THE ROAD TO RUIN: PARASITIC SUBURBS AND THE FUNDAMENTAL LAW OF TRAFFIC CONGESTION

Martin Farnham

Peter Kennedy

Levi Megenbir

Department of Economics

University of Victoria

January 2018

ABSTRACT

We examine the strategic relationship between a city and a satellite suburb. Living costs are lower in the suburb but its residents must commute to the city to utilize an amenity that cannot be made available in the suburb. The amenity provides a service only when consumed jointly with a local public good, financed by property taxes. The road between the city and the suburb is subject to congestion. In the first-best solution, welfare is always increasing in the size of the road but a larger road can nonetheless lead to more congestion even though road use is priced correctly. If the city and the suburb instead act independently and non-cooperatively then two types of distortion arise: a free-rider incentive for residents to live in the suburb so as to avoid paying city taxes; and a congestion externality associated with unpriced road use. If other living costs are fixed then these distortions mean that too many residents live in the suburb, and the public good is underprovided relative to the first-best solution. In addition, a larger road can reduce or raise congestion, depending on whether the public good and private consumption are substitutes or complements respectively, but always *reduces* welfare either way. If living costs are endogenous and at least partly unpriced then the population of the suburb could in principle be too small – despite the aforementioned free-riding and unpriced road congestion – and a larger road could be welfare-improving, but only up to a point. As the suburb expands with increasing road size, it eventually becomes too big, and further increases in road size have a negative impact on welfare even if traffic congestion continues to fall.

1. INTRODUCTION

In 1972 a journalist named Clark Taylor penned an article for *Society* magazine that described the fiscal crisis facing Newark, New Jersey. The growth of suburbs outside the city was drawing the regional population into bedroom communities, beyond the reach of city taxation powers. Yet the residents of those bedroom communities routinely commuted to Newark for work, study and entertainment, where they utilized city services financed by taxes paid by the residents of Newark. Taylor called these bedroom communities "parasitic suburbs".

The urban landscape has changed a great deal since 1972 but parasitic suburbs continue to pose a problem for many cities across North America. The parasitic suburb problem is intertwined with, but importantly distinct from, the "flight from blight" phenomenon that has hollowed-out many inner-city areas and led to an income-based stratification of the urban landscape. Flight from blight is primarily driven by Tiebout sorting and has been widely studied in the literature on suburbanization. (See Nechyba and Walsh (2004) and Glaeser and Kahn (2004) for reviews).

The parasitic suburb problem is more narrow in scope. It arises when some valuable feature of the city cannot be reproduced in the suburbs, meaning that suburban residents cannot easily cut themselves off from the city entirely. Typically, this unique feature of the city is geographical or historical in nature. For example, a port facility – like the one in Newark – and its associated employment opportunities cannot be easily reproduced in an inland suburb. Similarly, a picturesque harbor or historical center draws non-residents to the city for recreation because the suburbs cannot offer a close substitute.

Crucially, a unique city amenity like a port or harbor usually provides a valuable service only when complemented with the provision of local public goods (such as police and fire services, and infrastructure like local roads, street lighting and walkways). If these local public goods are financed by taxes on city residents alone, then suburban residents are able to free-ride on the services they provide. Therein lies the source of the parasitic suburb problem.

Our purpose in this paper is two-fold. First, we wish to elucidate the nature of the parasitic suburb problem in the context of a non-cooperative game between city and suburban residents, and to characterize its implications for the equilibrium distribution of a regional population.

We study this game in the context of a simple variation on the monocentric city model. The standard monocentric model [Alonso (1964), Mills (1967) and Muth (1969)] explains the spatial distribution of the population around a city center in terms of the tradeoff between commuting costs and land rents. Land rents in equilibrium must be lower at distances further from the city center in order to offset higher commuting costs.

It is widely recognized that the standard monocentric city model does not fully capture the richness and complexity that characterizes suburbanization in practice, especially with respect to stratification by income and the development of multiple urban centers.[1] However, an analytical framework based on a single center is well-suited to our setting precisely because there is something unique about the city in the problem we examine. Moreover, we are less concerned here with the specific composition of the city and suburban populations, and more interested in the fiscal implications of the game between the city and the suburb.

We focus on the tradeoff between commuting costs and tax liability as a determinant of location choices. Suburban residents must commute to enjoy the city amenity but they also avoid the city taxes that finance the provision of the city-based local public good. The household location choice itself is binary – a household locates either in the city or the suburb – and commuting costs are endogenous via the impact of road congestion. The city tax rate, and provision of the local public good, are chosen by the city to maximize the welfare of city residents.

A second goal of our paper is to link suburbanization to the "fundamental law of traffic congestion", otherwise known as the Pigou-Knight-Downs paradox [Arnott and Small (1994)]. This "fundamental law" states that an increase in road capacity does not relieve congestion if the congestion is not priced because the larger road simply attracts more users in equilibrium.[2] We derive a necessary and sufficient condition under which

---

[1] Bartolome and Ross (2003 and 2004) integrate elements of Tiebout sorting into a monocentric city model and produce a richer set of outcomes with respect to the spatial distribution of households by income. See Anas and Xu (1999) for a model with multiple "centers".

[2] Mogridge (1997) provides a review of the theory and the broadly supportive evidence from London-based studies. Duranton and Turner (2011) provide compelling evidence in the context of the US highway

the law holds in the context of the parasitic suburb setting. More generally, we show that traffic congestion can rise or fall after an expansion of the road between the city and the suburb depending on whether private consumption and the city-based public good are complements or substitutes. In either case, an exogenous road expansion *reduces* equilibrium welfare because it facilitates greater free-riding on the city-provided public good.

The rest of our paper is organized as follows. Section 2 describes our model. Section 3 presents the regional planning problem as a benchmark against which the non-cooperative game between the city and the suburb is assessed. Section 4 characterizes the distribution of the regional population in the non-cooperative equilibrium, and examines the traffic congestion paradox. Section 5 introduces endogenous living costs. Section 6 provides some concluding remarks. All proofs are included in the Appendix.

2. THE MODEL

The region has two distinct communities: the city, and a satellite suburb. The populations of the city and the suburb are denoted N_C and N_S respectively. All residents of the region are identical in every respect except for their domicile. The total population for the region as a whole is fixed, so we normalize values to express the city and suburban populations as fractions of the whole. Thus, $N_C = 1 - N_S$.

The city has a non-reproducible amenity that provides a service to residents of the region when consumed in conjunction with a local public good, denoted *G*. All residents utilize the amenity and the associated public good but suburban residents must travel to the city to do so.

The public good is financed through property taxes on regional residents. Only city residents pay the tax in the non-cooperative game but we allow for a more general specification in which all regional residents are taxed so as to contrast the non-cooperative equilibrium with the first-best outcome. The tax levied on city residents is denoted t_C; the tax levied on suburban residents is denoted t_S. The regional budget must balance so $(1 - N_S)t_C + N_S t_S = G$.

system. In a related vein, Baum-Snow (2007a and 2007b) provides evidence of a causal link between US highway construction and suburbanization.

