, a final check remains: commuting patterns must be utility maximizing, as defined by Equation (8.46). First, note that $T(x) = x$ for $x \in [-x_0, x_0]$. Workers within the mixed-use area cannot do better by commuting outward since this both lowers the wage and makes them incur commuting costs. They cannot do better by commuting to a job closer to the center if the increase in the wage is not greater than the commuting costs incurred: $-\frac{\mathrm{d}w(x)}{\mathrm{d}x} \leq \tau$. Using Equations (8.59) and (8.69), we see that this condition is satisfied if $0 \leq x_0$, which must hold. Regarding workers living in the purely residential area $[-\overline{x}, -x_1] \cup [x_1, \overline{x}]$, they will work in the closest purely commercial area, and if they change jobs within this area, by Equation (8.59), the change in the wage they obtain is exactly offset by the change in the commuting costs they incur. For concreteness, the following commuting function is utility maximizing and supports the equilibrium:

$$T(x) = \begin{cases} \dfrac{x_1(x+x_1) - x_0(\overline{x}+x)}{\overline{x}-x_1}, & x \in [-\overline{x}, -x_1], \\ x, & x \in [-x_0, x_0], \\ \dfrac{x_1(x-x_1) + x_0(\overline{x}-x)}{\overline{x}-x_1}, & x \in [x_1, \overline{x}]. \end{cases} \tag{8.71}$$

This expression implies that the worker living furthest away in the purely residential area at $\overline{x}$ commutes to the job in the purely commercial area closest to home at x_1, with workers living more centrally commuting to more centrally located jobs up until the worker living at x_1 who commutes to x_0. Thus, in Equation (8.71) there is no cross-commuting.

The equilibrium depends on the value of $\frac{\tau(1+\lambda)}{\gamma}$, which increases with the rate τ at which commuting costs increase with distance and with the land requirement of firms λ, and decreases with the rate γ at which productivity spillovers decay. If this combination of parameters is within some intermediate range ($\frac{N}{2} < \frac{\tau(1+\lambda)}{\gamma} < N$), we have the richest equilibrium configuration: there is an area close to the (endogenous) city center where there is mixed land use and where workers reside locally; on each side of this mixed land use area, there is an area in pure commercial use; and beyond each of these two commercial areas there is an area of pure residential use, from where workers commute to the nearest commercial area.

For firms, being clustered together in a purely commercial area increases their productivity through spillovers (with the gain from proximity being greater the higher the spillover decay γ), but forces them to compensate their workers for commuting costs (which are greater the higher the cost per unit of distance, τ, and the more space that each firm takes up, λ). If $\frac{\tau}{\gamma} < \frac{N}{2(1+\lambda)}$, productivity spillovers dominate commuting costs and the mixed-use area does not exist. In this case, there is a central commercial area surrounded by two residential areas. Figure 8.5 plots the bid-rent functions for land in residential use and in commercial use, $\Psi(x, \underline{u})$ and $\Phi(x)$, in this equilibrium configuration where commercial and residential uses are fully separated. This is like a monocentric model, except that now the configuration of a central area where all firms are located surrounded by land in residential use is an equilibrium outcome instead of an initial assumption. If we let firms produce without using land ($\lambda = 0$), then from Equations (8.69) and (8.70), $x_0 = x_1 = 0$ and we get exactly the monocentric city outcome with all firms clustered at one central point.

Finally, if $N \leq \frac{\tau(1+\lambda)}{\gamma}$, we are at the opposite extreme, and commuting costs dominate productivity spillovers and the pure commercial and residential areas do not exist. In this case, the entire city is under mixed use and every worker lives where they work.

It is worth noting that, unlike in the standard monocentric model presented in Section 8.2, when firms generate spillovers for each other the equilibrium may not be efficient. This is because, in choosing its location within the city, each firm takes into account the spillovers it receives from other firms but not the spillovers it creates for other firms. Thus, to calculate optimum land use patterns, we must replace γ by 2γ in the calculations above. Replacing γ (for the market equilibrium) by 2γ (for the optimum) in Equations (8.69) and (8.70) immediately implies the following. If $\frac{N}{2} < \frac{\tau(1+\lambda)}{\gamma} < N$, the market equilibrium involves a configuration like the one shown in Figure 8.4, but the

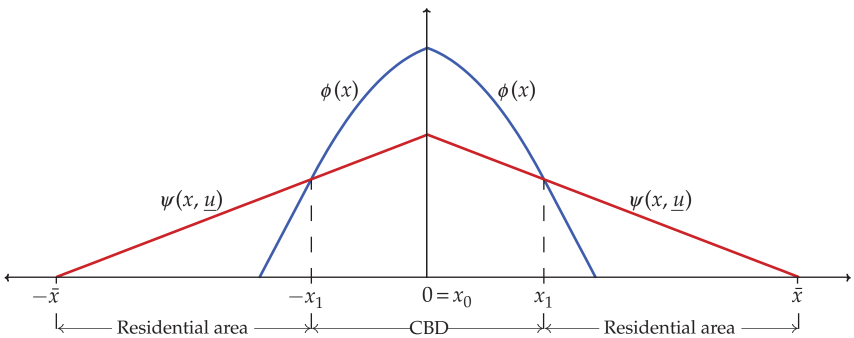

Figure 8.5 The monocentric equilibrium in Ogawa and Fujita (1980).

optimum configuration involves a monocentric city with a purely commercial district from $-\frac{\lambda}{2}N$ to $\frac{\lambda}{2}N$ surrounded by purely residential land. If $N < \frac{\tau(1+\lambda)}{\gamma} < 2N$, the market equilibrium involves mixed commercial and residential use throughout the city, but the optimum configuration involves a pattern like that in Figure 8.4. In both cases, the market delivers commercial development that is too dispersed. Only if $\frac{\tau(1+\lambda)}{\gamma} \leq \frac{N}{2}$ (monocentric configuration) or if $2N \leq \frac{\tau(1+\lambda)}{\gamma}$ (completely mixed configuration) do market equilibrium and optimum coincide.

The above framework can be extended and many of its simplifying assumptions relaxed. For instance, Fujita and Ogawa (1982) replace the linear decay of spillovers by an exponential decay. Lucas and Rossi-Hansberg (2002) also use an exponential functional form for commuting costs and further allow firms to substitute between land and labor and workers to choose their consumption of land. Many of the general features follow, but, without linearity, not only is the framework much more difficult to solve but the complexity of possible equilibria explodes.[34]

Explicitly modeling the location choice of both firms and workers in a city where they both compete for land is very useful for several reasons. First, it shows that the same method used to solve the monocentric model and its extensions to multiple groups can be applied in a context where firm location is endogenous. To do so, we construct the bid-rent function of land used by firms and, as before, ensure that land is allocated to the highest bidder. Second, the gradients that were present in the monocentric model still apply in this context. However, since multiple production centers can arise, these gradients are no longer necessarily monotonic. For instance, in the richer configuration described above, land rent gradients typically have local peaks both at the city center and at the center of commercial districts, with their relative height depending on parameters. Also, new gradients arise, in particular for wages, which also tend to be higher at denser locations. Third, we see that the monocentric model can be an equilibrium of a model where firms choose their location freely. This arises when the benefits of proximity for firms are strong relative to commuting costs. Finally, we are also able to explore alternative patterns of land use in the city. The richer configuration of the Ogawa and Fujita (1980) framework, while complex, may be a realistic stylized description of many cities: a central area where businesses and homes coexist and most people tend to work locally, with the next ring having a stronger commercial component, and finally an outer ring of residential areas from where people commute longer distances into the city. While the distribution of built-up land between commercial and residential uses shown in Figure 8.1 for Paris is more nuanced than the theoretical outcome in panel (a) in Figure 8.4, one can nevertheless see a very central area in Paris with more mixed use and then further out two peaks of commercial land (pointing downward, since the share of land in commercial use is plotted at the top in the bottom right panel in Figure 8.1).

[34] See also Helsley (1990) and Ota and Fujita (1993).

8.5. LAND USE REGULATION

In most countries and cities, land use is not solely the outcome of the market forces highlighted above. Zoning and other restrictions on land use and property transactions usually play a fundamentally important role in the determination of whether a parcel of land is developed, how it is developed, and who ends up using it. A comprehensive review of zoning and other restrictions on land use is beyond the scope of this chapter. We refer to Chapter 19 in this handbook for more details. See also Fischel (2000) for an introductory conceptual overview of the issues surrounding land use regulations. In developing countries, an important additional aspect of regulation is the insecurity that frequently surrounds land property and housing tenure. See Chapter 21 in this handbook for a review of work on this issue and other particular aspects of urbanization in developing countries. In what follows, we provide only a succinct account to show how land use regulations can be incorporated into the models we have explored so far and what type of issues they raise.

At a broad level, land use regulations limit the type and intensity of land development. They limit the type of development by specializing land use and separating users. They limit the intensity of development by imposing constraints on the size of buildings, limiting the floor-to-area ratio, limiting the fraction of land that can be developed for each parcel, or simply by making the application process for development extremely demanding and lengthy.

The main case for separating users relies on the existence of negative externalities that certain users may impose upon others.[35] This argument was first formalized by Stull (1974). He considers a city with land divided between a central commercial area and surrounding residential areas, as in Figure 8.5. Rather than considering this configuration as an equilibrium outcome where firms and residents compete for land everywhere in the city, as we did when constructing Figure 8.5, Stull (1974) considers the allocation of land from the point of view of a city developer who is constrained by such a monocentric configuration and has to decide where to place the limit between commercial and residential uses so as to maximize the aggregate value of land in the city. In the absence of any externalities across land uses, the limit between commercial and residential uses that maximizes aggregate land prices is at the point where the bid-rent curves for commercial and residential land intersect.

Consider now introducing an externality across uses. For instance, commercial activity may generate noise or pollution that reduces the utility of nearby residents. Suppose that this externality affects only the residents' utility and decays as one moves away from

[35] This is sometimes referred to as "Euclidian zoning" in the United States following the 1926 Supreme Court case *Euclid v. Ambler*, in which the Euclid community in suburban Cleveland opposed Ambler Realty, a developer of manufacturing space. See Fischel (2004) for an insightful history of zoning in the United States.

commercial land. Then this externality depresses the willingness to pay of residents to live close to the CBD and lowers their bid-rent curve. Firms are not directly affected by the externality so, absent any other changes, the point where the bid-rent curves for commercial and residential land intersect moves outward and commercial activity expands at the expense of residents. However, firms and workers interact not only in the land market but also in the labor market. The expansion of firms and the reduction in the number of residents increases the wage, lowering the bid-rent curve for commercial land and raising the bid-rent curve for residential land. In the end, wages will be higher and commercial land prices lower than in the absence of the externality. Residential land prices may be higher or lower depending on the magnitude of wage changes relative to the externality (they could be lower close to the CBD and higher toward the edge of the city). A city developer achieves a higher aggregate land value by constraining commercial land use, placing the limit between uses at a point where the commercial bid-rent curve is strictly higher than the residential bid-rent curve.

This example may seem of limited interest since most CBDs are no longer dominated by manufacturing in developed countries. However, it is easy to see that the same argument applies to any nonconforming use generating negative externalities in its neighborhood anywhere in a city.

We have just seen an example where zoning is justified on the basis of negative externalities across users. A similar argument can be made on the basis of positive externalities within uses. Consider the model in Section 8.4. We have already seen that in the presence of density externalities across firms the equilibrium may differ from the optimum because firms do not take into account the spillovers they generate for other firms by locating closer to them and they may end up too dispersed in equilibrium. For instance, we saw that if $N < \frac{\tau(1+\lambda)}{\gamma} < 2N$, the market equilibrium involves mixed commercial and residential use throughout the city, but the optimum configuration involves full separation between suburban residential zones and central commercial area.[36] Firms do not cluster enough because they do not internalize the effect of their own spillovers on other firms. Through zoning, one could achieve the optimal configuration.

While the sort of externalities examined here provides a strong justification for specializing land use and containing some types of users, some questions remain. First, it is obviously hard to know how much land should be devoted to manufacturing and where it should be located. Second, while the case for separation between dirty manufacturing and residential areas is strong, separating, as many cities do, residential from commercial areas is less obvious. Even the case for separating office space from housing is far from clear-cut (and far from being universally practiced). Third, even if separation is desirable, it is unclear whether this is best achieved by zoning to prevent the externality occurring in the first place or by having nuisance laws that force appropriate compensation ex post for

[36] The comparison between the equilibrium and the optimum configurations in this framework can be found in Imai (1982). See also Rossi-Hansberg (2004).

negative externalities. The latter can be optimal if the "compensation" is equal to the optimal Pigovian tax associated with the externality that is generated.

Turning to regulations that restrict the intensity of development, they could also be justified on efficiency grounds by a similar type of externality argument. For instance, a minimum lot size regulation in a nice neighborhood will prevent the construction of low-end housing which could affect the aesthetic quality of the neighborhood.[37] Restricting the intensity of residential development in certain areas could then be a way to specialize land use even further among subtypes of users in a desirable way. While this argument is logically consistent, it seems hard to believe that the negative externalities associated with mixing different forms of residential development are significant enough to justify the prevalence and importance of these regulations.

A more reasonable possibility is that residents value the (low) density around them. More specifically, assume residents value the open space that surrounds them as in Turner (2005). Although it is hard to fully operationalize the notion of open space empirically, there is little doubt that city residents value it highly.[38] In their property descriptions, realtors strongly underscore, when possible, nice views and proximity to parks and nature. They also keenly emphasize privacy using a positive adjective such as "secluded" or "oasis." This is not anecdotal. Parks and open spaces are quantitatively important in cities. Even a very intensely developed city such as New York boasts that more than a quarter of its land area is devoted to open spaces, parks, and other recreational areas. The data underlying Figure 8.1 indicate that open space represents 14% of the land in the 5 km ring around the center of Paris. The proximity to undeveloped land is highly appealing when residents value open space. Hence, the presence of open space begs new residential development, which of course reduces it. In equilibrium, this is self-defeating and there will be too much development. In turn, this may justify regulations that limit the intensity of development (Turner, 2005).[39] In practice, these regulations can take many forms, such as a maximum share of development for a parcel, maximum height,

[37] Restrictions on development impose an upper limit in the vast majority of cases. Although we ignore them here for the sake of brevity, lower limits such as minimum density are also used in some cases. They can find their justifications either as measures countervailing other inefficiencies that make the intensity of development suboptimally low or as measures necessary to achieve a "critical mass" to justify some amenities such as a neighborhood park, convenience stores, or public transport. Note also that preventing development in some locations is an implicit incentive to have it somewhere else as in the case of urban growth boundaries.

[38] See, for instance, Irwin (2002) and Geoghegan (2002).

[39] Strange (1992) studies this type of feedback effect in a model where density in neighboring city blocks imposes a negative disamenity, which in turn affects housing production decisions. Turner (2005) also provides some interesting results for the dynamics of development when residents have a taste for open space, showing, for instance, that remote locations will be developed before less remote residential areas and provides a consistent explanation for the leapfrogging that is often observed in the development process. A limitation to these results is that suburban development is often the work of nonatomistic developers who may internalize the externalities of the model, at least partially.

or "rights to lights" that allow neighbors to appeal against high-rise developments in large cities.

While a theoretical case can be made to restrict development and keep some land undeveloped in the form of parks and other green spaces, too little is known to provide firm guidance for urban planning. For instance, low density can occur through uniform development at a very low level with, say, one house per hectare over a 100 ha tract. The same level of density in this 100 ha tract can also be the outcome of a high-density development hosting 100 households over a 1 ha parcel surrounded by 99 ha of green space. Extant models and our knowledge of what residents value are not detailed enough for us to be able to deal with those issues at this stage. One might be tempted to use land values as a guide for making decisions. Although, in practice, land value appreciations seem grossly underused by the planning process, we know that in theory land prices provide an unbiased guide to investment in public projects only under restrictive conditions.[40] In the absence of solid knowledge of those issues, there are of course suspicions that many existing land use regulations might be widely off the mark.

