

9 Skilled Cities and Efficient Urban Transport

Stef Proost and Jacques-François Thisse

Abstract

This chapter surveys the fundamental forces that drive the formation and size of cities. We discuss the different types of agglomeration economies generated by a dense web of activities, with special emphasis on the benefits associated with the clustering of highly skilled workers. The distribution of activities within cities results from the trade-off between commuting and housing costs. We show that in this trade-off commuting costs are the cause and land rent is the consequence. The land rent capitalizes the advantages associated with proximity to particular urban locations such as employment centres. We identify the main sources of inefficiency in various urban policies implemented in Europe. Special attention is paid to the regulation of the land market and the pricing of urban transport.

9.1 Introduction

The main distinctive feature of a city is the very high density of activities and population, which allows agents to be close to one another. Households and firms seek spatial proximity because they need to interact for a variety of economic and social reasons on a daily basis. For example, individuals want to be close to each other because they like to meet other people, learn from others, and have a broader range of opportunities. Hence, *the main reason for the existence of cities is to connect people*. This need is gravitational in nature in that its intensity increases with the number of agents set up nearby, and decreases with the distance between them. Contrary to an opinion widespread in the media, and despite the Internet and other new communication devices, face-to-face contact remains important, at least for a certain number of human and economic activities. To understand why this is so, one has to remember that information transferred through modern communication tools must be structured according to clearly defined schemes and codes known to all. Only formal and precise information can be transmitted this way. In contrast, information that is difficult to codify can often be conveyed only through face-to-face contact.

In the industrial era, cities have enabled transport costs between large and connected production plants to decrease substantially. Today, cities are the cradles of new ideas that benefit firms of very different sizes. But this is not new; cities are – and have been for centuries – the source of productivity gains as well as technological and cultural innovations (Hohenberg and Lees, 1985; Bairoch, 1985). To a large extent, it is fair to say that the agglomeration of economic activities in cities is the geographical counterpart of social and economic development. However, these positive effects come with negative ones: congestion, segregation, pollution, and crime. European cities are much older than American ones. While European cultural heritage is an advantage for economic and social development, it is also a major constraint when organizing and managing mobility within cities. This should not conflict with the fact that wealth is increasingly created in cities, a fact that holds for the European Union (EU) and, more generally, for developed and emerging countries alike. And, although no EU-level urban strategy exists (at least not yet), there is a growing recognition that many large European cities face similar social and cohesion problems.

The city has a spatial extension because economic agents consume land, which implies that consumers travel within the city. Therefore, an urban space is both the substratum of economic activity and a private good (land) that is traded among economic agents. The worldwide supply of land vastly exceeds the demand for land. As a consequence, the price of land should be zero. Yet, we all know that for reasons that do not depend on the quality of the housing structure, housing costs vary enormously with the size of cities. Therefore, the price of land reflects the scarcity of ‘something’ that differs from land per se.

The main objective of urban economics is to explain why cities – understood here as metropolitan areas that extend beyond the core city limits – exist and how they are organized; that is, to explain why jobs are concentrated in a few employment centres and how consumers are spatially distributed within the city according to their incomes and preferences. Central to the workings of a city is the functioning of its land market, which allocates both economic agents and activities across space, as well as the quality of the transport infrastructure used by commuters and shoppers. Equally important are various types of social networks that operate within very short distances. For example, informational spillovers affect positively the productivity of the local R&D sector, whereas neighbourhood effects are often critical to sustaining criminal activities in particular urban districts. To understand cities, we must view them not simply as places in space but as anchored systems of market and nonmarket interactions.

Looking at cities through the lens of microeconomics sheds new light on issues that are often poorly understood otherwise. Many prosperous regions are city-regions or regions that accommodate a dense network of medium-sized cities; an example is the Randstad in the Netherlands. This is backed up by casual evidence: among the top 10 NUTS-2 regions of the EU in terms of gross

domestic product (GDP) per capita, 8 are formed by or organized around a major capital city. Figure 9.1 shows the range of the distribution of regional GDP per capita within each EU country; in most cases, the top position is occupied by the capital regions. In the US, the 20 largest metropolitan areas produce about half of the American GDP. This suggests that interregional systems should be studied in relation to urban systems.

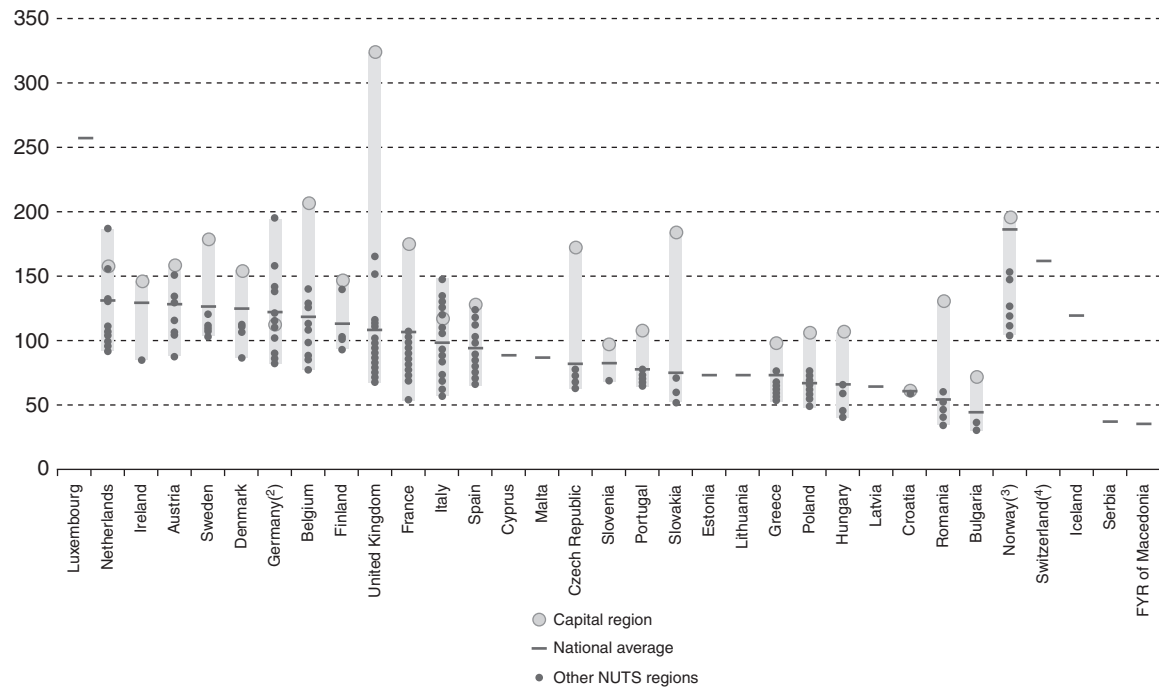
In this chapter, we start by analyzing the fundamental forces that drive the formation and size of cities, that is, the agglomeration economies generated by a dense web of activities and the trade-off between commuting and housing costs. Afterwards, we discuss more specific issues with a special emphasis on residential segregation and urban transport.

9.2 Agglomeration Economies

Humans have a strong drive to form and maintain lasting relations with others. Cities may thus be viewed, at least in the first order, as the outcome of the interplay between social interactions and competition for land. Isolation allows an individual to consume more land but makes interactions with others more costly. To study this trade-off, Beckmann (1976) assumes that the utility of an individual depends on the average distance to all individuals and on the amount of land bought on the market. In equilibrium, the city exhibits a bell-shaped population density distribution supported by a similarly shaped land rent curve. In other words, *the natural gregariousness of human beings turns out to be a sufficient motivation for them to gather within compact areas*. However, while relevant, this explanation is not sufficient to explain the existence of urban agglomerations with millions of inhabitants.

It is well known that consumers in large metropolises pay high rents, have a longer commute, live in a polluted environment, and face high crime rates. So why would they choose to live in such places? It is because they get much better pay in large cities than in small towns. But why do firms in larger cities pay higher wages to their employees? If firms do not bear lower costs and/or earn higher revenues in large cities, they should rather locate in small towns or the countryside where both land and labour are much cheaper. The reason why firms set up in large cities is now well documented: *the productivity of labour is higher in larger cities than in smaller ones*. Or to put it bluntly, after controlling for unemployment and participation, wages and employment (both levels and rates) move together. This does not mean the demand for labour is upward-sloping. Instead, the reason for this urban wage premium is found in what economists call ‘agglomeration economies’.

Whereas economists have long acknowledged the benefits associated with integrating international markets, it took them much longer to understand that there are similar benefits associated with dense and thick markets – such as



(¹) The light grey shaded bar shows the range of the highest to lowest region for each country. The dark grey bar shows the national average. The light grey circle shows the capital city region. The dark grey circles show the other regions.

(²) Only available for NUTS level 1 regions.

(³) Only available at national level.

(⁴) 2012.

Source: Eurostat (online data codes: nama_10r_2gdp and nama_10_pc)

Figure 9.1 The distribution of GDP per capita within EU countries (Eurostat, 2015).

those in large cities. Starting with the highly influential work of Glaeser et al. (1992), Henderson et al. (1995), and Ciccone and Hall (1996), research on city size, employment density, and productivity has progressed enormously during the last two decades. An ordinary least squares (OLS) regression of the logarithm of the average wage on the logarithm of the employment density across cities yields an elasticity that varies from 0.03 to 0.10 (Rosenthal and Strange, 2004). However, this result could be explained by the fact that some econometric problems have not been properly addressed.

First, using a simple reduced form omits explanatory variables whose effects could be captured by employment density. For example, overlooking variables that account for differences in, say, average skills or amenities is equivalent to assuming that skills or amenities are randomly distributed across cities and are taken into account in the random term. This is highly implausible. One solution is to consider additional explanatory variables but, in so doing, we face a familiar situation of adding an endless string of control variables to the regressions. Instead, if we use city/region and industry fixed effects, we can control for the omitted variables that do not vary over time. However, time-varying variables remain omitted.

Second, the correlation of the residuals with explanatory variables – which also biases OLS estimates in the case of omitted variables – can also result from endogenous location choices. Indeed, shocks are often localized and thus have an impact on the location of agents, who are attracted by cities that benefit from positive shocks and repelled by those suffering negative shocks. These agents' relocation has an impact on cities' level of economic activity and, consequently, on their density of employment. Employment density is correlated with the dependent variable and, therefore, with the residuals. To put it differently, there is reverse causality: an unobserved shock initially affects wages and thus density through the mobility of workers, not the other way around. This should not come as a surprise; once it is recognized that agents are mobile, there is a two-way relationship between employment density and wages. The most widely used solution to correct endogeneity biases, whether resulting from omitted variables or reverse causality, involves using instrumental variables. This consists of finding variables that are correlated with the endogenous explanatory variables but not with the residuals.

Caution is therefore needed when measuring the impact of employment density on labour productivity. Using advanced econometric methods and taking into account additional explanations of workers' productivity (such as nonobservable individual characteristics or the impact of previous individual locational choices on current productivity), urban economists have obtained more accurate estimations of agglomeration gains. There is now a broad consensus that, everything else being equal, the elasticity of labour productivity with respect to current employment density is about 0.03 (Combes et al., 2012). This

elasticity measures the static gains generated by a higher employment density. For example, doubling the employment density in Greater Paris would generate an increase in labour productivity that would be twice as large as it would be in the least populated 'départements' of France.

9.2.1 *The Nature and Magnitude of Agglomeration Economies*

Increasing returns are crucial to understanding the formation of the space-economy. The most natural way to think of increasing returns is when a plant with minimum capacity has to be built before starting production. This gives rise to overhead and fixed costs, which are typically associated with mass production. In this case, scale economies are *internal* to firms. Increasing returns may also materialize in a very different form, in which they are *external* to firms but specific to the local environment in which firms operate. Their concrete manifestation can vary considerably from one case to another, but the basic idea is the same: *each firm benefits from the presence of other firms*. In other words, even when individual firms operate under constant returns, there are increasing returns in the aggregate. In a nutshell, the whole is greater than the sum of its parts.

Duranton and Puga (2004) have proposed gathering the various effects associated with agglomeration economies into the following three categories: sharing, matching, and learning.

