UC Office of the President
Recent Work

Title
Discovering the Efficiency of Urban Sprawl

Permalink
https://escholarship.org/uc/item/02r068dg

ISBN
9780195380620

Author
Anas, Alex

Publication Date
2011-12-12

Peer reviewed

CHAPTER 6

DISCOVERING THE EFFICIENCY OF URBAN SPRAWL

ALEX ANAS

> The hypotheses of individuals are tested, the commitments shared by his group being presupposed; group commitments, on the other hand, are not tested, and the process by which they are displaced differs drastically from that involved in the evaluation of hypotheses.
>
> —Thomas S. Kuhn (1977, xxi)

Sprawl for Planners and Economists[1]

Planners see U.S. urban areas as too spread out and land use densities as too low. Cars and highways are blamed for locking in this pattern of land use. Remedies aim to contain suburban expansion, or transit-oriented developments are proposed to

increase compactness. Urban economists studied traffic congestion in the mono-centric model of land use and also concluded that sprawl is excessive and that urban growth boundaries could be beneficial. In contrast, recent theory showed that with mobile jobs or more than one city, more sprawl often occurs when the congestion externality is optimally remedied, and that second-best policies may require expansive, not restrictive, growth boundaries. A computable general equilibrium model of Chicago with endogenous road congestion shows that the interdependent suburbanization of jobs and residences keeps the average trip time remarkably stable by improving residence-to-job proximity in a sprawling city with growing population and gross product. Indeed, national travel data from 1990 and 2000 show that a U.S. urban area that is twice as big as another has, on average, only 10 percent longer commutes. A simple model of Beijing's congestion shows that some beneficial transit-oriented developments work by increasing sprawl, not by reducing it.

Planners view sprawl as the aggregate extent or some other feature of urban land use, while economists would like to explain sprawl as arising from inefficiencies in resource allocation caused by unpriced traffic congestion or other market failure. So planners view sprawl from a descriptive perspective, whereas economists view it from a normative one. Haphazard land development is one possible definition of sprawl.[2] But although discontinuous development or leapfrogging gives an appearance of haphazardness, it arises in perfect foresight models of land development without market failures (Ohls and Pines 1975; Fujita 1976; Moore and Wiggins 1990). Irwin and Bockstael (2007) write about sprawl as fragmentation in land development at the urban fringe caused by externalities among heterogeneous land uses. Galster et al. (2001) attempt a much more complex definition of sprawl involving the concepts of *density*, *continuity*, *concentration*, *clustering*, *centrality*, *nuclearity*, *mixed uses*, and *proximity*.

The simplest and most commonly held notion is that there is excess urban sprawl because population or job densities are below optimal levels on average, causing the aggregate urban land use to be above optimal. Inefficiencies in resource allocation can occur due to a variety of market failures causing departures from perfect competition, or because of tax-induced distortions. Examples of prominent market failures in urban economies are unpriced traffic congestion, uncompensated agglomeration effects, large-lot suburban zoning, unmitigated neighborhood and neighbor externalities that may result in flight from the central city, and deviations of infrastructure pricing from marginal cost. Property taxation and the

    2. The Random House Dictionary definition is "to be stretched or spread out in an unnatural or ungraceful manner." and www.dictionary.com defines sprawl as "haphazard growth or extension outward, especially that resulting from real estate development on the outskirts of a city."

subsidization of homeownership in the income tax code are examples of tax-induced distortions. A difference between market and optimal sprawl can arise due to any of these causes, since they all affect land use directly or indirectly, but it is not at all obvious that optimal sprawl measured as total land area is less than market sprawl as is commonly believed. The only possible way to see what the separate or combined effects of the many potential causes of excess sprawl might be is to build a well-calibrated empirical general equilibrium model that treats *all* relevant effects.

In this article, by *sprawl* I will refer only to the aggregate land area in urban use. By *excess sprawl*, I will refer to the inefficient aggregate land area that occurs because of the market failures, minus the optimal aggregate land area that would occur if the market failures were corrected. I will focus on only one market failure, that arising from the negative externality of unpriced road congestion.

The chapter is organized as follows. The next section is a brief review of how unpriced traffic congestion causes excessive travel, and how Pigouvian taxation would restore efficiency by eliminating the excessive travel. Then, I ask whether unpriced traffic congestion causes not only excessive travel but also excessive urban sprawl. First, we see why monocentric theory that is a standard tool in urban economics reaches the conclusion that sprawl from un-priced traffic congestion is indeed excessive as is commonly believed. Then, we see how this conclusion has been reversed in recent theoretical models of job dispersion and polycentric urban land use that, in a number of ways, relaxed the strict and unrealistic assumptions of the monocentric model. These new results, based on polycentric theory, have helped discover the efficiency of urban sprawl, hence the title of the chapter: that excess urban sprawl can often be negative, not positive as many planners and policy makers have believed. While the truth lay in the dark, urban economists kept looking for it under the monocentric lamppost, thus having no reason to disagree with the belief of many planners that there is too much sprawl.

Next, I turn to turns to key empirical facts about urban areas that any theory of the metropolis that explains job dispersion and polycentric forms would need to face in order to be relevant, and presents a statistical analysis of the city size's effect on average commutes. The key fact is that American urban expansion has not resulted in a large increase in commuting time, often presumed to increase with sprawl and congestion. I argue that this is primarily because of the simultaneous decentralization of jobs and housing just as the theoretical models, reviewed earlier, explained. I also report simulations from an empirical computable general equilibrium (CGE) model of the Chicago Metropolitan Statistical Area (MSA) to explain the relationship between sprawl, commute times, public transit, and employment decentralization. The simulation results are consistent with the empirical facts and indeed explain them well. The Chicago simulations confirm the benign effect of sprawl on travel times under a scenario of economic and population growth with and without highway expansion. In these simulations, sprawl increases a great deal, but travel times per person remain stable as observed in the U.S. data over the decades. The observed trend is explained by the shortening of the average distance between homes and jobs (and

homes and shops) in response to an increase in congestion per mile. It is the simultaneous and interdependent suburbanization of jobs and residences that makes stable travel times possible.

Next, I provide an assessment of some of the policy and regulatory remedies planners have often proposed under the presumption that there is excessive sprawl. I explain, based again on the same theoretical articles that were reviewed, why restrictive urban growth boundaries (UGBs) are poor substitutes for congestion pricing and why expansive growth boundaries may be second-best policies when optimal sprawl is larger than market sprawl. These boundaries would improve efficiency by helping urban areas expand at the expense of agriculture. I also comment on transit-oriented development (TOD) and the New Urbanism as possible antidotes to highway-oriented expansion. Although TOD and the New Urbanism play an important role serving niche markets, these planning and urban design movements are not necessarily antisprawl policies. Indeed, results from an empirical model of Beijing show that some TOD proposals might improve efficiency by increasing sprawl. The tools of the New Urbanism become more relevant if used to support polycentrism and job dispersion in the midst of urban expansion and continuing sprawl. In the last section, I conclude briefly by remarking on how the discovery of the efficiency of sprawl should change our approach to understanding and managing urban land use.