A second set of reasons for controlling the intensity of development is given by possible externalities arising from commuting. This activity has been viewed so far as free of externalities. This is obviously counterfactual. Having more cars on the road slows down traffic. This is a solidly established fact (see, e.g., Small and Verhoef, 2007), which implies that the unit commuting cost, τ, used in Section 8.2 should be treated not as a constant but as a function of the number of commuters. The main complication is that the number of commuters is not the same everywhere in the city. In a monocentric city, all commuters need to enter the CBD, but only one commuter uses the last segment of the road at the urban fringe. More generally, commuting costs at a given point will depend on the number of commuters that live beyond this point. Hence, land use determines commuting costs, which determine land use. This problem was studied by Solow (1972). Commuting costs at location x depend on the number of city residents that leave beyond x, which we denote N_x so that the equivalent of the Alonso–Muth condition (8.16) will include $\frac{\partial\tau(x, N_x)}{\partial x}$ instead of simply τ in the case where commuting costs are linear with distance. Hence, in equilibrium, $\frac{\mathrm{d}P(x)}{\mathrm{d}x}$, the gradient of the bid-rent curve at location x, depends on the density at location x and a function of the cumulative of housing density beyond x. Since in equilibrium the price of housing $P(x)$ also determines the quantity of housing consumed by a resident and hence residential density, solving for the allocation of land amounts to solving a nonlinear second-order differential equation, for which closed-forms solutions are available only in particular cases as shown by Solow (1972).

[40] See Kanemoto (1988) for further discussions. Changes in land prices locally provide a biased measure of welfare changes when mobility is imperfect, when residents are heterogeneous, or when lot size is endogenous.

More generally, road congestion leads to inefficient land use. Zoning has been alleged to provide a solution to this problem. Pines and Sadka (1985) show that an urban planner can implement optimal land use by controlling parcel size (and the intensity of development). Should it turn out to be impossible to control parcel size, Pines and Sadka (1985) show that an urban growth boundary can improve patterns of land use in a second-best world. However, it is unclear why an indirect instrument such as the regulation of lot size and the intensity of development should be used instead of a more direct tool such as a congestion toll. Such a toll is analyzed by Kanemoto (1980), who considers a model in the spirit of Solow (1972) with congestible roads. The danger with indirect instruments is that they often do not include important other margins of adjustment. For instance, the results obtained by Pines and Sadka (1985) and others, who explore the possibility of using land use regulations instead of congestion tolling, were obtained by typically assuming that the number of trips per household and their destination are fixed. Regulations of land use by benevolent and highly sophisticated urban planners may manage to impose the "right number" of residents in each location, but nothing guarantees that they will drive the "right amount." On the other hand, it is true that existing congestion tolls are subject to difficult acceptance issues. They also form fairly crude instruments: typically a cordon instead of, optimally, a charge at each location that depends on the state of the traffic.[41] Even if congestion tolling is not available, there are other instruments such as parking pricing that could substitute for congestion tolls (Arnott and Inci, 2006). Parking pricing seems easier to use, more direct, and more flexible than comprehensive zoning.

In turn, road congestion is raising the issue of the provision of roadway, which has been ignored so far. Roads and parking spaces are quantitatively important in cities. The data underlying Figure 8.1 show that, while, as already mentioned, parks and other public spaces represent around 14% of land within 5 km of the center of Paris, the roadway, parking areas, and other transport infrastructure occupy around 18%. These figures are likely higher in the central parts of American cities, which tend to have wider streets and more land fully dedicated to parking lots.[42] The use of most roads is not excludable, and as a result, roads are publicly provided. Following the pioneering work of Solow and Vickrey (1971), roads have been explicitly considered in land use models by Kanemoto (1980) and Pines and Sadka (1985). The provision of roads differs from the provision of other local amenities such as parks and green spaces discussed above because roads are not directly enjoyed by residents. Instead, they provide a link between locations, and one road might affect different locations alongside it differently. As just argued, the issue of congestion also looms

[41] See Small and Verhoef (2007) for further discussion of congestion tolling. There has been some recent progress toward more time-dependent pricing for roads, but this is still limited to a handful of roads in the United States and a few cities elsewhere in the world such as Singapore, Stockholm, and London.

[42] See Manville and Shoup (2003) for further discussion of these issues and the difficulty of providing accurate numbers for them.

very large. Despite these differences with other local public projects, many of our conclusions regarding roads are the same: we know too little to provide good guidance for policy, and land prices will provide an imperfect guide for decisions.[43]

Although land use regulations are often justified by efficiency motives and curbing externalities, in practice zoning is often motivated by other reasons and is exclusionary in nature.[44] This may be because of peer effects. For instance, residents of rich suburban areas may want to maintain some exclusivity for their neighborhood, wish to remain among themselves, or want socially selected children to attend local schools. While this is perhaps part of the explanation, a fiscal externality is also likely to be at play.

In many countries, including the United States, a large share of local public goods, including education, is financed through property taxation. This can generate a fiscal externality which can be curbed using exclusionary zoning. To understand this, it is worth going back to Tiebout's (1956) original model of fiscal federalism. In his model, a heterogeneous population will sort into homogeneous jurisdictions that efficiently provide local public goods. In each jurisdiction, local public goods will cater to the tastes and incomes of local residents. To obtain this efficient outcome, a number of stringent conditions must be met, including having residents being able to vote with their feet, the absence of public good spillovers across jurisdictions, and the availability of lump-sum taxes. In practice, lump-sum taxes are not available for a variety of reasons. Instead, local public goods are financed through property taxation. The tax paid on a property is usually roughly proportional to its value. This creates a problem because property taxation makes it possible for poorer resident to free-ride by moving to a rich jurisdiction to consume a high level of public goods but pay only low taxes by consuming a small quantity of

[43] Skepticism about land instruments to fund roads was first raised by Mohring (1961) and Solow and Vickrey (1971). After Mohring and Harwitz (1962), transport economics is more upbeat about the use of tolls to implement an optimal provision of roads. See Small and Verhoef (2007) for further discussion and the exposition of a number of "self-financing" theorems for roads.

[44] For instance, the municipality where one of the two authors of this chapter lives imposes strict requirements on the maximum share of "impervious" (i.e., developed) surface for each parcel. This regulation is ostensibly justified on environmental grounds, but it seems hard to believe that 75–80% of land in a close suburb should remain unbuilt or unpaved to avoid floods and other environmental damage when the central part of the city is much more densely built. More likely, this forces residents to consume a lot of land and thus selects mostly rich residents willing to finance high-quality primary and secondary schools as argued in what follows. The municipality where the other author lives, on the basis of similar environmental concerns, limits the total amount of impervious surface by excluding most parcels from development instead of limiting development within each individual parcel. Unlike the previous example, this reduces the number of detached houses built on large lots, but also raises house prices and selects residents. This alternative regulation preserves large contiguous natural spaces instead of leading to many large private gardens. However, it generates a large price gap between parcels on which development is allowed or not contingent on a local policy decision, which opens the potential for corruption.

housing. That is, a decentralized provision of public goods financed by property taxation induces the poor to chase the rich. To avoid this, the rich impose exclusionary zoning regulations.[45]

Then, exclusionary zoning, to the extent that it restores a Tiebout equilibrium, could promote efficiency, if not equity. There are worries, however, that exclusionary zoning may go beyond implementing a Tiebout equilibrium and may instead maintain land development at inefficiently low levels within the urban fringe. As eloquently illustrated by Fischel (2001), local residents may be incentivized to restrict development locally to maximize property values. In many countries, including the United States, land use regulations are local decisions taken by local officials elected by homeowners. Fischel (2001) notes that although some new property developments might be expected to be beneficial to incumbent property owners, the latter may nonetheless rationally resist those changes because there is a risk that things may not go according to plan.[46] Risk aversion is heightened by the fact that homeowners usually have most of their assets vested in their house. It is also possible that gains from new property developments are highly unequally distributed, with some residents ending up losing from them. Appropriate compensation schemes are difficult to set up. As a result, the status quo may naturally arise as a political economy equilibrium.

A more direct argument is that restricting housing supply may lead to higher prices. Note that this argument requires reneging on another assumption of the Tiebout model: perfect mobility. If the demand for locations is flat (i.e., residents are indifferent between locations after being appropriately compensated for differences in accessibility), housing must be efficiently provided for property values to be maximized. Put differently, under perfect mobility and in the absence of local preferences, residents have an incentive for optimal land use regulations since this is what maximizes property values. Overly restrictive regulations will entail the inefficient use of capital in housing and reduce values.[47] If the demand for locations is not perfectly elastic (if, for instance, residents have a preferred location, all else being equal), incumbent residents have an incentive to restrict entry and limit housing supply locally, in effect acting as monopolies. Ortalo-Magné and Prat (2014) and Hilber and Robert-Nicoud (2013) propose some versions of this political economy argument. If it is right, overly restrictive regulations in developed areas would be a powerful force explaining excessive urban sprawl in undeveloped areas. We return to this issue in Section 8.8.

[45] See Fischel (1987) for more regarding this type of argument.
[46] Breton (1973) provides an early version of this argument.
[47] Overly restrictive regulations may also imply negative feedbacks through agglomeration effects as suggested by Chatterjee and Eyigungor (2014).

8.6. EMPIRICAL PRICE AND DEVELOPMENT GRADIENTS

We now turn our attention to empirical work on land use patterns, beginning with estimations of the gradients predicted by the monocentric model. Even before the work of Alonso (1964), Mills (1967), and Muth (1969), research had taken an interest in some of the predictions of the monocentric model. Clark (1951) is usually credited for being the first to show a decline in population density as one moves away from the CBD for a diverse cross section of cities. The popularity of Clark's work is to some extent due to its simplicity and its weak data requirements. Studies in this tradition first decide on a city center before drawing concentric rings around it. They count the population within each ring and regress it on the distance to the center. Unsurprisingly, the population density in most cases decreases smoothly with the distance to the center. This sort of regression usually yields a high R^2 that authors often interpret as strongly supportive of the monocentric model. However, this high R^2 is mostly due to the inherent smoothing associated with the ring approach. Another approach is to consider small areas within a city, such as tracts, and regress tract density on their distance to the center. The R^2 associated with this type of regression is much lower as there are often areas of fairly high density that are located relatively far from the main center.[48]

Under the assumption of a linear relationship between the logarithm of the population density and distance, Mills (1972) showed that the density gradient could be estimated knowing only the population of the main city, its area, and the population of the entire metropolitan area. This "two-point" approach lowers the data requirements even further, perhaps at the cost of highly noisy estimates.

The large literature that followed Clark's work is generally supportive of negative population density gradients. See McDonald (1989) for an early review and Bertaud and Malpezzi (2003) for further evidence from world cities. Kim (2007) documents a gradual flattening of the density gradient in US cities over the twentieth century. Interestingly, exceptions to negative gradients include cities in formerly socialist economies such as Moscow (Bertaud and Renaud, 1997) and South African cities under apartheid (Selod and Zenou, 2001), where the market mechanisms at play in the monocentric model was heavily constrained. Even in countries where cities tend to follow a broadly monocentric pattern, several interesting features emerge. First, the density gradient typically becomes weaker far from the center. Second, other variables such as the distance to secondary centers and various geographical landmarks also often have some explanatory power.

[48] Gradients also tend to become much flatter far from the CBD in American cities. Mechanically, the fit of a weak negative gradient will be poor, with R^2 going to zero as the slope also goes to zero.

There is also work that attempts to measure the gradient of the price of housing as illustrated, for instance, by Yinger (1979) or Coulson (1991).[49] For reasons to be explained below, this literature has often struggled to provide evidence of negative gradients for the unit price of housing. Perhaps because of the mixed success of the approaches looking at unit housing price gradients, much of the literature has focused instead on land price gradients after Mills' (1969) pioneering effort. The works of Cheshire and Sheppard (1995) and more recently Ahlfeldt (2011) are perhaps the most advanced. This literature is too large to review extensively here. See McMillen (2006) and McMillen (2010) for surveys. Although the findings are generally supportive of a negative land price gradient, McDonald and Bowman (1979) raise some doubts.

In contrast to the abundance of work looking at density, housing price, and land price gradients, little to no work has been devoted to the predicted gradients of the capital intensity of housing development and of housing consumption per household. The work of McMillen (2006) is a lone exception that examines the floor-to-area ratio in metropolitan Chicago and shows a strong declining trend as one moves away from the CBD.[50] As far as we know, there is no work looking at housing consumption per household in relation to the distance to the center. Figure 8.1 provides a crude but sharp illustration for the case of Paris through the shares of land in residential use occupied by single-family and multifamily buildings at various distances from the city center. Only 3% of land in residential use within 5 km of the center of Paris is occupied by single-family homes. In contrast, single-family homes occupy 50% of residential land between 5 and 10 km from the center, a figure that rises to 79% between 10 and 20 km from the center, and to 87% between 20 and 30 km from the center.

Despite a large body of work that has developed over more than 60 years since the work of Clark (1951), the empirical knowledge accumulated on the monocentric urban model and its extensions remains limited. The first reason is that, until recently, data have been difficult to collect. Assembling data about tracts in cities or property prices often needed to be done manually. As a result, much of the literature is limited in scope and focuses on one particular city or perhaps a small number of cities. To make matters worse, these cities are often acknowledged to have been chosen for particular reasons, including strong priors about how monocentric or multicentric these places might have been. Looking at broad cross sections of cities is needed to avoid such sample selection issues.

[49] It is important to remember that the monocentric model makes a prediction for the price of housing per unit not for property prices. As one moves away from the CBD, the unit price of housing is expected to fall, but the size of properties is expected to increase. The net effect of these two forces on overall property prices is theoretically ambiguous.

[50] See also Clark (1967) for early evidence.

It is also important to assess how much cross-city heterogeneity there is within one country. The work of Combes et al. (2012) is of particular interest in that respect. They estimate land price gradients for a broad cross section of cities in France and find a lot of heterogeneity. The elasticity of land prices with respect to distance varies from about −50% for cities in the first decile to essentially zero for some cities in the last decile. Casual observation also suggests tremendous heterogeneity in urban land use across cities of different countries, with commentators often referring to the "American city" as synonymous for urban sprawl with extremely low suburban densities and flat gradients there. Instead, the "Asian city" is often taken as synonymous with extremely high density in the core. The behavior of gradients over time is also of interest as we expect growth and technological progress in the last 200 years to have affected cities in a major way. Unique in the literature, McMillen (1996) uses comprehensive land price data over 150 years in Chicago to document a flattening of the land price gradient. Hopefully, future work will make use of data that are now broadly available (at least in some countries) about property prices, urban land use, and the population to document the evolution of land use and its heterogeneity both within and between countries.

Data availability is not the only issue. The data and approaches that have been used in the literature so far are problematic in several respects. Let us first discuss measurement issues. The first problem concerns the definition of centers or subcenters. Older studies often assign a center in a somewhat arbitrary manner on the basis of history or casual evidence.[51] More recent studies tend to compute a centroid or use a peak of density for residents or, better, employment. While the latter approach is probably sufficient to determine a CBD for some applications, matters are more complicated when the analyst wishes to consider several subcenters. The most convincing approach to detect centers and subcenters was proposed by McMillen (2001). It is inspired by two related features from theory. First, subcenters will be associated with a concentration of jobs. This concentration will attract residents, and as a result, subcenters will affect land and housing prices in nearby locations. Consistent with these two features, the approach developed by McMillen (2001) is in two steps. First, a smooth employment density function is estimated nonparametrically. Candidate subcenters are locations with positive and significant residuals in this employment regression. Then in the second step, a semiparametric rent function is estimated. Subcenters will be those among the candidates identified in the first

[51] In Figure 8.1, we assigned the cathedral of Notre Dame as the center of Paris. Although the French government uses this location to measure all distances to Paris, it is hard to argue that the area around this cathedral constitutes a CBD. The historical business district of Paris is about 4 km to the west and the current main business center of La Defense is 9 km to the west. At the same time, we can see in Figure 8.1 that the fraction of land that is built up and the percentage of residential development accounted for by multifamily dwellings both peak at about the location of Notre Dame, indicating that it may be a reasonable choice for the city center in terms of residential uses. Once again, Paris provides a good example of the complexity of land use allocations in practice.

step of the analysis that provide significant explanatory power for rents. Despite its appeal, this approach is likely to be sensitive to the smoothing parameters since most cities will look fully monocentric with enough smoothing.[52] The significance of any subcenter in the second-step rent regression will also depend on the physical extent of the study area as the effect of small subcenters becomes more difficult to detect in larger areas.[53]

Another measurement problem with density gradients is that density is usually measured at the area level in the empirical work that seeks to estimate density gradients, whereas the model makes predictions at the parcel level. This is a worry because a smaller fraction of land may be used for residential purpose further from the CBD. This may explain a negative population density gradient irrespective of what happens to density at the parcel level.[54] A simple solution here is to rely on data from a land registry or any other source that is able to identify residential parcels. The difficulty is that such data are far less widely available than area population data at the tract level.[55]

While measuring population density involves some unexpected difficulties, measuring unit housing prices is distinctly harder. The chief reason is that we usually observe only the price of a house—that is, the product of the unit price of housing and the number of housing units that this house provides. The standard solution—used, for instance, in Coulson (1991)—is to use house characteristics and introduce them in the regression that estimates the housing price gradient, such as

$$\log P_i(x) = F(x) + X_i\alpha + \epsilon_i\,, \tag{8.72}$$

where $P_i(x)$ is the price of house i located at distance x from the CBD, $F(x)$ is a function of distance x, X_i is a collection of house characteristics, and ϵ_i is the error term. The mapping from the Alonso–Muth condition (8.4) to the empirical Equation (8.72) which aims to estimate it is not immediate. It is, however, possible to rewrite Equation (8.4) in first difference for a given house relative to another house nearby. After linearizing, we obtain $\Delta\log P(x) \approx \tau\,\Delta\log x + \Delta\log h$. Approximating h, the number of housing units of a property, by a vector of characteristics and using the same hypothetical house as a comparison

[52] Recall that the so-called optimal smoothing is based only on a rule of thumb that trades off noise reduction against systematic changes in curvature.