1. **Sharing** refers primarily to local public goods provided to consumers and producers. When seeking a reason for the existence of cities, one that comes most naturally to mind is the variety and quality of public services, as well as the availability of efficient and large infrastructures. This includes local public goods that contribute to enhancing firms' productivity, such as facilities required by the use of new information and communication technologies and various transport infrastructures, but it also includes public services that increase consumers' well-being. The large number of people and firms facilitate the provision of public goods. These public goods could hardly be obtained in isolation because they would then be supplied at a level below the critical mass permitting the goods to deliver their full impact. In other words, the efficiency of many public services rises when they are supplied to a dense population of users.

Sharing also refers to the supply of intermediate or business-to-business services available in large markets. Even though firms outsource a number of activities to countries where labour is cheap, they also use specialized services available only where they are produced, namely in big cities.

2. **Matching** means that the quality of matches between workers and firms in the labour market is higher in a thick market than in a thin one because of the greater number of opportunities for agents when they operate in a denser

labour market. But the strength of this effect remains an open question. However, sticky workers living in small cities operate in markets with few potential employers, thereby allowing firms to exploit their monopsony power and to pay lower wages.¹ In contrast, workers living in large cities do not have to move to search for jobs provided by other potential employers, which makes them more prone to change jobs. Consequently, workers having the same individual characteristics will earn higher wages in larger cities than in smaller ones because firms have less monopsony power in thicker than in thinner labour markets (Manning, 2010).

3. **Learning** asserts that different agents own different bits of information. Therefore, getting the agents together allows informational spillovers that raise the level of knowledge, thus improving firms' and workers' productivity. Spillovers stem from specific features of knowledge; in particular, knowledge is a nonrivalrous and partially excludable good. The role of information in modern cities has long been emphasized by economic historians. In the words of Hohenberg and Lees (1985), 'urban development is a dynamic process whose driving force is the ability to put information to work. After 1850, the large cities became the nurseries as well as the chief beneficiaries of an explosion in knowledge-centred economic growth.' Cities are the places where people talk. Of course, much of this talk does not generate productivity gains. However, the higher the number of people, the more likely the talk will lead to innovations, increasing productivity. For example, D'Costa and Overman (2014) find that rural workers with past experience in cities enjoy a wage growth premium, thus suggesting that people build knowledge and experience in cities.

Education generates an externality – the knowledge spillovers from skilled workers to other skilled workers – that did not attract much attention until recently. Moretti (2004) has convincingly argued that the social productivity of education exceeds its private productivity. In other words, acquiring human capital enhances not only the productivity of the worker who acquires it but also the productivity of others because we learn from others. What is important for the economic performance of cities is that skilled workers seem to benefit more from the presence of other skilled workers than unskilled workers. Evidently this effect is stronger in the case of regular, easy contacts between skilled workers. For example, Charlot and Duranton (2004) find that workers in larger and more educated cities exchange information more than in cities populated by less-skilled workers. These authors show that such communications explain between 13% and 22 per cent of the urban premium paid to high-skilled workers. In the same spirit, Bacolod et al. (2009) observe that the urban wage premium associated with large cities stems from cognitive skills rather than motor skills. Therefore, everything seems to work as if the marginal productivity of a worker endowed with a certain type of skill increased with the number of skilled workers working or living around him or her. It is

no surprise, therefore, that specific workers tend to sort across space according to their skills.

In the US, Moretti (2012) observed that college graduates living in the richest cities, which are typically knowledge-based metropolitan areas, earn wages 50 per cent higher than college graduates living in the bottom group of cities. In France, about half the spatial disparities in income are explained by the different locations of skilled and unskilled workers (Combes et al., 2008), while between 85 and 88 per cent of spatial wage disparities in the United Kingdom are explained by individual characteristics (Gibbons et al., 2014). The concentration of human capital and high-value activities in large cities is a marked feature of developed and emerging economies. In other words, *spatial inequalities tend more and more to reflect differences in the distribution of skills and human capital across space*.² This has significant implications for the organization of the space-economy: cities specializing in high-tech activities attract highly skilled workers, who in turn help make these places more successful through other agglomeration economies and better amenities. To put it differently, *workers tend to be spatially sorted by skills* (Behrens et al., 2014, Davis and Dingel, 2015). The downside of spatial sorting is the existence of stagnating or declining areas that display high unemployment rates, or are specialized in industries that pay low wages and are associated with a small number of local businesses.

To a large extent, this evolution is enabled by the low transport and communication costs prevailing today. Although these reduced costs allow standardized activities to be located in remote, low-wage countries, big cities remain very attractive for those activities where access to information and advanced technologies is of prime importance. Firms operating in industries that undergo rapid technological changes must be able to react quickly to market signals and to design specialized and sophisticated products that require a skilled labour force, especially when competition is intensified by low transport costs. In a knowledge-based economy where information moves at an increasingly rapid pace, the value of knowledge and information keeps rising. This eventually increases the need for proximity for activities involving firms' strategic divisions, such as management, marketing, finance, and R&D, as well as specialized business-to-business (advertising, legal, and accounting services) and high-tech industries.

If the existence of informational and knowledge spillovers is indisputable, measuring their magnitude is hard as the spillovers are not observable. Different strategies have been proposed to figure out their importance. One of the most original approaches is that of Arzaghi and Henderson (2008), who study the networking effects among geographically close advertising agencies in Manhattan. Advertising is an industry in which creativity matters greatly and where new ideas are quickly obsolete. The authors find there is an extremely rapid

spatial decay in the benefits of having close neighbours. They also show that firms providing high-quality services locate close to other high-quality firms because they do not want to waste resources on discovering which neighbours have valuable information or on establishing communication links with low-quality firms.

It is worth stressing that the geographical concentration of similar firms, such as advertising agencies on Madison Avenue in New York, shows the strength of the various agglomeration effects. Indeed, industrial organization provides evidence that competition between similar firms is a strong dispersion force that tends to push them away from one another (d'Aspremont et al., 1979, Tirole, 1988).

All the agglomeration effects discussed above may be intrasectoral as pointed out by Marshall (1890) or intersectoral as argued by Jacobs (1969). Regardless of their origin, the existence of these positive effects on firms' productivity is unquestionable. However, several issues remain unclear (Puga, 2010, Combes and Gobillon, 2015). First, different industries agglomerate for different reasons. Therefore, what is the relative importance of the various types of agglomeration economies in cities that specialize in different activities? Second, are agglomeration economies stronger in high-tech industries than in traditional sectors that are typically less information-based? Third, the geographical distribution of human capital explains a large share of spatial inequalities. However, it is not clear how much of the human capital effect is explained by the distribution of individual workers, and how much by the presence of human capital externalities across highly skilled workers. Last, how does city size affect the nature and magnitude of agglomeration economies? For example, in a specialized city, a negative shock to the corresponding industry affects its workers negatively. In contrast, in a city endowed with many different types of industries, workers may expect to find a job in firms belonging to other industries. In other words, a diversified – and probably large – city acts as an insurance device. For example, large French cities have been less affected by the Great Recession than other territories (Borzic and Le Jeannic, 2014). In the same vein, unplanned interactions allow firms belonging to one sector to benefit from the presence of another firm located in the same city.

In a recent comprehensive study, Faggio et al. (2014) give a qualified answer to these questions. They confirm the presence of the various effects discussed above but stress that *agglomeration is a very heterogeneous phenomenon*. For example, low-tech industries do benefit from spillovers but high-tech industries benefit more. Both intrasectoral and intersectoral external effects are at work, but they affect industries to a different degree. Firm size also matters: agglomeration effects tend to be stronger when firms are smaller. In other words, specialized and vertically disintegrated firms would benefit more from spatial proximity than larger ones. Despite the wealth of valuable new results, if we want

to design more effective policies for city development and redevelopment, we need a deeper understanding of the drivers behind the process of agglomeration. Furthermore, the interactions across agents are driven by the accessibility of an agent to the others. Although geographers and transport economists consider employment density as a rather crude proxy for accessibility, the question of how to define and measure accessibility in econometric studies of agglomeration economies has been mostly ignored.

Another striking difference across cities lies in their ability to innovate. Based on the success of Silicon Valley, the conventional wisdom among political decision-makers holds that the presence of large university labs is necessary (and, hopefully, sufficient) for a city to become the cradle of new high value added products and processes. Armchair evidence shows that this is not a sufficient condition for boosting innovation productivity. A few decades ago, the experience of Italian industrial districts led some analysts to maintain that a dense network of small and medium-sized businesses (SMEs) was the best environment for innovating. A recent study by Agrawal et al. (2014) suggests a possible reconciliation between these two tenets. Studying American MSA-level patent data during the period 1975–2000, these authors find that the combination of large research labs and a wide array of SMEs is likely to be the friendliest environment for innovation. Although studying the various aspects of entrepreneurship and innovation within the framework of urban economics seems a very promising avenue for research, more work is called for before making solid policy recommendations.

9.2.2 *Cities as Consumption Centres*

The usual cliché is that big cities are bad for consumers. But the authors of anti-city pamphlets forget two things: (i) all over the world, free people vote with their feet by moving to cities; and (ii) cities are not just efficient production centres but are also great places for consumption, culture, and leisure (Glaeser et al., 2001). Consumers living in large cities enjoy a wider range of goods, services, and contacts as the number of shops, cultural amenities, and opportunities for social relations all increase with city size. Even if dating on the Internet tends to be more and more pervasive, one day the two parties have to meet physically. While the steady decline in transport costs and the progressive dismantling of tariff barriers have vastly improved access to foreign goods, models in industrial organization show that the concomitant increase in competition incentivizes both incumbent and new firms to restore their profit margins by supplying higher-quality goods as well as more differentiated products. Because both taste and income ranges are wider in bigger cities, and because larger cities also allow for more varieties to cover their fixed production costs, more goods and services are available in such markets (Picard and Okubo,

2012). In sum, consumers living in larger cities enjoy a broader range of goods and business-to-consumer services.

Even though, as shown below, housing is more expensive in large cities than in small ones, tradable goods need not be more expensive. Since a larger city provides a larger outlet for consumption goods, there is more entry, which intensifies competition; such cities also attract the most efficient retailers that benefit from agglomeration economies and better logistics. Again, as suggested by industrial organization theory, market prices tend to be lower in larger than in smaller cities and the number of varieties of a base product is greater. Calculating the first theoretically based urban price index for 49 US cities, Handbury and Weinstein (2015) show that prices will fall by 1.1 per cent when the population doubles, while the number of available products will increase by 20 per cent. These consumption benefits become even more pronounced once it is recognized that the hierarchy of public services is often the mirror image of the urban hierarchy. In particular, the congregation of a large number of people facilitates the provision of public services that could not be obtained in isolation. Health care and educational facilities are good cases in point.

Notwithstanding many qualifications, the empirical evidence strongly suggests that cities are likely to remain one of the main engines of modern economic growth. Agglomeration economies are not disappearing but their nature and concrete form are changing. But even so, if agglomeration economies are that strong (at least in some sectors), why do cities have a finite size and why are there so many of them? As we are going to see, agglomeration economies have their dark side that restricts the process of city growth and leads to the emergence of a system of cities.

9.3 The Trade-Off between Commuting and Housing Costs

In addition to the idea of agglomeration economies, two other fundamental concepts lie at the heart of urban economics: (i) *people prefer shorter trips to longer trips*, and (ii) *people prefer having more space than less space*. Since activities cannot be concentrated on the head of a pin, they are distributed across space. The authoritative model of urban economics is the monocentric city model developed by Alonso (1964), Mills (1967), and Muth (1969). Treading in these authors' footsteps, economists and regional scientists alike have developed the monocentric model in which a single and exogenously given central business district (CBD) accommodates all jobs. In this context, the only spatial characteristic of a location is its distance from the CBD. The main purpose of this model is to study households' trade-off between housing size – approximated by the amount of land used – and their accessibility to the CBD, measured by the inverse of commuting costs.

Commuting and housing are the two main consumption items in household budgets. Housing and transport represent respectively 26 per cent and 17 per cent of French household expenditures (INSEE). In Belgium, they account for 26 per cent and 13 per cent, respectively (Statistics Belgium). The expenditure share for transport, which takes into account some outlays unrelated to commuting, disregards consumers' time costs and the disutility associated with commuting. The opportunity cost of time spent in commuting accounts for three to six weeks of work a year for a Manhattanite and, on average, four weeks of work for a resident in Greater Paris. These are big numbers, which confirm that commuting costs and traffic congestion are issues that policy-makers have neglected for far too long. Commuting is also perceived as one of consumers' most stressful and unpleasant activities (Kahneman and Krueger, 2006).