The round-trip cost of commuting to the city for a suburban resident is $v(N_S, \sigma)$, where $\sigma > 0$ measures the size of the road connecting the city and the suburb. We assume that $v(N_S, \sigma)$ is increasing in N_S and decreasing in σ. Total commuting cost (TCC) for the region is $N_S v(N_S, \sigma)$, and this is assumed to be strictly convex in N_S. This assumption on TCC places only a weak restriction on $v(N_S, \sigma)$ itself; namely, it can be convex or concave in N_S but it cannot be strongly concave at high values of N_S. (Specifically, $\partial v / \partial N_S$ cannot have an elasticity with respect to N_S that is less than -2). We also assume that the marginal benefit of a larger road – in terms of reduced commuting cost – is increasing in the volume of traffic; that is, $\partial^2 v / \partial \sigma \partial N_S < 0$. This simply means that a road expansion has a bigger impact on commuting cost if the road carries more traffic.[3]

Each resident in the region incurs living costs, in addition to any taxes or commuting costs. These living costs (which include the cost of housing) are non-discretionary and they generally differ between the two communities. Living costs are denoted l_C and l_S for the city and suburb respectively. We assume initially that l_C and l_S do not depend on the distribution of the regional population between the city and the suburb. This allows us to derive some sharp analytical results as a benchmark against which the implications of endogenous living costs can then be assessed (in Section 5).

All residents have income *m*. Living costs, commuting costs and taxes are all financed out of this income, and the residual is spent on a private good *x*, with unit price. Thus, for a city resident we have

(1) $$x_C = m - t_C - l_C$$

and for a suburban resident we have

(2) $$x_S = m - t_S - l_S - v(N_S, \sigma)$$

Disposable income for the region as a whole – defined as the total income available for expenditure on *G* and *x* – is denoted *Y*, and given by

(3) $$Y = (1 - N_S)(m - l_C) + N_S[m - l_S - v(N_S, \sigma)]$$

[3] Our assumptions on TCC are consistent with the functional forms typically assumed in the traffic congestion literature. See Downs (2004) for a widely-respected review and synthesis of the literature on congestion from both economic and traffic-engineering perspectives.

Preferences for each agent are represented by a utility function $u(G,x)$ that is strictly quasiconcave in *G* and *x*.

3. THE REGIONAL PLANNING PROBLEM

We first consider a setting in which a regional planner allocates residents between the city and the suburb, sets the level of the public good, and assigns taxes to the two communities to finance that public good. The objective of the planner is Pareto efficiency. In general, the solution to this planning problem is a Pareto frontier, constituting a continuum of Pareto-efficient solutions that differ according to the distribution of utility between city and suburban residents. We focus on a single point on this frontier: where the utilities of city and suburban residents are equal. This allows for a direct comparison with the non-cooperative equilibrium in which utilities are equalized through free location choices. We will henceforth refer to this equal-utility point on the Pareto frontier as the first-best solution. Note that G is necessarily the same for all residents, so the equal-utility condition implies that $x_S = x_C$.

Since commuting costs arise when any residents live in the suburb, the solution to the planning problem when $l_C \leq l_S$ is obvious: all residents should live in the city. In contrast, when $l_C > l_S$, commuting costs must be balanced against the living-cost savings achieved by locating some residents in the suburb. In that case the planning problem can have an interior solution with a positive suburban population, and it is on that case that we focus.

It does not matter whether we choose the utility of a suburban resident or that of a city resident as the maximand for the planning problem, so we arbitrarily pick that of a city resident. Thus, the planning problem is

(4) $$\max_{G,N_S,t_C,t_S} u(G, m - l_C - t_C)$$

subject to $x_S = x_C$ and $G = (1 - N_S)t_C + N_S t_S$. The first of these constraints is the equal-utility constraint; the second is the regional budget constraint. Using (1) and (2), we can write the first of these constraints as

(5) $$t_S = t_C + l_C - l_S - v(N_S, \sigma)$$

which we then substitute for t_S in the second constraint to obtain

(6) $$G = (1 - N_S)t_C + N_S[t_C + l_C - l_S - v(N_S, \sigma)]$$

Substituting (6) into the maximand in (4) then yields a straightforward maximization problem in two variables, t_C and N_S. The corresponding first-order conditions are

(7) $$\frac{\partial u}{\partial G} \Big/ \frac{\partial u}{\partial x} = 1$$

and

(8) $$v(N_S, \sigma) + N_S \frac{\partial v}{\partial N_S} = l_C - l_S$$

respectively. Strict quasiconcavity of $u(G, x)$ and strict convexity of $N_S v(N_S, \sigma)$ in N_S ensure that these conditions are necessary and sufficient for a global maximum.[4]

Equation (7) is a standard Samuelson condition for the efficient provision of a public good. It states that the sum of the marginal rates of substitution (MRS) between *x* and *G* across all residents of the region should be set equal to the rate at which *x* can be exchanged for *G* (the marginal rate of transformation, or MRT). In our setting, the number of regional residents has been normalized to one, and $\text{MRT} = 1$, so the Samuelson condition takes on a very simple form. It is nonetheless useful to maintain this interpretation of (7) because it will prove helpful in understanding some later results.

Equation (8) characterizes the optimal distribution of the population between the suburb and the city. The RHS is the marginal social benefit of allocating one more resident to the suburb, equal to the living-cost difference between the two communities. We will henceforth refer to this living-cost difference as the *city living-cost premium* (CLCP). The LHS is the marginal social cost of commuting (MSCC) for that additional suburban resident. This MSCC has two parts. The first part is the marginal private cost of commuting (MPCC) for that resident, equal to $v(N_S, \sigma)$. The second part is the marginal external cost of commuting (MECC), imposed on all other suburban residents via

[4] If $N_S v(N_S, \sigma)$ is not strictly convex then the optimum is a corner solution with either everyone living in the suburb ($N_S = 1$) or no one living in the suburb ($N_S = 0$). If $u(G, x)$ is not strictly quasiconcave then either $G = 0$ or $x = 0$ at the optimum.

increased congestion of the road; this is equal to $N_S(\partial v/\partial N_S)$. This MECC will play a prominent role in our analysis, so it will be useful at this point to assign it a label,

(9) $$\tau(N_S,\sigma) \equiv N_S \frac{\partial v}{\partial N_S}$$

A key property of the first-best solution is that this MECC forms part of the tax that suburban residents pay. In particular, if we substitute (8) into (5) we obtain the optimal suburban tax in relation to the optimal city tax:

(10) $$t_S^* = t_C^* + \tau(N_S,\sigma)$$

This states that suburban residents pay a tax premium over city residents, where the premium is equal to the MECC imposed on all other suburban residents. Thus, the tax premium is a standard Pigouvian congestion tax.

Since all tax revenue is used to finance the public good, the tax rule in (10) means that city residents pay less than their *per capita* share of the public good. This does not mean that city residents are advantaged relative to their suburban peers; utilities are equalized across the region by construction. In particular, note from (5) that the tax difference between city and suburban residents is

(11) $$t_S^* - t_C^* = l_C - [l_S + v(N_S^*,\sigma)]$$

That is, the tax difference is equal to the difference in non-discretionary costs for city and suburban residents. All residents are therefore left with the same residual income available for expenditure on x, and of course all residents benefit equally from G.

Now consider the relationship between regional welfare and the size of the road connecting the city and the suburb, as summarized in Proposition 1.