[53] See Redfearn (2007) for a more local approach that is not subject to this criticism. Another possible criticism is that this procedure is inspired by theory but not directly led by it. This criticism is less convincing because no existing theory of land use is detailed enough to go beyond the broad features used by McMillen (2001). In addition, one might be wary of using one particular theory to define subcenters as more precise definitions will likely be driven by specific features of models.

[54] Cheaper land further from the CBD may, of course, explain a greater prevalence of nonresidential use in more complex versions of the monocentric model where several types of users could be competing for suburban land. It remains that a negative density gradient at the area level is not the same thing as the negative density gradient at the parcel level predicted by the simplest models.

[55] The work of Mieszkowski and Smith (1991) is an exception using parcel-level data for the city of Houston.

point for all houses yields regression (8.72) for the particular case of a logarithmic specification for the distance effect. Taking a more general functional form for commuting costs in the theoretical model still yields Equation (8.72). Note that when the range of distances from the CBD becomes large, the approximation made in the linearization above which disregards the $\Delta \log x \Delta \log h$ term may not be warranted. The use of hedonics to condition out housing heterogeneity is not innocuous either, and several well-known problems may arise when using this tool. The most important problem is that, as predicted by the theory, house characteristics will be correlated with the distance to the CBD and it is unlikely that the vector of the observed house characteristics used by the analyst is both comprehensive and well specified. As a result, there may be missing house characteristics that are correlated with distance.[56]

Much attention in the literature has been devoted to the choice of the functional form for the distance function $F(x)$ in Equation (8.72) (and in the corresponding equation estimating density gradients). Following Clark (1951), the early literature typically used a dependent variable in log and the level of distance as the explanatory variable. As made clear by the model developed in Section 8.2, the price of land may be a complicated function of distance which depends finely on which assumption is made regarding the utility function, commuting costs, and residents' income. Nothing guarantees that the resulting gradient will be a negative exponential in distance. In general, it will not. An alternative is to use the log of distance instead of its level as explanatory variable.[57] Combes et al. (2012) argue that the fit is approximately the same for both specifications when estimating land price gradients for French cities. Many authors have used more comprehensive specifications by adding distance terms of higher order. These terms obviously increase the explanatory power of the regressions, sometimes by enough that the authors justify them by invoking specification tests.[58] McMillen (2010) defends the use of nonparametric or semiparametric estimation methods.

At some level, the usefulness of this debate about functional forms in the estimation of gradients is questionable because it forgets which proposition from theory is tested. An exception is the work of Brueckner (1986). He proposes using switching regression methods to capture discontinuities caused by vintage effects in housing construction such

[56] As possible alternatives, Epple et al. (2010b) and Combes et al. (2014) develop new methods to reconstruct the amount of housing offered by each property. Epple et al. (2010b) use land and property values as well as land area. Land values are a function of house prices. Then, one can estimate the capital-to-land ratio in housing from housing values per unit of land and retrieve the production function. The quantity of housing can then be recovered. Combes et al. (2014) use the first-order condition for profit maximization and free entry that gives the marginal product of housing with respect to capital (or land) and recover the quantity of housing by integrating.

[57] While negative exponentials impose a lot of curvature, linear specifications are at the other extreme since, as argued above, theory predicts convex gradients.

[58] Recall nonetheless that these tests weigh explanatory power against the number of explanatory variables, imposing somewhat arbitrary penalties for the latter.

as those described in Section 8.3.3. Hence, the objective is not to improve the fit of the regression for the sake of it but to capture an empirically important feature that the simplest version of the model misses.

The theory described in Section 8.2 makes a series of predictions under the assumption of a single city center. These concern accessibility, chiefly through the Alonso–Muth condition for land and housing prices and related conditions for the intensity of development, housing consumption, and the density of the population. This assumption of a single center is mostly for convenience and, as argued in Section 8.4, many of the key propositions regarding the importance of accessibility for housing prices and, in turn, land prices, the intensity of development, local population density, and the consumption of housing do not depend on the existence of a single center.

Much of the disagreement in the literature about the monocentric model is about whether cities are monocentric or not. They are not about whether accessibility is valued in accordance with the Alonso–Muth condition. These are two separate questions. The first question is really about an assumption about the geography of cities, whereas the second question is the analytical substance of the model. This said, while the monocentricity of cities in the monocentric model is an assumption, in the models with endogenous job locations reviewed in Section 8.4 it is just one possible outcome. This suggests that cities may be well approximated as monocentric or not depending on the conditions, making it valuable to assess the monocentricity of real cities. Unfortunately, much of the debate here has been misconstrued. The interesting question is not whether cities are monocentric or not. Strictly speaking, most cities are not monocentric and maybe no city is truly monocentric. For instance, McMillen (2001) provides convincing evidence that all the large American cities he looks at contain subcenters, sometimes a sizeable number of them. The prediction of a monotonic decline in land prices, housing prices, and population density in all directions as one moves away from the CBD will always be rejected provided one can work at a sufficiently high level of geographical resolution, be it only because of access to transit stops or major arterial roads. Adding to this, we cannot expect the forces highlighted by the monocentric model to be the only determinants of land and property prices, housing consumption, and population density. As shown by the review of the extensions of the monocentric model in Section 8.3, other factors need to be conditioned out. The world will never be as smooth as the simplest models.

Since monocentricity can always be rejected, the more interesting question is: How monocentric are cities? As argued above, the key difficulty here is that the R^2 of gradient regressions do not provide a good metric by which to measure this, given the granular aspect of population patterns. Reasonable metrics that allow us to measure how monocentric cities are independently of the level of resolution at which population and prices are measured or independently of how the data are smoothed are yet to be developed.

Turning to the second question, about accessibility, it has been only half answered by the literature. We know that indeed distances to the center, subcenters, and other landmarks nearly always matter greatly to explain house and land prices. There is also a body of literature which has developed a number of accessibility indices that weigh the distances of each location to various subcenters. See Anas et al. (1998) for a review. That accessibility matters is necessary for the accessibility versus price trade-off highlighted by the Alonso–Muth condition to be true. However, this does not guarantee that accessibility is valued as suggested by this condition, where, all else being equal, the difference in property prices between two blocks should be equal to the difference in transport costs involving not only commuting but also all other errands.

As highlighted in Section 8.2, the monocentric model also makes a number of predictions about some aggregate quantities and land rent at the center.[59] As argued in Section 8.3, these predictions are theoretically less robust than those that rest on the Alonso–Muth condition. Like with the monocentricity question above, the issue is not whether these predictions hold exactly—they will not—but how far we are from them in reality and what accounts for the difference between these simple theoretical predictions and reality. To our knowledge, the predictions about urban aggregates like the proportionality between total land rent and total commuting costs have never been seriously assessed. The proposition that land rent at the center is proportional to city population is assessed by Combes et al. (2012), who provide a measure of urban costs based on equilibrium land prices at the city center in France. Their preferred estimate for the elasticity of the price of land at the city center with respect to the city population is 0.72, and they attribute the difference between 0.72 and unity to greater decentralization in larger cities. The same article also provides results for regressions attempting to explain land price gradients with a variety of city characteristics. The results are not particularly encouraging. None of the explanatory variables that Combes et al. (2012) is robustly significant. Models of heterogeneous residents in the spirit of those explored in Section 8.3.2 predict that the parameters that govern income distribution will also determine land price gradients. Combes et al. (2012) find no evidence of that. Obviously, more work will be needed on this before definitive conclusions are reached.

A recent contribution by Ahlfeldt et al. (2012) estimates structurally a model of internal city structure. The city is a collection of blocks. Although these blocks are taken as given, the intensity of development of each block is endogenous. Residents consume housing, residential amenities, and a composite good. In turn, amenities in a block depend on the fundamentals of this block and amenities in neighboring blocks, which are discounted by a negative exponential function of the distance to the block under consideration. Production requires floor space and labor. As in the model described in Section 8.4, firms benefit from agglomeration economies that are measured by employment density in neighboring blocks discounted by distance just like with the consumption externalities. Productivity is also allowed to vary idiosyncratically across blocks.

[59] As argued in Section 8.2, it also predicts that the ratio of the land and housing gradients should give the amount of housing. This prediction has not been tested to the best of our knowledge.

When choosing their residence, workers not only consider (endogenously determined) housing prices, local amenities, and the distance to work, but they also receive an idiosyncratic utility shock for each possible commute between blocks. These shocks introduce some heterogeneity in residential and employment choices. Residents choose the block that gives them the highest utility. A key technical difficulty is that the highest utility is given by a potentially extremely hard-to-compute order statistic regarding the distribution of commuting shocks. To keep the problem tractable, Ahlfeldt et al. (2012) use the structure developed in international trade by Eaton and Kortum (2002), which relies on the Fréchet distribution. The main reason for doing so is that the maximum of a sequence of Fréchet-distributed variables is also Fréchet distributed.

An interesting property of the framework of Ahlfeldt et al. (2012) is that it predicts a gravity pattern for commuting flows where the logarithm of the number of commuters between two blocks is expected to be proportional to the travel time between them, after conditioning out features of the origin and destination blocks. Obviously, commuting flows between nearby residential blocks will be nil when there is no employment in these blocks. Ahlfeldt et al. (2012) provide some evidence consistent with this feature for Berlin commuters. The full estimation of their model requires recovering two parameters for commuting costs (the base utility cost of commuting per unit of distance and the idiosyncratic dispersion around it), four agglomeration parameters (the intensity and the spatial decay for both consumption and production externalities), and the productivity and amenity fundamentals of each block. There exists a unique mapping from observed data on commuting, land rents, the intensity of development, employment, and the number of residents for each block and knowledge of key aggregate shares in consumption, production, and construction to the unknown parameters and fundamentals of the model.

The first key result of Ahlfeldt et al. (2012) regards the estimation of rather large but extremely localized agglomeration effects for both production and consumption. They also take advantage of the fact that they are able to observe Berlin at three points in time: before its division and during and after reunification. Assuming unchanging fundamentals for each block in expectation, they are able to replicate with their model the changes in the employment and residential composition of blocks depending on where they are located.

This article is important for several reasons. First, it constitutes a pioneering effort to implement the methods of structural estimation to understand what happens within cities. These methods have been developed in the last 10–20 years in other fields of economics and only now are beginning to be used to analyze land use issues.[60] Second, the

[60] See Chapter 2 in this volume on structural estimation in urban economics. Structural methods have been used to explore housing and local public finance issues, but little to none of that work has an explicit spatial dimension. See Epple et al. (2010a) for a model that considers localized amenities within a multijurisdiction framework. Very recently, quantitative versions of the monocentric have been developed to explore a range of applied questions such as the welfare effects of a simple congestion toll (Brinkman, 2013).

work of Ahlfeldt et al. (2012) allows an economic interpretation of the various gradients observed in cities. Like Combes et al. (2012), it is an attempt to bring empirical work closer to theory, where gradients and other spatial patterns are no longer examined for their own sake but to learn about more fundamental issues. Third, the weaker points of the model or the dimensions where results are less satisfactory become more clearly apparent. For instance, the modeling of commuting behavior in Ahlfeldt et al. (2012) is somewhat *ad hoc* since it is not clear what drives the heterogeneity of bilateral commutes for residents. This is an avenue for future research. Some of the more provocative conclusions of Ahlfeldt et al. (2012) regarding local agglomeration effects in production and consumption will also be examined through other lenses. The robustness of the approach may also be assessed by implementing it for other cities. More generally, the interaction between theory and empirical work leads to their mutual questioning and paves the way for progress in our knowledge.

A few recent contributions notwithstanding, there is still a great paucity of empirical work that takes theory seriously. The main thrust of land use models in the spirit of Alonso (1964), Mills (1967), and Muth (1969) lies in the Alonso–Muth condition that highlights the key trade-off between prices and accessibility. This condition remains to be assessed empirically. The key extensions of the monocentric models explored in Section 8.3 have received scant empirical attention. As just argued, close to nothing is known regarding the effect of income heterogeneity on the various gradients. Similarly, this literature on urban land use has not really taken zoning constraints into account. The work of Cheshire and Sheppard (1995) that integrates planning together with a broader notion of accessibility is a lone exception.[61] As described in Chapter 16 in this handbook, there is a significant literature that deals with the durability and decay of housing, but only little of it is explicitly tied to broader and more formal land use models.[62]

8.7. PATTERNS OF RESIDENTIAL SORTING WITHIN CITIES

The empirical work discussed in the previous section relates to the framework developed in Section 8.2 which considered homogeneous agents. As argued in Section 8.3.2, many of the predictions of the monocentric model with homogeneous residents carry through to heterogeneous residents. Models with heterogeneous residents also make further important predictions regarding the sorting of residents depending on their observable characteristics.

[61] A key difficulty here is that land use regulations are hard to measure and raise simultaneity problems that have turned out to be extremely hard to deal with. See Chapter 19 in this handbook for further developments on these two issues.

[62] The work of Brueckner and Rosenthal (2009) is an exception that is discussed further below.

In the simplest case where residents differ only along one dimension (e.g., income), perfect segmentation across income groups is expected. When residents differ along two or more dimensions, only partial sorting is predicted. For instance, a household with high commuting costs may choose to live far from the CBD because it also demands a lot of land. That household may end up being a neighbor of a household with lower commuting costs and a lower demand for land. Income has been of particular interest in the literature because it is a key dimension along which households differ within cities and because it generates an interesting tension. We have seen that, from a theoretical standpoint, the effect of income on distance to the CBD is ambiguous. Higher household income is expected to increase the demand for housing and thus push households further from the CBD. However, since commuting costs are also paid in time, higher income and the associated higher value of time make households want to live closer to the CBD. As shown in Section 8.3.2, richer households will want to live further from the CBD than poorer households if the income elasticity of the demand for housing/land is greater than the income elasticity of commuting costs.

Before we discuss the predictive power of this theoretical proposition, it is important to know what the basic facts are. Looking at the historical record for the United States, LeRoy and Sonstelie (1983) argue that until the second half of the nineteenth century, richer residents were living in more central locations. That pattern reversed afterward with the emergence of the streetcar, followed by the rise of the car. LeRoy and Sonstelie (1983) then argue that some regentrification of central cities occurred after 1970. Their model attributes these changes to the fact that, early on, the streetcar and the automobile were available only to richer households but allowed considerable time savings on commutes. This could have led to the suburbanization of richer households. Eventually, the automobile became available to all households but the poorest ones. This allowed a much broader range of households to decentralize, and rising land prices in the suburbs may have driven richer households to return to central locations. Hence, these broad patterns are consistent with a simple monocentric framework in which the demand for land is relatively inelastic.