The Externality of Unpriced Congestion

That roads, the automobile, and urban sprawl are joined at the hip is widely understood. Dunphy (1997), Glaeser and Kahn (2004), Downs (2004), and Nechyba and Walsh (2004) have all suggested that car-dependence is responsible for sprawl, while the web page of the Sierra Club blames sprawl for car dependence:

> It is difficult to imagine large increases in suburbanization without this rise of the automobile, even if other causes have contributed to the sprawling of cities in the presence of the automobile. (Nechyba and Walsh 2004, 182)
>
> Sprawl spreads development out over large amounts of land; puts long distances between homes, stores, and job centers; and makes people more and more dependent on driving in their daily lives.…Sprawl lengthens trips and forces us to drive everywhere. The average American driver currently spends the equivalent of 55 eight-hour workdays behind the wheel every year. (Sierra Club)

Viewing reality initially through the narrow prism of the standard model of urban economics, the perfectly competitive monocentric model of land use without traffic congestion, urban economists in the 1960s and 1970s concluded that a reduction in

the cost of commuting per mile would indeed cause cities to sprawl outward but without excess (e.g., Wheaton 1974).[3]

Indeed, Burchfield et al. (2006) recently used satellite photography to estimate that a measure of developed land in the lower forty-eight states grew from 1.3 percent in 1976 to 1.9 percent in 1992.[4] The implied annualized growth rate is 2.48 percent, 2.5 times the population growth rate in the same period. Income growth causes suburbanization to proceed via several channels. The demand for larger housing and private yards rises, drawing households to locate peripherally where land is cheaper and larger lots more affordable. Car ownership increases with income, as documented by Ingram (1998), and the demand for discretionary trips also increases, as documented by Nelson and Niles (2000), contributing to a higher demand for owning cars. But none of these trends, in and of themselves, implies any excessive sprawl.

To understand urban sprawl as excess, that is, as resource misallocation, we should look at the results from a monocentric model *with* traffic congestion. Such models were in vogue in the late 1970s and early 1980s, and a well-articulated recent version is due to Wheaton (1998). Unpriced congestion causes a negative externality distorting how markets allocate resources. The time delay and fuel consumption externalities of congestion remain unpriced, and thus trips are too cheap and travel excessive. Marginal cost pricing as originally conceived by Pigou (1932) is the well-known remedy and has been advocated for traffic by Walters (1961) and by Vickrey (1963). In the absence of such pricing, road planners who respond to the excessive travel demand might well have overbuilt roads as observed by Kraus et al. (1976) and by Wheaton (1978). The extra road capacity may have induced even more socially excessive car travel.
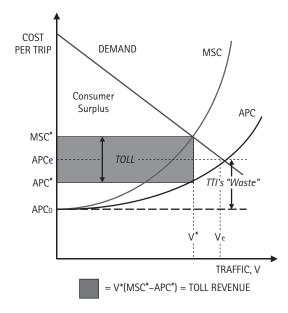
Travel misallocation due to unpriced congestion and its Pigouvian remedy are shown in figure 6.1.[5] The diagram applies to traffic on a single one-mile-long road. Once enough traffic is on the road, the average private cost (*APC*) of each driver increases convexly with the additional traffic, and the marginal social cost (*MSC*) is above the *APC*. The vertical difference between the two curves is the external

3. In the model all jobs are pinned at the city center, the CBD (central business district), while workers reside on private lots spread out in a circle around the CBD and commute. The rent on land falls with distance from the CBD, the city reaching a border where the urban and farming rents are equal. Lots get bigger and residential density falls with distance from the CBD. Single-centered cities had begun to evolve into polycentrism about 60 years before Alonso (1964) and Strotz (1965) formulated this model.

4. They measure sprawl by the average share of undeveloped land in the adjacent square km around each of the 8.7 billion 30×30 m cells of residential use in the U.S. and the change in 1976–1992. By this measure, New Jersey, the most developed state grew from 18% to 21% developed, and Florida, the fastest growing, from 4.4% to 9%.

5. The cost of a trip (vertical axis in figure 6.1) includes the monetary value of travel time plus the monetary cost of driving (fuel etc).

Figure 6.1 Congestion pricing

congestion cost a car imposes on all other cars. The equilibrium traffic flow when congestion is unpriced occurs at the intersection of the demand curve and the *APC*, but the socially optimal traffic occurs at the intersection of demand and the *MSC*. To eliminate the excess traffic, $Ve - V^\star$, a toll is charged on top of the *APC*. This toll is equal to the vertical gap $MSC^\star - APC^\star$ at the optimum traffic. When such a toll is charged, welfare measured as the sum of consumers' surplus and toll revenue is maximized.[6] If the toll is charged, then each car bears the cost of the delays it imposes on all of the other cars. But here we use tolling as a device for finding the social optimum, not as a policy. We are merely interested in assessing whether sprawl is excessive, not in advocating congestion pricing.

How More Sprawl Can Be Efficient

The diagrammatic analysis of figure 6.1 shows that there is excess traffic on the mile-long road, but does this imply that there will be excessive use of land as well?

---

6. A proof written in the form of an appendix is not included here, but is available from the author.

The Monocentric Model

As mentioned, one answer to this question has been given under the assumptions of the monocentric model. Wheaton (1998) expressed the essence of this model: "Because driving and location are equivalent, tolling congestion is the same as regulating density.…Simulations suggest optimal cities should have densities that are orders of magnitude greater than market cities" (258). In the monocentric setting everyone commutes to the CBD, and indeed "driving and location are equivalent." Since the Pigouvian toll levied per mile would depend on the gap between the *MSC* and the *APC*, congestion and the toll would be higher where more traffic flows per unit of road capacity. It is assumed in the model that all workers located at different distances from the CBD arrive at work at about the same time. Thus, congestion increases as commuters, traveling on a radius, approach the CBD. The farther away one's residence is from the CBD, the more one travels, and the higher, therefore, the level of cumulative Pigouvian tolls one must be charged on the miles traversed. In the standard monocentric model closed in total population, jobs are counterfactually pinned in the CBD and there is no public transit, no nonwork trips, and so on. There is only one margin for the commuters to exercise toll avoidance: by moving to a residence closer to the CBD. This will reduce their after-toll travel cost. Since all commuters can adjust only in this way, the land nearer the CBD becomes pricier, and all rent smaller lots as they relocate toward the CBD. The equilibrium with tolls is socially optimal, absent other externalities. Thus, indeed, in the context of the monocentric model closed in population, densities per acre increase, lot sizes decrease, and the congestion externality falls to its socially optimal level. The optimum monocentric city is more compact than the one with unpriced congestion, and total sprawl measured as the aggregate area is smaller. Urban economics in the 1970s and 1980s went on to show that land allocation in the monocentric city is such that the optimal toll revenue just pays for the land used by roads.[7]

The devotion of urban economists to the monocentric model has been remarkably durable. Often, results from the monocentric model are boldly extrapolated to reality and may even shape thinking on some policy issues. Brueckner (2000), in an influential policy-oriented article written for a broad audience, asserted that the congestion externality, when unpriced, would cause cities to occupy too much land. This conclusion that efficient urban forms are more compact may be psychologically pleasing. But, in science, a powerful and elegant result is only as strong as the assumptions on which it rests. The Achilles' heel of the monocentric model is that it disregards margins of adjustment that are important in reality. Recent polycentric theory and models with dispersed jobs showed that optimum urban form is often more spread out than if congestion remained unpriced. In many circumstances,