9.3.1 *The Monocentric City Model*

Ever since the early 1970s, urban economics has advanced rapidly and shows no sign of abating. The reason for this success is probably that the monocentric city model is based on a competitive land market.³ This assumption can be justified on the grounds that land in a small neighbourhood in any location belonging to a continuous space is highly substitutable, thus making the competition for land very fierce. By allocating a plot of land near the CBD to some consumers, the commuting costs borne by other consumers are indirectly increased as they are forced to set up farther away. Hence, determining where consumers are located in the city is a general equilibrium problem. In equilibrium, identical consumers establish themselves within the city so as to equalize utility. In such a state, the land rent at a particular location is equal to the largest bid for that location. Since people are willing to pay more to be closer to their workplace in order to save time and money on commuting costs, the urban land rent decreases with the distance from the CBD. In turn, since land is cheaper, the population density decreases with distance from the CBD because consumers can afford to buy more land. In sum, *the land rent reflects the differential in workers' accessibility to jobs*.

To illustrate, consider a featureless plain with a dimensionless CBD located at $x = 0$ and a population of consumers who share the same income and the same preferences $U(z, s)$ where z is the consumption of a composite good, chosen as the numéraire, and s the amount of space used. In this set-up, the essential quality which the CBD possesses is physical proximity, or accessibility, to all parts of the urban area. For this reason, consumers compete to be as close as possible to the workplace, but the amount of land available in the vicinity of the CBD is too limited to accommodate the entire population. How, therefore, do consumers distribute themselves across locations? This is where the land market comes into play. The formal argument is disarmingly simple. Denoting

by $R(x)$ the land rent prevailing at a distance x from the CBD and by $T(x)$ the commuting cost borne by a consumer residing at x , the budget constraint of this consumer is given by $z(x) + s(x) \cdot R(x) = I(x) \equiv Y - T(x)$, where consumers have, by assumption, the same income Y .

Let $V(R(x), I(x))$ be the indirect utility of a consumer at x . Since the highest utility level attainable by consumers is invariant across locations, the derivative of $V(R(x), I(x))$ with respect to x must be equal to zero:

$$V_R \cdot R'(x) + V_I \cdot I'(x) = 0.$$

Using Roy's identity and the equality $I'(x) = -T'(x)$, we obtain the Alonso-Muth equilibrium condition:

$$s(x) \cdot R'(x) + T'(x) = 0.$$

Since a longer commute generates a higher cost, this condition holds if and only if the land rent $R(x)$ is downward sloping. As a consequence, *a marginal increase in commuting costs associated with a longer trip ($T'(x) > 0$) is exactly compensated for by the income share saved on land consumption*. In other words, people trade bigger plots for higher commuting costs. If commuting costs were independent of the distance ($T'(x) = 0$), the land rent would be flat and constant across locations. In other words, *commuting costs are the cause and land rents the consequence*.

Furthermore, the lot size occupied by a consumer must increase with the distance from the CBD. Indeed, although a longer commute is associated with a lower net income $Y - T(x)$, the spatial equilibrium condition implies that the utility level is the same across all consumers. As a consequence, in equilibrium, the consumer optimization problem yields a compensated demand for land that depends on the land rent and the endogenous utility level. The utility level is treated as a given by every consumer who is too small to affect it. With housing being a normal good, a lower price for land therefore implies higher land consumption. In other words, as the distance to the CBD increases, the lot size rises, whereas the consumption of the composite good decreases.

Note that housing costs are higher than the land rent because the former account for the quality of housing, which may be higher in the suburbs as units are often older and smaller in the core city. For example, Albouy and Lue (2015) show that, in US metropolitan areas, rent differences due to housing quality are considerable but smaller than differences due to location. All of this implies that *housing rents, unlike the land rent, do not necessarily decrease with the distance to the CBD*.

The monocentric city model also explains how the development of modern transport methods (cars and mass transport) has generated both suburbanization and a flattening of the urban population density, an evolution known as urban sprawl. The monocentric city model has thus produced results that are

consistent with some of the main features of cities. However, it remains silent on why there is a district where all jobs are concentrated. So we are left with the following question: *Do cities emerge as the outcome of a trade-off between agglomeration economies and commuting/housing costs?*

9.3.2 *Why Do Employment Centres Emerge?*

The first answer to this question was provided by Ogawa and Fujita (1980) in a fundamental paper that went unnoticed for a long period of time.⁴ They combine consumers and firms in a full-fledged general equilibrium model in which goods, labour, and land markets are perfectly competitive. Informational spillovers act as an agglomeration force. Indeed, the value of a firm's location depends on its proximity to other firms because informational spillovers are subject to distance-decay effects. As before, workers are keen to minimize commuting costs. The clustering of firms increases the average commuting distance for workers, which in turn leads workers to pay a higher land rent. Therefore, firms must pay workers a higher wage as compensation for their longer commutes to work. In other words, the dispersion force stems from the interaction between the land and labour markets in firms' optimization programme. The equilibrium distribution of firms and workers is the balance between those opposing forces. Note the difference with the monocentric model in which the CBD is given: interactions among agents make the relative advantage of a given location for an agent dependent on the locations chosen by other agents.

Ogawa and Fujita show that, in equilibrium, *the city may display different configurations*, implying that the city may be polycentric. First, when commuting costs are high in relation to the distance-decay effect, the equilibrium involves a full integration of business and residential activities. To put it differently, land use is unspecialized. As commuting costs fall, two employment centres, themselves flanked by a residential area, are formed around an integrated section. Eventually, when commuting costs are low enough, the city becomes monocentric. In this configuration, land use is fully specialized. This seems to concur with the evolution in the spatial organization of cities that has been observed since the beginning of the revolution in transport. Activities were dispersed in pre-industrial cities when people moved on foot, whereas cities of the industrial era were often characterized by a large CBD. Modern cities retain a large CBD, but city centres now accommodate land-intensive activities performed in offices rather than factories that are big consumers of space. Other forces, such as traffic congestion and the development of new information and communication technologies, foster the emergence of secondary employment centres.

Although the process of nonmarket interaction between firms (or workers) is typically bilateral, firms care only about their role as 'receivers' and neglect

their role as ‘transmitters’. A comparison of the equilibrium and optimum densities shows that the former is less concentrated than the latter. This suggests that, from the social standpoint, the need to interact results in an insufficient concentration of activities around the city centre. Therefore, contrary to general belief, firms and consumers would not be too densely packed.

In the next two subsections, we briefly discuss housing and residential segregation. Both are fundamental issues that would require longer developments.

9.3.3 *Land Capitalization and Housing*

The choice of a residence implies differential access to the various places visited by consumers. Therefore, it should be clear that the same principle applies when consumers are sited close to locations endowed with amenities and/or providing public services such as schools and recreational facilities. As a consequence, if the general trend is a land rent that decreases as the distance from the CBD increases, the availability of amenities and public services at particular urban locations within the city affects this trend by generating rebounds in the land rent profile (Fujita, 1989). For example, everything else being equal, if the quality of schools is uneven, the price of land is higher in the neighbourhood of the higher-quality schools. Likewise, dwellings situated close to metro stations are more expensive than those farther away. All of this has a major implication: *in a city, the land rent value at any specific location capitalizes (at least to a certain extent) the benefits (and sometimes the costs) associated with the distance to the workplaces, as well as the accessibility of various types of facilities and amenities.* This value is created by community growth through actions taken by firms, households, and local governments, but not much value (if any) is created by landlords.

As a first approximation, the value of a residential property may be viewed as the sum of two components: the value of the land on which the structure sits, plus the value of the structure. The value of the residential structure has to belong to the agent responsible for its construction. In contrast, the land rent value depends on the proximity to jobs and on public service providers financed by local or federal governments. Therefore, a *laissez-faire* policy allowing the landlord to capture the land rent is like an implicit transfer from the collectivity to the landlord. Evidently, for the land capitalization process to unfold, land prices must be free to react to consumers’ residential choices.

Stiglitz (1977) has shown that the land capitalization process is a very powerful instrument with which to finance the provision of public goods: the aggregate land rent equals the level of public expenditure if, and only if, the population size maximizes the utility level of the city’s residents. Under these circumstances, public services can be financed by taxing the land rent. When there are too many consumers, this leads to higher land rents, generating a total

land rent that exceeds public expenditure. In contrast, when public expenditure exceeds the aggregate land rent, the population is below the optimal size. On this occasion, it is worth recalling that the gigantic transformation of Paris under the direction of Georges-Eugène Haussmann in the second half of the nineteenth century was financed by 'the money ... borrowed against future revenues that would result from the increased property values created by the planned improvements' (Barnett, 1986). What was possible then should be possible today, allowing our cities to finance – at least up to a certain threshold – the investments made to improve urban life.

Equally important, a better understanding of the land market allows shedding light on an ongoing heated debate in many European countries, namely rent control and land-use planning. Contrary to a belief shared by the media and the public, past and current rise in housing costs in many European cities is driven mainly by excessive, rather than insufficient, regulation of the housing and land markets. Public policies typically place a strong constraint on the land available for housing. By instituting artificial land rationing, these policies reduce the price elasticity of housing supply; they also increase the land rent and inequality that go hand in hand with the growth of population and employment. For example, the evidence collected by Glaeser and Gyourko (2003) in the US suggests that 'measures of zoning strictness are highly correlated with high prices', while Brueckner and Sridhar (2012) find large welfare losses for the building height restrictions in Indian cities. The beneficiaries of these restrictions are owners of existing plots and buildings. Young people and new inhabitants, particularly the poorest, are the victims of these price increases and crowding-out effects, which often make their living conditions difficult. In a detailed study of the causal effect of land use regulation in the US, Turner et al. (2014) find that the implications of regulatory constraints for land prices and welfare can be decomposed in three parts: (i) how land in specific plots is used, (ii) how land nearby is used, and (iii) the overall supply of developable land. Due to lost residential land, the first effect has a negative and substantial impact on welfare, while the third one induces losses for residents that are almost offset by land owners' gains. The estimates are not precise enough to determine the sign of the second effect.

By restricting population size, the implementation of urban containment hurts new residents by reducing their welfare level or motivates a fraction of the city population to migrate away. In addition, such policies prevent the most productive cities from fully exploiting their potential agglomeration effects. Admittedly, environmental and esthetic considerations require green space. However, the benefits associated with providing such spaces must be measured against the costs they impose on the population. For example, housing land in southeast England was worth 430 times its value as farmland (Cheshire et al., 2014). Under such circumstances, the land rent level also reflects the

‘artificial scarcity’ of land stemming from restrictive land use regulation. It is worth stressing here that, in many EU countries, land made available for housing depends on municipal governments. Therefore, it is hardly surprising that decisions regarding land use vary with political parties (Solé-Ollé and Viladecans-Marsal, 2012).

High housing prices make the city less attractive. This may deter young entrepreneurs and skilled workers from settling there, which weakens the city’s economic engine. Freezing rents – one of the most popular instruments used by political decision-makers in Europe – renders the housing supply function more inelastic. Subsidizing tenants does not work either because the money transferred to the tenants tends to end up in landlords’ pockets when the elasticity of the housing supply is weak. *Providing affordable housing through the adoption of market-savvy land and construction policies is one of the keys to the future economic success of cities.*

Housing markets play a critical role in the workings of a city with an important impact on the global economy. Hsieh and Moretti (2015) show that lowering regulatory constraints in highly productive cities like New York, San Francisco, and San Jose to the level of the median city would expand those cities’ workforce and increase GDP in the US by 7.5 per cent. Various housing supply constraints act as impediments to a more efficient spatial allocation of labour, which lowers the income and welfare of all US workers. Even though European workers are less mobile than Americans, these effects could very well be important within European countries too. *Smart land and housing policies are a key instrument for regional and urban development.* How to design such policies is an issue that cannot be underestimated, but going into further discussions would take us beyond the scope of this chapter. Note, however, that the literature on housing and cities is blossoming in the US but is still in its infancy in Europe.

9.3.4 *Residential Segregation*

The allocation of land across consumers may be viewed as the outcome of a competitive process in which consumers bid against others to occupy a specific plot. In other words, everything works *as if* consumers had a *bid rent function* that specifies their willingness to pay for one unit of land at different locations. The market then selects consumers who offer the highest bid. The bid rent function reflects the trade-off between commuting and housing costs at the individual level, while the land rent is the maximum of the individual bid rent functions (Fujita, 1989). This simple mechanism shows that residential segregation leads to the sorting of individuals within cities through the respective values of their bid rents. For example, when consumers share the same preferences but have different incomes, the same income group will occupy the city

neighbourhood where it outbids other groups. This has a far-reaching consequence: consumers' income differences translate into spatial segregation. To put it differently, *consumers are sorted within the city through the working of the housing/land market*. As a consequence, the city is segmented into neighbourhoods in which consumers have similar characteristics.