More recently, Lee and Lin (2013) have assembled a long-run dataset of neighborhood income and population in major US cities. They confirm that average neighborhood income declined with distance to the CBD in 1880. This income gradient changed sign as early as 1930, and became very steep by 1940. Since 1960 the income gradient has flattened, a process that continues today. These patterns of decentralization and recentralization of wealthier households are much less marked for coastal cities. Lee and Lin (2013) argue that this is due to the better amenities of central locations in coastal cities as stability in neighborhood income strongly correlates with measures of natural amenities.

Brueckner and Rosenthal (2009) also provide detailed evidence for contemporary American cities. Using information at the tract level, they show that there is a tendency for tracts located further from the densest tract to be richer. Importantly, they show that this tendency is not uniform since tract income flattens and even declines beyond a certain distance from the center. They also show that the income gradient is particularly steep close to the center. The major income gap occurs between central cities and their close suburbs. For large metropolitan areas, the gap between the richest and poorest tracts is about 50%.

This said, this broad characterization is not an absolute norm, far from it. A few American cities have a rich central part. New York is certainly an example. Glaeser et al. (2008) report U-shaped curves plotting incomes as a function of distance to the CBD for New York, Chicago, and Philadelphia. Importantly, as highlighted by Brueckner et al. (1999), European cities tend to have rich cores and poor peripheries, with, again, some exceptions such as Brussels. Adding to this, many cities also exhibit important directionality patterns. In most European cities, western suburbs tend to be richer than eastern suburbs.

It is also important to note that residential areas are far from homogeneous. Epple and Platt (1998) document that in the Boston area median income in the richest municipality was four times as high as that in the poorest municipality. However, nearly 20% of the households in the richest municipality had an income below that of the 20% of richest residents in the poorest municipality. This is indicative of considerable overlap in the income distributions across both municipalities. Hence, while the forces that push toward sorting by income are certainly present, they cannot be overwhelming since spatial sorting by income is far from perfect.

To assess the ability of the monocentric model with heterogeneous residents to explain residential location patterns by income, one needs to estimate both the income elasticity of the demand for housing/land and the income elasticity of the cost of commuting. The first attempt to compare these two elasticities was made by Wheaton (1977), who raised some early skepticism regarding the performance of simple monocentric models along that dimension.

In his work, Wheaton (1977) assumes a utility function with constant elasticity of substitution. He then estimates the coefficients of that utility function using household data collected before the construction of the Bay Area Rapid Transit in the San Francisco metropolitan area. Note that commuting appears both as a disutility and as an expenditure. Wheaton (1977) estimates the income elasticity of the demand for land and the income elasticity of commuting costs to be roughly equal at around 0.25. This suggest that the net effect of income on location within the city stemming from the trade-off highlighted by the monocentric model is likely too small to explain actual residential patterns.

Glaeser et al. (2008) revisit the issue. They first argue that since commuting is mainly paid in time, the income elasticity of commuting costs should be 1 or close to it. Then, they focus their attention on estimating the income elasticity of the demand for land. They first regress the logarithm of parcel area on the logarithm of household income and estimate an elasticity of the demand for land between 0.1 for single detached houses and 0.3 when some land consumption is ascribed to households living in apartments. Given that their estimate for the income elasticity of the demand for land is well below the unit income elasticity of commuting costs that they assume, they conclude that the monocentric model with heterogeneous residents strongly predicts that richer households should live in the urban cores of US cities. This is obviously counterfactual. As we discuss below, Glaeser et al. (2008) incorporate transit into their analysis to explain why the poor live close to the urban core of American cities. However, before we discuss transit and other alternative explanations, it is worth noting that the conclusions of Glaeser et al. (2008) regarding the relevance of the monocentric model for income sorting are perhaps overly strong. First, their estimate of the income elasticity of the demand for land is likely to be downward biased.[63] But more importantly, the income elasticity of commuting costs is likely to be substantially below 1.

While the cost of time is an important component of the cost of commuting, it is not the only one. Consider 1 h of urban driving at a typical speed of 40 km/h. For a car consuming 12.5 l of gasoline per 100 km, this correspond to about \$5.[64] Adding a depreciation of 12.5 cents per kilometer, the variable cost of the vehicle is \$10 for 1 h.[65] Consider now a low-skilled worker making \$10 an hour driving this car. This is slightly less than half the median wage. Estimates from transport economics suggest that time in vehicle is valued at about half the wage (Small and Verhoef, 2007). The total cost of driving for this worker is thus \$15 per hour but the cost of time is only a third of this. For this low-skilled worker, the elasticity of the cost of commuting with respect to the wage is only 0.33, which is far from unity. Even for a worker making close to the median wage of \$20 per hour, the wage elasticity of commuting costs is still 0.5. For highly paid workers, workers making close to twice the median wage, \$40 per hour, this elasticity is still 0.66 and maybe not very different from the elasticity of the demand for housing.[66]

[63] There are three reasons for this. First, the price of land is missing from their regression of the logarithm of parcel area on the logarithm of household income, and cheaper land in the suburbs is likely to increase demand for land by wealthier households. Second, they focus on instantaneous as opposed to permanent income. When they use education as an instrument for permanent income, the income elasticity of the demand for land rises above 0.5. Third, the consumption of land might be mismeasured if wealthy households choose to live in particularly low density areas with parks and a lot of open space. Considering total land per household instead of lot size raises the income elasticity of the demand for land to close to 0.5.

[64] We use 2014 prices close to \$4 per gallon. A gasoline consumption of 12.5 l per 100 km is equivalent to about 19 miles per gallon, which is slightly better than the fuel economy of a Ford Taurus.

[65] This depreciation would apply to a car worth \$25,000 fully depreciating over 200,000 km. This price tag is slightly below that of a Ford Taurus.

[66] This calculation neglects to consider that this driver may drive a more expensive car. With a car worth \$50,000 and a 20% lower fuel efficiency, we return to an elasticity of 0.55.

Hence, the relevance of the monocentric model with heterogeneous agents to explain patterns of location by income is still very much an open question.[67] Given a lack of decisive evidence, the literature has proposed a range of complementary explanations. In the spirit of LeRoy and Sonstelie (1983), Glaeser et al. (2008) emphasize the importance of transit in explaining why the poor live close to the urban core of American cities. While the majority of American households can count on a car for each adult, about one-third have only one car and about 10% have no car at all. This suggests that a sizeable minority of American adults must rely on public transport to go to work and do many of their daily errands. Transit is usually present in the core of most American cities, but is often absent from their suburbs. Glaeser et al. (2008) show a strong connection between poverty rates in urban tracts, access to public transport, and closeness to the city center.

While these correlations are interesting, they do not establish any form of causality. Glaeser et al. (2008) buttress the case in favor of public transport by comparing cities with subways and cities without subways. Unsurprisingly in both types of cities, patterns of transit usage differ considerably. Transit usage declines fast with the distance to the center in cities without subways, whereas it increases over the first 5 km and remains on a high plateau for about 10 km in cities with subways. In both types of cities, transit usage is closely negatively mirrored by local median incomes. Given that most subways have been in operation for three generations or more, the difference between subway and nonsubway cities has some degree of exogeneity.

Brueckner et al. (1999) focus on the difference between typical American cities, where the poor live in the urban core, and typical European cities, where the core is occupied by the rich, and propose an amenity-based explanation. They develop an extension of the monocentric model with heterogeneous residents who need to commute not only to the CBD but also to enjoy localized amenities. Their main comparison is between Paris and Detroit, and they argue that amenities in Paris are centrally located, whereas, if anywhere, they are located in the suburbs in Detroit. Depending on the assumptions made regarding the demand for land, many equilibrium configurations are possible. When the pull of central amenities is strong enough and the demand for land is not too income elastic, a Paris-type equilibrium is possible, with rich residents located in the core. When the demand for land is not too inelastic, peripheral amenities can induce rich residents to locate in the outer suburbs. Their analysis also allows amenities to be endogenous. They naturally occur where the rich live, and this allows for multiple equilibria since, for instance, the rich can live in the core and generate

[67] Better estimates of the demand for land would need to solve the three issues raised by Glaeser et al. (2008). Getting better estimates of commuting costs will be difficult since commuting arguably enters the utility function directly. Survey data seem to suggest that commuters have a strong dislike for long commutes but may not want to live exactly where they work (Krueger et al., 2009).

endogenous amenities such as restaurants which are enough to retain them even though the exogenous amenities such as open space are at the edge of the city.

A first shortcoming of this analysis is that it does not appear to rule out much. Many configurations can be rationalized as an equilibrium outcome that depends on the subtle interplay of features that are hard to observe, such as the presence of amenities and the shape of the demand for land. The model of Brueckner et al. (1999) also remains devoid of explicit empirical evidence. This said, casual evidence nonetheless suggests that there is some explanatory power to this type of explanation. Amenities may also explain an interesting asymmetry observed in many European cities, where, as already mentioned, western suburbs are often much richer than eastern suburbs. This appears to be closely related to winds blowing from the West in those cities. While wind direction no longer constitutes an important issue in these cities, it certainly did at the peak of the industrial era, when a windward location helped the affluent avoid factory emissions and smells.

A third extension of the monocentric model concerns the possible role of the age of the housing stock. Brueckner and Rosenthal (2009) show that, in American cities, controlling for the age of the housing stock explains away the positive income gradient as one moves away from the center. Older constructions, which are located more centrally in American cities, offer lower-quality housing and thus end up being allocated to poorer residents in equilibrium. This is an important element of the filtering mechanism explored in Section 8.3.3. Assessing causality is hard because, to some extent, the state of the housing stock is an endogenous variable. To limit the scope of this problem, Brueckner and Rosenthal (2009) use as an instrument for the age in 2000 of the housing stock in city tracts the same variable in 1980. Using information on the current age distribution of the housing stock and likely reconstruction patterns, they offer 20-year predictions regarding the distribution of the population by income in American cities. These predictions entail further gentrification of urban cores.[68]

While all the explanations examined so far build on the monocentric model, the last one puts much less emphasis on relative locations and the trade-off between accessibility and land prices. It instead relies on public finances and the provision of various public goods in the spirit of Tiebout (1956). The two main ideas are that central cities started engaging in costly social policies and significant redistribution in the 1960s and the 1970s. See, for instance, Inman's (1995) vivid account of these issues in Philadelphia. These policies led well-off residents to flee to suburban municipalities. In turn, poorer urban cores were afflicted by a variety of social ills ranging from crime to failing schools. This led to a further exodus of the middle class from the blight of central cities. This conjecture is consistent with the documented fact that the major income gap is between central cities and surrounding municipalities. This conjecture also provides a natural explanation for the

[68] See Chapter 16 in this handbook for more on these issues.

differences between Europe, where municipalities often have little fiscal autonomy, and the United States, where education is locally financed and fiscal autonomy is much greater.

There has been quite a lot of work over the years that has thought to evaluate the "flight from blight" conjecture. Among the social ills that have been explored, one may mention crime (Cullen and Levitt, 1999), the degradation of the housing stock (discussed in Chapter 16 in this handbook), racial preferences (Boustan, 2010), and related changes in the school system (Baum-Snow and Lutz, 2011). De Bartolome and Ross (2003) propose a model that embeds aspects of Tiebout fiscal competition into a monocentric framework with two groups of residents, rich and poor. There has been, however, no empirical work to our knowledge that runs a clear horse race between the role of fiscal decentralization and the trade-off between commuting costs and land prices to explain patterns of location by income within cities.

Before closing this section, we need to return to the considerable overlap in income in most areas regardless of how far they are located from the city center. This problem is tackled head on by Epple and Platt (1998) in a nonspatial framework with multiple Tiebout communities. In addition to the usual Tiebout mechanism of catering to the needs to local residents, Epple and Platt (1998) also introduce idiosyncratic preferences for locations. When they quantitatively estimate their model, they find that idiosyncratic preferences explain most of the location decisions of households. This is a disappointing result in that the idiosyncratic preferences introduced by Epple and Platt (1998) are a residual that, in the end, only measures our ignorance. At the same time, this result is consistent with the literature described above starting with Wheaton (1977) which suggests that the difference in the accessibility-land price trade-off for the rich and the poor is unlikely to lead to overwhelming patterns of segmentation.

Bayer and McMillan (2012), using a method developed in Bayer et al. (2011), took the work of Epple and Platt (1998) further by considering explicitly job accessibility and the heterogeneity of the housing stock as determinants of residential choices. Interestingly, they find that the dispersion of jobs and the heterogeneity of the housing stock act as strong brakes on the tendency for households to segregate by race, education, and income. Although they do not explicitly consider the trade-off of the monocentric model and pay minimal attention to the income elasticity of the demand for land, they find that a reduction in commuting costs would lead to large increases in segregation. This is because lower commuting costs make access to jobs a less relevant consideration for households when choosing a residence.

Overall, the evidence indicates a lack of predictive ability of the monocentric model for patterns of household location in cities by income. In itself, this is not a failure of the monocentric model. First, the literature has yet to come up with reliable estimates of the income elasticity of the demand for land and the income elasticity of commuting costs. In cities where the majority of jobs are well outside the urban core, it is also empirically problematic to equate accessibility to jobs (or even general accessibility) to the distance to the city center. In other words, better work is needed to reduce uncertainty. Second, the best estimates at hand of the two key elasticities for the demand for land and commuting costs are fairly close. This suggests that the monocentric model predicts only weak income effects in location choices. The fact that this matches the empirical reality may not be a huge success, but at least it does not constitute a failure. Third, just like in the attempt to understand geographical patterns of housing prices and population density, the forces highlighted by the monocentric model are arguably not the only ones. This was true for the main housing price versus accessibility trade-off. This is even more so when we consider how this trade-off changes with income. What might be more worrying here is the limited success of the literature at explaining patterns of location choices by income using any explanation. There is good evidence about the effects of the housing stock but disturbingly little regarding public finance. Identifying the effects of local amenities also raises considerable challenges.

8.8. PATTERNS OF RESIDENTIAL LAND DEVELOPMENT

The patterns of residential development in cities have changed substantially over time. These changes are most frequently described as urban sprawl. The term "sprawl" is used with very different meanings, often confusing patterns of residential development characterized by low density and scatteredness with possible causes of such patterns (e.g., car-based commuting) and with possible consequences (e.g., loss of environmental quality). Nevertheless, the most commonly emphasized characteristics of urban sprawl are low-density spread-out development and scattered development (Galster et al., 2001). Figure 8.1 shows just how pervasive is leapfrog development even in an extremely dense city such as Paris. Even as centrally as between 5 and 10 km from the center, 26% of the land remains undeveloped, 18% in the form of urban parks but 8% as unbuilt natural spaces. Between 10 and 20 km from the center, parks take up 15% and unbuilt natural spaces take up 31% of the land.

A first important change in residential patterns has been the gradual displacement of residents away from city centers. A simple measure of the suburbanization of the population compares the population in central cities with the population in the suburban parts of a metropolitan area. Between 1940 and 2000, the share of residents in metropolitan areas of the United States who lived in the central city fell from 56% to 32% (Boustan and Shertzer, 2012). This decline was strongest in the 1940s, but continued over the second half of the twentieth century.

This displacement of the population toward the suburbs has been most often characterized in terms of flatter population density gradients. The work on density gradients beginning with Clark (1951), discussed in Section 8.6, finds that urban population density gradients have declined over time. According to Mills (1972), most US cities have seen their population density gradients decline since 1880. The sharpest decline occurred during the 1940s, but it has continued at a steady pace since then (Edmonston, 1975). The decline in density gradients has also been observed in many other countries, even if it has been faster and more pronounced in the United States (Mieszkowski and Mills, 1993).