PROPOSITION 1. Let $\{G^*, N_S^*, t_C^*, t_S^*\}$ denote the solution to equations (6) – (8), and let u^* denote the utility of a regional resident at the optimum. If living costs are fixed, then

(a) N_S^* is increasing in σ

(b) G^* is increasing in σ if G is normal

(c) u^* is increasing in σ

(d) t_C^* is decreasing in σ if x is normal

(e) $t_S^* N_S^*$ is increasing in σ if G is normal

The intuition behind these results is as follows. An exogenous expansion of the road reduces the marginal cost of locating a resident in the suburb, since $\partial^2 v / \partial N_S \partial\sigma < 0$. The marginal benefit to that suburban resident – equal to the CLCP avoided – is unchanged because living costs are fixed. Additional residents should therefore be moved to the suburb. The road expansion also causes an increase in disposable regional income because lower living costs are extended to a greater number of residents, and because commuting costs *may* fall for existing suburban residents (under conditions we describe below in Proposition 2).

These impacts of the road expansion are illustrated in Figure 1. The figure depicts the first-best solution for N_S at two different road sizes, σ^0 and $\sigma^1 > \sigma^0$, reflected in the curves labeled $MSCC^0$ and $MSCC^1$ respectively. The optimal value of N_S in each case occurs where MSCC is just equal to $l_C - l_S$; these optimal suburban populations are labeled N_S^0 and N_S^1 for σ^0 and σ^1 respectively. Commuting cost for a suburban resident at these two solutions is v^0 and v^1 respectively, as determined by the curves labeled $MPCC^0$ and $MPCC^1$.

The increase in *Y* due to the increase in σ is equal to the sum of the two shaded areas. The area labeled ΔY_B is the increase in *Y* due to the extension of lower living costs to an additional $(N_S^1 - N_S^0)$ residents. The area labeled ΔY_A is the change in total commuting cost for existing suburban residents. The figure depicts a case where this commuting cost declines but we will see in Proposition 2 below that the opposite outcome is also possible. However, overall, *Y* must rise regardless.

The increase in *Y* means that the provision of *G* rises, if *G* is normal. Utility also rises for all residents because tax adjustments allow the higher Y to benefit everyone equally, regardless of whether or not they benefit directly from a fall in living costs or commuting costs. The exact nature of those tax adjustments depends critically on how the increase in Y is spent. In particular, if x is normal then some of the additional Y is optimally spent on *x*. Since any increased spending on *G* benefits everyone equally, any increased spending on *x* must also be the same for all residents (or else utilities would not remain equal across the region). Furthermore, since city residents gain no additional disposable income directly (either via a reduction in living costs or a reduction in commuting cost), the increased spending on *x* for these residents must be financed by a tax reduction. That is, t_C must fall.

Because the city tax rate falls, and the city population also falls, total tax revenue collected from the city must fall. Given that spending on *G* does not fall if *G* is normal, total tax collected from the suburb ($t_S N_S$) must rise. This does not necessarily require an increase in t_S because there are now more suburban residents subject to the tax; thus, the impact on t_S itself is indeterminate.[5]

We now turn to the question of whether or not a larger road reduces congestion.

PROPOSITION 2. Let α denote the elasticity of *MECC* with respect to N_S:

$$\alpha \equiv \frac{\partial \tau(N_S,\sigma)}{\partial N_S} \frac{N_S}{\tau(N_S,\sigma)}$$

Let $b(N_S,\sigma) \equiv \partial v / \partial \sigma$ denote the marginal benefit of a road expansion in terms of reduced commuting cost for a *given* level of N_S, and let

$$\beta \equiv \frac{\partial b(N_S,\sigma)}{\partial N_S} \frac{N_S}{b(N_S,\sigma)}$$

denote the elasticity of this marginal benefit function with respect to N_S. If living costs are fixed, then $v(N_S^*,\sigma)$

(a) is decreasing in σ if $\alpha > \beta$;

(b) is increasing in σ if $\alpha < \beta$; and

(c) is independent of σ if $\alpha = \beta$.

[5] However, we do know that *existing* suburban residents enjoy no saving in living costs from the road expansion, so we know that their additional consumption of *x* (if *x* is normal) must be financed by a reduction in t_S or a reduction in v; thus, at least one of these values – and possibly both – must fall.

This result is best explained with the aid of two diagrams. Recall that Figure 1 depicts the first-best solution for N_S at two different road sizes, σ^0 and $\sigma^1 > \sigma^0$. Proposition 2 relates to the behavior of $\tau(N_S, \sigma)$ in that figure in response to a change in σ. The size of $\tau(N_S, \sigma)$ determines the spread between the *MSCC* and *MPCC* curves for each value of σ. In particular, compare the vertical distances labeled *ab* and *cd*; these distances measure $\tau(N_S^0, \sigma^0)$ and $\tau(N_S^1, \sigma^1)$ respectively. The relative size of these two values depends on two opposing effects. On one hand, $\tau(N_S, \sigma)$ is increasing in N_S for any given value of σ, which means that the spread between the *MSCC* and *MPCC* curves grows as N_S rises. On the other hand, $\tau(N_S, \sigma)$ is decreasing in σ for any given value of N_S, since $\partial^2 v / \partial N_S \partial\sigma < 0$, which means that the spread between the *MSCC* and *MPCC* curves is smaller for σ^1 than for σ^0 at any given value of N_S. The net effect is determined by the relative responsiveness of these two effects with respect to N_S, as measured by α and β respectively. In the case depicted in Figure 1, $\alpha > \beta$; thus, $v^1 < v^0$. Figure 2 depicts the converse scenario. In this case $\alpha < \beta$, and so the negative impact of σ on $\tau(N_S, \sigma)$ outweighs the positive impact of N_S; thus, $v^1 > v^0$.

It may seem counter-intuitive that a road expansion could *optimally* lead to more congestion, as measured by commuting cost, but it is important to note that congestion *per se* is not the target of the planner; the planner acts to maximize regional welfare. This key distinction is highlighted in Figure 2, where the increase in disposable regional income is illustrated as the difference between the area labeled ΔY_B and the area labeled ΔY_A, the latter being negative in this case (in contrast to Figure 1). There is still a *net* increase in *Y*, and in regional welfare, despite the higher commuting cost because there are now more residents able to enjoy the lower living costs available in the suburb.[6]

[6] It may appear from Figure 2 that *v* could rise so much as to turn the net change in *Y* negative. This outcome cannot occur if second-order conditions for a maximum are satisfied. These conditions require that $2\partial v / \partial N_S + N_S \partial^2 v / \partial N_S^2 > 0$, which in turn places a lower bound on the spread between *MSCC* and *MPCC* in the figure.

An Example

To provide an additional sense of the mechanism behind Proposition 2 it is useful to consider a specific example. Suppose

$$v(N_S, \sigma) = N_S^{\gamma}(1 + \frac{N_S^{\delta}}{\sigma}) \tag{12}$$

where $\gamma \geq 1$ and $\delta < \gamma$. This function possesses all of the properties we have so far assumed about $v(N_S, \sigma)$, and is the actual function underlying the construction of Figures 1 and 2. It is straightforward to calculate α and β for this function, and to show that

$$\beta - \alpha = \frac{\delta\gamma\sigma}{\gamma\sigma + (\gamma + \delta)N_S^{\gamma}} \tag{13}$$

Thus, α is greater than or less than β according to whether $\delta < 0$ or $\delta > 0$ (since $\delta < \gamma$); that is, the impact of a road expansion on congestion hinges on a single parameter. In the knife-edge case where $\delta = 0$ it is possible to derive closed-form solutions for N_S^* and v^* for any given σ. These are

$$N_S^* = \left(\frac{(l_C - l_S)\sigma}{(1+\gamma)(1+\sigma)} \right)^{\frac{1}{\gamma}} \tag{14}$$

and

$$v^* = \frac{l_C - l_S}{1+\gamma} \tag{15}$$

respectively. Note that v^* is independent of σ in this knife-edge case; the road expansion has no impact on congestion.