7. The result holds assuming that the congestion technology, a function of road capacity and traffic flow, is constant returns to scale. Under decreasing returns, roads yield a profit that must be distributed to the population and under increasing returns there is a deficit that is made up by a head tax.

more urban sprawl may be key to lower or stable average travel costs, lower congestion, and economic efficiency.
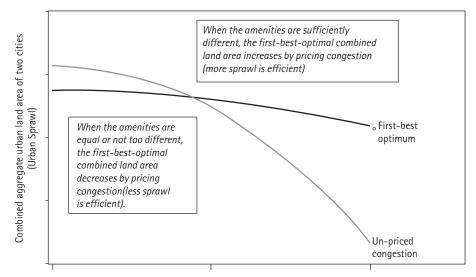
Two Asymmetric Monocentric Cities

In Anas and Pines (2008), a system of two monocentric cities with congestion is treated, one of them with more workers. The larger population is caused by a higher citywide amenity that can be enjoyed only by living in that city. A fixed population of identical workers is allocated within and between the cities. Each must work in her city of residence but can relocate both job and residence to the other city at zero cost. At equilibrium, with unpriced congestion, the more populous city (with the higher amenity) is the more congested. Aggregate land rents and congestion tolls are distributed equally in the combined population. Is it true in this two-city system that imposing congestion tolls in each of the two cities will always result in higher density in each city, lower total transport costs, and lower total land area as predicted by the one-city monocentric model? Two opposing effects exist, and their balance determines the outcome. On the one hand, congestion tolls induce commuters to relocate closer to their CBDs.[8] Doing so, one reduces one's lot but may increase other consumption. This intracity effect was unopposed in the single-city model. On the other hand, some commuters in the larger and more congested city can avoid the higher tolls on them by relocating to the smaller city. This is the intercity effect. Moving to the smaller city, one enjoys lower rents and congestion and a larger lot but lower amenity. Under optimal tolling, the aggregate commuting costs of both cities always decrease. If the intracity effect is more powerful, then the sum of the two cities' land areas also decreases. But if the intercity effect dominates, the reverse result is obtained, and at the optimum there is more urban sprawl, not less. The smaller city grows while the larger shrinks in workers and may also shrink in land, but the combined land area increases.

What determines which effect dominates? If the two amenities are identical, then the cities are also identical and nothing can be gained by moving from one city to the other. Only the intracity effect remains, and the optimal land use is less sprawled than when congestion is unpriced. Anas and Pines showed by a mix of proofs and precise numerical solutions that the intercity effect is stronger when, keeping all else constant, the consumer's elasticity of substitution between lot size and other goods is low enough. Otherwise, suburbanites in the higher amenity city prefer to move to their core, where they can substitute other goods for land and continue to enjoy the higher amenity. There are broad circumstances where the unpriced congestion causes too much population to be in the larger cities, with too few people living in the smaller ones. To reach optimal welfare and lower

8. They move from the congested suburbs to their city's core area where it is assumed there is no congestion (e.g., they can walk to work or use transit). In this model, the city cores are a transit oriented development.

Figure 6.2 When more sprawl can improve efficiency

congestion, workers must be shifted to smaller cities, increasing aggregate sprawl by doing so. More land should be in urban use in the optimal allocation than in the allocation with unpriced congestion. This result is shown in figure 6.2. The two curves, one corresponding to the first-best optimal land use in which congestion is priced and the other to unpriced congestion, show that the combined land area of the two cities decreases as the gap between the two citywide amenities is larger. But for a small amenity gap excess sprawl is positive, that is, the two optimally configured cities require less land than in the case of unpriced congestion. But for a larger amenity gap, the result is the opposite: the optimal allocation requires more, not less, sprawl.

Job Suburbanization

That more urban sprawl is more efficient also arises by recognizing that jobs can leave the CBD, as they have been doing for decades. Anas and Rhee (2007) studied a model with just two areas: the city and the suburbs. Initially, all jobs are in the city's center (CBD), and congestion is unpriced. They assumed that the productivity of a job did not depend on its location, similar to the assumption of Anas and Pines. But in this case, tolling can cause jobs to relocate to the suburbs, reducing average commuting distance. If this effect dominates, some jobs move to the suburbs, and this reduces congestion but increases sprawl. Once again, more urban sprawl, not less, is optimal. In this case, the efficiency of sprawl comes about as commuting cost and the congestion externality are lowered by job suburbanization, something we will see in more detail.

The Dispersed City

Anas and Rhee (2006) treat urban decentralization in a general equilibrium polycentric city model based on Anas and Xu (1999), in which jobs and residences are free to locate anywhere and mix everywhere. The city is divided into eleven distinct locations with idiosyncratic features that horizontally differentiate production and consumption. At the same time, the distance between any two areas causes them to be vertically differentiated in production and consumption because firms or residents at a location interact more with residents or firms at a closer location than at a more distant one. Each location is an imperfect substitute for any other in production, shopping, and for residences. The model is closed in population, income, and trade. Production, residences, and roads use land. At equilibrium, there are jobs and residences in any location, and a worker can choose any residence and work location pair based on both systematic and idiosyncratic preferences. Producers in any location are perfectly competitive and produce a composite product unique to that area and offer jobs there. Workers value access to their jobs and to all locations for shopping. Shopping in the model is realized by making a trip from one's home location to any location of production. Workers shop in all locations, since the products made in different locations are viewed as imperfect substitutes and workers have a taste for product variety. Producers locate so that they are accessible to workers and customers. If not, then all else being equal, producers would have to pay higher wages. In annular rings around the geometric center, area increases with distance from the center. It is assumed also that all travel follows the radial direction. The mutual interdependence of producers and consumers results in an equilibrium allocation in which production and residences are mixed in all locations, with employment and residential density and road congestion peaking at the geometric center. With distance from the center, job density declines more rapidly than does residential density. Rents decline with distance, and wages increase, reflecting the substitution of land for labor, which increases the marginal product of labor with distance. At the equilibrium, all 121 (11 by 11) commuting pairs are used by some workers, and each worker shops in all 11 areas. While the dominant commuting direction in equilibrium is toward the geometric center, there are reverse commuters whose residence is more central than their job, and commuters who reside on one side and work on the other side of the center.