Documenting patterns in the extent to which development is scattered or compact is more difficult. Burchfield et al. (2006) merge data based on high-altitude photographs from 1976 with data based on satellite images from 1992 to create a grid of 8.7 billion 30 m by 30 m cells that tracks the evolution of land use across the United States. These high-resolution data make it possible to observe the amount of open space in the neighborhood of houses in every US city. Thus, they can measure urban sprawl by calculating the amount of undeveloped land in the 1 km neighborhood around the average house in each city. Burchfield et al. (2006) find that residential development in 1992 was no more scattered than in 1976. For the United States as a whole, the proportion of open space in the square kilometer of land surrounding the average house was 42% in 1976 compared with 43% in 1992. While a substantial amount of scattered residential development was built between 1976 and 1992, overall residential development did not become any more biased toward such sprawling areas. On average, areas that were already densely built up in 1976 experienced little change, largely unbuilt areas in the neighborhood of earlier development saw some scattered development, while areas with initial scattered development had the highest rate of new development and became more densely built up in the process. As a result, the total amount of developed land grew substantially, but the proportions of sprawling and compact development remained substantially unchanged.

The nationwide finding that residential development in 1992 is no more scattered than development in 1976 also holds for most individual metropolitan areas. However, this stability over time coexists with large cross-sectional variation: there are very large persistent differences in the extent of sprawl across metropolitan areas. Angel et al. (2012) study the fragmentation of development in a global sample of 120 cities from around the world. In addition to the fragmentation index of Burchfield et al. (2006), they calculate four other measures helping characterize infill and leapfrog development. They find that the proportion of open space in the square kilometer of land surrounding the average development in cities in developed countries was 44% in 1990 and 39% in 2000, figures very similar to those reported by Burchfield et al. (2006) for the United States. Construction in cities from their world sample located in developing countries was more fragmented, with 48% and 43% of open space in the immediate square kilometer in 1990 and 2000, respectively.

Of the two aspects of sprawl we have discussed, falling densities is most easily related to the monocentric model. Lower density is equivalent to a larger amount of developed land for a given population in a city. Brueckner and Fansler (1983) study the comparative statics for the physical size of the city predicted by a closed-city version of the monocentric model. They measure the physical size of the city based on the surface area of urbanized areas in the United States as defined by the 1970 census. These are constructed by aggregating contiguous census blocks with high population density. For data availability reasons, they focus on 40 small urbanized areas contained within a single county. Their empirical analysis regresses the area of each of these 40 cities on the population, average household income, median agricultural land value per acre for the county, and two alternative proxies for commuting costs (the percentage of households owning one or more cars, which is thought to be higher where commuting costs are low, and the percentage of commuters using urban transit, which is thought to be lower where commuting costs are low). Cities are found to occupy more land when they host a larger population, when average income is higher, and when agricultural land is less valuable. All of these findings conform with the comparative statics of the closed-city version of the monocentric model (the open-city version treats the population as endogenous, but makes similar comparative-statics predictions for income and agricultural land rents). Neither of the two proxies for commuting costs is statistically significant. Several articles have replicated the analysis of Brueckner and Fansler (1983) with data for more cities or other time periods. McGrath (2005) studies large metropolitan areas between 1950 and 1990, while Paulsen (2012), in addition to studying all metropolitan areas regardless of the size for several time periods, also uses data on developed land derived from satellite imagery.

One possible reason why the proxies for commuting costs in Brueckner and Fansler (1983) do not perform well is endogeneity. In a city where residents own more cars, they may be able to travel longer distances more easily, and this can lead to more land being developed. However, it is also possible that the city expands for other reasons and residents buy more cars in response to the larger distances that must be traveled. Most other measures of commuting costs, such as the availability of roads, are subject to similar concerns. Baum-Snow (2007a) addresses these endogeneity concerns while studying another measure of sprawl aiming to capture suburbanization. We saw in Section 8.2.3 that the monocentric model predicts that lower commuting costs are associated with a lower central city population relative to the suburban population.[69] Baum-Snow regresses the change between 1950 and 1990 in the logarithm of the central city population on a measure of the change in commuting costs, while controlling for a change in the logarithm of the population for the entire metropolitan area and the radius of the central city. His measure of the change in commuting costs is the change in the number of rays of interstate highways that converge toward the central city. Of course, it is possible that the rays of interstate highway going through the city center are as much a consequence as a cause of suburbanization. To tackle this identification issue, Baum-Snow (2007a) uses as an instrument for actual highways crossing cities those that were part of the 1947 interstate highway

[69] We derived the result in the context of the open-city version of the monocentric model. Baum-Snow (2007a) uses a closed-city version of the monocentric model to motivate his analysis. In the closed-city version, where the population is constant, when a fall in commuting costs flattens the land and house price gradients, each resident consumes more housing and land. This expands the city boundary outward and also (unlike in the open-city version of the model with an endogenous population) reduces density. Suburbanization then follows from the relocation of some former central city residents to the suburbs.

plan. The instrument is relevant because planned rays of interstate highways are a strong predictor of rays that were actually built. At the same time, a reasonable case can be made for exogeneity because the 1947 highway planners were interested in linking US cities together but were not trying to address future commuting patterns. Baum-Snow (2007a) finds that an extra ray of interstate highways leads to a decline in the central city population of about 9%. Interestingly, this instrumental variable estimate is larger than its ordinary least squares counterpart, perhaps because more highways were built in cities that suburbanized less. Glaeser and Kahn (2004) emphasize that the prevalence of cars for commuting has not only reduced commuting costs but has also eliminated the larger fixed cost associated with public transport use, further promoting decentralization.

Baum-Snow's (2007a) finding of a relative decline in the central city population in response to lower commuting costs is consistent with the monocentric model. What is more difficult to reconcile with the monocentric model is the fact that the central city population fell not only in relative terms but also in absolute terms in the United States. Between 1950 and 1990, the population of central cities fell by an average 17%, while the total metropolitan area population rose by 72%. One possible explanation is the simultaneous economy-wide increase in income. In the monocentric city model, it follows from Equation (8.24) that a rise in income affecting all cities equally leaves their populations unchanged. By Equation (8.23), land rent at the CBD is also unchanged. Since land rent at the city edge must still equal the rent paid for land when it is allocated to the best alternative use, if housing is a normal good, the economy-wide increase in income simply makes the house-price gradient flatter. This is consistent with an absolute decline in the central city population and a rise in the suburban population, as residents drift outward to consume more land. Margo (1992) studies the contribution of rising incomes to suburbanization. He first estimates an individual-level regression of the probability of living in a suburban area in 1950 as a function of income while controlling for other variables. He then applies the economy-wide increase in average income to those individual-level estimates and suggests that rising incomes account for about 40% of the suburbanization of US population between and 1950 and 1980. Note, however, that this sort of exercise does not include the equilibrium effects in the housing market.

Kopecky and Suen (2010) simultaneously look at the influence of commuting costs and income on suburbanization. They calibrate a monocentric model, extended to allow for two forms of transport (cars and public transport), and perform counterfactual simulations. These suggest that the increasing affordability of cars together with the fall in the fixed and time costs of commuting by car were the major driver of population suburbanization in the United States between 1910 and 1950. However, between 1950 and 1970, rising incomes played a greater role.

In addition to falling commuting costs and rising incomes, a number of different explanations have been proposed for the suburbanization of the population in the United States. Cullen and Levitt (1999) show that increases in crime rates in the city and decreases in crime rates in the suburbs are associated with more residents leaving the city. They address the endogeneity of changes in crime rates by using as instruments for these lagged changes in the punitiveness of the state criminal justice system.

Boustan (2010) emphasizes another cause of suburbanization, which is specific to the United States: black immigration into northern cities from the South, followed by white flight toward the suburbs. Between 1940 and 1970, 4 million black migrants relocated from the South of the United States. Boustan (2010) estimates that each black arrival was associated with 2.7 white departures from central cities in the North. These may have been motivated by distaste for racial or income diversity. However, it is also possible that white departures had a different motive and, by lowering housing prices, encouraged black immigrants to settle in central cities. To address the endogeneity of black location choices, she uses as an instrument for actual black immigration the variation in local agricultural conditions in southern states to predict black emigration flows and assigns these flows to northern cities on the basis of settlement patterns from an earlier wave of black migration. She also argues that the fact that more than one white resident left for every black arrival indicates that white flight was not just a consequence of black immigration raising house prices. Baum-Snow and Lutz (2011) study the extent to which desegregation in public schools also contributed to suburbanization. They find that while desegregation had a very significant impact on sorting by race within cities, it generated only a small fraction of the decentralization of the urban population between 1960 and 1990.

Besides the decentralization of the residential population, another aspect of urban sprawl is the fragmentation of residential development. Burchfield et al. (2006) study the determinants of urban fragmentation by estimating the relationship between the percentage of open space in the square kilometer around an average new house built between 1976 and 1992 in each metropolitan area of the United States and a host of characteristics for each metropolitan area in 1976. The standard monocentric model can generate scattered development only through larger private open space (a lower capital intensity of development will result in shorter and smaller buildings on larger yards). This kind of development is more likely to take place in cities where commuting costs are lower because it is easier to get around by car. Burchfield et al. (2006) find that development is more scattered in cities built around the automobile (in particular those where streetcars were not widely used in the early twentieth century).

To generate scattered urban expansion in the sense of leapfrog development, leaving some parcels undeveloped while building up other parcels further away, we must either turn to dynamic models as covered in Section 8.3.3 or allow open space to provide some consumption value as in Turner (2005). Regarding dynamic models, we have seen that in the presence of long development lags, greater uncertainty can favor leapfrog development (Bar-Ilan and Strange, 1996). Since developers will likely see future local population growth as more uncertain in cities that have had more ups and downs in population growth rates over previous decades, Burchfield et al. (2006) use as a proxy for uncertainty the standard deviation of the decennial percentage population growth rates for 1920–1970. As expected, higher uncertainty leads to more sprawl. The value of open space will be greater if open space is likely to remain undeveloped for a long time and if local conditions, such as climate, make being outdoors more enjoyable. Burchfield et al. (2006) show that cities that have been growing more slowly in the past, where undeveloped land is less likely to be converted soon, have more scattered development. This is true even after controlling for contemporaneous population growth rates. Also, cities with a more temperate climate have more scattered development. As we discuss in the following section, employment is often much less centralized than what is assumed in the standard monocentric model. The models of endogenously polycentric cities reviewed in Section 8.4 tell us that more localized agglomeration spillovers will tend to endogenously produce an urban structure with employment centralized in a CBD, as in the monocentric model. Such cities, with strong agglomeration economies, will pay higher wages but also have higher housing and land prices. This will encourage more capital-intensive development. Cities specializing in sectors that, economy-wide, tend to be more concentrated close to city centers indeed tend to have more compact development.

Most models assume a homogeneous landscape. However, natural geography matters for spatial development patterns. Burchfield et al. (2006) examine several aspects of heterogeneity in the natural environment. Mountains are an important aspect of geography. These have been prominent, for instance, in Los Angeles, where the mountains bordering the city have limited further expansion of its sprawling suburbs (a situation described locally as "sprawl hits the wall"). However, in studying the effect on sprawl of mountains more generally, one needs to be careful with two features. First, it is important to focus on mountains in the vicinity of earlier development where they truly act as a barrier to further expansion. Second, one needs to be careful to separate large-scale from small-scale terrain irregularities. Mountains and hills tend to have opposite effects. When an expanding city hits a mountain range, further scattered development in the urban fringe becomes very costly. Thus, high mountains in the urban fringe encourage infilling and lead to increasingly compact residential patterns. On the other hand, small-scale irregularities in the urban fringe have the opposite effect. This is perhaps because rugged terrain encourages scattered development as steep hillsides, where development is more costly, alternate with flat portions, where development is less costly.

Another physical feature with important effects on sprawl is aquifers. Most households in the United States get their water through the nearest municipal or county water supply. However, extending water systems to service new scattered development in the urban fringe requires substantial infrastructure investments, the cost of which is typically borne by developers through connection fees and is ultimately reflected in housing prices. In places where water-yielding aquifers are pervasive, developers can instead sink a well at a fraction of the cost of connecting to the municipal or county water supply. The presence of aquifers is a particularly interesting dimension of underlying heterogeneity in the physical landscape because of the way it interacts with scale economies: wherever aquifers underlie the urban fringe, household water can be obtained without the large increasing returns associated with public water systems, and this facilitates scattered development.

A final set of determinants of scattered development are of political nature. Two of the main dimensions emphasized in public discussions—competition between municipalities of different sizes and the extent of municipal fragmentation—do not matter for sprawl in practice. Municipal boundaries matter, but for different reasons. Burchfield et al. (2006) find that a disproportionate share of 1976–1992 development happened in unincorporated areas that were close to existing development but just beyond the municipal boundaries at the beginning of the period, where land use regulation was laxer. This development is also more dispersed than that on incorporated land. Finally, one of the common complaints about urban sprawl is that as development spreads, municipal services such as roads, sewers, the police, and fire protection are more expensive. Indeed, when a smaller fraction of local expenditures is financed through transfers from other levels of government as opposed to local taxes, development tends to be less scattered. This suggests that when local taxpayers are held accountable for infrastructure costs, they respond by insisting on more compact patterns of development that require less infrastructure spending.

In their study of fragmentation in a global sample of cities, Angel et al. (2012) find that cities that are smaller and had more buildable land around them have more fragmented development, which is consistent with an initial phase where a city grows in a fragmented way, followed by gradual infilling of undeveloped land as the city continues to grow. Like Burchfield et al. (2006) for the United States, they also find that more geographical constraints lead to more compact development, while aquifers providing well water encourage fragmentation. One of their more interesting findings is that cities in countries with greater car ownership have more compact development, after income differences have been controlled for. Greater car ownership can facilitate commuting over long distances, which would seem to favor leapfrog development. However, public transport relies on stations and bus stops and is likely to encourage development to be concentrated around these. By enabling easy travel between any two points, car ownership can in fact favor infilling. Angel et al. (2012) suggest that their result arises because the latter effect dominates.

If urban sprawl has received so much attention in the popular press, it is mostly because it is presumed to have important consequences. Most prominently, sprawling residential development is presumed to encourage car use. Bento et al. (2005) study the effects of various measures of urban form and centralization on transport mode choices and distance traveled. They find that individual measures of urban form have only a modest impact on the probability of commuting by car and distance traveled. However, when multiple aspects of urban form are considered simultaneously, the effects are much larger. For instance, they calculate that changing Atlanta to look like Boston in terms of the central location of its population, the shape of the city, and the availability of public transport would reduce the total distance traveled by car by 25%. Glaeser and Kahn (2004) suggest that by making commutes more car dependent, sprawling cities may hurt poorer households, who are much less likely to be able to own a car for every working adult.

In addition to favoring car use, sprawl is also sometimes thought to reduce social interactions (Putnam, 2000) and to encourage unhealthy behavior increasing the prevalence of obesity (Ewing et al., 2003; Giles-Corti et al., 2003; Saelens et al., 2003; Frank et al., 2004). Regarding social interactions, empirical studies show that social interactions are, in fact, greater in suburban areas and in areas with lower residential density (Glaeser and Gottlieb, 2006; Brueckner and Largey, 2008). Regarding obesity, people living in sprawling neighborhoods tend to be heavier than those living in neighborhoods where development is compact and there are plenty of shops and amenities within walking distance. However, using panel data tracking the characteristics of people's residential locations and their weight over time, Eid et al. (2008) show that this correlation between sprawl and obesity is fully accounted for by sorting. Once they take advantage of the panel dimension of their data to control for unobserved propensity to be obese, the correlation between obesity and sprawl vanishes: changes in neighborhood characteristics do not lead to changes in weight. People in sprawling neighborhoods are heavier because the same characteristics that make them obese (e.g., a distaste for walking) make them prefer to live in sprawling neighborhoods (e.g., so that they can get around by car and not walk).

8.9. EMPLOYMENT DECENTRALIZATION AND PATTERNS OF BUSINESS LOCATION CHANGES WITHIN CITIES

Section 8.8 discussed the causes and manifestations of residential decentralization in cities. Decentralization, however, did not only affect patterns of residential locations. It was also followed and accompanied by the decentralization of employment and economic activity.