Implementation of the Regional Planning Solution

A regional government with the authority to set *G* and impose taxes on city and suburban residents could in principle implement the first-best solution even if it has no direct control over resident location choices. To see this, recall that we have constructed the tax rules in (9) and (10) to ensure that utilities are equated across the region. This means that no resident has an incentive to relocate. Moreover, the equilibrium supported by these taxes is stable: any movement into or out of the suburb would reduce or raise suburban utility respectively – due to the change in congestion – thereby creating an incentive for the relocating resident to move back again.

However, while a regional government does not need authority over location choices to achieve the first-best solution here, it does need the political capacity to set higher taxes for suburban residents despite the fact that all tax revenue is used to fund the public good, and all regional residents utilize the public good equally. Thus, on the face of it, suburban residents subsidize city residents in terms of funding the public good, and this arrangement may be difficult to implement politically, especially given that suburban residents must commute to the city to utilize the public good.

A more serious obstacle to implementation of the first-best solution arises if there is no regional government at all. In particular, if the city government has no jurisdiction outside the limits of the city then it cannot tax suburban residents, and it cannot prevent city residents from joining the suburb to avoid paying for the public good altogether. We examine the equilibrium under this non-cooperative (parasitic suburb) scenario next.

4. NON-COOPERATIVE EQUILIBRIUM: THE PARASITIC SUBURB

The governance structure we examine here is one in which the city and the suburb act independently, and non-cooperatively. The city taxes its own residents to finance provision of the public good but it has no authority to tax suburban residents, who are nonetheless able to utilize the public good. Moreover, we assume that neither the city nor the suburb can impose a toll on the road between them because neither party has jurisdiction over it. This means that two distinct distortions arise in this setting: a free-rider incentive that drives residents to the suburb to avoid the city tax, compounded by a road congestion externality that puts too much traffic on the road.

The interaction between the city and the suburb is modeled as a one-shot, simultaneous-move game: the city sets t_C optimally (from the perspective of its own residents) based on a correct expectation of the equilibrium population it induces, and all residents correctly anticipate this equilibrium tax and choose their locations accordingly.

The city government acts to maximize the utility of a representative city resident, charging each resident a tax $t_C = G/(1-N_S)$ to finance the public good. The associated first-order condition is

(16) $$(1-N_S)\frac{\partial u}{\partial G}\Big/\frac{\partial u}{\partial x}=1$$

Like equation (7) from the planning problem in section 3, equation (16) is a standard Samuelson condition but with only $(1-N_S)$ users of the public good taken into account; the city has no incentive to account for the benefits that accrue to suburban residents. Let $G(N_S)$ denote the solution to (16) for any given N_S.

Residents of the region allocate themselves between the city and the suburb until utilities in the two communities are equal. Since all residents have equal access to *G*, this requires

(17) $$l_C+t_C=l_S+v(N_S,\sigma)$$

Setting $t_C=G(N_S)/(1-N_S)$ in (17) and solving for v yields

(18) $$v(N_S,\sigma)=l_C-l_S+\frac{G(N_S)}{1-N_S}$$

This condition characterizes the equilibrium suburban population, denoted $\hat{N}_S$. The LHS is the marginal private cost of locating in the suburb, equal to the equilibrium commuting cost. The RHS is the marginal private benefit of locating in the suburb. This comprises two terms: the CLCP avoided, plus the city tax avoided.

Comparing (18) with the first-best solution in (8), we have

(19) $$v(\hat{N}_S,\sigma)-v(N_S^*,\sigma)=N_S^*\frac{\partial v}{\partial N_S}+\frac{G(\hat{N}_S)}{1-\hat{N}_S}$$

This states that equilibrium commuting cost exceeds first-best commuting cost, due to two separate distortions. The first RHS term is the MECC; this is the increase in commuting cost imposed on other suburban residents as a group when one more resident locates in the suburb. In the absence of a congestion tax on road use, this cost is ignored by that marginal resident. The second RHS term reflects the free-rider incentive to locate in the suburb; this term is equal to the city tax avoided. These two distortions together mean that $\hat{N}_S>N_S^*$. We summarize these results in Proposition 3.

PROPOSITION 3. If living costs are fixed, then
(a) the equilibrium suburban population is higher than first-best; and
(b) the equilibrium commuting cost is higher than first-best.

The distortions associated with free-riding and unpriced road congestion also have an impact on the city tax and the amount of public good provided. Naturally, the precise impact depends on preferences but we can state some informative sufficient conditions, described in Proposition 4.

PROPOSITION 4. If living costs are fixed and *G* and *x* are both normal, then
(a) the equilibrium provision of *G* is lower than first-best; and
(b) the equilibrium city tax is higher than first-best if *G* and *x* are complements.

There are two effects behind these results. The first is a wealth effect. Recall from (10) that suburban residents pay a higher tax than city residents in the first-best solution: each suburban resident pays a Pigouvian congestion tax in addition to the tax that city residents pay. This means that city residents effectively receive a transfer of wealth from suburban residents in the first-best solution (though utilities are nonetheless equal across the region). This wealth transfer does not occur in the non-cooperative equilibrium. If *G* and *x* are both normal goods then this loss of wealth for city residents calls for a lower level of *G* and a smaller consumption of *x* relative to the first-best solution. Since $dx_C = -dt_C$ (because m and l_C are fixed), this reduction in the consumption of *x* stems from a higher tax.

The second effect behind the results in Proposition 4 is a price effect. In the first-best solution each city resident contributes

$$t_C^* = \frac{G - N_S \tau(N_S, \sigma)}{(1 - N_S) + N_S} = G - N_S \tau(N_S, \sigma) \tag{20}$$

towards *G*, where $\tau(N_S, \sigma)$ is the Pigouvian congestion tax paid by each of the N_S suburban residents. Thus, the *marginal price* of *G* for each city resident in the first-best setting is $g^* \equiv \partial t_C^* / \partial G = 1$.[7] In contrast, the marginal price of *G* for each city resident in the non-cooperative setting is $\hat{g} \equiv 1/(1-N_S) > 1$ because suburban residents are free-riders and make no contributions. If *G* is a normal good (and hence, *not* a Giffen good), then this higher price for *G* calls for a city-optimal level of *G* that is lower than the first-best level. This price effect on *G* reinforces the wealth effect described above.

The price effect on t_C is less definite, and so result (b) in Proposition 4 is more conditional. If *G* and *x* are substitutes then the higher price for *G* calls for a higher city-optimal consumption of *x*, financed by a *lower* city tax. In that case, the price effect and wealth effect operate in opposite directions and the overall effect on t_C cannot be ascertained without more specific information on preferences. Conversely, if *G* and *x* are complements then the higher price for *G* calls for a lower city-optimal consumption of *x*, and hence a higher city tax; in that case the price effect and wealth effect operate in the same direction, and the overall impact on t_C is unambiguously negative.

This relationship between t_C and the complementarity between *G* and *x* also turns out to be crucial with respect to how equilibrium congestion and public good provision are related to road size. We examine this question next.

