Transport subsidies, system choice, and urban sprawl

Jan K. Brueckner

Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine, CA 92697, United States

Received 10 September 2004
Available online 22 March 2005

Abstract

This paper analyzes the effect of transport subsidies on the spatial expansion of cities, asking whether such subsidies are a source of undesirable urban sprawl. Even though the cost-reducing effect of transport subsidies is offset by a higher general tax burden (which reduces the demand for space), the analysis shows that subsidies nevertheless lead to spatial expansion of cities. If the transport system exhibits constant returns to scale, the subsidies are inefficient, making the urban expansion they entail undesirable. The paper also studies transport “system choice,” with the city portrayed as selecting its transport system from along a continuum of money cost/time cost choices. The analysis shows that subsidies inefficiently bias choice in the direction of a high money cost/low time cost option. Lastly, the paper considers system choice in a city with rich and poor groups, showing that the rich favor a system with a high money cost and low time cost, but that their choice, if implemented, leads to a city whose spatial size is smaller than optimal. Thus, rich control of system choice does not lead to urban sprawl.

© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Urban sprawl has become a major policy issue in the U.S. in recent years. Sprawl critics argue that the spatial growth of cities has been excessive, consuming inordinate amounts of farmland and other open space while generating long commutes, traffic congestion and extra pollution. Public policy has responded to these concerns, with many states and localities adopting measures to curb the spatial growth of cities. Oregon, New Jersey,
Maryland, Tennessee and a number of other states have adopted explicit anti-sprawl ordinances. Over 240 sprawl measures appeared on 1998 election ballots nationwide, and many were approved by the voters. Among existing measures, perhaps the best known is Portland’s “urban growth boundary,” which prohibits development outside a specified ring drawn around the city.

Urban economists have long understood that several fundamental forces drive the spatial growth of cities. These forces include the growth in household income (which raises the demand for living space), the reduction in commuting costs due to transport improvements (which eases suburban access) and rising city populations. With each of these forces strongly evident in American cities, their spatial growth over recent decades comes as no surprise. But while growth in response to such fundamental forces cannot be faulted on efficiency grounds, market failures or other distortions might bias the growth process in an upward direction, causing cities to expand too much.

Several such factors have been identified as potential culprits. The first, a market failure, arises because developers fail to account for the amenity value of the open space surrounding cities in their development decisions. The second distortion arises through the system of local public finance, which often fails to charge developers for the full cost of the infrastructure (sewers, parks, roads etc.) necessitated by their developments. Both of these failures make development look artificially cheap, thus encouraging its occurrence.

The property tax, which may cause cities to expand by reducing the intensity of land development, is another culprit, as is the tax subsidy to homeownership, which raises the demand for space by owner–occupiers. Finally, the market failure associated with unpriced road congestion is a likely contributor to overexpansion of cities. Because of the congestion externality, the perceived cost of commute travel is less than its social cost, which leads to overly long commute trips and cities that are too spread out.1

The present paper focuses on a closely related issue: the effect of transport subsidies on the spatial expansion of cities. In the U.S., road networks as well as public transit systems are subsidized, with revenues from user fees falling well short of the combined operating and capital costs of the systems. Such subsidies, which are common in other countries as well, reduce the cost of travel within cities, potentially encouraging their spatial expansion. This conclusion is not immediate, however, because transport subsidies must be supported by general tax revenue. As a result, while a subsidy reduces the direct cost of using the transport system, it raises the general tax burden, reducing disposable incomes. The first effect causes a city to expand, while the second (by reducing the demand for space) causes it to contract. In an urban context, the net effect is not clear a priori, and one purpose of the analysis in the paper is to determine its sign.2


2 In a nonspatial setting with homogeneous consumers, simultaneously subsidizing the price of a good and levying a lump sum tax to cover the expenditure generates an unambiguous increase in consumption, which corresponds to the substitution effect. Two aspects of the urban model, however, prevent a direct application of this argument. First, while the benefits of the transport subsidy depend on location, the lump sum tax is locationally invariant, as seen below. In addition, the two policy instruments affect land rent in the model, but do so differentially.
indeed positive: transport subsidies contribute to the spatial expansion of cities. To provide a simple analysis, this result is demonstrated in a model where road congestion is absent.\footnote{This absence reflects the assumption that the transport system is uncongestible. Note that if congestion was present in the model and optimal tolls were levied, the toll revenue could be used to defer any fixed cost associated with the transport system.} The model also omits the other distortionary effects discussed above in order to provide a clean focus on the issue at hand.

Whether or not the spatial expansion due to subsidies is inefficient depends crucially on the nature of the transportation technology. The analysis in the paper assumes that the transport system exhibits constant returns to scale, with total costs proportional to total passenger miles. In this case, transport subsidies are inefficient, which implies that the resulting spatial growth of the city represents undesirable overexpansion. If, on the other hand, the transport technology exhibits increasing returns, then subsidies are warranted, and (provided they are not too large) their effects on city size cannot be viewed as undesirable. In attempt to determine which view is appropriate, the discussion reviews empirical work on returns to scale in transportation, noting that some of the evidence points to increasing returns. But the discussion concludes by arguing that, from one perspective, transport systems necessarily exhibit constant returns, making subsidies undesirable and their city-expanding effects inefficient. Note that this inefficiency compounds the effects of unpriced road congestion, which is omitted from the model. Thus, transport subsidies may amplify the effect of unpriced congestion in producing undesirable urban sprawl.

While the analysis just described treats the transport system as given, the discussion in Section 3 analyzes “system choice.” The city is portrayed as selecting its transport system from along a continuum of money cost/time cost choices. A road network has a high money cost, which includes the cost of automobile operation, but a low time cost, while a slow but resource-efficient public transit system has the opposite characteristics. The analysis first investigates the effect of transport subsidies on system choice, showing that subsidies inefficiently bias choice in the direction of a high money cost/low time cost option. This conclusion provides support for the common allegation that U.S. cities have overinvested in freeways at the expense of public transit, with the culprit being money cost subsidies. The analysis shows that adding a system choice dimension to the model leaves the previous conclusions about subsidies and urban sprawl unaffected.