Although some American core cities have rich enclaves, high-income residents in US urban areas tend to live in the suburbs. This pattern is often reversed in the EU. Brueckner et al. (1999) have proposed an amenity-based theory that predicts a multiplicity of location patterns across cities. Europe's longer history provides an obvious reason why its core cities offer more amenities, such as buildings and monuments of historical significance, than do their US counterparts. When the centre has a strong amenity advantage over the suburbs, the bid rent function can be used to show that the rich are likely to live in central locations. When the centre amenity advantage is weak or negative, the rich are likely to live in suburbs. In other words, superior amenities make the core city rich, while weak amenities make it poor.

In the same vein, when the urban space is not featureless, the rich can afford to set up in locations with better amenities, which may be exogenous or endogenous, and with more transport options than the poor. In particular, decentralizing the supply of schooling may exacerbate initial differences between people by allowing the rich to afford better education for their children. This in turn tends to increase differences in human capital among young people and worsen income inequality between individuals and neighbourhoods within the same city. Besides income and preferences, spatial segregation as an equilibrium outcome can also be based on culture, race, and language. However, it is worth stressing that, *regardless of the attributes that determine the bid rent that consumers are willing to pay for particular locations, the above sorting mechanism keeps its relevance*. Furthermore, through nonmarket interactions, the gathering of people sharing the same characteristics may generate different types of externalities. As in the foregoing, we end up with more homogeneous districts, but more heterogeneity between districts (Bénabou, 1993).

What makes spatial segregation a robust outcome is that, even in the absence of externalities, similar people competing on the land market will choose independently to be close to each other. This segmentation is the unintentional consequence of decisions made by a great number of consumers acting in a decentralized environment. The bid rent mechanism suggests that 'causation runs from personal characteristics to income to the characteristics of the neighbourhood in which people live' (Cheshire et al., 2014). This probably explains why many public policies that promote social mixing within cities fail to reach their objective.

Whether and how neighbourhood effects have an impact on individual characteristics is an important topic, as European cities tend to become more

polarized and segregated. Topa and Zenou (2015) stress the importance of understanding the causality links and of distinguishing between neighbourhood effects and network effects. Neighbourhood effects mean that better access to jobs increases the employment prospects of the poor. This can be addressed by housing, transport, or neighbourhood regeneration policies. For example, distressed urban areas can be more or less isolated. This helps to explain why place-based policies, like the French enterprise zone programmes, may increase the employment rate of the poor in well-connected areas, but not in rather isolated areas (Briant et al., 2015). Network effects have to do with the poor quality of the socio-economic group to which they belong. In this case, transport policy is useless and specific social integration and human capital policies are needed. Topa and Zenou (2015) point to empirical evidence for Sweden and Denmark that suggests ethnic enclaves can have positive effects on labour market outcomes and the education level of immigrants, especially for the unskilled. The dark side is that such enclaves seem to have a positive impact on crime, as growing up in a neighbourhood with many criminals around has a long-term effect on the local crime rate.

To sum up, even though urban land use patterns reflect a wide range of possibilities, the way the bid rent functions vary with places' and residents' characteristics allows us to understand what kind of residential pattern emerges. The bid rent function, because it relies on a fundamental principle that guides consumers' spatial behaviour, is likely to be useful in designing market-savvy policies fostering less segregation.

9.4 More Cities or Bigger Cities?

Agglomeration economies explain why human activities are concentrated in cities. However, because commuting and housing costs rise along with the population size, they – along with negative externalities generated by the concentration of people in small areas – act as a strong force to put a brake on city growth. In accordance with the fundamental trade-off of spatial economics, *the size of cities may then be viewed as the balance between these systems of opposite forces*. Finding the right balance between agglomeration economies and diseconomies is at the heart of the urban problem.

Not all cities are alike. The existence of very large cities in different parts of the world at different time periods is well documented (Bairoch, 1985). Cities have very different sizes and form an urban system that is hierarchical in nature: *there are few large cities and many small cities*, together with an intermediate number of medium-sized cities. The stability of the urban hierarchy over decades or even centuries is remarkable (Eaton and Eckstein, 1997, Davis and Weinstein, 2002). All cities provide private goods that are nontradable (e.g., shops) and a variable range of public services (e.g., schools, day care

centres). To a certain extent, the urban system reflects the administrative hierarchy of territorial entities. Because public services are subject to different degrees of increasing returns, cities accommodate a variable number of governmental departments and agencies, hospitals, universities, museums, and the like. More importantly, cities have a different industrial composition. In the past, cities produced a wide range of goods that were not traded because shipping them was expensive. Once transport costs decreased sufficiently, medium-sized and small cities became specialized in the production of one tradable good. This increased specialization often leads to significant labour productivity gains but makes cities vulnerable to asymmetric shocks. Today, only a few urban giants accommodate several, but not all, sectors.

Unlike specialized cities, *diversified* cities are better equipped when confronted with asymmetric shocks. Besides spillover effects between sectors, the coexistence of different sectors may also reduce the uncertainty associated with the initial phases of the product cycle (Duranton and Puga, 2001). For example, the preliminary stages in the development of a new technology or product require repeated contacts among those involved, which are much easier when these people are in close proximity. Information becomes a spatial externality because, as it circulates within the local cluster of firms and workers, it inadvertently contributes to aggregate productivity. However, as shown by Helsley and Strange (2014), potentially beneficial clusters do not necessarily emerge, while the co-agglomeration that does occur in diversified cities may not be that which creates the greatest productive benefits.

Henderson (1974, 1988) has developed a compelling and original approach that allows us to describe an urban system that involves an endogenous number of specialized cities trading goods. The second-generation models explore both the sorting of workers and the composition of population across cities, which are consistent with recent empirical evidence (Behrens et al., 2014, Eeckhout et al., 2014). Davis and Dingel (2015) observe that in the US the hierarchy of skills is highly correlated to the urban hierarchy. Specifically, these authors have proposed a new modelling strategy that suggests that ‘the most skilled individuals in the population live only in the largest city and more skilled individuals are more prevalent in larger cities’. What makes these new models especially appealing is their ability to capture what we know from urban economics about the role of human capital externalities in the formation of cities.

However, in this strand of literature, cities produce the same good or, equivalently, different goods traded at zero cost. These models do not recognize that cities are anchored in specific locations and embedded in intricate networks of trade relations that partially explain the cities’ size and industrial mix. In sum, they put aside the fact that *location matters because trade is costly* (see Chapter 8). Allowing for a large number of potentially asymmetric

locations while trade between any two locations is subject to bilateral costs, Allen and Arkolakis (2014) and Redding (2015) explore a new line of research whose aim is to assess the role of locations in the geographical distribution of activities. This approach relies on the calibration of Ricardian spatial models that permit the study of counterfactuals. For example, the analysis undertaken by Allen and Arkolakis (2014) suggests that about 20 per cent of the spatial variation in income across the US can be explained by pure locational effects. These new models bring together ideas borrowed from trade and urban economics, but they use particular functional forms whose effects remain unexplored. Therefore, it seems fair to say that the dust has not yet settled.

It is also worth noting the work of Desmet and Rossi-Hansberg (2013), who decompose the determinants of the city size distribution into the following three components: efficiency, amenities and frictions. Higher efficiency and more amenities lead to larger cities but generate greater frictions (congestion). This model may be used to simulate the effects of reducing variations in efficiency and amenities, which makes it a relevant tool for designing regional and urban policies. Averaging the level of the above three components across cities and allowing the population to relocate leads to large population relocations but generates low welfare gains in the US. Using the same model for China, the authors find much bigger welfare gains.

The number of large metropolitan areas in the US is proportionally much higher than in the EU. Therefore, it is tempting to follow *The Economist* (13 October 2012), which argues that European cities are too small and/or too few for the European economy to benefit fully from the informational spillovers lying at the heart of the knowledge-based economy. A more rigorous analysis has been developed by Schmidheiny and Südekum (2015). Using the new EC–OECD functional urban areas dataset, they show that, unlike the US urban system, the EU city distribution does not obey the Zipf Law. The reason for this discrepancy is that the largest European cities are ‘too small’. Undoubtedly, many European governments were not – and several of them are still not – aware of the potential offered by their large metropolitan areas to boost economic growth. Both in Europe and the US, ‘*urbaphobia*’ has led governments to design policies deliberately detrimental to their large metropolises. In this respect, France is a good (or bad) case in point. For a few decades, Paris was considered ‘too big’ and public policies were designed to move activities toward other French regions. By French standards, Paris is big. Yet, on the international marketplace, Paris competes with a great many comparable or larger cities. However, in view of the productivity shown by the dense network of large/medium and well-connected German cities, it is not clear whether new and large metropolitan areas (10+ million people) are necessary to enhance European competitiveness.

9.5 The Organization of Metropolitan Areas

As the spread of new cities in Europe came to an end long ago, for a long time the European landscape has been dominated by a wide array of monocentric cities. European cities, probably because they were smaller than their American counterparts, undertook a structural transformation illustrated by the emergence of polycentric metropolitan areas. Indeed, the burden of high housing and commuting costs may be alleviated when secondary employment centres are created. Such a morphological change in the urban structure puts a brake on the re-dispersion process and allows big cities to maintain, to a large extent, their supremacy (Glaeser and Kahn, 2004). Among other things, this points to the existence of a trade-off between within-city *commuting costs* and between-cities *transport costs*, which calls for a better coordination of transport policies at the urban and interregional levels.

Urban sprawl and the decentralization of jobs have given rise to metropolitan areas that include a large number of independent political jurisdictions providing local public goods to their residents and competing in tax levels to attract jobs and residents. A few facts documented by Brühlhart et al. (2015) suggest the magnitude of this evolution. Metropolitan areas with more than 500,000 inhabitants are divided, on average, into 74 local jurisdictions, while local governments in the OECD raise about 13 per cent of total tax revenue. Therefore, a cost-benefit analysis of an urban agglomeration cannot focus only on the core city. Indeed, the metropolitan area is replete with different types of externalities arising from its political fragmentation. As a consequence, what matters is what is going on in the metropolitan area as a whole.

The efficient development of a metropolitan area requires a good spatial match between those who benefit from the public goods supplied by the various jurisdictions and the taxpayers (Hochman et al., 1995). This is not often the case because a large fraction of commuters no longer live in the historical centre. In other words, the administrative and economic boundaries of jurisdictions usually differ within metropolitan areas. Since constituencies are located inside the jurisdictions, local governments tend to disregard effects of economic policies that are felt beyond the political border, an issue that also arises at the international level. In addition to spillovers in the consumption of public goods, this discrepancy is at the origin of business-stealing effects generated by tax competition, which are studied in local public finance. However, this literature has put aside the spatial aspects that play a central role in the working of metropolitan areas. For example, the huge Tiebout-based literature leaves little space for urban considerations.

To the best of our knowledge, urban economics is not used as a building block in models studying the workings of a metropolitan area. Thus, research needs to be developed that recognizes the importance of the following aspects

of the problem. First, agglomeration economies within core cities represent a large share of metro-wide agglomeration economies. This in turn implies that the CBD still dominates the metropolitan area's secondary business centres and attracts cross-commuters from the suburbs. As a consequence, agglomeration economies being internalized (even if partly) in wages, the economy of the CBD generates some wealth effects that go beyond the core city to positively impact suburban jurisdictions. Moreover, owing to the attractiveness of the CBD, the core city's government is incentivized to practise tax exporting through the taxation of nonresident workers. As a consequence, the structure of the metropolitan area is inefficient as firms and jobs are too dispersed for agglomeration economies to deliver their full potential (Gaigné et al., 2016).

Second, suburbanites who work in the CBD benefit from public services provided in the core city but do not pay for them. This is a hot issue in cities like Berlin, Brussels or Hamburg, which are also legal regional entities. Third, the metropolitan area is formed of local labour markets that are poorly integrated and coexist with pockets showing high and lasting unemployment rates. Fourth, and last, as cities grow, spatial segregation and income polarization tend to get worse. While the social stratification of cities seems to be less of a political issue in the US, it ranks high on the agenda of many EU politicians and is a major concern for large segments of the European population.