Road Size, Congestion and Welfare in the Non-Cooperative Equilibrium

Recall from Proposition 2 that in the first-best world, a larger road may or may not reduce commuting cost (depending on the elasticities α and β), but it is welfare-improving regardless. In the non-cooperative equilibrium, commuting cost could also be higher or lower with a larger road, but for quite different reasons. Moreover, the impact of higher road capacity on welfare is *unambiguously negative*.

To reach these results, we begin with condition (18), which characterizes the equilibrium suburban population. Differentiating both sides of (18) with respect to σ and collecting terms yields

[7] Note that the solution to (8) for N_S in the planning problem does not depend on preferences; it is a function only of the physical congestion properties of the road. This means that $\tau(N_S, \sigma)$ is independent of preferences, and hence independent of *G*, at the optimum. Thus, $\partial\tau / \partial G$ does not appear in g^*.

(21)
$$\frac{dN_S}{d\sigma} = \frac{\dfrac{\partial v}{\partial \sigma}(1-N_S)^2}{G + (1-N_S)\dfrac{\partial G}{\partial N_S} - (1-N_S)^2 \dfrac{\partial v}{\partial N_S}}$$

We know that $\partial v / \partial \sigma < 0$, so the sign of $dN_S / d\sigma$ hinges on the sign of the denominator in (21). If the equilibrium is *stable*, then this term is positive. To see this, recall the equilibrium condition from (18). Stability of the equilibrium means that the LHS of (18) crosses the RHS from below, as depicted in Figures 3 and 4 (which illustrate the two possible relationships between t_C and N_S). Thus, stability requires

(22)
$$\frac{\partial v}{\partial N_S} > \frac{G + (1-N_S)\dfrac{\partial G}{\partial N_S}}{(1-N_S)^2}$$

and this in turn means that the denominator in (21) is positive. Thus, without further proof, we can state the following result.

PROPOSITION 5. If living costs are fixed and the equilibrium is stable then a larger road induces a higher suburban population.

We can now derive a key result regarding the relationship between road capacity and equilibrium congestion. From (18) we have

(23)
$$\frac{dv}{d\sigma} = \frac{1}{1-N_S}\left(\frac{\partial G}{\partial N_S} + \frac{G}{1-N_S}\right)\frac{dN_S}{d\sigma}$$

Recall that the marginal price of *G* for city residents is $\hat{g} = 1/(1-N_S)$. Hence, we can express $dv / d\sigma$ in terms of an elasticity:

(24)
$$\frac{dv}{d\sigma} = g^2 G(1+\varepsilon)\frac{dN_S}{d\sigma}$$

where ε is the own-price elasticity of *G*. This means that the sign of $dv / d\sigma$ hinges on whether the demand for *G* is elastic or inelastic. Since there are only two goods in this setting, we can in turn relate the sign of $dv / d\sigma$ directly to the complementarity between *G* and *x*. We state this as Proposition 6.

PROPOSITION 6. If living costs are fixed and G is normal then a larger road
(a) reduces equilibrium commuting cost if G and x are substitutes;
(b) raises equilibrium commuting cost if G and x are complements; and
(c) has no impact on equilibrium commuting cost if G and x are cross-price neutral.

These results can be explained as follows. A larger road reduces commuting cost for suburban residents, but has no direct effect on city residents. This gives city residents an incentive to relocate to the suburb, and equilibrium is restored only when utilities are equalized across the region. There are two mechanisms through which that equalization occurs. The first is road congestion. As city residents relocate to the suburb they bring higher traffic volumes with them, and commuting costs start to rise again. In the absence of any other equalization mechanism, commuting cost would necessarily rise back to its pre-expansion level. This is the "fundamental law of traffic congestion", as it applies to this setting. However, a second mechanism is at play here.

As city residents relocate to the suburb, the remaining city residents must carry a higher burden in terms of funding G; the price of the public good for city residents rises in *per capita* terms. If G is normal then this price rise leads the city to reduce provision of G. This in itself has no bearing on the relative appeal of city and suburban living since all residents of the region are affected equally by a change in G. What matters for relocation incentives is the impact on city taxes. If G and x are *substitutes* then a higher price for G means an increased demand for x by city residents, which must be financed by a tax cut. This city tax cut reduces the relative appeal of the suburb and thereby dampens the rise in road congestion. Equilibrium is therefore re-established before commuting cost rises back to its pre-expansion level. Thus, the road expansion causes equilibrium congestion to fall.

Conversely, if G and x are *complements* then a higher price for G reduces the demand for x. The city-optimal response is to raise the city tax and apply the funding to G. This tax increase boosts the appeal of the suburb and amplifies the incentive for city residents to relocate. Commuting cost must therefore rise beyond its pre-expansion level in order to offset this additional tax-induced incentive to locate in the suburb.

Recall from Section 3 that whether or not a road expansion reduces commuting cost in the first-best world depends critically on the behavior of MECC in response to changes in σ and N_S. In the non-cooperative equilibrium, MECC plays no role at all in terms of relocation incentives precisely because it is *external*; the congestion externality is not priced. On the other hand, city tax changes are a key determinant of post-expansion commuting costs in the non-cooperative setting but not in the first-best world; in that world suburban residents must pay the tax too. Thus, the potential for commuting costs to rise or fall in response to the road expansion exists in both the first-best and non-cooperative settings, but for very different reasons.

The non-cooperative and first-best settings are also very different in terms of the welfare implications of the road expansion. To examine welfare change in the non-cooperative setting we focus on the utility of a pre-expansion suburban resident. (Since utilities are equal across the region in equilibrium, both before and after the road expansion, we know that all residents experience the same utility change regardless of their pre- or post-expansion location). Recall that the utility of a suburban resident is $u(G, m - p_S - v)$. Differentiation with respect to σ yields

(25) $$\frac{du}{d\sigma} = \frac{\partial u}{\partial G}\frac{\partial G}{\partial N_S}\frac{dN_S}{d\sigma} - \frac{\partial u}{\partial x}\frac{dv}{d\sigma}$$

Using (23) for $dv/d\sigma$ and (16) for $\partial u/\partial x$, this reduces to

(26) $$\frac{du}{d\sigma} = -\frac{\partial u}{\partial G}\frac{dN_S}{d\sigma}\left(\frac{G}{1-N_S}\right)$$

Since $dN_S/d\sigma > 0$ if the equilibrium is stable (by Proposition 5), we have the following result without further proof.

PROPOSITION 7. If living costs are fixed, *G* is normal and the equilibrium is stable then welfare is decreasing in road size.

To understand this result, imagine for the moment that there is no public good. In particular, suppose the city amenity provides an essential service to city and suburban residents without the need for *G*. In that case, a lower cost-of-living creates the only incentive for residents to locate in the suburb; there are no city taxes to avoid. Then the equilibrium condition before and after the road expansion is simply $v(N_S, \sigma) = l_C - l_S$; thus, commuting cost must remain unchanged. The absence of congestion pricing means that all of the potential benefits of the road expansion are dissipated via an increase in the number of commuters. Utility therefore remains unchanged.

Now reintroduce the public good, and suppose it is normal. The ability of residents to free-ride by locating in the suburb means that *G* is under-provided relative to first-best (by Proposition 4). The road expansion makes that under-provision worse by drawing more people to the suburb, thereby raising the number of free-riders in equilibrium. The impact on welfare is unambiguously negative.[8]

5. ENDOGENOUS LIVING COSTS

We now extend our analysis to allow living costs to depend on community size. The most obvious source of this dependence is a housing-price effect: higher demand for housing in a land-limited area puts upward pressure on its price. This housing-price effect adds another mechanism – in addition to taxes and commuting cost – through which utilities become equalized across the region via equilibrium location choices.