Several properties of this model are important to our understanding of sprawl and its effects. Suppose that all the jobs are counterfactually forced to the center, mimicking a monocentric city. Beginning from such a state, sprawl is unleashed by letting jobs disperse freely among all locations to reach a new general equilibrium. The authors did not include an alternative use for land at the urban fringe and assumed instead extension to a natural boundary, with all the available land always being utilized. Hence, sprawl is measured by its economic cost on trips, as daily average travel time per worker (DATT) and by job and residential density. When the city is monocentric, all trips go to the center, and DATT is 63 minutes (44

commuting, plus 19 for shopping trips). When CBD jobs disperse to all zones, the DATT increases from 63 to 69 minutes (by 9.5 percent), commute time decreasing from 44 to 40 minutes (by 9.1 percent) and shopping time increasing from 19 to 29 minutes (by 52.6 percent). The mild decrease in average commute time occurs because the dispersion of jobs out of the center brings jobs closer to residences. Job decentralization reduces average commute time. The sharper increase in the time of shopping trips is composed of two main effects. One effect is the substitution for shopping of the time saved from commutes. The second effect is that as producers disperse, the location-variety of products increases because products are indexed by the horizontally differentiated locations. This is appreciated by recalling that as real cities grow new shopping areas with distinct personality and attraction appear and location variety for shoppers improves. In the model, personal shopping trips per month are 9.25 in the monocentric city, rising to 15.29 in the dispersed. In the dispersed city 65 percent more trips are made to explore, as it were, the new product-locations, but these are shorter on average. Thus, job dispersion induces more travel time to shops because of the proliferating variety of production/shopping destinations, but this is beneficial and not in itself a sign of excess mobility. On the commuting side the model shows that job sprawl does not increase the cost of commuting, but decreases it.

How does the spatial equilibrium of such a dispersed city differ from the socially optimal allocation? Tolls are levied on an initial situation with too much land in roads in the suburbs and too little near the center. This initial misallocation of roads, as mentioned earlier, is a likely result of the unpriced congestion. When congestion tolls are levied, roads are allowed to revert to the optimal pattern, aggregate tolls replacing a head tax in paying for the rents on roads. By varying the initial deviation of the road profile from the optimal one, the authors show that the travel cost of excess sprawl measured as the DATT is reduced by 7.2 to 9.6 minutes per day per worker (about 10 percent or so). As for net residential densities, tolling in the base case causes them to increase from about 8.6 percent in the center to about 1 percent in the fringe. Net job densities also increase from about 12 percent in the center to about 1 percent in the fringe. Thus, at least by the measures of DATT and density, the dispersed city with unpriced congestion is too sprawled, but the welfare loss from this misallocation is only 0.33 percent of income.

Agglomeration Economies

In Anas and Rhee (2007) and Anas and Pines (2008), a job's productivity is independent of its location and wages are exogenous. But in Anas and Rhee (2006) the substitution of labor for land and the output price determine a job's productivity. Hence, jobs in the last model have a strong tendency to disperse. In reality jobs may be bound to each other by positive externalities, offsetting the incentive to relocate a job out of a high-agglomeration center, such as the CBD. If the productivity loss from leaving the agglomeration is high enough, the job would not be moved. Such a condition would strengthen the old argument that the response to congestion

tolling would be for consumers to move their residence closer to their job. There are several counterarguments, however.

First, in Anas and Rhee (2006), jobs have a strong incentive to disperse out of the CBD, because by moving to a more peripheral and lower-rent location, they can substitute land for labor, increasing land per worker and thus labor productivity. Any loss in the proximity advantages of being in the center is balanced against the productivity gains of dispersing. Second, Anas and Kim (1996) is a model that is similar to Anas and Rhee (2006), but it includes not only congestion but also a positive consumption externality in shopping. All else being equal, consumers in the model prefer to shop where there is a bigger concentration of producers-retailers.[9] When congestion is low enough, the resulting equilibrium job concentration is monocentric, as all producers locate at the geometric center, since travel is cheap enough and being centrally located increases agglomeration benefits. As congestion rises, the cost of travel rises, so producers disperse into peripheral subcenters increasing their proximity to labor and to customers while sacrificing the agglomeration benefits that become less important given the higher congestion. Historically, falling transport costs, the Internet, and the growing importance of intercity linkages have weakened agglomeration economies. It is unlikely that this trend will reverse. Agglomeration economies are not going to become so important again, as to bring back relevance to monocentric analysis.

Empirical Facts and Results from a CGE Model

Models mentioned so far depart from the monocentric assumptions. Although they are realistic, they are not empirical. I will now review how data support the issues treated by the polycentric models discussed previously. In particular, we will focus on what the data tell about (1) the decentralization of jobs; (2) suburban and reverse commuting and nonwork trips; (3) how congestion varies by city size; and (4) commute time variation by MSA size. We will then see how a CGE model of the Chicago MSA generates results that support these data facts.

Job Decentralization

Mieszkowski and Mills (1993) expressed the decades-long decentralization trend as follows:

> In the 1950s, 57 percent of MSA residents and 70 percent of MSA jobs were located in central cities; in 1960, the percentages were 49 and 63; in 1970, they were

---

9. This is just one of many ways to treat agglomeration economies.

Table 6.1 Commuting patterns in the United States and Canada

Commuting Patterns		United States	Canada
Residence	**Workplace**	**2000 Census**(%)	**2001 Census (%)**
Central city	*Central city*	27.5	46.1
Central city	*Suburb*	8.9	7.5
Suburb	*Central city*	20.2	16.2
Suburb	*Suburb*	43.4	30.2
		100.0	100.0

Source: Anas and Rhee (2007).

> 43 and 55; in 1980, they were 40 and 50; in 1990, they were about 37 and 45. The United States is approaching the time when only about one-third of the residents within an MSA will live in central cities and only about 40 percent of MSA jobs will be located there. (135).

These numbers show that population and jobs have decentralized apace: in 1990 the shares remaining in central cities were at 65 percent of 1950 levels.

Commuting Patterns and Nonwork Travel

Table 6.1 shows the commuting pattern in the United States and Canada. In the United States, suburb-to-suburb commuting dominates (43.4 percent share), while reverse commuting from homes in central cities to jobs in the suburbs is an importantly large 8.9 percent. Those commuting from suburban homes to jobs in the central city are only 20.2 percent. Table 6.2 shows the growing importance of nonwork trips (i.e., discretionary shopping trips and other personal business trips). In 1995, commutes were only 20 percent of all trips. Growth in vehicle miles traveled (VMT) per capita for shopping and other family and personal business has been strong and much higher than the growth in the number of trips. From 1969 to 1995, VMT per capita increased by 60 percent, largely because of the rise in discretionary trips, much less due to any lengthening of commutes. Aggregate VMT closely tracks the rise in GDP (Sorensen, 2009).

Variation of Congestion by Urban Area

The Texas Transportation Institute (TTI) measures the congestion of a commute as the commute time during the peak periods (6:00 to 9:00 A.M. and 4:00 to 7:00 P.M.) divided by the time the commute would take if traffic flowed at free-flow (uncongested) speeds (sixty miles per hour on freeways and thirty-five miles per hour on principal arterials).[10] In figure 6.1, this ratio corresponds to *APC*e/*APC*0. The TTI

10. Texas Transportation Institute, "National Congestion Tables, Table 1. What Congestion Means to You, 2007"; "Table 4. Congestion Trends – Wasted Hours (Annual Delay per Traveler, 1982 to 2007)."