The last part of the analysis considers system choice in a city where subsidies are absent but the transport system serves two distinct income groups. It is shown that the rich group favors a high money cost/low time cost option, but that their choice, if implemented, leads to a city whose spatial size is smaller than optimal. Thus, if the rich group is able to skew investment decisions toward freeways and away from public transit, the result is not spatial overexpansion of the city, as some observers might argue. The reason is that the resulting transport system is wrong from the point of view of the poor, which depresses their demand for space, retarding the city’s spatial expansion.

Before proceeding to the analysis, it is useful to gain a sense of the magnitude of the transport subsidies present in the U.S. and Europe. In the case of public transit, data on 6000 public transit agencies compiled by the American Public Transit Association show...
that total fares collected in 2001, which amounted to $8.89 billion, covered only 38% of total system operating costs and 25% of combined operating and capital costs. For highways, Federal Highway Administration data show that user fees, including gasoline taxes, license fees and related charges, accounted for 60% of highway disbursements, which include both capital and maintenance expenditures (disbursements were $129.9 billion in 2001).

Although such aggregate data are not readily available for Europe, information provided by the Transportation Research Board (2001) shows that public transit fares account for a higher proportion of operating costs than in the U.S. Figs. 2–4 in this publication, which show subsidy rates for individual European transit systems, indicate that fares cover on average about 50% of operating costs. While no aggregate highway data are available, the dramatically higher level of gasoline taxes in Europe (see Pucher (1988)) suggests that user fees more than cover the cost of European roads.

2. Subsidies and urban sprawl

2.1. Key assumptions of the analysis

As just noted, both public transit systems and highways are subsidized in the U.S. The goal of the analysis in this section is to explore the effect of these subsidies on the spatial sizes of cities, abstracting from the question of system choice. As a result, the analysis assumes that the city relies on a particular type of transportation system, which is unrealistically assumed to carry all traffic. The system might consist solely of a road network, with the city offering no public transit option. Or the city might rely entirely on a bus or rail system to transport its residents, with automobile travel ruled out. This stylized view, which unrealistically eliminates the mode-choice decision faced by real-world commuters, is adopted in order to generate a simple and clearcut analysis. Presumably, the lessons of the analysis would generalize to a more realistic model where several different transport options are available to the city’s residents.

Travel in the city consists solely of commute trips, with all workers commuting to jobs at the employment center (CBD), incurring both time and money costs. As noted in the introduction, transport congestion is absent. Although the most general approach to generating time costs relies on a model with a labor–leisure choice, the resulting complexity limits the usefulness of such an approach. Instead, the present analysis relies on a convenient assumption that generates time costs in a tractable fashion while still capturing the spirit of a more general model. In particular, the analysis assumes that, rather than reducing the hours available for both work and leisure, commuting instead reduces only work hours. Thus, an extra commuting hour robs the worker of an hour of work time, lowering income by the wage per hour.

---

4 See the APTA website at <http://www.apta.com/research/stats/overview/natsum.cfm>.
5 See the FHWA website at <http://www.fhwa.dot.gov/ohim/hs01/hf10.htm>. It should be noted that this 60% figure somewhat understates the magnitude of user fees, given that a portion the resulting revenue is devoted to non-highway uses, such as mass transit.
Accordingly, let \( y \) denote “full income,” which represents the CBD income earned per period by a worker with a zero commute time. Then, let \( x \) denote distance from a worker’s residence to the CBD, and let \( \theta \) denote the fraction of the available work hours consumed by each round-trip mile of commuting. The parameter \( \theta \) is an inverse measure of the speed of travel allowed by the transport system. Pre-tax income for a worker living at distance \( x \) then equals \( y - \theta xy = y - \tau x \), where \( \tau = \theta y \) is the time cost per mile of commuting. These time costs do not play a central role in the analysis presented in this section of the paper, but instead enter crucially in the discussion of system choice in Section 3.6

The money cost of commuting from distance \( x \) is given by \( kx \), making total commuting cost equal to \((s + k)x\). In practice, these money costs embody subsidies, which are the focus of the present analysis. For example, in the case of transit riders, fares cover only a portion of operating and capital costs. For automobile commuters, gasoline taxes and other user fees cover only a portion of the costs of building and operating a road network.

To gauge the extent of the subsidy embodied in \( k \), the full cost of transportation must be specified. The analysis will proceed under a strong and potentially controversial assumption regarding these costs. In particular, the full money cost of transportation incurred within the city is assumed to be proportional to total passenger miles of travel, indicating constant returns to scale. Thus, letting \( t \) denote the constant of proportionality, the full money cost for an individual living at distance \( x \) equals \( tx \) (\( t \) times individual passenger miles), and the total money cost of transportation in the city is found by appropriately aggregating this expression across all residential locations. In the presence of a subsidy, the money cost \( k \) paid by commuters is some fraction \( \alpha \) of the full cost, with \( k = \alpha t \). Thus, while the full money cost of commuting is \( tx \) for a worker living at \( x \), the worker pays only \( \alpha tx \). The difference, \((1-\alpha)tx\), must be covered by general tax revenue. Note that total commuting cost is then \((s + \alpha t)x\) for a worker living at \( x \).

For public transit users, \( \alpha \) represents the percentage of operating and capital costs covered by fares (which are implicitly proportional to distance traveled under the present formulation). For automobile commuters, by contrast, full transport costs represent a combination of the private cost of vehicle operation and the cost of road construction and maintenance. Because the user fees paid by drivers cover only a portion of these latter road costs, a particular \( \alpha < 1 \) again emerges. Recall, however, that only one of these transport modes is available to commuters in the stylized model under consideration.

The above proportionality assumption is controversial because, with total costs proportional to passenger miles, the transport technology exhibits constant returns to scale, even though some empirical evidence points to the existence of increasing returns. Nevertheless, the analysis will proceed to evaluate both the desirability and the effect of existing transport subsidies under the constant returns assumption. Once the analysis is

\[ \text{In the above formulation, commute time is valued at 100\% of the wage, contradicting empirical evidence showing that valuation occurs at some fraction of the wage (see Calfee and Winston (1998) for a recent discussion). While the analysis would be unaffected if time cost were arbitrarily set at a fraction of } \tau x, \text{ such a modification is hard to justify on theoretical grounds. Indeed, the proper way to generate the fractional valuation observed in the empirical literature is to use a model with an explicit labor–leisure choice along with the added feature that commute time yields leisure benefits, but at a lower rate than pure leisure time. However, as is well known, adding a labor–leisure choice to urban models generates severe complications.} \]
complete, the discussion considers how the conclusions would change under increasing returns, while appraising the defensibility of the constant returns assumption.