The political fragmentation of metropolitan areas has other unsuspected consequences. For example, establishing new malls on the city outskirts benefits suburbia but diverts consumers from visiting downtown retailers. This in turn leads to a contraction of the central commercial district through the exit of retailers, which makes this shopping area even less attractive. The overall effect is to further reduce the number of customers, which cuts down the number of retailers once more. By making the core city less attractive, this amenity-destructive process is likely to be damaging the productivity of the metropolitan area (Ushchev et al., 2015).

9.6 Managing Traffic and Congestion

People travel within metropolitan areas for a wide array of reasons, such as commuting to work, dropping children off at schools, shopping in the CBD or suburban malls, and attending various family and social events. Even trade is much localized, thus implying a large flow of local shipments.

The origin and destination of a trip, as well as the choice of a transport mode, are decisions users make. Economists study these decisions in a supply-and-demand context. The supply side is given by the transport infrastructure (roads, rail, airports), the transport service (bus, metro, taxi), and a price charged to the users (road user charge, parking fees, public transport prices). Users also supply personal inputs to their trips: cars, fuel, bicycles, insurance, and most

importantly, their own time. On the demand side, for every origin-destination pair, people travel for different reasons and have different opportunity costs of time. Since the supply of infrastructure is limited, the precise timing of trips also matters. It is, therefore, the total user cost of a trip (including money, time and discomfort) that ultimately determines an individual's demand for trips by time of day, by mode of travel, and by route.

Most American cities (exceptions include New York, Washington D.C. and San Francisco) rely on car transport, whereas public transport accounts for a significant fraction of trips in most European cities. This duality is reflected in the topics studied in the academic literature. In the US, where road pricing seems to be banned from public debate, there is more focus on the pricing of parking and optional varieties of road pricing like pay lanes. In the EU, even though some European cities have pioneered new congestion pricing schemes, national and local governments alike favour other policies such as high gasoline and diesel prices, as well as investments and subsidies in public transport.

Urban transport issues can be studied from a short-run or a long-run perspective. In the short run, the origins and destinations (residences, workplace and shops) as well as the transport infrastructures (roads, rail and subway) are exogenous, and thus policy options are restricted to pricing (road pricing, parking and rail tickets) and regulation (speed limits, pedestrian zones). Passengers can react via the number and timing of trips, as well as the type of transport mode. In the long run, locations are endogenous, as is the city size. By implication, users of the transport system have more options because they may change destinations (workplace, school, shopping) and origins (residence). The set of policy options is also much wider in the long run: one can add transport infrastructure and regulate the use of land (housing permits, type of activities). Most of transport economics focuses on the case where locations are given: how the current infrastructure is used (choice of mode, network equilibrium) and how the policy-maker can improve the use of existing infrastructures. Several types of externalities exist, thus there is no satisfactory market mechanism to guarantee the best use of existing capacity. In addition, most road infrastructure can be accessed freely.

In what follows, we first consider the case in which locations and infrastructure are exogenous and focus mainly on passenger transport. To be precise, we first define and discuss the estimation of external costs associated with transport trips for given origins and destinations. We then look at public policies that can be used to address the various market failures associated with the supply and demand for trips. In the next section, we discuss the policy issues when locations and transport infrastructure are endogenous. This will bring us back to the core question in urban economics of how to understand the organization of cities and the location of different activities.

Table 9.1 *External costs by order of magnitude*

External costs	Costs in euro cents	
	Cars ^a	Public transport ^b
Climate costs	0.8	2.1 (bus)
Environmental costs	4.3	21.4 (bus)
Accident costs	0.3	
Congestion costs	0.6 to 242.6	0 to 576.3 (bus)
Wear-and-tear infrastructure costs	0.8	2.7 (bus)

^a By passenger-kilometre.

^b By vehicle-kilometre.

9.6.1 *External Costs of Urban Transport*

Urban transport accounts for some 20 per cent of total passenger-kilometres, where a passenger-kilometre is defined as one passenger who is carried one kilometre. In European cities, cars are the dominant transport mode (70 per cent), while public transport (rail, metro and bus) accounts for the remaining share. External costs of urban transport are difficult to measure because they result from decisions made by a myriad of individuals who do not pay the full cost that their decisions impose on other users. One therefore has to rely on indirect measurements using connected markets (e.g., the variation of housing values as a function of traffic nuisances) or constructed markets (experiments and surveys). In the *Handbook of External Costs* published by the European Commission (2014), five types of external costs are considered: climate costs, environment costs, accident costs, congestion costs, and wear and tear on infrastructure. In Table 9.1, we document the relative importance of these costs for cars and public transport (PT) in the EU. Although the emission of greenhouse gases is proportional to the type and quantity of fossil fuels used, an open question remains about how to evaluate the damage generated by one ton of greenhouse gases, which is the same across industries, power generation, and the residential sector. In Table 9.1, the climate damage generated by one vehicle-kilometre is evaluated at €25/ton of CO₂. In industry, the cap on greenhouse emissions has resulted in prices for tradable pollution permits varying between 5 and €30/ton of CO₂. As the place and sector of climate emissions does not matter, efficient pollution policies call for reducing greenhouse gases where it is cheapest. Above all, Table 9.1 confirms *the sizable impact and variability of congestion costs compared with other external costs*.

Road congestion costs are the most important external costs generated in urban areas, but they also vary substantially across space and time. The marginal external cost generated by traffic congestion is the additional time,

schedule delay, and resource costs borne by other road users when one additional user decides to travel by car. This type of external cost is poorly understood by the general public, probably because car drivers experience their own time loss. Drivers internalize this time loss, but they do not take into account the additional time loss that others incur. In the simplest formulation, the average time cost of a road trip is given by $AC(X) = a + bX$, where X is the volume of traffic on a given road. If the total time cost is given by $TC(X) = aX + bX^2$, the marginal social cost (MSC) is given by $MSC(X) = a + 2bX$. The marginal external cost is then equal to $MSC(X) - AC(X) = bX$. Since the road capacity is constant over the day, the marginal external cost is expected to vary greatly with the intensity of the traffic flow.

For PT, positive density economies arise when the frequency of service increases with demand. Higher frequency decreases the expected waiting time for passengers who arrive at the bus stop randomly and decreases the schedule delay time for nonrandomly arriving passengers. PT by bus also contributes to congestion on the road. Because an additional passenger has to fit into the fixed capacity of the PT vehicles, there is also a negative discomfort external cost.

9.6.2 *The Difficult Road to First-Best Pricing of Congestion*

First-best pricing means that all transport activities are priced (or subsidized) so that the marginal user cost equals the marginal resource cost plus the marginal external costs. As there are different types of external costs, this requires different types of instruments. The easiest external cost to internalize is damage to the climate because this cost is proportional to the consumption of fossil fuel. A fuel excise tax on gasoline and diesel is sufficient to provide the right incentive to save fuel and, therefore, to reduce carbon emissions. That said, the most important marginal external cost of car use is congestion.

Congestion

Ever since the pioneering works of Vickrey (1969), economists have agreed that *the ideal instrument to tackle urban road congestion is congestion pricing*. The idea is easy to grasp. Many road transport externalities are strongly place- and time-dependent and, therefore, can hardly be tackled by using instruments such as fuel taxes or vehicle ownership or purchase taxes, whereas congestion pricing is based on the ongoing traffic. The first successful implementation of a congestion charge was in Singapore (1976). European cities that have introduced similar pricing instruments include London (2003), Stockholm (2007), Milan (2012) and Göteborg (2014).

There have been heated debates in a large number of cities about adopting congestion pricing. The application of road pricing is currently limited to only a few cities. Thus, the question to why implementing such a welfare-enhancing

instrument fails is challenging. Of course, implementation of the pricing system and the transaction costs can eat away 10 to 20 per cent of toll revenues but technology is making big progress on this front. De Borger and Proost (2012) analyse the political economy of road pricing by means of a model of policy reform. They show that the majority of population may vote against road pricing *ex ante* because the expected gain of almost all drivers is negative, whereas some of the drivers may support this policy after implementation.

Congestion pricing has been studied intensively in transport economics. Two lessons can be drawn. First, the design of the road pricing system is very important for the magnitude of the net welfare effect. For example, Stockholm was more efficient than London because the Stockholm system had lower transaction costs and more finely differentiated charges for different times of the day. Indeed, time differentiation is crucial for capturing the full gains of congestion pricing. In the more detailed bottleneck model where homogeneous drivers trade off queuing costs and schedule delay costs by selecting a departure time, an appropriate toll scheme with strong time differentiation can transform all queuing costs into revenue. The result would be an unchanged user cost for the total trip and an unchanged total number of car trips, but departure times would be better distributed, and the local government would end up with extra tax revenues (Arnott et al., 1993). A simple differentiation of peak/off-peak times, as in London, foregoes a large part of these gains and has to rely mainly on reducing the total number of peak trips to alleviate congestion. A more finely tuned pricing scheme narrows the gap between social benefits and toll revenues. This is important for the political acceptability of peak pricing. For example, in London, toll revenues may be a factor five higher than the net benefits, which generate strong lobbying against peak pricing or on the way of sharing the collected toll revenues. More generally, *smart pricing of a bottleneck can transform queuing into toll revenue, bring about important time and productivity gains, and be a sensible alternative to building new and expensive transport infrastructures.*

A second striking feature is that only a small proportion (35 per cent or less) of the suppressed car trips were replaced by PT; the rest of the trips disappeared due to more car sharing, combining trips, or simply foregoing the trip (Eliasson et al., 2009). Having only one or more of the lanes as toll lanes can be effective only if there is a sufficient difference in time values among users, and it does require a careful design of the tolls (Small and Verhoef, 2007).

First-best pricing of public transport is comparatively easy to implement because every passenger has to enter a bus or metro and can be asked to pay. The resource costs and external costs of PT are complex but are known and vary strongly as a function of the density of demand and occupancy of the vehicle. For an almost empty bus, the cost of an additional passenger is limited to the additional time cost for the driver, plus the delay for the existing

passengers and the other road users. There is also a positive externality when additional passengers increase the frequency of the bus service and decrease the expected waiting time at the bus stop. In most urban areas, the largest external cost of public transport is probably the discomfort imposed by additional passengers in the peak period when the PT-system is close to capacity. First-best pricing would then require higher prices in the peak than in the off-peak time.

Finally, it is worth stressing the importance of a major trade-off between boosting the productivity of the Metropolitan Area (MA) through a stronger concentration of jobs in the CBD and reducing congestion and the emission of greenhouse gases in the central city. Indeed, whereas subsidizing commuters may boost the productivity of the MA by fostering a better exploitation of agglomeration economies, road pricing policies, which aim to reduce the discrepancy between the social and the private costs of commuters' trips, have the nature of a tax. Arnott (2007) shows that, in the presence of agglomeration economies, the optimal congestion toll should be lower than the marginal external congestion cost, unless other instruments (e.g., subsidies to firms) are used to correct the agglomeration externalities. Moreover, if fine-tuned road pricing implies only small shifts in working hours, then agglomeration externalities are not very much affected. Note also that the recommendation of a bigger CBD need not necessarily conflict with the objective of lower emissions within the MA. Indeed, when the central city provides a denser and a more energy-efficient public transport system than the suburban jurisdictions, increasing the attractiveness of the CBD may be justified both for economic and environmental reasons, as car use typically increases at the expense of public transport when secondary business centres grow. Clearly, more work is called for to understand how to design better policies under the above trade-off.

Parking

Besides traffic, parking is another major source of urban congestion. Parking space in a city takes up a lot of valuable urban land that could be used for housing and economic activities. A car is parked 95 per cent of the time and often requires a parking spot at both the origin and the destination. Parking supply is divided into parking available for everybody and long-term contracted parking.

One of the main changes over the last 20 years has been the privatization of enforcement of on-street parking. Enforcement has become much more effective and the net revenues have increased. New technologies allow the regular update of prices for on-street parking. For example, in San Francisco, sensors keep track of the occupancy rate per block, which allows for the regular adjustment of parking fees. On-street parking is often provided for free, which worsens the unpriced congestion externalities. There have been many proposals to

abolish these fringe benefits. A well-known example is the cash-out parking proposal where employers are forced to offer the option of receiving the cash equivalent of free parking instead of free parking itself. As parking largely determines the role of cars in urban transport (compare Los Angeles and New York), more research is needed on the effect of parking pricing and regulation in EU cities.