Table 6.2 Trips and growth in vehicle-miles by trip purpose

Trip purpose	Person-trip distribution (%)				Percent growth in vehicle miles traveled per capita, 1969–1995
	1977	1983	1990	1995	
Non-work					
Shopping	17.1	18.0	18.9	20.2	+166.24
Other family and personal business	14.0	17.4	22.7	25.6	+171.69
Social and recreational	24.4	27.7	24.8	24.9	+9.89
Other	21.4	14.1	12.1	9.0	+14.24
All non-work	76.9	77.2	78.4	79.7	
Work and related	23.1	22.8	21.6	20.3	+59.60
All purposes	100.0	100.0	100.0	100.0	+46.05

Source: From Nelson and Niles (2000).

reports that, by this measure, the average U.S. commuter in 2007 "wasted" 36 hours per year (or 8.64 minutes per day) and 24 gallons of fuel per year, or about 25 cents daily. Among urban areas of more than 1 million people, Los Angeles is at the top of TTI's list with 70 hours and 53 gallons of waste per commuter, and the Buffalo MSA at the bottom with 11 hours and 7 gallons. By the same measure of "wasted time," TTI tables show that between 1982 and 2007 congestion increased by 263 percent in the 301 urban areas with less than 250,000 people and by 143 percent in the 14 urban areas of greater than 3 million.

The TTI's definition of these magnitudes as "waste" is misleading, since the optimal level of congestion is not zero, and so waste cannot be calculated by comparing actual congestion experienced relative to the utopian zero congestion.[11] Also, since congestion is a negative externality, the calculation of "waste," acceptable to economics, should reflect the cost the commuter imposes on others. In figure 6.1, all "wasted" or excess time and fuel would be $(Ve)(APCe) - (V^*)(APC^*)$: the aggregate cost of travel not including tolls when congestion is unpriced minus when congestion is optimally priced. Figure 6.1 makes it clear that to calculate this we need to know the demand curve. But the "waste of congestion" is best calculated as the welfare loss from forgoing optimal congestion pricing.

Commuting Time by MSA/CMSA

Changes over the decades would indicate whether congestion is building up. Gordon and Richardson have documented that the average commute time in the United States has been stable.[12] McGuckin and Srinivasan (2003) observe from census tabulations[13] that the average one-way commute time was 21.70 minutes in 1980, 22.40 in 1990, but added a surprising 3.1 minutes in the 1990s to rise to 25.50 in 2000. They claim that about one-third of the increase reflects that commutes of more than 100 minutes (one way) were coded as 99 minutes in the earlier censuses, but in the 2000 census the coding limit was raised to 200 minutes. Table 6.3 shows the commute time changes between 1990 and 2000 juxtaposed against the changes in workers for nested groupings of MSA/CMSAs. The percent increase in commute time over the decade is higher the higher is the percent increase in workers. This is what one would expect congestion to do. But the average commute times are still quite low and the increases not large and not much higher than the increase in workers. This is despite the few highway additions during the decade.

We isolated the forty-nine most populous MSA/CMSAs (by workers) for further analysis. In this sample, the MSAs as a group gained 11 percent in workers,

11. Aggregate "*TTI Waste*" = $(Ve) [APCe–APCo]$. Others often refer to this as "the cost of congestion" which is a more reasonable name than is "waste".

12. Gordon, Kumar, and Richardson (1989); Gordon and Richardson (1996).

13. See http://www.fhwa.dot.gov/ctpp/jtw/contents.htm Last viewed on February 14, 2010.

Table 6.3 Changes in commute time (all modes) and changes in worker population

	All MSAs	Most populous 49 MSAs	Top 27 of the 49 MSAs ranked by the percent change in commute time	Top 17 of the 27 MSAs ranked by the percent change in commute time
1990 one-way (min.) trip-weighted mean commute per worker	22.40	24.74	24.79	22.70
2000 one-way (min.) trip-weighted mean commute per worker	25.50	27.89	28.37	26.95
Trip-weighted mean commute time change per worker 1990–2000	13.80%	12.74%	14.47%	18.75%
Increase in workers 1990–2000	11.5%	10.44%	13.06%	20.60%

Source: From McGuckin and Srinivasan (2003).

average commuting time across all of these forty-nine MSAs increased by 13.6 percent, and transit ridership share dropped 27.2 percent in the decade. We estimate the cross-sectional elasticity of commute time with respect to MSA workers and the percent using transit. The regression is

$$\ln\left(\begin{array}{c} \textit{Average one way} \\ \textit{commute time}\,(\textit{min}\,/\,\textit{day}) \end{array}\right) = a + b_1 \ln\left(\begin{array}{c} \textit{Workers} \\ > 16\,\textit{years} \end{array}\right) + b_2 \ln\left(\begin{array}{c} \%\ \textit{transit} \\ \textit{ridership} \end{array}\right)$$

where b_1, b_2 are the desired elasticities. The results are shown in table 6.4 and figure 6.3. There is no great change in the elasticities between 1990 and 2000. An MSA/CMSA with 10 percent more workers has only about a 1 percent longer commute. If 10 percent more workers commuted by public transit, the average commute would be about 0.2 percent longer, but transit share's effect is not significant.

Comparing MSA/CMSA groups over time in table 6.3 suggests a more powerful response than the one observed in the two cross sections: the forty-nine most populous MSAs added 10.44 percent more workers, and trip-weighted average commute time rose 12.74 percent. In the seventeen of these forty-nine MSAs that showed the most commute time increase, workers increased by 20.60 percent and commute times by 18.75 percent. The cross-sectional relationship of the regression is not affected much by the changes over time. Why, then, did average commute time increase by about two minutes from 1990 to 2000 (after taking out the one minute that, according to McGuckin and Srinivasan [2003], is due to census count limits)? An explanation based on income is proposed in a recent article by Lee et al. (2009)
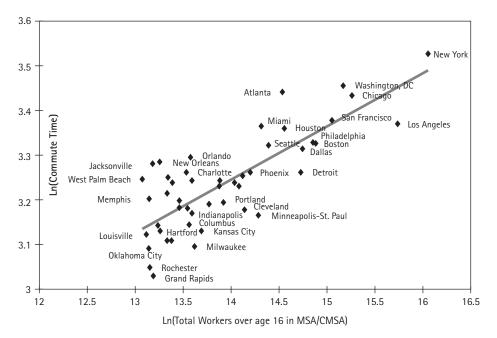

Figure 6.3 Commute time and MSA/CMSA size (2000)

Table 6.4 Elasticity of MSA commute times with respect to workers and transit use

Year	Sample means (49 most populous MSAs by workers)			Constant	Elasticity		R-squared (%)
	Commute (minutes)	*Workers (millions)*	*Transit (share,%)*		**ln (Workers)**	**ln (%Trans.)**	
1990	22.50	1.38	5.37	+1.5697 (10.32*)	+0.1117 (10.13*)	—	68.6
				+1.7723 (8.99*)	+0.0947 (6.21*)	+0.0261 (1.58**)	70.2
2000 (% change 1990–2000)	25.57 (+13.6%)	1.532 (+11.0%)	3.91 (−27.2%)	+1.5792 (8.55*)	+0.1190 (8.97*)	—	63.2
				+1.7129 (6.96*)	+0.1083 (5.81*)	+0.0152 (0.83***)	63.7