Other components of living costs may also be endogenous, and many of these may not be priced. For example, crowding in areas to which residents have open access – such as local roads, parks and shopping malls – typically raises the time-cost of using these services or reduces the quality of the services provided. Similarly, unpriced air and noise pollution can reduce quality of life. In addition, the increased anonymity that typically accompanies community growth can erode trust between community members and thereby raise transaction costs. Also important is free-riding *within* a community – in areas such as local charity, maintenance of public spaces, volunteer services, and so on – that arguably gets worse as the population grows.

[8] The only scenario in which equilibrium welfare can possibly rise after the road expansion is one in which *G* is inferior, but not Giffen. In that case, the wealth effect associated with the absence of congestion-tax revenue can mean that *G* is *over-provided* relative to first-best (recall the discussion following Proposition 3). Provided *G* is not Giffen, the road expansion still induces a higher rate of free-riding via relocation to the suburb but this can actually be welfare-improving because it reduces the over-provision of *G*.

In principle, if these unpriced externalities associated with population size are sufficiently large, and the city population itself is sufficiently large, then they could reverse the direction of inefficiency we have identified in Proposition 3. In particular, the suburban population could be too small in equilibrium. On the other hand, if the suburb is sufficiently large, then these additional externalities could *exacerbate* the excessive size of the suburb and the associated under-provision of G.

To explore this issue further, we assume that living costs are increasing in population size for both the city and the suburb, and that any external impacts of these effects are unpriced. Aggregate living costs for the city population as a whole are $L_C(N_S) = (1 - N_S)l_C(N_S)$, where $\partial l_C / \partial N_S < 0$. Thus, when a city resident relocates to the suburb, these costs *fall*. There are two components to that cost reduction:

$$\frac{\partial L_C}{\partial N_S} = -l_C(N_S) + (1 - N_S)\frac{\partial l_C}{\partial N_S} \tag{27}$$

The first RHS term is a private benefit to the resident who leaves the city. The second RHS term is an external benefit bestowed on all other city residents, and this benefit is ignored in the private location decision.

Similarly, aggregate living costs for the suburban population as a whole are $L_S(N_S) = N_S l_S(N_S)$, where $\partial l_S / \partial N_S > 0$. When a city resident relocates to the suburb, these costs *rise*. Again, there are two components to that cost increase:

$$\frac{\partial L_S}{\partial N_S} = l_S(N_S) + N_S \frac{\partial l_S}{\partial N_S} \tag{28}$$

The first RHS term is a private cost to the new suburban resident. The second RHS term is an external cost imposed on all other suburban residents, and again this cost is ignored in the private location decision.

The presence of these external effects associated with population change mean that the net private benefit of moving from the city to the suburb (in terms of reduced living costs) differs from the net social benefit of doing so, quite beyond any distortion associated with road congestion and free-riding on the public good. The direction of this difference depends critically on the size of the city relative to the suburb. In particular,

$$l_C(N_S) - l_S(N_S) > (<) -\frac{\partial L_C}{\partial N_S} - \frac{\partial L_S}{\partial N_S} \quad \text{iff} \quad N_S > (<) \overline{N}_S \tag{29}$$

where $\overline{N}_S$ solves

(30) $$\overline{N}_S \frac{\partial l_S}{\partial N_S} + (1 - \overline{N}_S)\frac{\partial l_C}{\partial N_S} = 0$$

At $N_S = \overline{N}_S$ the external cost imposed on suburban residents when one more resident moves to the suburb is exactly offset by the external benefit bestowed on city residents. Thus, the change in living costs for that marginal suburban resident, and the change in living costs for society as a whole, are one and the same.

The relationship between private and social net benefits when a resident moves from city to suburb is illustrated in Figure 5, together with the *MSCC* curves corresponding to two different road sizes, σ_0 and $\sigma_1 > \sigma_0$. The first-best suburban populations are labeled N_S^0 and N_S^1 for road sizes σ^0 and σ^1 respectively. The dashed horizontal line labeled $l_C^F - l_S^F$ depicts the CLCP in the fixed living-cost case, with the associated first-best suburban populations labeled N_S^{0F} and N_S^{1F} for σ^0 and σ^1 respectively. Note that $l_C^F - l_S^F$ must cross $l_C(N_S) - l_S(N_S)$ at $\overline{N}_S$ since $\partial l_S / \partial N_S = 0$ and $\partial l_C / \partial N_S = 0$ in the fixed living-cost case.

A positive link between living costs and population size has three key implications for our results. First, the impact of a larger road on the suburban population is more muted than when living costs are fixed. This holds true for both the first-best and non-cooperative settings. In both settings, there is an additional mechanism through which utilities are equalized between city and suburban residents as the suburban population expands. Consequently, the relationship between road size and commuting cost is more likely to be negative, all other things equal.

Second, the relationship between road size and welfare in the non-cooperative setting is no longer unambiguously negative. In particular, if the link between living costs and population size is sufficiently strong then the relationship between σ and equilibrium welfare has an inverted-U shape. At small values of σ, the equilibrium suburban population ($\hat{N}_S$) is small, and hence $\hat{N}_S < \overline{N}_S$. This means that the net private benefit of moving from city to suburb (in terms of reduced living costs) is *smaller* than the net social benefit of doing so (recall Figure 5). This under-valuation of living-cost savings could be large enough to more-than-offset the distortions associated with free-riding and unpriced congestion that otherwise lead city residents to over-value a move to the suburb. If so, the positive impact of a larger road on the suburban population is welfare-improving.

That impact on welfare must eventually turn negative as the under-valuation of living-cost savings falls to the point where it becomes an *over-valuation*, as $\hat{N}_S$ grows beyond $\overline{N}_S$ (again, recall Figure 5). At $\hat{N}_S > \overline{N}_S$, the living-cost effect augments the free-riding and congestion distortions, and welfare must decline as σ rises. Thus, the overall relationship between σ and equilibrium welfare is non-monotonic – if the living cost effect is large enough – and the turning point of that relationship must occur at a suburban population somewhere below $\overline{N}_S$.

The third implication of endogenous living costs also relates to the under-valuation of living-cost savings at low values of N_S. If that under-valuation is large enough then the equilibrium suburban population can be too *small* relative to first-best, at low values of σ; that is, Proposition 3 can be reversed. However, this can only occur if $N_S^* < \overline{N}_S$, and only then if the living cost effect is large relative to the free-riding and congestion distortions.

An Example

A simple example may shed more light on the implications of endogenous living costs. Suppose commuting cost is given by

(31) $$v(N_S, \sigma) = N_S(1 + \frac{1}{\sigma})$$

This is the same function used for the example in Section 3 – see equation (12) – but with $\delta = 0$ and $\gamma = 1$. Recall from that example that setting $\delta = 0$ means that first-best commuting cost is independent of σ when living costs are fixed. Imposing that restriction here allows us to focus directly on the implications of endogenous living costs. Setting $\gamma = 1$ allows the derivation of closed-form solutions.

In a similar vein, suppose that utility is log-linear:

(32) $$u(G, x) = \theta \log(G) + \log(x)$$

where $\theta > 0$. This specification means that *G* and *x* are both normal, and cross-price neutral. Recall from Proposition 6 that this in turn means that equilibrium commuting cost is independent of σ when living costs are fixed. Again, this allows us to focus directly on the implications of endogenous living costs.