As already mentioned, Glaeser and Kahn (2001) report that only 24% of jobs in US metropolitan areas in 1996 were located within 3 miles (about 5 km) of the CBD. Consistent with this, Baum-Snow (2014) reports that about 24% of workers in 2000 worked in the central city of one of the largest 100 metropolitan areas in the United States. Second, although residential decentralization has been stronger (Glaeser and Kahn, 2001), the decentralization of jobs has been considerable. Baum-Snow (2014) reports that for the 100 largest US metropolitan areas, the employment share of central cities went from 61% in 1960 to 34% in 2000. Third, the decentralization of jobs started after the decentralization of residents, but the high correlation between employment and residential decentralizations across cities makes it difficult to separate the two. Fourth, since jobs were initially more centralized than residents, they have become even more centralized in relative terms.[70] Fifth, although job decentralization affected most sectors, it has been relatively stronger for manufacturing than for services (Baum-Snow, 2014), and less-skilled jobs have decentralized more than high-skilled and managerial jobs (Rossi-Hansberg et al., 2009). Finally, there is also considerable heterogeneity in the patterns of employment decentralization. Some cities such as Los Angeles have become extremely decentralized, whereas others such as New York have been far less affected.[71]

While these broad facts are well established, little is known about the details of the spatial patterns of decentralized employment. As reported above, there is evidence of the emergence of new subcenters. Following a pattern initially and vividly described by Garreau (1991) and, more formally, by Henderson and Mitra (1996), a fraction of CBD employment went to subcenters that, as a first approximation, may be viewed as miniature replicas of the main CBD.

Using the subcenter detection method proposed by McMillen (2001), McMillen and Smith (2003) examine the existence of subcenters in large US cities inspired by the model of Fujita and Ogawa (1982) discussed in Section 8.4. For 62 large American metropolitan areas with a mean population of 2.1 million, they detect an average of 2.7–4.4 subcenters per metropolitan area depending on the stringency that they apply to the detection of subcenters. Importantly, they find that cities with more population and a higher travel time index have more subcenters. These two variables alone account for 80% of the variation in the number of subcenters in their sample.[72]

Urban employment decentralization is not only about the emergence of subcenters. Employment has also decentralized in a "diffuse" manner. In fact, subcenter formation and diffuse employment decentralization should not be seen as a binary dichotomy. Reality is about a continuum ranging from small isolated facilities, to groups of several offices in a strip-mall, to small industrial parks with a couple hundred workers, to full-fledged business subcenters with tens of thousands of employees. Although diffuse employment is hard to define precisely and measure empirically, casual evidence suggests it perhaps accounts for a large part of decentralized employment.

As argued in Section 8.4, firms face a complicated trade-off when they choose to decentralize. The equilibrium layout of the city in Ogawa and Fujita (1980), as captured

[70] Put differently, the gradients of employment density were initially steeper than those of residential density. Both types of gradients have become much flatter. Since residential density has flattened more, employment gradients have become relatively steeper.

[71] The findings of Glaeser and Kahn (2001) are also suggestive of a role of institutional fragmentation with more decentralization in more fragmented metropolitan areas, a weak role of city age, with more recent cities being more decentralized, and no role for city demographics.

[72] While these results are interesting, they should be taken as a description of the data but not be given a causal interpretation. The number of subcenters is arguably determined simultaneously with the metropolitan population and with a travel time index. Appropriate identification of causal effects is still missing here.

in Equations (8.69) and (8.70), features three key parameters: the spatial decay of interactions, the intensity of land use in production, and commuting costs. By moving away from central locations, firms lose some agglomeration benefits, the more so the steeper the spatial decay of interactions. At the same time, commercial real estate in central locations is more expensive. Hence, by moving away from the center, firms lower their land costs. This represents a larger saving the higher the land intensity of production. Finally, if firms concentrate in a CBD, they force their workers to commute to this center instead of working locally. To offset those commuting costs, firms must pay higher wages if they are centrally located. This complicated trade-off is summarized by the bid-rent curve for land in commercial use. If agglomeration economies are very local, production is not very land intensive, and commuting costs are not too high, the bid-rent curve for commercial land features a central peak putting it above the bid-rent curve for land in residential use, as in Figure 8.5. If agglomeration economies spread out more easily, production is more land intensive, or commuting is more costly, firms become decentralized, first through the emergence of subcenters as in Figure 8.4 and then through the diffusion of employment throughout the city.

The model of Ogawa and Fujita (1980) features a single production sector. However, its comparative statics are informative of which sectors are more likely to become decentralized: those with less localized agglomeration economies and those that use land less intensively. Thus, manufacturing plants would be among the first to spread out, while, for instance, financial services are likely to remain concentrated in the CBD. Commuting costs are likely to matter more for comparisons across cities and countries than across sectors, with lower commuting costs favoring decentralization. Decentralization raises further accessibility issues when the new location is difficult to access with public transport. In this respect, we expect major differences between North America and the rest of the world, where the car is less prevalent in commutes. Job decentralization, especially in its more diffused forms, is arguably much easier when the vast majority of workers commute by car.

The standard monocentric model characterize a number of important gradients. However, by assuming a single central location for all firms, it features a single wage for each occupation. As we saw in Section 8.4, once we allow for the endogenous decentralization of jobs, an immediate new feature is the existence of wage gradients. Early evidence consistent with the existence of wage gradients was provided by Madden (1985) and McMillen and Singell (1992). It is summarized in Anas et al. (1998). As noted by Manning (2003), extant work on wage gradients mostly establishes that workers with longer commutes receive higher wages. This is consistent with decentralized firms offering lower wages to their workers and capturing the gains from better access. However, in a monocentric framework, a positive correlation between wages and commuting distance is also consistent with the workers that receive higher wages choosing to live further from the CBD. As shown in Section 8.3.2, this is a standard prediction of the monocentric model extended to multiple income groups. Finally, as underscored by Manning (2003), a positive correlation between commuting distance and wages is also consistent with spatially dispersed firms having some monopsony power. With this type of labor market friction, employers need to offer higher wages to workers who commute from further away irrespective of their location.[73] Manning (2003) provides a range of evidence consistent with the latter feature.

Rather than examine the wage received by workers depending on where they live, evidence for wage gradients should be sought using information about the wages offered by employers depending on how far they are from their workforce and how their workers are distributed in the city. Since the location choices of both employers and workers are likely to be simultaneously determined, this is a challenging problem. Even with a plausible identification strategy, getting significant results will be difficult. The models with decentralized employment we have examined assume that, all else being equal, workers commute to the closest workplace. In reality, workers do not go to the closest job. They commute much more, perhaps a full order of magnitude more than indicated by the work-residence assignment that would minimize aggregate commuting. This fact is usually referred to as "wasteful" or "excess" commuting and was originally evidenced by Hamilton (1982) (see Ma and Banister, 2006, for a more recent survey). The adjective "wasteful" is probably misleading since there are many reasons why workers do not work for the employer closest to their residence. These reasons include preferences for specific local amenities, schools in particular, the fact that labor is far from being a homogeneous commodity, more complex location optimization for dual earners, and the costs associated with changing residence after changing job. Interestingly and despite strong job decentralization, the commuting time of workers has remained fairly stable (Gordon et al., 1991; Levinson and Wu., 2005), a fact dubbed the "commuting time paradox." Proper integration of these features into a consistent theoretical framework that would form the basis for subsequent empirical work is still missing.

Aside from wage gradients, a second prediction of employment decentralization is a flattening of land and housing price gradients as employment decentralizes. As already mentioned, McMillen (1996) documents a flattening of the land price gradient for Chicago over a long period. Further evidence about declining gradients is discussed in Anas et al. (1998). The issue with declining gradients is that they could also be caused by lower commuting costs in the absence of decentralization.

Another implication of employment decentralization is that greater job dispersion will make it harder for some workers to access jobs. In particular, diffuse job decentralization makes it harder for workers with no car to reach potential employers once these

[73] If the labor market for each firm is thin enough, its workers will be widely dispersed across the city. Then, it could even be that firms further away from the CBD need to offer, on average, higher wages because of worse accessibility. This would then imply an inverse wage gradient.

are located far from the CBD. This is the spatial mismatch hypothesis, first proposed by Kain (1968). According to this hypothesis, in the US context, the departure of jobs from central cities made it much harder for minorities to find employment since these minorities have traditionally resided in the more central parts of cities. Whether this increased disconnect between residence and jobs is a major explanatory factor for the plight of minorities in the United States is still an open question. There is no doubt that, as reported above, a decrease in the employment share of central cities from 57% to 47% over 20 years is a major change. These figures probably understate the true extent of the change for minority workers since, as already stated, low-skilled manufacturing jobs decentralized much more than high-skilled jobs, which in many cases grew in central cities (Gobillon et al., 2007; Rossi-Hansberg et al., 2009). Given that minority residents are overwhelmingly over represented in central cities, their access to jobs diminished, perhaps considerably so.

A key complication in the empirical analysis of the spatial mismatch hypothesis is that a variety of mechanisms are likely to be at play. As noted above, greater distance may reduce both the efficiency of the search for jobs and its intensity. Poor accessibility may also make workers less likely to accept job offers since the wage net of commuting costs is more likely to fall below some reservation level. Finally, minority workers may also suffer from various forms of discrimination in predominantly white suburban areas ranging from customer-driven discrimination in service jobs to redlining and statistical discrimination afflicting workers coming from deprived central areas with high crime rates.[74]

Because of this variety of mechanisms, it is hard to get a sense of the importance of the spatial mismatch hypothesis from the studies that examine individual channels.[75] Some empirical studies of the spatial mismatch hypothesis have taken a more aggregated approach. In one of the most careful explorations in this literature, Raphael (1998) estimates two sets of regressions for a cross section of neighborhoods in the San Francisco metropolitan area. In the first, he regresses neighborhood labor market participation rates on measures of job accessibility and other neighborhood characteristics as controls. The key finding is that the coefficient on the share of black residents drops after accessibility has been controlled for. Hence, much of the lower employment rate of blacks in San Francisco neighborhoods appears to be accounted for by the poor accessibility of these neighborhoods. To confirm this, Raphael (1998), in the second set of regressions, uses employment rates across neighborhoods, but this time, by race. His decomposition of racial employment rate differentials also suggests an important role for accessibility. Overall, he concludes that about 20% of the employment rate differentials can be attributed to differences in job accessibility. The key identifying assumption is that residential choices are exogenous. To minimize problems, he looks only at young workers, below the age of

[74] See Zenou (2009) for a theoretical exposition of these various channels.

[75] The various threads of the literature about specific channels are discussed in Gobillon et al. (2007).

19 years, who disproportionately live with their parents. This, however, may not help much with the main identification challenge if employability of youths is highly correlated with the employability of their parents, who choose their residential location on that basis.

Even if we ignore identification problems, it is unclear what this type of result implies in terms of policy. One might want to bring the residents to the jobs, bring the jobs to the residents, or help central city residents access suburban jobs. While a full discussion of the evidence on those issues would take this chapter too far, note that the literature on place-based policies is often skeptical about bringing jobs to central city residents as argued in Chapter 18 in this handbook. The results from the rare cases where residents from poor neighborhoods were relocated to richer areas are not supportive of large employment benefits either (see, e.g., Katz et al., 2001). Improving access might be more promising provided major transit investments with poor returns are avoided. In this respect, subsidizing private transport is sometimes suggested (Glaeser and Kahn, 2004), but this possibility does not receive much interest from policy makers.

Turning to the causes of employment decentralization, several conjectures have been proposed in the literature. As with residential decentralization discussed in Section 8.8, changes in transport obviously loom large (Anas et al., 1998; Glaeser and Kahn, 2004; Baum-Snow, 2007a). Most extant empirical work focuses on the effect of transport on residential decentralization. This work is discussed in Section 8.8. Baum-Snow et al. (2013) study the effect of transport networks on both the decentralization of the population and the decentralization of specific types of economic activity in China. They use historical transport networks as an exogenous source of variation. They find that railroads and roads had a strong effect on population decentralization. Production in industries with high weight-to-value ratios did not respond to transport improvements, while production in lighter industries did. Baum-Snow (2014) studies residential and employment decentralization for the United States. He uses some plausible exogeneity associated with the construction of US interstate highways to identify the effect of new radial roads on both employment and residential decentralization. He finds that new roads had a much larger effect on residential decentralization and uses a model to back out the importance of agglomeration effects that retain firms in the more central locations of cities.

In addition, the same social ills that have pushed many residents outside central cities in the United States may also have been responsible for the departure of businesses. These ills were briefly discussed in Section 8.8.

There is a factor that is nonetheless uniquely important to firms: advances in communication and computing technologies have made it much easier for firms to separate their activities across various sites. In particular, firms have been able to separate administrative functions from production, and within administrative functions high-end and front-office activities from back-office work. Such separation is expected to occur when the cost of separating activities and the possible loss of agglomeration for the separated activity is more than offset by lower production costs in a cheaper location. A large share of these separations involves relocating back-office work or production to different countries (Markusen, 2002) and different cities (Duranton and Puga, 2005; Davis and Henderson, 2008; Liao, 2012). In many cases however, close coordination between activities is still needed and production or back-office labor moves to the fringe of the city, while management remains in the CBD.

This hypothesis was first articulated by Ota and Fujita (1993) and later by Rossi-Hansberg et al. (2009). It is also discussed in Anas et al. (1998). Unfortunately, there is no empirical work that goes beyond documenting broad descriptive facts to substantiate the importance of the telecommunication revolution as an explanatory factor for the decentralization of employment.

Unlike what the material in Section 8.8 and in this section may implicitly suggest, cities are not only about change. Much of the recent literature actually suggests there is a lot of persistence in urban patterns. Bleakley and Lin (2012) document the persistence and continuous development of many cities in the United States located along a fall line. These cities initially specialized in the portage of merchandise traveling by boat but remained and prospered even after their main economic justification disappeared. In a different vein, Davis and Weinstein (2008) show that Japanese cities recovered in population terms and redeveloped the same economic specializations after World War II even though many of those cities were heavily bombed.

In an interesting exploration of the evolution of subcenters in Los Angeles, Redfearn (2009) shows that the most important predictor of clusters of decentralized jobs in 2000 in a location is the presence of decentralized jobs in the same location in 1980. This strong stability in the location patterns of firms is surprising against a backdrop of a population growth of more than 25% during the period, ethnic change, and strong deindustrialization. Even more interesting, as noted by Redfearn (2009), old highway networks are more relevant in explaining the current spatial distribution of employment than is the modern highway system.

Brooks and Lutz (2013) present related evidence also for Los Angeles. They examine to what extent streetcars, whose importance culminated in the 1910s and which were completely gone by the early 1960s, still affect land use patterns. They document strong negative and persistent correlations between the distance to old streetcar tracks and contemporaneous population density or property prices. These correlations became stronger over time and hold even after conditioning out contemporary transit and measures of locational quality.

This suggests an important role for durable housing and persistence in zoning patterns in explaining why urban decentralization did not change cities even more than it did. Further indirect evidence supportive of this idea is proposed by Siodla (2014), who documents changes in the city of San Francisco after the fires that followed the 1906 earthquake. These fires destroyed large parts of the city. He shows that, at the border of fires, the density subsequently increased by 40% on the side that burnt down relative to the side that did not burn.

The bottom line of any conclusion on residential and employment decentralization is that much remains to be done. A first set of priorities is to document facts. For employment decentralization, the best account is still arguably the one provided by Glaeser and Kahn (2001). It is nearly 15 years old and much of it relies on county-level information. With lots of more data available at much finer spatial scales, a more precise and up-to-date documentation of the patterns of employment decentralization should be within reach. It is also disconcerting that the overwhelming majority of the little we know about employment decentralization concerns only one country, the United States, which is arguably an outlier.

While the drivers of residential decentralization have begun to be explored, much remains to be done. There is very little on employment decentralization and on the relationship between employment and residential decentralization. While the exploration of some drivers of urban decentralization will benefit from clever identification strategies relying on transport innovations or exogeneity in institutional change, some broader issues will probably need to rely on modeling assumptions to be able to back out key parameters from observable moments of the data. There is an emergent body of work in that direction. One could cite the articles by Ahlfeldt et al. (2012), Baum-Snow (2014), or Brinkman et al. (2012) as examples of the current frontier. A lot of the evidence also points at the persistence created by the durability of commercial and residential properties and the persistence of zoning. A better understanding of how houses and buildings are built, maintained, and torn down and how zoning regulations affect this process are clearly important avenues for future research.[76]

8.10. CONCLUSION

This chapter aimed to selectively review the large literature on urban land use. The theoretical literature is well developed and has provided numerous important insights. It mostly builds on the monocentric structure initially developed by Alonso (1964), Mills (1967), and Muth (1969). Even in the simple version presented in Section 8.2, the monocentric model is an outstanding piece of theory which delivers subtle and nontrivial results from the interplay of fairly straightforward and natural assumptions. Higher housing prices in some locations arise to offset better accessibility. In turn, higher housing prices are reflected in higher land prices, and affect housing consumption, the intensity of housing development, and the number and characteristics of residents locally.