2.2. Urban spatial model

To explore the effect of transport subsidies on the spatial equilibrium of the city, this section develops a spatial model, relying on the above assumptions on commuting costs. The model represents a standard application of the monocentric city framework, as presented by Wheaton (1974) and others.

Without loss of generality, the city is assumed to be linear with unit width. In effect, the city occupies a rectangular island, and for simplicity, the CBD is assumed to be located at one end. Rental income from the urban land accrues to absentee landowners, who live outside the city. Since the closed-city version of the urban model is appropriate for the analysis at hand, the city’s population $n$ is viewed as fixed.\(^7\)

City residents consume land, denoted by $q$, and a numeraire nonland good, denoted by $c$. Without loss of generality, housing capital is suppressed, with land itself representing the housing commodity. With land rent denoted by $r$, the budget constraint for a consumer living at distance $x$ is $c + rq = y - T - (\tau + xt)x$, where $T$ is a head tax levied on all urban residents. This tax, which is independent of residential location, covers the subsidy-induced deficit in the operation of the transport system. The consumer chooses $c$ and $q$ to maximize the utility function $v(c, q)$ subject to this constraint, and $r$ adjusts so that the realized utility level is uniform across locations. The two conditions $v_q/v_c = r$ and $v(y - T - (\tau + xt)x - rq, q) = u$ must then be satisfied, where $u$ is the uniform utility level. These conditions determine $r$ and $q$ as functions of the parameters in the equations, with the solutions written $r(y - T, \tau + xt, x, u)$ and $q(y - T, \tau + xt, x, u)$. Note that the relevant income and commuting cost parameters are full net-of-tax income, $y - T$, and the combined time and money cost per mile of commuting, $\tau + xt$.

The urban equilibrium conditions impose three requirements, two of which are standard. First, urban land rent at $\bar{x}$, the edge of the city, must equal the agricultural rent, $r_a$. Second, the city population $n$ must fit inside $\bar{x}$. These conditions are written

$$r(y - T, \tau + xt, \bar{x}, u) = r_a$$

(1)

$$\int_0^{\bar{x}} \frac{1}{q(y - T, \tau + xt, x, u)} dx = n.$$  

(2)

Note that since $1/q$ in Eq. (2) is population density, and since the city is linear with unit width, the integral aggregates the population out to $\bar{x}$ and equates it to $n$.

---

\(^7\) When the goal is to analyze a policy change (such as raising a transport subsidy) that affects all cities in the economy, the closed-city model is appropriate given that the policy change will not induce intercity migration.
The third condition is a balanced-budget condition stating that the transport deficit is just covered by revenue from the head tax. This condition is written

$$nT = (1 - \alpha) \int_0^\infty \frac{tx}{q(y - T, \tau + \alpha t, x, u)} \, dx.$$  \hfill (3)

Note that the integral in Eq. (3) is total money commuting cost in the city (money cost at distance $x$ times population density $1/q$, aggregated over the city). Since only a fraction $\alpha$ of this cost is paid by commuters, the remaining share must be covered by the head tax. It is convenient to rewrite Eq. (3) as the two separate conditions

$$T = (1 - \alpha) t\tilde{x}$$  \hfill (4)

$$\tilde{x} = \frac{1}{n} \int_0^\infty \frac{x}{q(y - T, \tau + \alpha t, x, u)} \, dx,$$  \hfill (5)

where $\tilde{x}$ is the average commuting distance in the city. Eq. (4) thus indicates that $T$ must equal the subsidized portion of average money commuting costs.

Eqs. (1), (2), (4) and (5) constitute four conditions that solve for the unknowns $u$, $\bar{x}$, $T$ and $\tilde{x}$. For the purposes of the ensuing analysis, however, it is helpful to view this solution as occurring recursively. First, Eqs. (1) and (2) are solved to yield solutions for $u$ and $\bar{x}$ conditional on $T$, which may be written

$$u = F(y - T, \tau + \alpha t)$$  \hfill (6)

$$\bar{x} = G(y - T, \tau + \alpha t)$$  \hfill (7)

(the solutions’ dependence on $n$ and $r_a$ is suppressed). These solutions are substituted into Eqs. (4) and (5), which then determine $T$ and $\tilde{x}$ as functions of the underlying parameters $y$, $\tau$, $t$ and $\alpha$, with the latter subsidy parameter being the one of interest. Finally, the resulting $T$ solution is substituted back into Eqs. (6) and (7), which then determine $u$ and $\tilde{x}$ as functions of the underlying parameters. The advantage of this approach is that the signs of the derivatives of the $F$ and $G$ functions are known from the standard comparative static analysis of the monocentric city, as presented by Wheaton (1974). The following discussion relies, in particular, on the inequalities $G_1 > 0$ and $G_2 < 0$, which show that $\tilde{x}$ rises with an increase in income and falls with an increase in commuting cost per mile. As seen below, this information can be exploited to show the effect of the transport subsidy on the spatial equilibrium of the city.

### 2.3. The effect of the transport subsidy

With the above preparation, it is possible to appraise the effect of the transport subsidy on the urban equilibrium. Two issues are of interest. First, is the subsidy inefficient? Second, what is the subsidy’s effect on the city’s spatial size, as represented by $\tilde{x}$? If the answer to the first question is affirmative, while the second answer shows that the subsidy
raises $\bar{x}$, then the analysis will have established that transport subsidies contribute to undesirable urban sprawl.

The efficiency question can be answered easily by appealing to the standard welfare analysis for a monocentric city. A social planner chooses the spatial pattern of land consumption in the city to minimize resource usage subject to a fixed utility requirement for the residents. Resource usage is equal to the sum of total nonland consumption, total commuting cost, and the opportunity cost of the urban land. The optimality conditions for this problem are the same as the equilibrium conditions (1) and (2), modified so that $x=1$ and $T=0$ hold and $y$ is set at a level consistent with the target utility (see, for example, Fujita (1989, Ch. 3)). Thus, since the optimality conditions require urban residents to face the full money cost $t$ of commuting at each location, equilibrium in the presence of a transport subsidy is inefficient. Total resource usage in the city could be reduced, while maintaining the equilibrium utility level, if the subsidy were removed. This inefficiency result recapitulates well-known conclusions regarding subsidies and returns to scale, providing an illustration of these results in the context of an urban spatial model.