Now suppose that

$$l_S(N_S) = k_S + \lambda N_S \tag{33}$$

where $k_S > 0$ is a constant, and $\lambda > 0$ reflects the strength of the link between living costs and population size. Similarly, suppose

$$l_C(N_S) = k_C + \lambda(1 - N_S) \tag{34}$$

where $k_C > k_S$. Note that λ is common between the two cost functions. This symmetry between the city and the suburb is unlikely to hold in a real setting but for the purposes of the example it serves to allow the strength of the living-cost effect to be captured by a single parameter. It also implies that $\overline{N}_S = \frac{1}{2}$.

It is straightforward to derive the first-best suburban population and associated commuting cost in this setting. These are given by

$$N_S^* = \frac{\sigma(k_C - k_S + 2\lambda)}{2(1 + \sigma + 2\sigma\lambda)} \tag{35}$$

and

$$v^* = \left(\frac{1+\sigma}{\sigma}\right) N_S^* \tag{36}$$

respectively.

There are two points of interest here. First, differentiating (35) with respect to λ reveals that $\partial N_S^* / \partial \lambda > 0$ if $N_S^* < \overline{N}$ and $\partial N_S^* / \partial \lambda < 0$ if $N_S^* > \overline{N}$. This reflects the fact that the effects of λ on L_C and L_S are exactly offsetting at $\overline{N}$. The second point of interest is that $\partial v^* / \partial \sigma = 0$ when $\lambda = 0$, but $\partial v^* / \partial \sigma < 0$ when $\lambda > 0$. That is, when living costs are endogenous, a larger road means less congestion at the optimum. This reflects the first implication we identified above: the impact of a larger road on the suburban population is more muted than when living costs are fixed.

The same is true for the non-cooperative equilibrium. The equilibrium suburban population and associated commuting cost are given by

$$\hat{N}_S = \frac{\sigma(k_C - k_S + \theta(m - k_S) + \lambda)}{(1+\theta)(1+\sigma+\sigma\lambda)+\sigma\lambda} \tag{37}$$

and

$$\hat{v} = \left(\frac{1+\sigma}{\sigma}\right)\hat{N}_S \tag{38}$$

respectively. As with v^*, $\partial\hat{v}/\partial\sigma = 0$ when $\lambda = 0$, but $\partial\hat{v}/\partial\sigma < 0$ when $\lambda > 0$. Again this reflects the fact that the decline in the CLCP – but in this case, the *private* CLCP – that comes with suburban growth acts to moderate that growth as σ rises.[9]

Next consider the impact of σ on equilibrium welfare. The expression for utility at the equilibrium is too messy to report here, but it is straightforward to show that it reaches a maximum at a value $\sigma = \overline{\sigma}$, where $\overline{\sigma} > 0$ if and only if

$$\lambda > \overline{\lambda} \equiv \left(\frac{1+\theta}{1+2\theta}\right)\hat{t}_C \tag{39}$$

where

$$\hat{t}_C = \frac{\theta(m - k_C)}{1+\theta} \tag{40}$$

Thus, if $\lambda > \overline{\lambda}$ and σ is not too large then an increase in σ is welfare-improving. Conversely, if $\lambda < \overline{\lambda}$ then welfare is declining in σ for all $\sigma > 0$. Note the importance of $\hat{t}_C$ for the size of $\overline{\lambda}$. This tax rate is the city tax in the non-cooperative equilibrium when living costs are *fixed*. Why is it relevant here? The living-cost effect can reverse the otherwise-negative welfare effect of a larger σ only if it is strong enough to more-than-offset the congestion and free-riding distortions. In this simple example, congestion is neutral with respect to σ, so everything hinges on the size of the free-riding distortion, and that distortion is directly proportional to the equilibrium city tax.

Finally, consider the impact of endogenous living costs on the relationship between $\hat{N}_S$ and N_S^*. In particular, are there conditions under which the suburban

---

[9] Note from (37) that the equilibrium suburban population is increasing in income. In contrast, the first-best suburban population is not; see (35). Thus, one characteristic of this simple example is that income growth makes the parasitic suburb problem worse.

population can be too *small*? Using (35) and (37), it is straightforward to show that there exists a value $\tilde{\sigma}$ such that $\hat{N}_S < N_S^*$ if and only if $\sigma < \tilde{\sigma}$, but where $\tilde{\sigma} > 0$ if and only if

$$\lambda > \tilde{\lambda} \equiv \left(\frac{1+\theta}{2\theta}\right)\left(\hat{t}_C + \hat{\tau}\right) \tag{41}$$

where

$$\hat{\tau} = \frac{k_C - k_S + \theta(m - k_S)}{1+\theta} \tag{42}$$

is the *MECC* at the equilibrium when living costs are fixed.[10] That is, if $\lambda > \tilde{\lambda}$ then the suburb is too small relative to first-best but only if the road is small. Conversely, if $\lambda < \tilde{\lambda}$ then $\hat{N}_S > N_S^*$ for all $\sigma > 0$.

The condition in (41) can be explained as follows. For the suburb to be too small in equilibrium, the living-cost effect must be strong enough to more-than-offset the combined distortion associated with free-riding and unpriced road congestion. This combined distortion is in turn proportional to the sum of the equilibrium city tax and the unpriced equilibrium MECC. In the simplest setting where $\theta = 1$ (such that *G* and *x* have equal utility weight), $\tilde{\lambda} = \hat{t}_C + \hat{\tau}$.

6. CONCLUSION

We have examined the strategic relationship between a city and a satellite suburb. Living costs are lower in the suburb but its residents must commute to the city to utilize an amenity that cannot be made available in the suburb. The amenity provides a service only when consumed jointly with a local public good. The characteristics of the city-suburb relationship depend critically on the type of governance structure in place. A regional government with authority to impose taxes on city and suburban residents can in principle achieve the first-best outcome in terms of public good provision and the allocation of regional residents between the city and the suburb. In such a setting, welfare is always increasing in the size of the road connecting the city and the suburb. A larger road can nonetheless lead to more congestion even though road use is priced correctly.

[10] This is independent of σ only because *G* and *x* are cross-price neutral in this example.

The relationship is very different if the city and the suburb act independently and non-cooperatively. In that case, two types of distortion arise: a free-rider incentive for residents to live in the suburb so as to avoid paying the city tax that finances the public good; and a congestion externality associated with unpriced road use. If living costs are fixed then these distortions mean that too many residents live in the suburb and the public good is underprovided relative to the first-best solution. In addition, a larger road can reduce or raise congestion, depending on whether the public good and private consumption are substitutes or complements respectively, but always reduces welfare either way.

The properties of the non-cooperative equilibrium are less clear-cut if living costs are endogenous, and where at least some of those costs are not priced (as for example, with crowding *within* the city or the suburb). In such a setting, the population of the suburb could in principle be too small – despite the aforementioned free-riding and unpriced road congestion – and a larger road could be welfare-improving, but only up to a point. As the suburb expands with increases in road size, it eventually becomes too big, and further increases in road size have a negative impact on welfare even if traffic congestion continues to fall.