(t-scores in parentheses;*significant at 1% or lower;**significant at 12%; ***insignificant)

and a different one based on demographic changes in Kirby and LeSage (2009). Using urbanized area averages, Lee et al. study the claim that the roughly 10 percent higher average commute time in the decade is associated with a 3 percent increase in average incomes. They conjecture that the increase in incomes has caused an increase in nonwork trips during morning hours, which increased congestion and therefore average commute times. This is what would be predicted in the Anas and Xu (1999) and Anas and Rhee (2006) models. But although average income is insignificant in their 1990 regression, but highly significant and ten times more powerful in their 2000 regression, it explains only a small part of the increase in commuting time in part because average incomes increased by only about 3 percent in the decade. The change in the slope of average income could be the result of the top coding of commutes in 1990 but not in 2000 (as McGuckin and Srinivasan [2003] have pointed out). Although Lee et al. (2009) are aware of this, they cannot control for it in their regressions. The great change in the slope of income that they find could have been caused by the high colinearity between the average income of the urbanized area and the number of multiworker households (.73 correlation in 2000). Lee et al. (2009) show that the total change in all of their independent variables together explains only about 11.7 percent of the total change in average commute times during the decade. Kirby and Le Sage (2009) counter the explanation offered by Lee et al. with a study of their own based on census tract level aggregation. They suggest that changes in household demographics may be responsible. An alternative possible explanation is that the increase in average commute times is due to new traffic bottlenecks that might have emerged as population increased during the decade, as Downs (2004) has claimed, while traffic capacity did not increase much. Although employment continued to suburbanize during the decade much as it did in earlier decades, new bottlenecks due to the population increase's nonlinear effect on congestion could have caused commuting times to have increased by 10 percent or so despite the adjustment in job locations. Statistical analysis, although clearly useful, has its limits as an explanatory tool, which underscores the importance of using theoretically structured CGE models to understand in a much better and complete way the interplay between congestion, urban form, income, nonwork trips, and other factors.

Results from a CGE Model

The reciprocal effects of decentralization on congestion and of congestion on decentralization were studied in longitudinal simulations with RELU-TRAN, a CGE model of the Chicago MSA (Anas and Liu 2007) similar in structure to but more elaborate than Anas and Rhee (2006). The main results[14] are (1) if, counterfactually, all traffic flowed uncongested, the share of CBD jobs would increase by 14 percent, indicating that the rise in congestion in the central area over time has

---

14. These results will soon be reported in detail in another article.

caused job decentralization as demonstrated by the theoretical models of dispersed employment of Anas and Rhee (2006, 2007) or the endogenous subcentering models of Anas and Kim (1996) reviewed earlier; (2) rough Pigouvian congestion tolls on major roads cause Chicago's CBD jobs to increase by 7 percent, returning to the CBD half the employment that was drained by congestion; and (3) when MSA workers increase over thirty years, in the context of export-driven growth, regional gross product rises and more residential sprawl occurs in the suburbs. Despite the substantial expansion in land, the travel time per worker (for commuting and shopping) remains little changed. Since existing road capacity is not increased in this simulation, the growth causes suburban congestion per mile to increase. If consumers did not change their shares of location choices, trips, and trip lengths, their travel times would increase. This does not happen because in the new equilibrium jobs and residences adjust by relocating closer to each other on average, trips getting shorter to make up for the higher congestion per mile. Aggregate car VMT increases because workers increase, but VMT per trip decreases.[15] The higher congestion per mile also causes transit ridership to increase in some places, but the effect is offset by car trips shortening as jobs and residences spread to the suburbs apace. These adjustments enable cars to compete effectively with public transit where it is available, in the face of increased average congestion per mile. The result: more urban sprawl does not increase the cost of travel as measured by VMT or travel time per car trip. Rather, sprawl serves as a subtle safety-valve mechanism that enables the cost of travel to remain resilient and stable while the region adds population and its appetite for mobility rises with income.

The Cost of Drastic "Sprawl Remedies": The Urban Growth Boundary

Planners have been prone to recommend drastic land use controls to contain what they think is excessive and wasteful sprawl. For example, New Jersey and Maryland have favored aggressive restrictions on urban land use expansion. Portland, Oregon, and Boulder, Colorado, have growth boundaries. Greenbelts have been used in London and are a part of the United Kingdom's Town and Country Planning System (Cheshire and Sheppard, 2002), as they are also in Oregon's land use law, they have been used in Seoul, Korea (Lee and Linneman, 1998; Son and Kim, 1998), and in Moscow, New Delhi, Ottawa, and Tianjin.

Urban economists in the late 1970s and early 1980s discovered not only congestion tolling and its effects on the monocentric city but also that an urban growth

---

15. Also, VMT per worker increases, all else being constant, when the workers' incomes increase, because more shopping trips occur to buy normal goods, and also when road capacity additions are made.

boundary that constricts the radius of a monocentric city with unpriced congestion would be a second-best policy if tolls were not an available instrument. Kanemoto (1977), Arnott (1979), and Pines and Sadka (1985) illuminated this issue. The intuition behind this strong result is simple. Since tolls in the monocentric city work by inducing residents to move closer to the CBD, a UGB that is not too stringent would do the same, albeit imperfectly. Bento et al. (2006) were the first to study antisprawl policies, and they used the monocentric model. But what about the effectiveness of the UGB policy in the polycentric models or in the models with dispersed jobs discussed earlier.

Recall that in Anas and Pines (2008), the first-best policy of tolling congestion could cause the large city to shrink in area but the small one to expand, their aggregate land area increasing as the aggregate externality and aggregate transport costs are reduced. The authors went on to show that a modified UGB policy would indeed be a second-best policy in such cases, but only by placing a restrictive UGB around the larger city and an expansive one around the smaller city. Furthermore, such a system of coordinated UGBs achieves second-best efficiency by increasing, not by decreasing, the aggregate land area of both cities, just as that is also what the congestion tolls do in the first-best case. In Anas and Rhee (2007), the authors showed that when Pigouvian tolling of congestion causes jobs to move to the suburbs and this increases the suburban land area, then a second-best optimal policy is to impose an expansive growth boundary, again mimicking the effects of the first-best policy.

Anas and Rhee (2006) were the first to show that the restrictive UGB in the city with dispersed jobs and residences mixed throughout was not a second-best instrument in the absence of tolls. Indeed, their simulations suggested that an expansive UGB would be a second-best instrument. Only when the city's residents regarded the greenbelt carved out at the urban fringe by the UGB as a sufficiently valuable pure public good were the authors able to show that a restrictive UGB would create net benefits. In the same article, the authors evaluated just how harmful the restrictive UGB can be in a general equilibrium, in those cases when it is not a second-best instrument. The restrictive UGB makes an urban area more compact instead of pricing congestion. Then trips are shortened somewhat as the UGB squeezes jobs and residences into proximity, reducing some congestion by cutting trip distances, but congestion per mile rises. More important, aggregate land rents rise dramatically due to the artificial supply limitation created by the UGB, which works in the land market to correct the congestion inefficiency that arises in the travel market. The higher rents make production and consumption more expensive, and the deadweight losses can be huge, despite the fact that the rent increases are taxed and redistributed. This is only part of the story, however. In a dynamically growing city, the UGB indirectly confiscates the option values of the farmers who owned land that remains outside the UGB, adding to the deadweight loss.