The monocentric model faces three key challenges. The first is heterogeneity of users of land. While, as seen in Section 8.3, much attention has been devoted to broad trends

[76] A detailed review of the literature on this is beyond our scope here. About the production function for single family homes, we already mentioned Epple et al. (2010b) and Combes et al. (2014). See also Albouy and Ehrlich (2012). For larger buildings and differentials in construction costs, see Glaeser et al. (2005) and Gyourko and Saiz (2006). Finally, see Dye and McMillen (2007) for tear-downs.

such as the average income of residents in relation to the distance to the CBD, much has less effort has been devoted to understanding heterogeneity in parcel size, housing, and local residents in different locations. Existing models tend to predict extreme forms of sorting, while actual patterns of land use tend to be smoother. Understanding what drives this local heterogeneity and its implications is still part of the agenda for future research. With much recent progress elsewhere in the discipline in dealing with heterogeneity, this is a task that seems feasible. The second main challenge is to accommodate more fully the durability of housing. Here again, tractability issues loom large, although progress in computational economics is likely to help. The third modeling challenge seems much harder. For more than 30 years urban economists have struggled to accommodate more complex spatial structures that involve the endogenous formation of centers and patterns of mixed land use. As we showed in Section 8.4, these models are both fragile and difficult to manipulate.

Relative to theory, empirical work is less advanced. The first problem has been one of data availability. For a long time, the best empirical research could do was to combine aggregate statistics for urban tracts with maps and attempt to uncover some patterns predicted by theory, usually some sort of gradient. The data environment has changed dramatically in the last 15 years. It went from being extremely data poor to incredibly data rich. Landsat data recently released by the US Geological Survey offer land cover information for the entire world for pixels of 30 m by 30 m since the mid-1970s. This represents more than 500 billion observations for a single cross section.

Better data will certainly help produce better and more informative descriptive work. For many first-order questions, we still lack basic facts. Very little is known about how land is being used in cities. Models often assume that all noncentral land is residential, but commercial land, open space, and roads and parking may use three-quarters or more of the land of contemporary cities. As highlighted in Sections 8.8 and 8.9, much of our knowledge of urban decentralization and urban sprawl in the United States derives from a tiny number of articles. There is a wealth of detailed studies about small areas but very little regarding broad cross sections of cities and even less about their evolution. Also, the vast majority of what is known about land use concerns American cities. Much less is known about other cities from the developed world, although they appear to differ greatly from American cities in a number of aspects. Our knowledge of land use in developing cities is even more rudimentary.

However, data availability is not and should not be everything. The second major problem of empirical investigations of land use in cities is that, although they are often inspired by theory, often they are only loosely connected to the models they claim to draw from. For instance, empirical predictions of the monocentric model are examined qualitatively and not quantitatively. Gradients of property prices are then exhibited but without much notion of whether they are quantitatively consistent with the models. We certainly call for a tighter connection between theory and empirical work in future research.

Third, identification concerns are often ignored. To return to the example of property price gradients, alternative explanations for these gradients not based on the accessibility-price trade-off are usually ignored. Hence, although we would like theory to be taken more seriously in empirical work, the limitations of extant theory and the empirical implications of these limitations also need to be better understood and recognized. Hopefully this chapter will be helpful in this respect.

Fourth, land use is often studied in isolation. While land use is of clear interest on its own, it should also be studied in relation to other issues. The most obvious connection is with transport. Although the land use and transport nexus is at the core of the monocentric model, most of the empirical work on land use ignores transport or treats it superficially with low-quality data such as ad hoc measures of congestion. As made clear by Chapter 20 in this handbook and in Sections 8.8 and 8.9, recent advances in the transport literature are starting to provide useful insights into the drivers of land use. Land use and changes in land use are also expected to reflect technological change since the location choices of firms and households depend on existing technologies and the availability of substitutes for travel. Aside from some broad trends relating to the importance of the automobile and changes in what firms do and how they are organized, our knowledge here is still extremely sparse. As shown by Chapter 13 in this handbook, urban and real estate research is developing a more detailed and sophisticated understanding of the land and property markets. However, the spatial and land use implications of this better understanding of how sales of land and properties are transacted are still mostly unexplored.

To conclude, while in this chapter we have highlighted the deficiencies of extant theory, we believe that the main priority for future research on urban land use should be empirical work.

ACKNOWLEDGMENTS

We thank Frédéric Robert-Nicoud for his early comments and several novel insights and Frédéric Gilli for his help with the data. We are also grateful to Will Strange, Vernon Henderson, Dan O'Flaherty, and Steve Ross for their detailed comments. Gilles Duranton gratefully acknowledges funding from the Zell Lurie Real Estate Center at Wharton. Diego Puga gratefully acknowledges funding from the European Union's Seventh Framework Programme under ERC Advanced Grant agreement 269869.

REFERENCES

Ahlfeldt, G., 2011. If Alonso was right: modeling accessibility and explaining the residential land gradient. J. Reg. Sci. 51, 318–338.

Ahlfeldt, G.M., Redding, S.J., Sturm, D.M., Wolf, N., 2012. The Economics of Density: Evidence from the Berlin Wall. Processed, London School of Economics.

Albouy, D., Ehrlich, G., 2012. Metropolitan land values and housing productivity. Working paper 18110, National Bureau of Economic Research.

Alonso, W., 1964. Location and Land Use; Toward a General Theory of Land Rent. Harvard University Press, Cambridge, MA.

Anas, A., 1978. Dynamics or urban residential growth. J. Urban Econ. 5, 66–87.
Anas, A., 1990. Taste heterogeneity and urban spatial structure: the logitmodel and monocentric theory reconciled. J. Urban Econ. 28, 318–335.
Anas, A., Moses, L.N., 1979. Mode choice, transport structure and urban land use. J. Urban Econ. 6, 228–246.
Anas, A., Arnott, R., Small, K.A., 1998. Urban spatial structure. J. Econ. Liter. 36, 1426–1464.
Anenberg, E., Bayer, P., 2013. Endogenous Sources of Volatility in Housing Markets: The Joint Buyer-Seller Problem. Processed, Duke University.
Angel, S., Parent, J., Civco, D.L., 2012. The fragmentation of urban landscapes: global evidence of a key attribute of the spatial structure of cities, 1990–2000. Environ. Urban. 24, 249–283.
Arnott, R., Inci, E., 2006. An integrated model of downtown parking and traffic congestion. J. Urban Econ. 60, 418–442.
Arnott, R.J., Stiglitz, J.E., 1981. Aggregate land rents and aggregate transport costs. Econ. J. 91, 331–347.
Arnott, R., Davidson, R., Pines, D., 1983. Housing quality, maintenance and rehabilitation. Rev. Econ. Stud. 50, 467–494.
Asami, Y., Ohtaki, T., 2000. Prediction of the shape of detached houses on residential lots. Envir. Plann. B 27, 283–296.
Asami, Y., Fujita, M., Smith, T.E., 1991. On the foundations of land use theory: discrete versus continuous populations. Reg. Sci. Urban Econ. 20, 473–508.
Bar-Ilan, A., Strange, W.C., 1996. Urban development with lags. J. Urban Econ. 39, 87–113.
Baum-Snow, N., 2007a. Did highways cause suburbanization? Q. J. Econ. 122, 775–805.
Baum-Snow, N., 2007b. Suburbanization and transportation in themonocentric model. J. Urban Econ. 62, 405–423.
Baum-Snow, N., 2014. Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies. Processed, Brown University.
Baum-Snow, N., Lutz, B.F., 2011. School desegregation, school choice and changes in residential location patterns by race. Am. Econ. Rev. 101, 3019–3046.
Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2013. Roads, Railroads and Decentralization of Chinese Cities. Processed, Brown University.
Bayer, P., McMillan, R., 2012. Tiebout sorting and neighborhood stratification. J. Publ. Econ. 96, 1129–1143.
Bayer, P., McMillan, R., Rueben, K., 2011. An Equilibrium Model of Sorting in an Urban Housing Market. Processed, Duke University.
Becker, G.S., 1965. A theory of the allocation of time. Econ. J. 75, 493–517.
Beckmann, M.J., 1969. Distribution of urban rent and residential density. J. Econ. Theory 1, 60–67.
Behrens, K., Combes, P.P., Duranton, G., Gobillon, L., Robert-Nicoud, F., 2015. Household Sorting in the City. University of Pennsylvania, Wharton School.
Bento, A., Cropper, M.L., Mobarak, A.M., Vinha, K., 2005. The effects of urban spatial structure on travel demand in the United States. Rev. Econ. Stat. 87, 466–478.
Berliant, M., 1985. Equilibrium models with land: a criticism and an alternative. Reg. Sci. Urban Econ. 15, 325–340.
Bertaud, A., Malpezzi, S., 2003. The Spatial Distribution of Population in 48 World Cities: Implications for Economies in Transition. Processed, The Center for Urban Land Economics Research, University of Wisconsin.
Bertaud, A., Renaud, B., 1997. Socialists cities without land markets. J. Urban Econ. 41, 137–151.
Bleakley, H., Lin, J., 2012. Portage and path dependence. Q. J. Econ. 127, 587–644.
Boustan, L.P., 2010. Was postwar suburbanization "white flight"? Evidence from the black migration. Q. J. Econ. 125, 417–443.
Boustan, L., Shertzer, A., 2012. Population trends as a counterweight to central city decline, 1950–2000. Demography 50, 125–147.
Braid, R.M., 1981. The short-run comparative statics of a rental housingmarket. J. Urban Econ. 10, 286–310.

Braid, R.M., 2001. Spatial growth and redevelopment with perfect foresight and durable housing. J. Urban Econ. 49, 425–452.
Breton, A., 1973. Neighborhood selection and zoning. In: Hochman, H. (Ed.), Issues in Urban Public Economics. Institut International de Finance Publique, Saarbrücken, pp. 241–251.
Brinkman, J., 2013. Congestion, Agglomeration, and the Structure of Cities. Processed, Federal Reserve Bank of Philadelphia.
Brinkman, J., Coen-Pirani, D., Sieg, H., 2012. Estimating a Dynamic Equilibrium Model of Firm Location Choices in an Urban Economy. Processed, Federal Reserve Bank of Philadelphia.
Brooks, L., Lutz, B., 2012. From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly. Processed, Trachtenberg School of Public Policy and Public Administration, George Washington University.
Brooks, L., Lutz, B., 2013. Vestiges of Transit: Urban Persistence at a Micro Scale. Processed, Trachtenberg School of Public Policy and Public Administration, GeorgeWashington University.
Brueckner, J.K., 1980. A vintage model of urban growth. J. Urban Econ. 8, 389–402.
Brueckner, J.K., 1981. A dynamic model of housing production. J. Urban Econ. 10, 1–14.
Brueckner, J.K., 1986. A switching regression analysis of urban population densities. J. Urban Econ. 19, 168–198.
Brueckner, J.K., 1987. The structure of urban equilibria: A unified treatment of the Muth-Mills model. In: Handbook of Regional and Urban Economics, vol. 2. Elsevier, North-Holland, pp. 821–845.
Brueckner, J.K., 2000. Urban growth models with durable housing: An overview. In: Huriot, J.M., Thisse, J.F. (Eds.), Economics of Cities: Theoretical Perspectives. Cambridge University Press, Cambridge, pp. 263–289.
Brueckner, J.K., Fansler, D.A., 1983. The economics of urban sprawl: theory and evidence on the spatial sizes of cities. Rev. Econ. Stat. 65, 479–482.
Brueckner, J.K., Largey, A.G., 2008. Social interaction and urban sprawl. J. Urban Econ. 64, 18–34.
Brueckner, J.K., Rosenthal, S.S., 2009. Gentrification and neighborhood housing cycles: will America's future downtowns be rich? Rev. Econ. Stat. 91, 725–743.
Brueckner, J.K., Thisse, J.F., Zenou, Y., 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. Eur. Econ. Rev. 43, 91–107.
Brueckner, J.K., Thisse, J.F., Zenou, Y., 2002. Local labor markets, job matching, and urban location. Int. Econ. Rev. 43, 155–171.
Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2006. Causes of sprawl: a portrait from space. Q. J. Econ. 121, 587–633.
Capozza, D.R., Helsley, R.W., 1989. The fundamentals of land prices and urban growth. J. Urban Econ. 26, 295–306.
Capozza, D.R., Helsley, R.W., 1990. The stochastic city. J. Urban Econ. 28, 187–203.
Carrillo, P.E., 2012. An empirical stationary equilibrium search model of the housing market. Int. Econ. Rev. 53, 203–234.
Chatterjee, S., Eyigungor, B., 2014. Agglomeration Economies, Geography and the Value of Urban Land. Processed, Federal Reserve Bank of Philadelphia.
Cheshire, P., Sheppard, S., 1995. On the price of land and the value of amenities. Economica 62, 247–267.
Clark, C., 1951. Urban population densities. J. R. Stat. Assoc. Ser. A 114, 490–496.
Clark, C., 1967. Population Growth and Land Use. Macmillan, London.
Combes, P.P., Duranton, G., Gobillon, L., 2012. The Costs of Agglomeration: Land Prices in French Cities. Processed, Wharton School, University of Pennsylvania.
Combes, P.P., Duranton, G., Gobillon, L., 2014. The Production Function for Housing: Evidence from France. Processed, Wharton School, University of Pennsylvania.
Coulson, N.E., 1991. Really useful tests of the monocentric city model. Land Econ. 67, 299–307.
Couture, V., Duranton, G., Turner, M.A., 2012. Speed. Processed, University of Toronto.
Cullen, J.B., Levitt, S.D., 1999. Crime, urban flight, and the consequences for cities. Rev. Econ. Stat. 81, 159–169.

Cunningham, C., 2013. Estimating the Holdout Problem in Land Assembly. Processed, Federal Reserve Bank of Atlanta.
Davis, D.R., Dingel, J.I., 2013. The Comparative Advantage of Cities. Processed, Columbia University.
Davis, J.C., Henderson, J.V., 2008. The agglomeration of headquarters. Reg. Sci. Urban Econ. 38, 445–460.
Davis, D.R., Weinstein, D.E., 2008. A search for multiple equilibria in urban industrial structure. J. Reg. Sci. 48, 29–65.
De Bartolome, C.A.M., Ross, S.L., 2003. Equilibria with local governments and commuting: income sorting vs income mixing. J. Urban Econ. 54, 1–20.
Dixit, A., 1973. The optimum factory town. Bell J. Econ. Manag. Sci. 4, 637–651.
Duranton, G., Puga, D., 2005. From sectoral to functional urban specialisation. J. Urban Econ. 57, 343–370.
Duranton, G., Puga, D., 2014. The growth of cities. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth, vol. 2B. Amsterdam, North-Holland, pp. 781–853.
Dye, R.F., McMillen, D.P., 2007. Teardowns and land values in the Chicago metropolitan area. J. Urban Econ. 61, 45–63.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70, 1741–1779.
Edmonston, B., 1975. Population Distribution in American Cities. Lexington Books, Lexington.
Eid, J., Overman, H.G., Puga, D., Turner, M.A., 2008. Fat city: questioning the relationship between urban sprawl and obesity. J. Urban Econ. 63, 385–404.
Ellickson, B., 1977. Economic analysis of urban housing markets: a new approach. Report R-2024-NSFRand Corporation.
Epple, D., Gordon, B., Sieg, H., 2010a. Drs. Muth and Mills meet Dr. Tiebout: integrating location-specific amenities into multi-community equilibrium models. J. Reg. Sci. 50, 381–400.
Epple, D., Gordon, B., Sieg, H., 2010b. A new approach to estimating the production function for housing. Am. Econ. Rev. 100, 905–925.
Epple, D., Platt, G.J., 1998. Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes. J. Urban Econ. 43, 23–51.
Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., Raudenbush, S., 2003. Relationship between urban sprawl and physical activity, obesity, and morbidity. Am. J. Health Prom. 18, 47–57.
Fischel, W.A., 1987. The Economics of Zoning Laws: A Property Rights Approach to American Land Use Controls. Johns Hopkins University Press, Baltimore, MD.
Fischel, W.A., 2000. Zoning and land use regulations. In: Boudewijn, B., Geest, G.D. (Eds.), Encycolopedia of Law and Economics, vol. 2. Edward Elgar, Cheltenham, pp. 403–442.
Fischel, W.A., 2001. The Homevoter Hypothesis. Harvard University Press, Cambridge, MA.
Fischel, W.A., 2004. An economic history of zoning and a cure for its exclusionary effects. Urban Stud. 41, 317–340.
Frank, L.D., Andresen, M.A., Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. Am. J. Prev. Med. 27, 87–96.
Fujita, M., 1982. Spatial patterns of residential development. J. Urban Econ. 12, 22–52.
Fujita, M., 1989. Urban Economic Theory: Land Use and City Size. Cambridge University Press, Cambridge, MA.
Fujita, M., 2012. Thünen and the new economic geography. Reg. Sci. Urban Econ. 42, 907–912.
Fujita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Reg. Sci. Urban Econ. 12, 161–196.
Fujita, M., Thisse, J.F., 2013. Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge University Press, Cambridge.
Galster, G., Hanson, R., Ratcliffe, M.R., Wolman, H., Coleman, S., Freihage, J., 2001. Wrestling sprawl to the ground: defining and measuring an elusive concept. Hous. Pol. Deb. 12, 681–717.
Garreau, J., 1991. Edge City: Life on the New Frontier. Doubleday, New York, NY.
Genesove, D., Han, L., 2012a. Measuring the Thinness of Real Estate Markets. Processed, University of Toronto.
Genesove, D., Han, L., 2012b. Search and matching in the housing market. J. Urban Econ. 72, 31–45.
Geoghegan, J., 2002. The value of open spaces in residential land use. Land Use Pol. 19, 91–98.