With the first question above answered, consider now the effect of the transport subsidy on $\bar{x}$. To do so, let $\partial \bar{x}/\partial x$ and $\partial \bar{x}/\partial \bar{x}$ denote the effects of $x$ on $\bar{x}$ and $\bar{x}$ from the full solution to Eqs. (1), (2), (4) and (5). Unfortunately, given the complexity of this equation system, the general signs of these derivatives cannot be established. However, a more limited result can be derived, as follows. Using Eq. (4), let Eq. (7) be rewritten as $\bar{x}=G(y-(1-x)t\tilde{x}, \tau+xt)$. Differentiating Eq. (7) then yields

$$\frac{\partial \bar{x}}{\partial x} = G_1 \left[ t\tilde{x} - (1-x)t \frac{\partial \tilde{x}}{\partial x} \right] + G_2 t$$

The key observation is that, while both derivatives in Eq. (8) are generally ambiguous in sign, the sign of $\partial \bar{x}/\partial x$ is determinate for $x=1$. In this case, the $\partial \bar{x}/\partial x$ derivative drops out of Eq. (8), and $\partial \bar{x}/\partial x$ has the same sign as $G_1 \tilde{x} + G_2$. While the sign of this expression is not immediately apparent given $G_1 > 0$ and $G_2 < 0$, the Appendix shows that the sign is negative. As a result, starting in the no-subsidy case, where $x=1$, the introduction of a small transport subsidy (a marginal reduction in $x$) leads to an increase in $\bar{x}$.

Intuitively, a decrease in $x$ (an increase in the subsidy) has two opposing effects. By lowering commuting cost per mile, the higher subsidy tends to increase $\bar{x}$, causing the city to spread out. However, a higher subsidy requires an increase in the head tax $T$, which tends to reduce $\bar{x}$ by lowering net-of-tax income for the city residents. The relation between the magnitudes of these opposing effects is determined via the budget-balance condition (3), and this condition implies that the first effect dominates the second.\(^8\)

---

8 This dominance would be strengthened in the case of a favored city within a federal system. In this situation, funds supplied to the city by the federal government would partly cover the local transport subsidy, so that the local head tax $T$ could be lower than the value implied by Eq. (3). As a result, $\bar{x}$ would be even larger than in a balanced-budget setting.
It should be noted that this dominance appears to hinge on the fact that the net burden on the edge resident, who lives at $\bar{x}$, falls as the subsidy rises. Because of his long commute, this resident enjoys the highest absolute subsidy in the city but pays a head tax that recovers only the average amount of the subsidy. The edge resident’s disposable income thus increases as $\zeta$ falls, suggesting that his willingness to pay for land should rise. Higher land rent at $\bar{x}$ in turn implies an expansion of the city.

This discussion can be summarized as follows:

**Proposition 1.** Suppose that the transport system exhibits constant returns to scale, with total costs proportional to total passenger miles. Then, imposition of a transport subsidy is inefficient, reducing social welfare. In addition, starting from the no-subsidy case, introduction of a small subsidy leads to an increase in $\bar{x}$. Thus, transport subsidies tend to be associated with inefficient urban sprawl.

It should be noted the sign of $\partial u / \partial \zeta$ is ambiguous, indicating that the utility of the city’s residents can either rise or fall as the transport subsidy is increased. This conclusion is not inconsistent with the overall inefficiency of the subsidy, a verdict that takes into account the welfare of both consumers and absentee landlords.

While a general “global” analysis of the effect of subsidies would be difficult, accounting for presentation of a local result (where $\zeta=1$) in Proposition 1, such an analysis might be eased by imposing particular functional forms on preferences. For example, in the case of Leontief preferences, it is easily shown that Proposition holds globally, for arbitrary $\zeta$. In the case of other familiar preferences such as Cobb–Douglas, numerical analysis of the equilibrium conditions would be required to derive a conclusion. Such a task could be a subject for further work.

Another extension would be to assume that the subsidy is financed with funds collected via an income tax rather than the head tax $T$. Recognizing that time cost causes income to fall as $x$ increases, the two taxes are not equivalent in the present setting (in particular, the individual income tax burden would fall as $x$ increases).  

2.4. Appraising the constant returns assumption

The inefficiency of subsidies hinges crucially on the assumption that the transport system exhibits constant returns to scale. Although there are many ways of modeling non-constant returns, this point can be simply demonstrated using one plausible specification that embodies a form of increasing returns. In particular, suppose that total transport costs have two components. The first, a “variable cost” component, is proportional to total passenger miles, as above, while the second is a fixed cost that depends only on the spatial extent of the transport system. In the current model, suppose that this fixed cost equals $s\bar{x}$,

---

9 If property taxes, which are location dependent, were responsible for covering the transport deficit, then $T$ would again be replaced by a spatially varying tax. However, it seems reasonable and realistic to assume, in the context of a model with homogeneous consumers, that the deficit is covered by a head tax. In a related observation, a referee pointed out that, if labor supply were variable and an income tax regime were used, then the need to raise tax revenue to cover the transport deficit would depress labor supply, with a host of attendant effects on the urban equilibrium. However, analysis of such a model would be challenging.
capturing the notion that the distance \( \bar{x} \) to the edge of the city also equals the “length” of the transport network.\(^{10}\)

To generate the optimum under this modification, it is easy to see that commuters should face the full variable cost of transportation, with \( a = 1 \). However, the fixed cost of the transport system should be recovered through a lump sum charge that is unrelated to system use.\(^{11}\) For example, one way to support the optimum would be to charge absentee landowners a development tax of \( s \) per acre of land to recover the fixed cost of the transport system.\(^{12}\) Thus, within this simple framework, user fees should only cover the variable cost of the transport system, so that system use is effectively subsidized. But this outcome is now efficient rather than undesirable. A similar conclusion would emerge under other specifications that involve increasing returns.

Whether existing transport subsidies are inefficient, leading to an undesirable spatial expansion of cities, therefore depends on whether constant or increasing returns characterizes real-world transportation systems. Empirical evidence on returns to scale for urban bus and rail systems is surveyed by Small (1992), and he argues that the evidence demonstrates increasing returns when output is measured by passenger miles or a related construct. Small (1992) also surveys the evidence on returns to scale in road networks, arguing that the evidence is consistent with “mild” increasing returns when road capacity rather than passenger miles is viewed as the output measure.