Most economic research has focused on pricing, while most policy interventions focus on regulations and allocation of space. Optimizing transport flows requires the right combination of rules (for example, speed limits), prices, and the allocation of space (for example, bus lanes, and on-street parking).

9.6.3 *The Patchwork of Policy Instruments*

In practice, we are far from first-best pricing schemes in urban transport. When it comes to transport policies, the division of responsibilities among member countries and among regional and city authorities leads to a complex and knotty patchwork. The EU uses mainly regulation (car emissions, safety standards and the like), while taxation power belongs to the member states. Cities have limited authority: parking fees, local traffic regulations and subsidies for PT.

The main tax instrument used to tax externalities of road use is the *fuel tax*. Even though this tax was probably established to raise public income (the average total revenue is 1.4 per cent of the EU GDP), it is de facto the main tax instrument affecting the use of cars. If one considers the fuel tax on cars as the main instrument for correcting externalities, the tax should have the following second-best structure where the tax is equal to all external costs associated with the consumption of a litre of fuel (Parry and Small, 2005):

$$\text{Fuel tax/litre} = \text{carbon damage/litre} \\ + \gamma \cdot (\text{kilometre/litre}) \cdot (\text{other external costs/kilometre}).$$

The first term of this expression is the carbon damage that is proportional to the combustion of fossil fuel. When climate damage is assessed at €25/ton of CO₂, a low excise tax per litre (10 cents/litre) is sufficient. When there is no specific instrument used to price congestion, and other externalities are related to distance driven rather than to fuel consumed, the only way to ‘price’ these externalities is by adding an extra excise tax to the carbon tax for road use. This tax should equal the average other (nonclimate) externalities related to road use, which explains the term (kilometre/litre) (other external costs/kilometre) in the above expression. To compute the tax per litre, one needs information on the other external costs per car-kilometre and the number of kilometres per litre. Finally, one needs a correction factor (γ) that takes into account the share of fuel reduction accounted for by reduced road traffic, not by more fuel-efficient cars. Indeed, because congestion and accident externalities are related to distance

rather than fuel use, it is the amount of driving that the second component of the fuel tax aims to reduce, not the use of fuel itself.

To fix ideas, assume that a 20 per cent increase in fuel tax leads to a reduction in fuel consumption of 10 per cent, of which 5 per cent comes from a more fuel-efficient car and 5 per cent from less driving. Then, the factor (γ) equals 0.5. Assume, furthermore, that the other external costs per kilometre are on average €0.10 and that the car consumes 5 litres per 100 km (see Table 9.1 for other orders of magnitude). Under these circumstances, we obtain a second-best fuel tax equal to €0.10 per litre + $0.5 \cdot (20 \text{ kilometre per litre}) \cdot (\text{€} 10 \text{ per litre}) = \text{€} 1.10$ per litre.

It is worth stressing that *there is an inherent conflict in using the fuel tax to internalize both fuel-related externalities* (climate change) *and mileage-related externalities* (congestion, accidents). For climate-damage reasons, we want a car to be more fuel efficient (up to a marginal cost of €25/ton of CO₂). But to make car drivers take into account the other externalities, we want them to keep paying the same tax per kilometre. As the main objective of the fuel tax is probably to collect tax revenue, using this tax as an instrument to solve all problems amounts to squaring the circle.

It is not only the pricing of petrol that has gone wrong; the pricing of diesel fuel for cars is also problematic as low diesel excise taxes have led to the massive introduction of diesel cars in most of Europe. Diesel cars have a small carbon emission advantage but are more damaging to health when we rely on real world emission results rather than on the results of the test cycle (ICCT (International Council on Clean Transportation) (2014)). The US has taken another route and has almost no diesel cars.

One of the most effective additional instruments to control the environmental externalities of car use is the regulation of emissions of traditional air pollutants. The Auto-Oil programme of the EU regulated the emissions of new cars and the quality of fuel. This was efficient in tackling traditional pollutants (NO_x, SO₂, particulates). By installing additional equipment (catalytic converter, lower sulfur content of fuels) at relatively low cost, emissions could be reduced by a factor of 5 to 20 (Proost and Van Dender, 2012). As for petrol, the EU could benefit from the American and Japanese experience and technologies, which was not the case for diesel.

A complement to stricter air pollution regulations is the use of low-emission zones. In a low-emission zone, only the cleanest cars are allowed to move freely, while 'dirtier' ones have to pay a charge or, if they get caught, a fine. As air pollution damage is directly proportional to population density, it makes sense to have an additional instrument for dense urban areas. The EU ambient air quality regulation sets a maximum for the concentration of air pollutants and, when this maximum is exceeded, city or national governments have to take action. More than 50 German cities have experimented with different policy

measures. The overall conclusion was that improvements in public transport were not effective, but access restrictions for dirty cars were (Wolff, 2014). This type of instrument is less effective at present because, over time, all cars will comply with the latest EU emission standards. But as attention to conventional air pollution in cities is increasing and as the marginal cost of greening cars is growing, this instrument could again become more useful. It allows for the differentiation of requirements for urban road traffic and nonurban road traffic. Instead, one could think of banning diesel cars and even petrol cars in dense cities.

Using fuel-efficiency regulation for cars in order to reduce greenhouse gas emissions is costly as transport already has a high carbon tax in the form of a gasoline tax. One possible justification is the possible myopia or fuel-efficiency gap. If consumers systematically underestimate the future fuel costs, a fuel-efficiency regulation would help consumers and better signal the external costs. But the empirical evidence for consumer myopia is very weak for EU car buyers. Grigolon et al. (2014) have analysed car buyer behavior in the EU and found that consumers take 90 per cent of future fuel costs into account when they select a car. When this is combined with a fuel tax that is more related to mileage externalities than to fuel-related externalities, imposing more fuel-efficient cars is not an efficient policy measure. The EU is a world leader in terms of fuel-efficiency standards. If the aim is also to successfully transfer technology, we may need to reorient our technology standards toward less ambitious targets because other countries have less ambitious climate objectives and do not want to pay for elaborate super-efficient technologies (Eliasson and Proost, 2015). Note also that many countries have used vehicle purchase and ownership taxes as additional instruments to reduce CO₂ emissions. The Netherlands, Denmark, Sweden, and France have used vigorous policies to achieve significant carbon emission reductions but there is evidence that these policies are very costly and not effective.

9.6.4 *Public Transport Pricing*

In the EU, PT accounts for a significant share of commuters. In most European cities, the recovery of operational costs is low (below 50 per cent), while the peak demand is close to the rail and metro capacity. Implementing low prices for PT in cities is often presented as a good illustration of second-best pricing. But is such a recommendation well grounded? In the expression given below, the optimal PT price, $P_{PT}(peak)$, is equal to the social marginal cost of PT, $MC_{PT}(peak)$, corrected by the gap between the price, $P_{CAR}(peak)$, and the social marginal cost of car use, $MC_{CAR}(peak)$. Computing the social marginal cost of a PT trip is not simple. Indeed, it requires taking into account on-board scale economies (using available seats in the metro or bus) and negative

discomfort economies when vehicles are crowded. It must also account for the following positive economies: even when buses or metros are not full, it is optimal to raise frequency as this allows one to reduce waiting time (Mohring, 1972). In the absence of congestion pricing, the price of car use in the peak period is lower than its social marginal cost, so a subsidy for PT is efficient insofar as this subsidy is able to make car users switch to PT. For this, we need the fraction φ of new PT users who would, in the absence of the subsidy, be car users:

$$P_{PT}(peak) = MC_{PT}(peak) + \varphi \cdot [P_{CAR}(peak) - MC_{CAR}(peak)].$$

Parry and Small (2009) have found that a subsidy of close to 90 per cent of the average operational costs for urban rail transport is socially desirable when $\varphi = 0.5$, which seems to ground the proposal of strongly subsidized PT. These authors find the subsidy efficient for two reasons. First, there are important scale economies, which are the most important element to justify subsidies in the off-peak period. Second, there are significant unpriced car congestion externalities, which are the main reason for justifying subsidies in the peak period.

However, some empirical studies find values for φ that are smaller than 0.2 (van Goeverden et al., 2006). In this case, the optimal subsidy for the peak falls from 90 to 10 per cent, thus casting serious doubt on the relevance of subsidizing the use of PT. In a numerical study for London as well as Santiago de Chile, Basso and Silva (2014) compare the pricing of car and bus combined with other instruments (bus subsidies, dedicated bus lanes, and congestion pricing). They find that dedicated bus lanes can be a much more efficient instrument than PT subsidies and are, in terms of efficiency, almost as efficient as road pricing for Santiago de Chile. Results tend to be city-specific as they depend on the current modal shares and the ease of substitution.

Current marginal prices for PT in the EU are often zero as most users pay a monthly subscription price, which allows them to travel when and as much as they want, giving rise to massive congestion problems in PT systems of big EU cities (London, Paris). There is a need to look for more efficient pricing systems that account for the differences in cost between peak and off-peak trips and in function of area and distance travelled, and for the congestion levels of car transport. As long as attention is paid to who pays for the PT subsidies, there is not necessarily a conflict between more appropriate PT fares and redistribution policies (Mayeres and Proost, 2001).

In the last 20 years, the United Kingdom has experimented with privatized PT services. In London, bus services were tendered to private companies but one central bus authority remained as the decision-maker for schedules and prices. The end result was a significant reduction in costs. Outside of London, bus services were fully privatized with the private companies deciding on the number of companies offering services, scheduling and prices. There are only

limited and targeted subsidies. As each bus company offers services at different times of the day, there is a clear tendency to offer higher service frequencies. By offering a time schedule that closely matches the timetable of a competitor, a company could steal passengers from other companies, but this did not turn out to be in the interest of passengers. The end result was lower costs, higher prices, higher frequencies, and less competition (Gwilliam et al., 1985). Contracting out the operation is more common for buses than for rail and has led to important efficiency gains when the contracts are well designed (Gagnepain et al., 2013).

Opponents of congestion pricing put forward the argument that labour (therefore commuters) is already heavily taxed in the EU. Parry and Bento (2001) find that charging the full external congestion cost to commuters remains the best policy as long as additional tax revenues are used to reduce the existing labour tax. Moreover, in many EU countries, employers offer a company car as an untaxed fringe benefit, which amounts to subsidizing high-income commuters and leads to excessive car use, while some employers also pay for all public transport expenses of their employees (Harding, 2014). All of this shows *the need for a global assessment of commuting expenses in relation to income taxes*.

9.7 The Benefits of New Transport Infrastructure

9.7.1 Does New Infrastructure Reduce Congestion?

To the public and to many decision-makers, the answer seems obvious and positive. However, things are not that simple. First, when origins and destinations are given, more capacity leads to more car users. Hence, the time benefit of road extensions in the presence of unpriced congestion is reduced by this induced demand (Cervero, 2003). This already suggests that the standard approach to controlling congestion – forecast traffic growth and build enough roads to accommodate it – is likely to be ineffective. Second, Arnott (1979) shows that improving transport in a congested monocentric city leads to a new residential equilibrium in which congestion increases everywhere compared with the initial equilibrium. In other words, once it is recognized that consumers respond to changes in commuting costs, building new transport links loses a great deal of its appeal.

Duranton and Turner (2011) observe that those who argue in favour of a new transport infrastructure forget the simultaneity problem that we encountered when studying agglomeration economies: the supply of roads and the density of traffic are interdependent phenomena. When the number of vehicles on the road is given, additional capacity decreases the density of traffic and makes trips faster. However, a higher capacity attracts more traffic, and thus density

increases. All this implies that it is a priori unclear how the causality runs. This has led Duranton and Turner to study the congestion problem in American cities for the years 1983, 1993, and 2003, using modern econometric techniques. Their conclusions cast serious doubt on the merits of infrastructure-based congestion policies. First, Duranton and Turner confirm that new roads and public transport generate more traffic. What is less expected, but more important, is that in the absence of road pricing and for some types of roads, ‘new road capacity is met with a proportional increase in driving’. But where do the additional travellers come from? Again, the answer is not the one that comes immediately to mind: ‘the law of traffic congestion reflects traffic creation rather than traffic diversion’. New cars and new trucks share the responsibility for the extra trips almost equally. Last, whenever the road capacity is extended and road use is not appropriately priced, the road extension will attract PT passengers back to driving cars. This reduces frequency in the use of PT, a vicious circle that may lead to the disappearance of the PT alternative.