According to this revised understanding of the UGB, not only is it often necessary to allow an urban area to create more land sprawl in order to optimally reduce the excessive cost of travel caused by congestion, but often, optimal UGBs should be a set of

restrictive and expansive UGBs according to a city's size. Such controls would often manage congestion best by increasing urban sprawl, not by reducing it.

Transit-Oriented Development and the New Urbanism

The low-density sprawling suburbs fostered by the car have handicapped public transit investments in the United States. Table 6.5 shows the numbers on the dramatic decline of public transit ridership in the United States. Since 1960, the number of workers almost doubled, and private vehicles per person more than tripled, but transit commuters have decreased by 22.3 percent, and the national share choosing transit to commute fell under 5 percent by the year 2000. Northeastern cities were built in public transit corridors supported by streetcars and later by subways or trains, fostering a transit-dependent suburbanization (Warner 1978). The adoption of the car, supported by highway expansions, reduced transit use, siphoning development from transit corridors into outlying areas or to open spaces between transit lines (Barrett 1983). Later, the suburbanization of jobs has rendered the radial transit lines serving the CBDs sparsely utilized, unprofitable, and subsidized. These long-term processes have locked in a land use pattern that is costly to reverse. At best it would take massive economic change or a giant policy push to bring back transit on a large scale.

Ewing (1997), Cervero (1998), Dittmar and Ohland (2003), and Downs (2004) defend well the value of compact, high-density cities supported by rapid transit. But for transit-oriented development (TOD) to take hold in the United States, a considerable share of metropolitan jobs should revert to the CBD and the central city or to selected suburban employment subcenters. It is not clear whether TOD itself can do that. The simulations with RELU-TRAN discussed earlier showed that traffic congestion has driven these jobs away, but congestion pricing can bring some of them back. But for TOD to work, policies would have to be pursued thatcould concentrate housing, retail, and office space in high-density corridors. In Greater Copenhagen's "finger plan," for example, development is required to concentrate along railway networks and radial expressways. Large office buildings must be located within 600 meters of train stations to induce more walking, and green wedges between corridors may not be converted to urban use (Danish Ministry of the Environment 2007). Some TOD planners have argued in favor of traffic calming (using urban design to discourage or slow down driving in downtowns) and even banning cars from city cores. The tools advocated by TOD planners thoughtfully combine the ideas of constraining suburban development, improving transit in urban cores, and making driving less attractive but without pricing congestion.

Table 6.5 Trends related to public transit's decline in the United States

Percent change in	1960–1970	1970–1980	1980–1990	1990–2000	Cumulative 1960–2000
Urban population	+19.5	+11.6	+12.0	+18.9	+77.5
Workers	+18.9	+25.7	+19.1	+11.5	+98.4
Vehicles per worker	+44.3	+64.2	+17.4	+17.0	+225.6
Public transit commuters	–16.6	–7.8	–2.0	–0.0	–22.3
Mode choices in commuting (%)	**1960**	**1970**	**1980**	**1990**	**2000**
Private vehicle	69.5	80.6	85.9	88.0	87.9
Public transit	12.6	8.5	6.2	5.1	4.7
Walk and other	10.4	7.4	5.6	3.9	4.1
Work at home	7.5	3.5	2.3	3.0	3.3

Source: From McGuckin and Srinivasan (2003).

Urban economists have not sufficiently studied the economics of public transit within urban models and need to do more. Following the planners, they could study European and Asian settings where transit has been more successful. Anas and Timilsina (2010) recently studied congestion and carbon dioxide emissions from cars in a simple core-periphery model of Beijing where in 2005 only 20 percent of commuters drove at an average speed of fourteen to eighteen kilometers per hour, 35 percent used transit (mostly bus), and 45 percent walked or bicycled. In this situation, improving transit in the core does not provide an effective alternative to driving because the cross-elasticity for drivers to switch to bus is too low and because many more switch from walking to bus, offsetting a large part of the gains in congestion and car carbon dioxide emissions induced by the transit improvement. Improving transit in the suburbs and handicapping cars in the core (by reducing road capacity) are more effective instruments toward a carbon-neutral policy. These relatively successful TOD policies work because they reduce core congestion and emissions by shifting some population to the periphery, increasing sprawl. This is another example of how oftentimes results that are considered to be good arise because sprawl is increased, not because it is reduced.

The New Urbanism is an urban design movement, of which TOD is often considered a part. The movement advocates the introduction of high-density, mixed-use, pedestrian-friendly neighborhoods in city and suburban settings. There is clearly demand for such developments from a niche of people who would pay a price that would allow the developers to make a normal profit. Those choosing to live in such developments in the suburbs are trading off a larger private lot and lower residential density for the accessible neighborhood amenities and the positive externalities the developers may provide for them. So far, New Urbanism developments have not been numerous enough to measurably impact the broad sweep of low-density sprawl in the United States.

CONCLUSIONS

The data on the largest U.S. MSAs show that commute times increase only slightly with city size: the elasticity of the average commute time with respect to the number of workers was about 0.1 in 1990 and 2000. The CGE simulations of Chicago explain this by showing that with jobs and population sprawling to the suburbs, the average road distance between home and job and home and shop can become shortened as congestion increases on the average mile, so that travel time per consumer remains stable. These results suggest that urban sprawl itself has not been the cause of significant travel cost increases. Meanwhile, theoretical models of urban areas with polycentric and dispersed employment show that more sprawl, not less, is often needed to offset the negative externality of unpriced congestion and improve efficiency. Planners, like urban economists devoted to the monocentric model, have

long viewed sprawl as something that should be reduced. Such a bias leads to potentially drastic planning and policy remedies of which the restrictive urban growth boundary is the prime and most costly example. A higher level of sprawl and polycentric land use may indeed be optimal. To the extent that TOD and the New Urbanism are perceived as antisprawl tools, they may be wrongly promoted. But these tools of modern planning have an important role to play in serving niche markets. Planners may do better to view them as mechanisms that will promote efficient polycentric land uses.

REFERENCES

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

Anas, A., and I. Kim. 1996. "General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration." *Journal of Urban Economics* 40:232–256.

Anas, A., and Y. Liu. 2007. "A Regional Economy, Land Use and Transportation Model (RELU-TRAN), Formulation, Algorithm Design and Testing." *Journal of Regional Science* 47:415–455.

Anas, A., and D. Pines. 2008. "Anti-sprawl Policies in a System of Congested Cities." *Regional Science and Urban Economics* 38:408–423.

Anas, A., and H.-J. Rhee. 2006. "Curbing Excess Sprawl with Congestion Tolls and Urban Boundaries." *Regional Science and Urban Economics* 36:510–541.

———. 2007. "When Are Urban Growth Boundaries Not Second-Best Policies to Congestion Tolls?" *Journal of Urban Economics* 61:263–286.