Despite this evidence, one could argue that transport systems should in principle exhibit approximate constant returns to scale. To see the argument, note that Small attributes increasing returns in rail and bus systems to underutilized capacity.\(^{13}\) However, if system capacity is instead adjusted to match demand, then costs should increase roughly in step with passenger miles. Similarly, if highways were built to maintain a uniform traffic density throughout the network, then adding a passenger mile would require adding some fraction of a lane mile, again at an approximately constant cost. As with bus and rail systems, this argument hinges on keeping capacity utilization constant as passenger miles expands. If these arguments supporting constant returns are correct, then the negative efficiency verdict on subsidies contained in Proposition 1 may be realistic. Note that this verdict would probably remain relevant even in the presence of mild increasing returns.

\(^{10}\) Although a closed-form relationship between total transport costs and passenger miles cannot be derived in general for this modified model, such a derivation is possible if land consumption in the city is constant. If \( q \) is constant (taking the value unity for simplicity), then total passenger miles equals \( \int_0^\bar{x} x \, dx = \bar{x}^2/2 = m \), implying \( \bar{x} = \sqrt{2m} \). Total costs then equal \( tm + s\sqrt{2m} \), which increases less than proportionally with \( m \), indicating increasing returns.

\(^{11}\) The optimality conditions for the city are again modified versions of Eqs. (1) and (2). As before, \( a \) is set equal to unity and \( T \) is set equal to zero. In addition, \( r_u \) on the RHS of Eq. (1) is replaced by \( r_u + s \). The explanation is that, in addition to foregoing agricultural rent when a unit of land is converted to urban use, the transport system must also be extended at cost \( s \). At the optimal \( \bar{x} \), these losses are equated to the gain from urban development, as measured by urban rent.

\(^{12}\) With such a tax, the net return from urban land is \( r - s \), and the city expands until this net-of-tax return is equated to \( r_u \), satisfying the new optimality condition.

\(^{13}\) Small says that increasing returns arise “presumably because vehicles can be used more intensively as demand grows”.

which would not be sufficient to justify large transport subsidies like those observed in practice.\textsuperscript{14}

3. System choice

3.1. The effect of subsidies on system choice

The analysis in the Section 2 assumed that the city was served by a given type of transport system, as characterized by particular values of $\tau$ and $t$, the time and money cost parameters. A road network, for example, would entail a high $t$ (capturing both road and automobile costs) and a relatively low $\tau$, reflecting the speed and convenience of auto travel. Relative to a road network, a rail or bus system would involve a lower $t$ and a higher $\tau$, reflecting lower resource usage per passenger and higher time costs.\textsuperscript{15}

Although no city relies exclusively on one type of transport system, contrary to the stylized assumption used in the analysis, cities do favor particular transport modes. Some cities invest heavily in public transit systems, which may then account for an appreciable share of traffic, while others rely almost entirely on road networks to transport their residents. City choices thus affect the $t$ and $\tau$ parameters that govern transport costs.

To investigate this question of system choice within the context of the present model, let the discreteness of existing transport options be ignored. Instead, suppose that the city is portrayed as choosing its single transport system from among a continuum of possibilities, with the continuum capturing the trade-off between time and money costs. To capture this trade-off, recall that time cost parameter $\tau$ equals $\theta y$, where $\theta$, the fraction of work hours consumed by each round-trip mile of commuting, is an inverse measure of travel speed. The given trade-off can then be represented by the function $\theta = \phi(t)$, which relates the transport system’s travel speed to its money cost. The function $\phi$ satisfies $\phi^t < 0$, indicating that higher money costs are associated with lower time costs (higher speed), as in the above discussion of roads vs. rail or bus systems. Substituting $\phi$, the $\tau$ parameter can then be rewritten as $y\phi(t)$.

System choice involves selecting a particular $(t, \tau)$ pair from along the available continuum.\textsuperscript{16} Consider first the outcome of this choice in a city without a transport subsidy. The equilibrium in such a city is characterized by Eqs. (1) and (2) with $x=1$ and $T=0$, which yield the equilibrium solutions for $u$ and $\bar{x}$ given by Eqs. (6) and (7), again with $x=1$ and $T=0$. Assuming that the city’s transport system is chosen through a

\textsuperscript{14} If subsidies are inefficient, the question then arises as to why they exist. One answer is that subsidies may benefit urban residents (with $\partial u/\partial x > 0$ holding) even though they reduce overall welfare. Alternatively, in a more complete model, subsidies could be the result of lobbying by narrower interest groups, such as road builders or transport unions.

\textsuperscript{15} The analysis will assume a uniform subsidy parameter $\alpha$ even though the evidence presented in the introduction shows that subsidies may differ across modes.

\textsuperscript{16} Sasaki (1990) and LeRoy and Sonstelie (1983) analyze models where a city has two transport systems, one with a high money cost and low time cost and the other with the reverse characteristics. The present analysis differs by assuming the existence of just one system but allowing that system to be selected from a continuum of $(t, \tau)$ choices.
democratic process to serve the interests of its residents, the choice then maximizes the equilibrium utility level in Eq. (6). Since standard analysis shows that \( F_2 < 0 \), indicating that utility is decreasing in commuting costs, the goal is to choose \( t \) to minimize \( \tau + t = y\phi(t) + t \). The first-order condition for this choice is

\[
y\phi'(t) + 1 = 0, \tag{9}
\]

and the second-order condition for a minimum is \( \phi'' > 0 \), indicating that time cost falls at a decreasing rate as money cost rises. Note that since \( G_2 < 0 \), the transport system that minimizes \( \tau + t \) leads to the largest spatial size for the city among all possible systems. Finally, it is easy to see that condition (9) also characterizes the socially optimal system, where the interests of absentee landowners are considered along with those of city residents.

Now consider the system choice problem for a city with a transport subsidy, where \( z < 1 \) and \( T > 0 \). In order to provide a simple characterization of this choice, it is useful to impose a plausible myopia assumption. In particular, suppose that city residents do not understand the connection between system choice and their general tax burden, as reflected in the head tax \( T \) that covers the transport deficit. Thus, the city residents choose the transport system focusing on the transport costs that they face directly. This assumption is likely to be an accurate description of the public choice process.