In sum, work by Arnott, Duranton, Turner and others have a major implication that runs against standard policy recommendations: *when road pricing is not implemented, building new roads might not be the appropriate policy to reduce traffic congestion*. Therefore, congestion pricing is back to centre stage as the main tool to curb urban congestion. Despite the lack of enthusiasm of public policy-makers for this instrument, the impressive number of results obtained by urban transport economics should encourage governments and other authorities to evaluate new transport projects against smart pricing schemes.

Whenever we consider extending current road or PT infrastructure, we should keep in mind that new technologies may enhance the effective capacity of the existing transport system (Winston and Mannering, 2014). For example, the capacity of the current road infrastructure may be enhanced by software applications that facilitate ridesharing. In the long run, vehicle-to-vehicle communication may increase the capacity of a road network by coordinating conflicting traffic flows and by using the stock of cars more intensively, freeing urban space from parking. In the case of public transport, new technologies may also lead to a better use of existing capacity. For example, better software may generate ‘on demand’ collective transport. When there is a capacity shortage, pricing is crucial to using capacity optimally, while road pricing also stimulates the development of new technologies.

9.7.2 *The Wider Benefits of Urban Transport Projects and New Developments in Assessment Methods*

There is growing empirical evidence that big urban transport projects lead to changes in the city form. García-López et al. (2015) have looked into the effects

of highways on urbanization patterns in Spain. They have found that a highway leading from a central city caused an 8 to 9 per cent decline in the central city population between 1960 and 2011. In addition, a highway ramp fostered a 20 per cent population growth in the suburban municipalities where ramps were located. Finally, each additional kilometre closer to the nearest highway ramp increased municipal density growth by 8 per cent. This provides strong evidence for the role of highway capacity on the population distribution within the urban area.

It is, therefore, important to understand the full impact of a large transport project (or important traffic regulation) on the welfare of the metropolitan population, including efficiency as well as equity aspects. Planners typically have little faith in the efficiency or equity of market-determined outcomes, and advocate detailed land-use planning. Yet, as argued in the foregoing, *market forces drive land use to its most productive use if markets are corrected for the most important externalities*. However, care is needed in selecting which externalities to correct. For example, compact cities are often promoted to reduce carbon emissions generated by private transport. However, in the EU, we have seen that carbon is already overtaxed via the gasoline tax. What is more, 30 years from now, standard cars might well be electric battery cars. Consequently, climate considerations are not a good motivation for compact cities.

Economists have developed cost-benefit analysis (CBA) techniques that aim to assess transport projects, be they new infrastructures, new pricing or new regulations. In the EU member countries, they are now widely used, but not necessarily followed by decision-makers. CBA techniques have progressed over the last 50 years from the Dupuit consumer surplus measures to methods that correct for externalities, as well as for market imperfections and the opportunity cost of public funds. Quinet and Raj (2015) review the advances made in assessment methods and distinguish among three approaches: (i) the basic CBA method focusing on changes in the transport market, corrected for externalities and side effects on other markets; (ii) the econometric analysis of causality effects; and (iii) a detailed spatial modelling embedded in land-use planning models. For nonmarginal projects, such as large transport network extensions, there is a need to use them all.

Land-use planning models have been around for a long time. However, there is a need for operational models that integrate both land use and transport (LUTI). Indeed, new transport infrastructures often increase the demand for land, while there is often a new demand for infrastructures when new land is made available for urban activities. Given the long-run implications of decisions made about land use and transport infrastructure, the market alone cannot solve all problems. Accordingly, *cities need to be planned*. For this, different agents (developers, firms, governmental agencies) pursuing different, and sometimes conflicting, objectives must coordinate their actions. Furthermore,

coordination requires commitment on the part of some agents, which is not always possible. Finally, it would be futile to seek a model based on a unified theory of cities that would appeal equally to economists, geographers, architects and urban planners. Therefore, developing LUTI models is a formidable challenge. It is only recently that researchers have tried to build such models in line with the basic principles of urban economics (de Palma et al., 2015).

In principle, LUTI models help us understand the effect of one particular policy intervention and ultimately answer the important question of the ideal urban form. We begin to understand the different mechanisms that come into play: agglomeration economies, congestion, environmental externalities, as well as the impacts of policy instruments (land use, buildings regulation, transport, and parking pricing and capacity). However, our knowledge is still partial, as most studies focus on only one or two mechanisms and on only one instrument at a time. Moreover, most analyses focus on an ideal government planner, while in the real world, political authority is dissipated over sometimes overlapping jurisdictions. The new LUTI model developed by the Netherlands CPB (Bureau for Economic Policy Analysis) provides a good example of what can be accomplished in terms of a detailed understanding of the effects associated with a given policy (Teulings et al., 2014).

9.8 Where Do We Stand?

1. Cities – but not all of them – have been and still are the main engines of cultural, economic, and social development. By encouraging social interactions and the exchange of ideas, cities allow for a finer division of labour and the quick adoption of innovations. As new ideas are often a new combination of old ideas, connecting people remains crucial for the Schumpeterian process of innovation to unfold. As human capital is the main production factor in knowledge-based economies, ignoring the role played by cities often leads governments to design policies that are harmful (not on purpose, of course!) to the economic fabric of their countries.

Not all cities are equally affected by innovation and growth; inequality cuts through the urban system. History tells us that in each period of time there are vibrant as well as dormant cities. If anything else, the development of human capital should be the main target of urban policies. As accurately argued by Glaeser (2011), the oversupply of structures and infrastructures is the hallmark of stagnating and declining cities. Rather than spending billions of euros on large infrastructures and fancy buildings, local governments should facilitate movement in cities by means of congestion pricing and promote the supply of affordable housing.

What is more, *housing and transport markets are intimately intertwined with local labour markets*. As a consequence, European or national employment

policies that ignore the urban environment in which jobs are created and destroyed are unlikely to be able to deliver their full potential. Similarly, if international immigration policies must be coordinated at the EU level, migrants typically have an impact on particular local economies. Moreover, understanding how land capitalization works might help finance local public goods and services, thus alleviating the need to reduce city budgets because of macroeconomic fiscal constraints. In a nutshell, as Cheshire et al. (2014) argue, ‘urban policy informed by economic insights can help improve policy-making for individual cities and urban systems as a whole’, hence the whole economy.

2. All regions benefit from the agglomeration effects arising in large cities through interregional and interpersonal transfers. For example, in 2012, the Île-de-France (Paris) produced 30.1 per cent of French GDP but received only 22 per cent of the disposable income. In other words, 8 per cent of the GDP is redistributed toward other French regions. Greater London’s share of the GDP in the United Kingdom is 23.1 per cent, while its share of the UK’s disposable income is about 16.7 per cent. In Belgium, the contrast is even more striking. The NUTS-2 region Brussels-Capital produces 20.6 per cent of the Belgian GDP but receives only 10.3 per cent of disposable income; thus, more than 10 per cent is redistributed towards the other two regions of Belgium. Very much like some American cities, Brussels attracts high-income commuters as well as poor residents.

3. Urban policies are probably more important for economic growth and social cohesion than regional policies. This is in contrast with the EU’s role in designing regional policies and its absence from urban policies. Social tensions between urban neighbourhoods are strong and income discrepancies within large cities are wide, and both are growing. Investments in human capital and housing are needed to counter this evolution, but they will not be sufficient. Several aspects of urban policy suffer from the fragmentation of policy areas. This holds for public finance, spatial segregation and housing. Urban transport is characterized by many negative externalities, but the present policy orientations are far from optimal, as they do not address the most important externality, that is, congestion. Even though more work is called for, we understand better how cities work and what policies they need. By contrast, due to the relative absence of in-depth studies of the subject matter in Europe, we still have a fairly poor knowledge of what the effects of people’s mobility across the European space-economy are and could be in the future.

4. For the research agenda proposed in chapters 8 and 9 to be carried out, we need data that are often available in the US but not necessarily in the EU. First, for comparative studies across cities to be meaningful, member countries should agree on the same geographical definition of what a metropolitan area is, as in the US where the concept of ‘statistical metropolitan area’ is widely used. Similarly, local data about employment, transport,

GDP, human capital, physical attributes (buildings, roads), environmental quality (air quality, soil) and cultural amenities should be made available for more countries. European economists quite often study American cities rather than European cities because very good data are available in the US, but not in the EU. There is also a need for data at a fine spatial scale about what is going on within cities. For example, such data are needed to study how firms and households choose their locations. New technologies of data collection can help to overcome the data gaps and definitional problems in Europe.

Acknowledgements

We thank F. di Comite, G. Duranton, M. Lafourcade, F. Robert-Nicoud, J. Südekum, A. Turrini, K. Van Dender, E. Viladecans-Marsal, as well as R. Lindsey, J.-Cl. Prager, E. Quinet, K. Small, R. Evers, and C.-M. Fung for their comments on previous drafts. We are grateful to M. Ivaldi, W. Leininger, and M. Paasi for their encouragement and the organization of the COEURE and DG-REGIO workshops, and acknowledge the funding by the FP7/320300 grant under the FP7 SSH research programme.

Notes

1. This and the low value of land explain why many manufacturing firms have relocated their production plants from large to small cities.
2. See, for example, Glaeser and Maré (2001) and Moretti (2012) for the US; Combes et al. (2008) for France; Mion and Naticchioni (2009) for Italy; and Groot et al. (2014) for the Netherlands.
3. The best synthesis of the results derived with the monocentric city model remains the landmark book by Fujita (1989).
4. Only a limited number of papers have tackled the endogenous formation of employment centres. They are surveyed in Duranton and Puga (2004) and Fujita and Thisse (2013).

References

- Agrawal, A., Cockburn, I., Galasso, A., and Oettl, A. 2014. Why are some regions more innovative than others? The role of small firms in the presence of large labs. *Journal of Urban Economics*, **81**, 149–165.
- Albouy, D., and Lue, B. 2015. Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life. *Journal of Urban Economics*, **89**, 74–92.
- Allen, T., and Arkolakis, C. 2014. Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, **129**, 1085–1140.
- Alonso, W. 1964. *Location and Land Use*. Cambridge, MA: Harvard University Press.

- Arnott, R. 1979. Unpriced transport congestion. *Journal of Economic Theory*, **21**, 294–316.
- Arnott, R. 2007. Congestion tolling with agglomeration externalities. *Journal of Urban Economics*, **62**, 187–203.
- Arnott, R., de Palma, A., and Lindsey, R. 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *American Economic Review*, **83**, 161–179.
- Arzaghi, M., and Henderson, J. V. 2008. Networking off Madison Avenue. *Review of Economic Studies*, **75**, 1011–1038.
- Bacolod, M., Blum, B. S., and Strange, W. C. 2009. Skills in the city. *Journal of Urban Economics*, **65**, 136–153.
- Bairoch, P. 1985. *De Jéricho à Mexico. Villes et économie dans l'histoire*. Paris: Éditions Gallimard. English translation by C. Braider, *Cities and Economic Development: From the Dawn of History to the Present* Chicago: University of Chicago Press, 1988.
- Barnett, J. 1986. *The Elusive City: Five Centuries of Design, Ambition and Miscalculation*. New York: Harper and Row.
- Basso, L. J., and Silva, H. E. 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. *American Economic Journal: Economic Policy*, **6**, 1–33.
- Beckmann, M. J. 1976. Spatial equilibrium in the dispersed city. In: Papageorgiou, Y. Y. (ed), *Mathematical Land Use Theory*. Lexington, MA: Lexington Books, pp. 117–125.
- Behrens, K., Duranton, G., and Robert-Nicoud, F. 2014. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, **122**, 507–553.
- Bénabou, R. 1993. Workings of a city: Location, education, and production. *Quarterly Journal of Economics*, **108**, 619–652.
- Borzić, M., and Le Jeannic, T. 2014. En matière d'emploi, les métropoles ont davantage résisté à la crise. *INSEE Première No. 1503*. https://insee.fr/fr/themes/document.asp?reg_id=99&ref_id=ip1503.
- Briant, A., Lafourcade, M., and Schmutz, B. 2015. Can tax breaks beat geography? Lessons from the French enterprise zone experience. *American Economic Journal: Economic Policy*, **7**, 88–124.
- Brueckner, J. K., and Sridhar, K. S. 2012. Measuring welfare gains from relaxation of land-use restrictions: The case of India's building-height limits. *Regional Science and Urban Economics*, **42**, 1061–1067.
- Brueckner, J. K., Thisse, J.-F., and Zenou, Y. 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, **43**, 91–107.
- Brühlhart, M., Bucovetsky, S., and Schmidheiny, K. 2015. Taxes in cities. In: Duranton, G., Henderson, J. V., and Strange, W. (eds), *Handbook of Regional and Urban Economics*, vol. 5. Amsterdam: Elsevier, pp. 1123–1196.
- Cervero, R. 2003. Road expansion, urban growth, and induced travel: A path analysis. *Journal of the American Planning Association*, **69**, 145–163.
- Charlot, S., and Duranton, G. 2004. Communication externalities in cities. *Journal of Urban Economics*, **56**, 581–613.
- Cheshire, P., Nathan, M., and Overman, H. 2014. *Urban Economics and Urban Policy*. Cheltenham, U.K.: Edward Elgar Publishing.