Giles-Corti, B., Macintyre, S., Clarkson, J.P., Pikora, T., Donovan, R.J., 2003. Environmental and lifestyle factors associated with overweight and obesity in Perth. Australia. Am. J. Health Prom. 18, 93–102.
Glaeser, E.L., Gottlieb, J.D., 2006. Urban resurgence and the consumer city. Urban Stud. 43, 1275–1299.
Glaeser, E.L., Gyourko, J., 2005. Urban decline and durable housing. J. Pol. Econ. 113, 345–375.
Glaeser, E.L., Kahn, M., 2001. Decentralized employment and the transformation of the American city. In: Brookings-Wharton Papers on Urban Affairs, pp. 1–47.
Glaeser, E.L., Kahn, M.E., 2004. Sprawl and urban growth. In: Henderson, V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. North-Holland, Amsterdam, pp. 2481–2527.
Glaeser, E.L., Gyourko, J., Saks, R., 2005. Why is Manhattan so expensive? Regulation and the rise in housing prices. J. Law Econ. 48, 331–369.
Glaeser, E.L., Kahn, M.E., Rappaport, J., 2008. Why do the poor live in cities? The role of public transportation. J. Urban Econ. 63, 1–24.
Gobillon, L., Selod, H., Zenou, Y., 2007. The mechanisms of spatial mismatch. Urban Stud. 44, 2401–2427.
Gordon, P., Richardson, H.W., Jun, M.J., 1991. The commuting paradox: evidence from the top twenty. J. Am. Plann. Assoc. 57, 416–420.
Gyourko, J., Saiz, A., 2006. Construction costs and the supply of housing structure. J. Reg. Sci. 46, 661–680.
Hamilton, B.W., 1982. Wasteful commuting. J. Pol. Econ. 90, 1035–1051.
Helsley, R.W., 1990. Knowledge production in the CBD. J. Urban Econ. 28, 391–403.
Henderson, J.V., 1977. Economic Theory and the Cities. Academic Press, New York, NY.
Henderson, J.V., Mitra, A., 1996. The new urban landscape: developers and edge cities. Reg. Sci. Urban Econ. 26, 613–643.
Hilber, C., Robert-Nicoud, F., 2013. On the origins of land use regulations: theory and evidence from US metro areas. J. Urban Econ. 75, 29–43.
Imai, H., 1982. CBD hypothesis and economies of agglomeration. J. Econ. Theory 28, 275–299.
Inman, R.P., 1995. How to have a fiscal crisis: lessons from Philadelphia. Am. Econ. Rev. 85, 378–383.
Irwin, E.G., 2002. The effects of open space on residential property values. Land Econ. 78, 465–480.
Kain, J.F., 1968. Housing segregation, negro employment, and metropolitan decentralization. Q. J. Econ. 82, 175–197.
Kanemoto, Y., 1980. Theories of Urban Externalities. North-Holland, Amsterdam.
Kanemoto, Y., 1988. Hedonic prices and the benefits of public projects. Econometrica 56, 981–989.
Katz, L.F., Kling, J.R., Liebman, J.B., 2001. Moving to opportunity in Boston: early results of a randomized mobility experiment. Q. J. Econ. 116, 607–654.
Kim, S., 2007. Changes in the nature of urban spatial structure in the United States, 1890–2000. J. Reg. Sci. 47, 273–287.
Koopmans, T.C., Beckmann, M., 1957. Assignment problems and the location of economic activities. Econometrica 25, 53–76.
Kopecky, K.A., Suen, R.M.H., 2010. A quantitative analysis of suburbanization and the diffusion of the automobile. Int. Econ. Rev. 51, 1003–1037.
Krueger, A.B., Kahneman, D., Schkade, D., Schwarz, N., Stone, A.A., 2009. National time accounting: The currency of life. In: Krueger, A.B. (Ed.), Measuring the SubjectiveWell-Being of Nations: National Accounts of Time Use andWell-Being. NBER and University of Chicago Press, Chicago, pp. 9–86.
Landvoigt, T., Piazzesi, M., Schneider, M., 2011. The Housing Market(s) of San Diego. Processed, Stanford University.
Lee, S., Lin, J., 2013. Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income. Processed, Federal Reserve Bank of Philadelphia.
LeRoy, S.F., Sonstelie, J., 1983. The effects of urban spatial structure on travel demand in the United States. J. Urban Econ. 13, 67–89.
Levinson, D., Wu, Y., 2005. The rational locator reexamined: are travel times still stable? Transportation 32, 187–202.
Liao, W.C., 2012. Inshoring: the geographic fragmentation of production and inequality. J. Urban Econ. 72, 1–16.
Lucas Jr., R.E., Rossi-Hansberg, E., 2002. On the internal structure of cities. Econometrica 70, 1445–1476.
Ma, K.R., Banister, D., 2006. Excess commuting: a critical review. Transp. Rev. 26, 749–767.

Määttänen, N., Terviö, M., 2014. Income distribution and housing prices: an assignment model approach. J. Econ. Theory 151, 381–410.

Madden, J.F., 1985. Urban wage gradients: empirical evidence. J. Urban Econ. 18, 291–301.

Manning, A., 2003. The real thin theory: monopsony in modern labour markets. Lab. Econ. 10, 105–131.

Manville, M., Shoup, D., 2003. People, parking and cities. J. Urban Plann. Devel. 131, 233–245.

Margo, R.A., 1992. Explaining the postwar suburbanization of population in the United States: the role of income. J. Urban Econ. 31, 301–310.

Markusen, J.R., 2002. Multinational Firms and the Theory of International Trade. MIT Press, Cambridge, MA.

McDonald, J.F., 1989. Econometric studies of urban population density: a survey. J. Urban Econ. 26, 361–385.

McDonald, J.F., Bowman, H.W., 1979. Land value functions: a reevaluation. J. Urban Econ. 6, 25–41.

McGrath, D.T., 2005. More evidence on the spatial scale of cities. J. Urban Econ. 58, 1–10.

McMillen, D.P., 1996. One hundred fifty years of land values in Chicago: anonparametric approach. J. Urban Econ. 40, 100–124.

McMillen, D.P., 2001. Nonparametric employment subcenter indentification. J. Urban Econ. 50, 448–473.

McMillen, D.P., 2006. Testing for monocentricity. In: Arnott, R.J., McMillen, D.P. (Eds.), A Companion to Urban Economics. Blackwell, Oxford, pp. 128–140.

McMillen, D.P., 2010. Issues in spatial data analysis. J. Reg. Sci. 50, 119–141.

McMillen, D.P., Singell Jr., L.D., 1992. Work location, residence location, and the intraurban wage gradient. J. Urban Econ. 32, 195–213.

McMillen, D.P., Smith, S.C., 2003. The number of subcenters in large urban areas. J. Urban Econ. 53, 332–342.

Merlo, A., Ortalo-Magné, F., Rust, J., 2013. The Home Selling Problem: Theory and Evidence. Processed, University of Pennsylvania.

Mieszkowski, P., Mills, E.S., 1993. The causes of metropolitan suburbanization. J. Econ. Persp. 7, 135–147.

Mieszkowski, P., Smith, B., 1991. Analyzing urban decentralization: the case of Houston. Reg. Sci. Urban Econ. 21, 183–199.

Mills, E.S., 1967. An aggregativemodel of resource allocation in ametropolitan area. Am. Econ. Rev. Pap. Proc. 57, 197–210.

Mills, E.S., 1969. The value of urban land. In: Perloff, H.S. (Ed.), The Quality of the Urban Environment. Resources for the Future, Baltimore, MD, pp. 231–253.

Mills, E.S., 1972. Studies in the Structure of the Urban Economy. Johns Hopkins Press, Baltimore, MD.

Mirrlees, J.A., 1972. The optimum town. Swed. J. Econ. 74, 114–135.

Mohring, H., 1961. Land values and the measurement of highway benefits. J. Pol. Econ. 69, 236–249.

Mohring, H., Harwitz, M., 1962. Highway Benefits: An Analytical Framework. Northwestern University Press, Evanston, IL.

Montesano, A., 1972. A restatement of Beckmann's model on the distribution of urban rent and residential density. J. Econ. Theory 4, 329–354.

Mortensen, D.T., Pissarides, C.A., 1994. Job creation and job destruction in the theory of unemployment. Rev. Econ. Stud. 61, 397–415.

Muth, R.F., 1969. Cities and Housing. University of Chicago Press, Chicago, IL.

Ogawa, H., Fujita, M., 1980. Equilibrium land use patterns in a nonmonocentric city. J. Reg. Sci. 20, 455–475.

Ortalo-Magné, F., Prat, A., 2014. On the political economy of urban growth: homeownership versus affordability. Am. Econ. J. 6, 154–181.

Ota, M., Fujita, M., 1993. Communication technologies and spatial organization of multi-unit firms in metropolitan areas. Reg. Sci. Urban Econ. 23, 695–729.

Papageorgiou, Y.Y., Pines, D., 1990. The logical foundations of urban economics are consistent. J. Econ. Theory 50, 37–53.

Paulsen, K., 2012. Yet even more evidence on the spatial size of cities: urban spatial expansion in the us, 1980–2000. Reg. Sci. Urban Econ. 42, 561–568.

Pines, D., Sadka, E., 1985. Zoning, first-best, second-best, and third-best criteria for allocating land for roads. J. Urban Econ. 17, 167–185.
Pissarides, C.A., 2000. Equilibrium Unemployment Theory. MIT Press, Cambridge, MA.
Polinsky, A.M., Shavell, S., 1975. The air pollution and property value debate. Rev. Econ. Stat. 57, 100–104.
Polinsky, A.M., Shavell, S., 1976. Amenities and property values in a model of an urban area. J. Publ. Econ. 5, 119–129.
Postel-Vinay, F., Robin, J.M., 2002. Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica 70, 2295–2350.
Putnam, R.D., 2000. Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster, New York, NY.
Quigley, J.M., 1998. Urban diversity and economic growth. J. Econ. Persp. 12, 127–138.
Raphael, S., 1998. The spatial mismatch hypothesis and black youth joblessness: evidence from the San Francisco Bay Area. J. Urban Econ. 43, 79–111.
Redfearn, C.L., 2007. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. J. Urban Econ. 61, 519–541.
Redfearn, C.L., 2009. Persistence in urban form: the long-run durability of employment centers in metropolitan areas. Reg. Sci. Urban Econ. 39, 224–232.
Ricardo, D., 1817. On the Principles of Political Economy and Taxation. John Murray, London.
Rogerson, R., Shimer, R., Wright, R., 2005. Search-theoreticmodels of the labor market: a survey. J. Econ. Liter. 43, 959–988.
Rosenthal, S.S., 2014. Are private markets and filtering a viable source of lowincome housing? Estimates from a "repeat income" model. Am. Econ. Rev. 104, 687–706.
Rossi-Hansberg, E., 2004. Optimal urban land use and zoning. Rev. Econ. Dyn. 7, 69–106.
Rossi-Hansberg, E., Sarte, P.D., Raymond Owens, I., 2009. Firm fragmentation and urban patterns. Int. Econ. Rev. 50, 143–186.
Saelens, B.E., Sallis, J.F., Black, J.B., Chen, D., 2003. Neighborhood-based differences in physical activity: an environment scale evaluation. Am. J. Public Health 93, 1552–1558.
Samuelson, P.A., 1983. Thünen at two hundred. J. Econ. Liter. 21, 1468–1488.
Sattinger, M., 1993. Assignment models of the distribution of earnings. J. Econ. Liter. 19, 831–880.
Selod, H., Zenou, Y., 2001. Location and education in SouthAfrican cities under and after Apartheid. J. Urban Econ. 49, 168–198.
Siodla, J., 2014. Razing San Francisco: The 1906 Disaster as a Natural Experiment in Urban Redevelopment. Processed, Colby College.
Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge, New York, NY.
Solow, R.M., 1972. Congestion, density and the use of land in transportation. Swed. J. Econ. 74, 161–173.
Solow, R.M., 1973. On equilibrium models of urban location. In: Nobay, A.R., Parkin, M. (Eds.), Essays in Modern Economics. Longman, London, pp. 2–16.
Solow, R.M., Vickrey, W.S., 1971. Land use in a long narrowcity. J. Econ. Theory 3, 430–447.
Strange, W.C., 1992. Overlapping neighborhoods and housing externalities. J. Urban Econ. 32, 17–39.
Strange, W.C., 1995. Information, holdouts, and land assembly. J. Urban Econ. 38, 317–332.
Stull, W.J., 1974. Land use and zoning in an urban economy. Am. Econ. Rev. 64, 337–347.
Sweeney, J.L., 1974a. A commodity hierarchymodel of the rental housing market. J. Urban Econ. 1, 288–323.
Sweeney, J.L., 1974b. Quality, commodity hierarchies, and housing markets. Econometrica 42, 147–167.
von Thünen, J.H., 1826. Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie. Perthes, Oxford, Hamburg, English Translation: The Isolated State: Pergammon Press, 1966.
Tiebout, C.M., 1956. A pure theory of local expenditures. J. Pol. Econ. 64, 416–424.
Turnbull, G.K., 1988. Residential development in an open city. Reg. Sci. Urban Econ. 18, 307–320.
Turner, M.A., 2005. Landscape preferences and patterns of residential development. J. Urban Econ. 57, 19–54.
Wasmer, E., Zenou, Y., 2002. Does city structure affect job search and welfare? J. Urban Econ. 51, 515–541.

Wheaton, W.C., 1977. Income and urban residence: an analysis of consumer demand for location. Am. Econ. Rev. 67, 620–631.
Wheaton, W.C., 1982. Urban spatial development with durable but replaceable capital. J. Urban Econ. 12, 1–21.
Wheaton, W.C., 1990. Vacancy, search, and prices in a housingmarketmatching model. J. Pol. Econ. 98, 1270–1292.
White, M.J., 1976. Firm suburbanization and urban subcenters. J. Urban Econ. 3, 323–343.
Wildasin, D.E., 1986. Spatial variation of the marginal utility of income and unequal treatment of equals. J. Urban Econ. 19, 125–129.
Yinger, J., 1979. Estimating the relationship between location and the price of housing. J. Reg. Sci. 19, 271–286.
Zenou, Y., 2009. Urban Labor Economics. Cambridge University Press, Cambridge, MA.