System choice then involves selecting \( t \) and the associated \( \tau \) to maximize Eq. (6), with \( T \) viewed as parametric. The goal is thus to minimize \( y\phi(t) + zt \), and the first-order condition is

\[
y\phi'(t) + z = 0. \tag{10}
\]

Using \( \phi'' > 0 \) and \( z < 1 \), the \( t \) value that solves Eq. (10) is larger than the value that solves Eq. (9). Thus, the transport subsidy inefficiently biases system choice toward transport systems with high money cost and low time cost. This conclusion is natural since, by sheltering consumers from the full money cost of travel, the subsidy encourages a higher \( t \) in pursuit of a lower time cost. Concretely, this result means that a transport subsidy inefficiently biases choice away from low-\( t \)/high-\( s \) public transit systems toward high-\( t \)/low-\( \tau \) highway networks. This conclusion could be used to buttress allegations that U.S. cities have overinvested in freeways at the expense of public transit.

It remains to consider how the addition of a system choice affects the results of Section 2. The answer is that the results are unchanged. To see this conclusion, note that with \( t \) now a decision variable that depends on \( z \) via Eq. (10), condition (8) is replaced by

\[
\frac{\partial \hat{\mathbf{x}}}{\partial z} = G_1 \left[ t\hat{\mathbf{x}} - (1 - z) \left( t \frac{\partial \hat{\mathbf{x}}}{\partial z} + \frac{\partial t}{\partial z} \hat{\mathbf{x}} \right) \right] + G_2 \left[ t + (y\phi' + t) \frac{\partial t}{\partial z} \right], \tag{11}
\]

Since \( y\phi' + z \) equals zero by Eq. (10) and since the second part of the term multiplying \( G_1 \) vanishes when \( z = 1 \), the right hand side of Eq. (11) once again has the negative sign of
\(G_1 \dot{x} + G_2\) when evaluated at \(x=1\). Thus, system choice leaves Proposition 1 unchanged, leading to the following conclusion:\(^{17}\)

**Proposition 2.** Suppose that all transport systems exhibit constant returns to scale, and that the city residents choose their system in myopic fashion, ignoring the effect on their general tax burden. Then, the conclusions of Proposition 1 continue to hold. In addition, the transport subsidy inefficiently biases system choice toward transport systems with high money cost and low time cost.

### 3.2. System choice in a city with two income groups

It is sometimes argued that urban sprawl is encouraged by a transport system choice made in the interests of high-income suburban residents. Investment in freeways, intended to facilitate commuting from the suburbs, combined with a corresponding deemphasis of public transit, is thought to spur excessive spatial growth of cities. The purpose of the analysis in this section is to appraise this argument. Since the transport subsidy plays no role in the argument, it is eliminated from the model by setting \(\alpha=1\) and \(T=0\).

The model now contains two income groups with different time costs, but the assumption that the city has a single type of transport system is retained. The groups thus disagree over the characteristics of the preferred transport system, and this disagreement is a key element in the analysis.\(^{18}\)

To make any headway in analyzing the two-group case, a further simplifying assumption is necessary. In particular, both groups are assumed to have identical Leontief preferences, with \(v(c,q)=\min\{c,q\}\). This assumption means that land consumption \(q\) is uniform (though endogenous) across locations for each income group, but different between groups, which greatly simplifies the analysis.

Group 0 represents the poor and group 1 the rich, with full incomes denoted \(y_0\) and \(y_1\). These two incomes generate different time costs, \(\tau_0=y_0\phi(t)\) and \(\tau_1=y_1\phi(t)\), with \(\tau_0<\tau_1\). The populations of the groups are \(n_0\) and \(n_1\).

Because of right-angled indifference curves, \(c=q\) will hold for both groups, so that utility can be measured directly by \(q\). In addition, using this equality to eliminate \(c_i\), the budget constraint for group \(i\) can be written \(q_i+r_i q_i = y_i - (\tau_i + t)x\), for \(i=0,1\). Note that both groups face the same money cost parameter \(t\), and that \(r_i\) is the land rent paid by group \(i\). Rearranging the budget constraint to solve for \(r_i\) yields

\[
 r_i = \frac{y_i - (\tau_i + t)x}{q_i} - 1. \tag{12}
\]

Each income group lives where it offers the highest bid for land, and these locations depend on the steepness of the two land–rent curves. If the poor curve is steeper, the poor live near center (where their bid is higher), with rich living in the suburbs, and the reverse location pattern holds if the rich curve is steeper. The slope of group \(i\)’s rent curve is

\(^{17}\) It can be shown that Proposition 2 holds without the myopia assumption under particular parameterizations of the model. For example, under the Leontief preferences considered in the next section, Proposition 2 holds even when consumers understand the connection between \(T\) and \(x\).

\(^{18}\) For an early analysis of a city with multiple income groups, see Hartwick et al. (1976).
\(- (\tau_i + t) / q_i\), which depends on the endogenous \(q_i\). It is assumed, however, that parameter values in the model are such that the poor curve is steeper, which requires that the poor’s lower land consumption offsets their smaller time cost, leading to a higher absolute value for the above slope expression. This assumption yields a realistic central location for the poor group.

Because land consumption is constant within each group, the total amount of land occupied by a group just equals its population times the relevant \(q\) value. Recalling that the city has unit width, the poor thus live between the CBD and \(\hat{x} = n_0q_0\), while the rich live between \(\hat{x}\) and \(\tilde{x} = n_0q_0 + n_1q_1\). The equilibrium conditions for the city require that urban land rent equals \(r_a\) at this \(\hat{x}\) value and that the poor and rich land rent functions intersect at \(\hat{x}\). Assuming \(r_a = 0\) for simplicity and using Eq. (12), these conditions are

\[
y_1 - (\tau_1 + t)(n_0q_0 + n_1q_1) = \frac{1}{q_1}
\]

\[
y_0 - (\tau_0 + t)n_0q_0 = y_1 - (\tau_1 + t)n_0q_0.
\]