Anas, A., and G. R. Timilsina. 2010. "Lock-In Effects of Road Expansion on CO_2 Emissions in a Core-Periphery Model of Beijing." Working paper, Department of Economics, State University of New York at Buffalo.

Anas, A., and R. Xu. 1999. "Congestion, Land Use and Job Dispersion." *Journal of Urban Economics* 45:451–473.

Arnott, R. 1979. "Unpriced Transport Congestion." *Journal of Economic Theory* 21:294–316.

Barrett, P. 1983. *The Automobile and Urban Transit: The Formation of Public Policy in Chicago, 1900–1930: Technology and Urban Growth*. Philadelphia: Temple University Press.

Bento, A. M., S. Franco, and D. Kaffine. 2006. "The Efficiency and Distributional Impacts of Alternative Anti-sprawl Policies." *Journal of Urban Economics* 59:121–141.

Brueckner, J. 2000. "Urban Sprawl: Diagnosis and Remedies." *International Regional Science Review* 23:160–179.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner. 2006. "Sprawl: A Portrait from Space." *Quarterly Journal of Economics* 121:587–633.

Cervero, R. 1998. *The Transit Metropolis*: *A Global Inquiry*. Washington, DC: Island Press.

Cheshire, P., and S. Sheppard. 2002. "The Welfare Economics of Land Use Planning." *Journal of Urban Economics* 52:242–269.

Danish Ministry of the Environment. 2007. *Spatial Planning in Denmark*. Copenhagen: Agency for Spatial and Environmental Planning (www.blst.dk).

Dittmar, H., and G. Ohland, eds. 2003. *The New Transit Town: Best Practices in Transit-Oriented Development.* Washington, DC: Island Press.

Downs, A. 2004. *Still Stuck in Traffic: Coping with Peak-Hour Traffic Congestion.* Washington, DC: Brookings Institution Press.

Dunphy, R. 1997. *Moving beyond Gridlock: Traffic and Development.* Washington, DC: Urban Land Institute.

Ewing, R. 1997. "Is Los Angeles-Style Sprawl Desirable?" *Journal of the American Planning Association* 63:107–126.

Fujita, M. 1976. "Spatial Patterns of Urban Growth: Optimum and Market." *Journal of Urban Economics* 3:209–241.

Galster, G., R. Hanson, M. Ratcliffe, H. Wolman, S. Coleman, and J. Freihage, 2001. "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept." *Housing Policy Debate* 12:681–717.

Glaeser, E., and M. Kahn. 2004. "Sprawl and Urban Growth." *Handbook of Urban and Regional Economics*, ed. J. V. Henderson and J.-F. Thisse, 4:2481–2527. Amsterdam: North-Holland.

Gordon, P., A. Kumar, and H. W. Richardson. 1989. "The Influence of Metropolitan Spatial Structure on Commuting Time." *Journal of Urban Economics* 26:138–151.

Gordon, P., and H. W. Richardson. 1996. "Beyond Polycentricity: The Dispersed Metropolis, Los Angeles, 1970–1990." *Journal of the American Planning Association* 63:95–106.

Ingram, G. K. 1998. "Patterns of Metropolitan Development: What Have We Learned?" *Urban Studies* 35:1019–1035.

Irwin, E., and N. E. Bockstael. 2007. "The Evolution of Urban Sprawl: Evidence of Spatial Heterogeneity and Increasing Land Fragmentation." *Proceedings of the National Academy of Sciences* 104:20672–20677.

Kanemoto, Y. 1977. "Cost–Benefit Analysis and the Second-Best Land Use for Transportation." *Journal of Urban Economics* 4:483–503.

Kirby, D. K., and J. P. LeSage. 2009. "Changes in Commuting to Work Times over the 1990 to 2000 Period." *Regional Science and Urban Economics* 39:460–471.

Krauss, M., H. Mohring, and T. Pinfold. 1976. "The Welfare Costs of Non-optimum Pricing and Investment Policies for Freeway Transportation." *American Economic Review* 66:532–547.

Kuhn, T. S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change.* Chicago: University of Chicago Press.

Lee, B., P. Gordon, H. W. Richardson, and J. E. Moore III. 2009. "Commuting Trends in U.S. Cities in the 1990s." *Journal of Planning Education and Research* 29:78–89.

Lee, C.-M., and P. Linneman. 1998. "Dynamics of the Greenbelt Amenity Effect on the Land Market: The Case of Seoul's Greenbelt." *Real Estate Economics* 26:107–129.

McGuckin, N., and N. Srinivasan. 2003. *Journey to Work Trends in the United States and its Major Metropolitan Areas 1960–2000.* Publication No. FHWA -EP-03-058. Washington, DC: U.S. Department of Transportation, Federal Highway Administration, Office of Planning.

Mieszkowski, P., and E. S. Mills. 1993. "The Causes of Metropolitan Decentralization." *Journal of Economic Perspectives* 7:135–147.

Moore, J. E., and L. Wiggins. 1990. "A Dynamic Mills Heritage Model with Replaceable Capital." *Papers in Regional Science* 68:23–41.

Nechyba, T. J., and R. P. Walsh. 2004. "Urban Sprawl." *Journal of Economic Perspectives* 18:177–200.

Nelson, D., and J. Niles. 2000. "Observations on the Causes of Non-work Travel Growth." Paper No. 00-1242, 79th Annual Meeting of the Transportation Research Board, January 9–13.

Ohls, J. C., and D. Pines. 1975. "Discontinuous Urban Development and Economic Efficiency." *Land Economics* 51:224–234.

Pigou, A. C. 1932. *The Economics of Welfare*. 4th ed. London: MacMillan.

Pines, D., and E. Sadka. 1985. "Zoning, First-Best, Second-Best and Third-Best Criteria for Allocating Land to Roads." *Journal of Urban Economics* 17:167–183.

Sierra Club. http://www.sierraclub.org/sprawl/transportation/. Accessed February 14, 2010.

Son, J.-Y., and K.-H. Kim. 1998. "Analysis of Urban Land Shortages: The Case of Korean Cities." *Journal of Urban Economics* 43:362–384.

Sorensen, P. 2009. "Moving Los Angeles." *Access*: *Transportation Research at the University of California* 35:16–24.

Strotz, R. H. 1965. "Urban Transportation Parables." In *The Public Economy of Urban Communities*, ed. J. Margolis, 127–169. Baltimore: Johns Hopkins University Press.

Vickrey, W. S. 1963. "Pricing in Urban and Suburban Transport." *American Economic Review* 53:452–465.

Walters, A. A. 1961. "The Theory and Measurement of the Private and Social Cost of Highway Congestion." *Econometrica* 29:676–699.

Warner, S. B., Jr. 1978. *Streetcar Suburbs*: *The Process of Growth in Boston (1870–1900)*, 2nd ed. Cambridge, MA: Harvard University Press.

Wheaton, W. C. 1974. "A Comparative Statics Analysis of Urban Spatial Structure." *Journal of Economic Theory* 9:223–237.

———. 1978. "Price-Induced Distortions in Urban Highway Investment." *Bell Journal of Economics and Management Science* 9:622–632.

———. 1998. "Land Use and Density in Cities with Congestion." *Journal of Urban Economics* 43:258–272.