- Ciccone, A., and Hall, R. E. 1996. Productivity and the density of economic activity. *American Economic Review*, **86**, 54–70.
- Combes, P.-P., and Gobillon, L. 2015. The empirics of agglomeration economies. In: Duranton, G., Henderson, J. V., and Strange, W. (eds), *Handbook of Regional and Urban Economics*, vol. 5. Amsterdam: Elsevier, pp. 247–348.
- Combes, P.-P., Duranton, G., and Gobillon, L. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, **63**, 723–742.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, **80**, 2543–2594.
- d'Aspremont, C., Gabszewicz, J. J., and Thisse, J.-F. 1979. On Hotelling's 'Stability in Competition'. *Econometrica*, **47**, 1145–1150.
- Davis, D. R., and Dingel, J. I. 2015. The comparative advantage of cities. Columbia University, mimeo.
- Davis, D. R., and Weinstein, D. E. 2002. Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, **92**, 1269–1289.
- D'Costa, S., and Overman, H. 2014. The urban wage growth premium: Sorting or learning? *Regional Science and Urban Economics*, **48**, 168–179.
- De Borger, B., and Proost, S. 2012. A political economy model of road pricing. *Journal of Urban Economics*, **71**, 79–92.
- de Palma, A., Bierlaire, M., Hurtubia, R., and Waddell, P. 2015. Future challenges in transport and land use modeling. In: Bierlaire, M., de Palma, A., Hurtubia, R., and Waddell, P. (eds), *Integrated Transport and Land Use Modeling for Sustainable Cities*. Lausanne: EPFL Press, pp. 513–530.
- Desmet, K., and Rossi-Hansberg, E. 2013. Urban accounting and welfare. *American Economic Review*, **103**, 2296–2327.
- Duranton, G., and Puga, D. 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, **91**, 1454–1477.
- Duranton, G., and Puga, D. 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J. V., and Thisse, J.-F. (eds), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier, pp. 2063–2117.
- Duranton, G., and Turner, M. A. 2011. The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, **101**, 2616–2652.
- Eaton, J., and Eckstein, Z. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics*, **27**, 443–474.
- Eeckhout, J., Pinheiro, R., and Schmidheiny, K. 2014. Spatial sorting. *Journal of Political Economy*, **122**, 554–620.
- Eliasson, J., and Proost, S. 2015. Is sustainable transport policy sustainable? *Transport Policy*, **37**, 92–100.
- Eliasson, J., Hultkrantz, L., Nerhagen, L., and Rosqvist, L. Smidfelt. 2009. The Stockholm congestion charging trial 2006: Overview of effects. *Transportation Research A*, **43**, 240–250.
- European Commission. 2014. *Update of the Handbook on External Costs of Transport*. Final Report for the European Commission: DG Move. Ricardo-AEA/R/ED57769. London, UK: Ricardo-AEA. <https://ifr.uni-kiel.de/de/forschung/handbook-external-costs-transport-2014.pdf>.
- Eurostat. 2015. Regional statistics. http://ec.europa.eu/eurostat/statisticsexplained/index.php/GDP_at_regional_level.

- Faggio, G., Silva, O., and Strange, W. C. 2014. Heterogeneous agglomeration. SERC Discussion Papers, SERCDP0152. Spatial Economics Research Centre (SERC), London School of Economics and Political Science.
- Fujita, M. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge, MA: Cambridge University Press.
- Fujita, M., and Thisse, J.-F. 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge University Press.
- Gagnepain, P., Ivaldi, M., and Martimort, D. 2013. The cost of contract renegotiation: Evidence from the local public sector. *American Economic Review*, **103**, 2352–2383.
- Gaigné, C., Riou, S., and Thisse, J.-F. 2016. How to make the metropolitan area work? Neither big government, nor laissez-faire. *Journal of Public Economics*, **134**, 100–113.
- García-López, M.-Á., Holl, A., and Viladecans-Marsal, E. 2015. Suburbanization and highways: When the Romans, the Bourbons and the first cars still shape Spanish cities. *Journal of Urban Economics*, **85**, 52–67.
- Gibbons, S., Overman, H. G., and Pelkonen, P. 2014. Area disparities in Britain: Understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics and Statistics*, **76**, 745–763.
- Glaeser, E. L. 2011. *Triumph of the City*. London: Macmillan.
- Glaeser, E. L., and Gyourko, J. 2003. The impact of building restrictions on housing affordability. *Federal Reserve Bank of New York Economic Policy Review*, **9**, 21–39.
- Glaeser, E. L., and Kahn, M. E. 2004. Sprawl and urban growth. In: Henderson, J. V., and Thisse, J.-F. (eds), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier, pp. 2481–2527.
- Glaeser, E. L., and Maré, D. 2001. Cities and skills. *Journal of Labor Economics*, **19**, 316–342.
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. 1992. Growth in cities. *Journal of Political Economy*, **100**, 1126–1152.
- Glaeser, E. L., Kolko, J., and Saiz, A. 2001. Consumer city. *Journal of Economic Geography*, **1**, 27–50.
- Grigolon, L., Reynaert, M., and Verboven, F. 2014. Consumer valuation of fuel costs and the effectiveness of tax policy: Evidence from the European car market. CEPR Discussion Paper No. DP10301.
- Groot, S. P. T., de Groot, H. L. F., and Smit, M. J. 2014. Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities. *Journal of Regional Science*, **54**, 503–523.
- Gwilliam, K. M., Nash, C. A., and Mackie, P. J. 1985. Deregulating the bus industry in Britain – (B) the case against. *Transport Reviews*, **5**, 105–132.
- Handbury, J., and Weinstein, D. 2015. Goods prices and availability in cities. *Review of Economic Studies*, **82**, 258–296.
- Harding, M. 2014. Personal tax treatment of company cars and commuting expenses: Estimating the fiscal and environmental costs. OECD Taxation Working Papers no. 20. Paris: OECD Publishing.
- Helsley, R. W., and Strange, W. C. 2014. Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, **122**, 1064–1093.

- Henderson, J. V. 1974. The sizes and types of cities. *American Economic Review*, **64**, 640–656.
- Henderson, J. V. 1988. *Urban Development. Theory, Fact, and Illusion*. Oxford: Oxford University Press.
- Henderson, J. V., Kuncoro, A., and Turner, M. 1995. Industrial development in cities. *Journal of Political Economy*, **103**, 1067–1090.
- Hochman, O., Pines, D., and Thisse, J.-F. 1995. On the optimal structure of local governments. *American Economic Review*, **85**, 1224–1240.
- Hohenberg, P., and Lees, L. H. 1985. *The Making of Urban Europe (1000–1950)*. Cambridge, MA: Harvard University Press.
- Hsieh, C.-T., and Moretti, E. 2015. Why do cities matter? Local growth and aggregate growth. Mimeo, <http://eml.berkeley.edu/~moretti/papers.html>.
- ICCT (International Council on Clean Transportation). 2014. *European Vehicle Market Statistics Pocketbook 2014*. Berlin: ICCT. <http://eupocketbook.theicct.org>.
- Jacobs, J. 1969. *The Economy of Cities*. New York: Random House.
- Kahneman, D., and Krueger, A. B. 2006. Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, **20**, 3–24.
- Manning, A. 2010. The plant size-place effect: Agglomeration and monopsony power in labour markets. *Journal of Economic Geography*, **10**, 717–744.
- Marshall, A. 1890. *Principles of Economics*. 8th edn. London: Macmillan, 1920.
- Mayeres, I., and Proost, S. 2001. Marginal tax reform, externalities and income distribution. *Journal of Public Economics*, **79**, 343–363.
- Mills, E. S. 1967. An aggregative model of resource allocation in a metropolitan area. *American Economic Review*, **57**, 197–210.
- Mion, G., and Naticchioni, P. 2009. The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, **42**, 28–55.
- Mohring, H. 1972. Optimization and scale economies in urban bus transportation. *American Economic Review*, **62**, 591–604.
- Moretti, E. 2004. Estimating the social returns to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, **121**, 175–212.
- Moretti, E. 2012. *The New Geography of Jobs*. Boston: Houghton Mifflin Harcourt.
- Muth, R. F. 1969. *Cities and Housing*. Chicago: University of Chicago Press.
- Ogawa, H., and Fujita, M. 1980. Equilibrium land use patterns in a nonmonocentric city. *Journal of Regional Science*, **20**, 455–475.
- Parry, I., and Bento, A. 2001. Revenue recycling and the welfare effects of road pricing. *Scandinavian Journal of economics*, **103**, 645–671.
- Parry, I., and Small, K. A. 2005. Does Britain or the United States have the right gasoline tax? *American Economic Review*, **95**, 1276–1289.
- Parry, I., and Small, K. A. 2009. Should urban subsidies be reduced? *American Economic Review*, **99**, 700–724.
- Picard, P., and Okubo, T. 2012. Firms' locations under demand heterogeneity. *Regional Science and Urban Economics*, **42**, 961–974.
- Proost, S., and Van Dender, K. 2012. Energy and environment challenges in the transport sector. *Economics of Transportation*, **1**, 77–87.
- Puga, D. 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science*, **50**, 203–219.

- Quinet, E., and Raj, A. 2015. Welfare and growth effects in transport and trade – a practitioner's point of view. Report prepared for the World Bank.
- Redding, S. J. 2015. Goods trade, factor mobility and welfare. Princeton University, mimeo.
- Rosenthal, S. S., and Strange, W. C. 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J. V., and Thisse, J.-F. (eds), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier, pp. 2119–2171.
- Schmidheiny, K., and Südekum, J. 2015. The pan-European population distribution across consistently defined functional urban areas. *Economics Letters*, **133**, 10–13.
- Small, K. A., and Verhoef, E. T. 2007. *Economics of Urban Transportation*. London: Routledge.
- Solé-Ollé, A., and Viladecans-Marsal, E. 2012. Lobbying, political competition, and local land supply: Recent evidence from Spain. *Journal of Public Economics*, **96**, 10–19.
- Stiglitz, J. 1977. The theory of local public goods. In: Feldstein, M. S., and Inman, R. P. (eds), *The Economics of Public Services*. London: Macmillan, pp. 273–334.
- Teulings, C. N., Ossokina, I. V., and de Groot, H. L. F. 2014. Welfare benefits of agglomeration and worker heterogeneity. CESifo Working Paper Series No. 4939.
- Tirole, J. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.
- Topa, G., and Zenou, Y. 2015. Neighborhood and network effects. In: Duranton, G., Henderson, J. V., and Strange, W. (eds), *Handbook of Regional and Urban Economics*, vol. 5. Amsterdam: Elsevier, pp. 561–624.
- Turner, M. A., Haughwout, A., and Van Der Klaauw, W. 2014. Land use regulation and welfare. *Econometrica*, **82**, 1341–1403.
- Ushchev, P., Sloev, I., and Thisse, J.-F. 2015. Do we go shopping downtown or in the 'burbs'? *Journal of Urban Economics*, **85**, 1–15.
- van Goeverden, C., Rietveld, P., Koelemeijer, J., and Peeters, P. 2006. Subsidies in public transport. *European Transport / Trasporti Europei*, **32**, 5–25.
- Vickrey, W. S. 1969. Congestion theory and public investment. *Papers and Proceedings of the Eighty-first Annual Meeting of the American Economic Association*, 251–260.
- Winston, C., and Mannering, F. 2014. Implementing technology to improve public highway performance: A leapfrog technology from the private sector is going to be necessary. *Economics of Transportation*, **3**, 158–165.
- Wolff, H. 2014. Keep your clunker in the suburb: Low emission zones and adoption of green vehicles. *Economic Journal*, **124**, F481–F512.