Eqs. (13) and (14) determine the equilibrium values of \(q_0\) and \(q_1\). Letting \(z_i = \tau_i + t\) denote time plus money cost per mile for group \(i\), the solutions can be written

\[
q_0 = \frac{y_0}{1 + n_0z_0 + n_1z_1},
\]

\[
q_1 = \frac{1}{1 + n_1z_1} \left[ y_1 - \frac{n_0z_1y_0}{1 + n_0z_0 + n_1z_1} \right].
\]

Since utility is measured by \(q\) (or \(c\)) consumption, consumer welfare in the city can be written \(n_0q_0 + n_1q_1\). Since utility can also be measured in units of consumption of the numeraire good \(c\), it follows that, under a Benthamite social welfare function, total welfare for both city residents and absentee landowners is just total \(c\) consumption in the city plus total land rent. It can be shown that the relevant social welfare expression is simply a function of the consumer welfare measure, \(n_0q_0 + n_1q_1\). Using Eqs. (15) and (16), the latter measure equals

\[
n_0q_0 + n_1q_1 = \frac{1}{1 + n_1z_1} \left[ n_1y_1 + \frac{n_0y_0}{1 + n_0z_0 + n_1z_1} \right].
\]

Using these solutions, two questions can be addressed. First, is the spatial size of the city inefficiently large if the transport system is chosen in the interests of the rich group? In other words, does rich control of the transport system choice create urban sprawl? Second,
how do the transports systems favored by the rich and poor differ? Which has the lower
time cost?

The answer to the first question is almost immediate. The key observation is that the
welfare measure \( n_0q_0+n_1q_1 \), which equals total land consumption in the city, is just equal
to \( \bar{x} \). Thus, the welfare-maximizing transport system leads to the largest possible spatial
size for the city, as in the one-group model discussed above (where satisfaction of Eq. (9)
maximizes \( \bar{x} \)). System choice according to any other criterion leads to a smaller city. One
such criterion would be maximizing of the welfare of the rich group, which requires
choosing the transport system to maximize \( q_1 \). The result of such a choice would be a
smaller-than-optimal city size.

This conclusion shows that, rather than leading to sprawl, rich control of transport
system choice inefficiently limits the spatial size of the city. The intuitive reason is that,
while the rich group’s choice maximizes their own land consumption, and hence leads to a
maximal spatial size for the city’s rich zone, the transport system chosen by the rich is
wrong from the perspective of the poor. This discrepancy reduces the poor group’s land
consumption, depressing the city’s overall spatial size despite the enlargement of the rich
area. While this result is intimately tied to the Leontief specification, it shows that rich
control of system choice need not lead to inefficient spatial expansion of the city. Even
though the conclusion would also hold by a continuity argument when substitution
between land and \( c \) is nonzero but sufficiently small, a proper generalization must be a task
for future work.

To compare the optimal \( t \) values for the two groups, answering the second question,
consider the solutions in Eqs. (15) and (17). First, maximizing \( q_0 \) requires choosing \( t \) to
minimize \( n_0z_0+n_1z_1 \), or equivalently to minimize the population-weighted average
\( \bar{z} = \beta z_0 + (1-\beta)z_1 \), where \( \beta = n_0/(n_0+n_1) \) is the poor population share. With \( \bar{z} = \\
(\beta y_0 + (1-\beta)y_1)\phi(t)+t = \bar{y}\phi(t)+t \), where \( \bar{y} \) is average income, the first-order condition is
\( \bar{y}\phi(t)+1=0 \). By contrast, inspection of Eq. (17) shows that maximizing this expression
requires minimizing an appropriately weighted average of \( z_1 = y_1\phi(t)+t \) and \( \bar{z} \). But
since \( \partial z_1/\partial t = y_1\phi^b-1 < \partial \bar{z}/\partial t \) given \( y_1 \geq \bar{y} \) and \( \phi < 0 \), it follows that \( \partial \bar{z}/\partial t = \bar{y}\phi(t)+1 > 0 \) must
hold at the solution (where a weighted average of the \( z_1 \) and \( \bar{z} \) derivatives equals
zero). As a result, the \( t \) value that maximizes \( n_0q_0+n_1q_1 \) is larger than the value that
maximizes \( q_0 \). It then follows that the \( t \) value that maximizes \( q_1 \) exceeds both these
values.\(^{19} \)

Thus, the rich group’s preferred transport system has a higher money cost and a lower
time cost (a lower \( \theta \)) than the preferred system of the poor. This result is natural given that
a high full income raises the loss from time spent in commuting, heightening the rich
group’s incentive to increase (via a higher money cost) the travel speed offered by the
transport system. Because of its lower full income, the poor group places less value on an
increase in speed, making a lower \( t \) and higher \( \theta \) optimal. Note that the \( t \) and \( \theta \) values

\(^{19} \) Note that, from the perspective of the poor, the optimal transport system reflects the city’s average income,
satisfying Eq. (9) with \( y = \bar{y} \). The optimality condition for the rich incorporates a larger income value, although one
different from \( y_1 \).
preferred by the two groups bracket the socially optimal values, which represent a compromise. Summarizing yields

**Proposition 3.** Suppose that consumers have Leontief preferences and that the poor group lives in the central part of the city. Then, if the rich group controls the city’s transport system choice, it chooses a system with a higher money cost and higher travel speed than the socially optimal system. Since this choice leads to a smaller-than-optimal spatial size for the city, rich control of system choice does not contribute to urban sprawl.

Future work could extend the framework analyzed in this section in a number of fruitful directions. First, the transport subsidy, which was excluded from the analysis, could be reintroduced. The subsidy could be supported by an income tax rather than a head tax, realistically generating different general tax burdens for the rich and poor. Analysis of system choice under such a subsidy regime might yield additional insights, although preliminary work suggests that even solving for the Leontief equilibrium with \( t \) held fixed is difficult. Another extension would allow the city to construct two transport systems with different values of \( t \) and \( \tau \), with each system designed to serve a particular income group. **Brueckner and Selod (2004)** carry out such an analysis, but full results are only reached through the use of numerical methods.

**4. Conclusion**

This paper has shown that transport subsidies may contribute to undesirable urban sprawl, while biasing system choice toward freeways and away from public transit. Although subsidies are desirable, and their spatial effects benign, when the transport system exhibits increasing returns to scale, the paper argues that returns to scale are constant when properly measured, making transport subsidies and their spatial impacts inefficient. If this view of the transport technology is correct, subsidies should be included among the culprits leading to inefficient spatial expansion of cities.

Sprawl critics frequently point to compact European cities as the preferred alternative to U.S.-style sprawl. This compactness is due in part to Europe’s lower reliance on transport subsidies, with fares covering a greater share of public transit costs and user fees exceeding the cost of road provision. If the constant returns assumption is valid, making subsidies inefficient, then Europe’s lower subsidies may well be desirable, and the resulting compactness of its cities may represent a socially preferable urban form.

Future work could be directed toward relaxing several of the model’s assumptions. The first useful modification would be to eliminate absentee landownership by assuming that the city is inhabited by resident landowners, who receive the urban land rent as income. Such a model generates an equilibrium solution for \( \bar{x} \) like the one in Eq. (8), and **Arnott et al. (1986)** and **Pines and Sadka (1986)** show that the solution continues to satisfy \( G_1 > 0 \) and \( G_2 < 0 \), implying that \( \bar{x} \) is decreasing in commuting cost per mile and increasing in the exogenous part of income. Since Eq. (8), which yields \( \frac{\partial \bar{x}}{\partial x} \), remains relevant in the
resident–landowner case, Proposition 1 will remain valid provided that the appendix derivation can be replicated for this new model, a task that could be undertaken in future research.

Another extension would be to investigate the effect of transport subsidies on the structure of a city where job locations are endogenous. Such an analysis could appeal directly to the results in Fujita and Ogawa (1982), whose model allows a variety of location patterns for business and residential land use. At one extreme is the monocentric equilibrium, where the city has a central business zone surrounded by residences, while the other extreme exhibits complete intermixing of businesses and residences, so that no commuting need occur. Intermediate cases include a duocentric city as well as the case of incomplete mixing, where purely residential and business zones coexist with mixed-use areas. Although the details are complex, the numerical analysis in the paper shows that, when transport cost declines from a high level, as would occur under a sufficiently large subsidy, land use ultimately shifts from a completely mixed equilibrium, with no employment concentration and no commuting, to the monocentric equilibrium (different transitional paths are possible, however). Since fixed land consumption by consumers and firms means that the offsetting increase in $T$ has no effect on the equilibrium, a simple conclusion emerges. In particular, by reducing transport cost, subsidies favor the emergence of centers of concentrated employment, where firms can benefit from agglomeration effects.

A final extension would be to carry out an analysis of the political economy of transport system choice that is more comprehensive than the one offered in Section 3. Brueckner and Selod (2004) present such an analysis by focusing on a model without subsidies where the urban residents exhibit a continuum of incomes. Time costs thus vary across the population, as in the two-group model analyzed above, and this variation generates disagreement over the features of the transport system, which is reconciled through a majority-voting process. The assumption that land consumption is fixed and uniform simplifies the analysis, and the results compare the socially optimal transport system to the one chosen under the voting process.

Acknowledgements

I have benefited from helpful comments from Richard Arnott, Fred Giertz, Marvin Kraus, Kangoh Lee, Dan McMillen, David Pines, Ken Small, Clifford Winston, a number of seminar participants, and two referees. Any shortcomings in the paper, however, are my responsibility.

Appendix

Proof of Proposition 1. Recall from Eq. (7) that $	ilde{x} = G(y - T, \tau + \alpha t)$. The proof shows that $G_1 \tilde{x} + G_2 \leq 0$ holds at $x=1$, or $G_2 \leq -G_1 \tilde{x}$. The first step is to totally differentiate Eqs. (1) and (2) to determine $G_1$ and $G_2$. Since all derivatives are evaluated at $x=1$, the arguments of
the \( r \) and \( q \) functions in Eqs. (1) and (2) are simply \( y \) and \( z = \tau + t \). Let the \( r \) and \( q \) derivatives with respect to these arguments be denoted \( r_y, r_z, q_y \) and \( q_z \).

Letting a bar over \( r \) or \( q \) indicate that the variable is evaluated at \( \bar{x} \), total differentiation of Eqs. (1) and (2) yields

\[
G_1 = \frac{1}{A} \left[ \bar{r}_u \int_0^\bar{x} \frac{q_y}{q^2} \, dx - \bar{r}_y \int_0^\bar{x} \frac{q_y}{q^2} \, dx \right] 
\]

\[
G_2 = \frac{1}{A} \left[ \bar{r}_u \int_0^\bar{x} \frac{q_z}{q^2} \, dx - \bar{r}_z \int_0^\bar{x} \frac{q_u}{q^2} \, dx \right] 
\]

where

\[
A = \frac{\bar{r}_u}{q} + \bar{r}_z \int_0^\bar{x} \frac{q_u}{q^2} \, dx < 0. 
\]

Since it is easily seen from differentiating the consumer equilibrium conditions that \( r_z = -xr_y \) and \( q_z = -xq_y \), \( G_2 \) can be rewritten as

\[
G_2 = \frac{1}{A} \left[ -\bar{r}_u \int_0^\bar{x} \frac{xq_y}{q^2} \, dx + \bar{x}r_y \int_0^\bar{x} \frac{q_y}{q^2} \, dx \right]. 
\]

Analysis of the consumer equilibrium conditions shows that \( q_y, r_u < 0 \). As a result,

\[
-\bar{r}_u \int_0^\bar{x} \frac{xq_y}{q^2} \, dx > -\bar{x}r_y \int_0^\bar{x} \frac{q_y}{q^2} \, dx, 
\]

noting that \( x \leq \bar{x} \) holds over the range of integration. Using Eq. (a4) and noting \( A < 0 \), it follows that

\[
G_2 < \frac{1}{A} \left[ -\bar{x}r_u \int_0^\bar{x} \frac{q_y}{q^2} \, dx + \bar{x}r_y \int_0^\bar{x} \frac{q_y}{q^2} \, dx \right] = -\bar{x}G_1 < -\bar{x}G_1, 
\]

establishing the desired inequality. The second inequality in \( a6 \) uses \( G_1 > 0 \) and \( \bar{x} > \bar{x} \).

References


\[20\] When \( y \) increases with utility held constant, which increases the vertical intercept of the budget line, the line itself must rotate clockwise to preserve the indifference curve tangency, raising \( c \) and reducing \( q \) (yielding \( q_y < 0 \)). When \( u \) rises, \( r \) must fall to allow the consumer to reach the higher indifference curve, yielding \( r_u < 0 \).