Our analysis has focused on the impact of expanded road capacity but it is important to stress that any non-price measure that attempts to reduce congestion will have similar effects on the region, including the provision of an alternative transportation service (such as light rail). The simple logic of the "fundamental law of traffic congestion" applies to any measure that attempts to relieve congestion without pricing the congestion externality itself [Arnott and Walsh (1994) and Downs (2004)]. In the parasitic suburb setting, an alternative transportation link would draw more residents to the suburb and exacerbate the free-riding problem. The ultimate source of the misallocation of residents between city and suburb – the combined effects of free-riding and unpriced congestion – cannot be eliminated via increased transportation capacity, regardless of its form.

There does of course appear to be one simple solution to the parasitic suburb problem: amalgamation. The homogeneity of residents in our model means that the first-best solution – and its implementation via a regional government – Pareto-dominates the non-cooperative equilibrium; city and suburban residents would all be better-off under amalgamation.

In practice, the benefits of amalgamation are rarely so universal. Quite apart from political issues associated with fiefdoms and turf-wars, there are two key impediments to welfare-improving amalgamation in practice. The first is heterogeneity across residents with respect to preferences or income. If residents sort themselves between city and suburb along these lines then Pareto-dominance of the first-best solution is no longer assured. Mutually beneficial amalgamation will typically require wealth transfers from the winning community to the losing community, and such transfers are difficult to orchestrate in practice.

A second impediment to amalgamation in practice arises if there is more than one suburb. Imagine a generalization of our simple model in which independent communities are located along a ring around the city, their locations determined by historical or geographic features.[11] In this context, amalgamation becomes a "treaty problem" very much like the one that underlies global environmental problems like climate change. Even if the first-best (amalgamation) solution Pareto-dominates the non-cooperative equilibrium, each community has an incentive to free-ride on an amalgamation agreement between the city and all *other* communities. A key insight from the literature on environmental treaties is that a stable treaty with universal membership (a grand cooperative coalition) rarely exists [Barrett (1994)]. If a message for urban governance can be gleaned from that literature, it is that amalgamation may need to be forced upon communities by a higher order of government.

[11] See Wu (2006) for an interesting model in which suburban sprawl is governed partly by geophysical features of the landscape.

APPENDIX

Proof of Proposition 1

(a) Differentiation of (8) with respect to N_S and σ yields

(A1) $$\frac{dN_S}{d\sigma} = -\left(\frac{\partial v}{\partial \sigma} + N_S \frac{\partial^2 v}{\partial N_S \partial \sigma}\right) \Bigg/ \left(2\frac{\partial v}{\partial N_S} + N_S \frac{\partial^2 v}{\partial N_S^2}\right)$$

The denominator is positive by strict convexity of $N_S v(N_S, \sigma)$ in N_S (and must be positive for conditions (7) and (8) to yield a maximum), and the numerator is positive since $\partial v / \partial \sigma < 0$ and $\partial^2 v / \partial N_S \partial \sigma < 0$.

(b) Suppose the planner makes no change to N_S in response to the increase in σ, and let $dv_0 < 0$ denote the resulting passive change in commuting cost. Then the passive change in disposable income is $dY_0 = -N_S dv_0 > 0$. An *optimal* response by the planner cannot yield a smaller change. Thus, $dY > 0$. If G is a normal good then it follows by definition that $dG > 0$.

(c) Again, suppose the planner makes no change to N_S in response to the increase in σ, and let $dv_0 < 0$ denote the resulting passive change in commuting cost. Then the planner can raise the tax on all suburban residents by $dt_S = -dv_0$ and use this revenue to fund an increase in the public good, $dG = -N_S dv_0$. Utility is still equated across regional residents, since x remains unchanged for all residents, and utility rises by $-(\partial u / \partial G) N_S dv_0$ for all residents. An *optimal* response by the planner can be no worse.

(d) Let $T = t_C N_C + t_S N_S$ denote total tax revenue collected from the region. Let the superscript $i = 0$ denote an optimal value when $\sigma = \sigma^0$, and let the superscript $i = 1$ denote an optimal value when $\sigma = \sigma^1 > \sigma^0$. Then

(A2) $$dT = [t_C^1(1 - N_S^1) + t_S^1 N_S^1] - [t_C^0(1 - N_S^0) + t_S^0 N_S^0]$$

Setting $dT = dG$ and solving for t_C^1 yields

(A3) $$t_C^1 = \frac{t_C^0(1 - N_S^0) + t_S^0 N_S^0 - t_S^1 N_S^1 + dG}{1 - N_S^1}$$

Since the utilities of a city and suburban resident are equal before and after the road expansion, it follows that

(A4) $t_S^0 = t_C^0 + l_C - l_S - v^0$

and

(A5) $t_S^1 = t_C^1 + l_C - l_S - v^1$

Making these substitutions into (A3) yields

(A6) $t_C^1 = t_C^0 + N_S^0(l_C - l_S - v^0) - N_S^1(l_C - l_S - v^1) + dG$

Expressing this in difference form yields

(A7) $dt_C = N_S^0 dv - dN_S(l_C - l_S - v_1) + dG$

Now note that the change in disposable regional income is

(A8) $dY = N_S^1(l_C - l_S - v^1) - N_S^0(l_C - l_S - v^0)$

Expressing this in difference form yields

(A9) $dY = -N_S^0 dv + dN_S(l_C - l_S - v_1)$

Thus, $dt_C = -dY + dG$. Now let $dG = \phi dY$. Then $dt_C = -(1-\phi)dY$. Since $dY > 0$, it follows that $dt_C < 0$ if $\phi < 1$, and we know that $\phi < 1$ if *x* is normal (that is, if at least some of the disposable regional income gain is spent on *x*).

(e) Let T_C denote total tax revenue collected from the city and let T_S denote total tax revenue collected from the suburb. Then by the balanced budget condition,

(A10) $dT_S = dG - dT_C$

Since $T_C = t_C(1 - N_S)$, it follows that

(A11) $dT_C = t_C^1(1 - N_S^1) - t_C^0(1 - N_S^0)$

Expressing this in difference form yields

(A12) $dT_C = dt_C(1 - N_S^1) - dN_S t_C^0$

Since $dt_C < 0$ by result (d) above, and $dN_S > 0$ by result (a) above, it follows that $dT_C < 0$. Thus, from (A10) we know that $dT_S > 0$ for any $dG > 0$, and we know from result (b) above that $dG > 0$ if *G* is normal. ♣

Proof Proposition 2

Differentiating $v(N_S, \sigma)$ with respect to σ yields

(A13) $$\frac{dv}{d\sigma} = \frac{\partial v}{\partial N_S}\frac{dN_S}{d\sigma} + \frac{\partial v}{\partial \sigma}$$

Substituting for $dN_S / d\sigma$ from (A1) and collecting terms yields

(A14) $$\frac{dv}{d\sigma} = \frac{\frac{\partial v}{\partial \sigma}\left(\frac{\partial v}{\partial N_S} + N_S \frac{\partial^2 v}{\partial N_S^2}\right) - N_S \frac{\partial v}{\partial N_S}\frac{\partial^2 v}{\partial N_S \partial \sigma}}{2\frac{\partial v}{\partial N_S} + N_S \frac{\partial^2 v}{\partial N_S^2}}$$

which can be then be expressed in terms of the elasticities from the text:

(A15) $$\frac{dv}{d\sigma} = -\left(\frac{\frac{\partial v}{\partial \sigma}\frac{\partial v}{\partial N_S}}{2\frac{\partial v}{\partial N_S} + N_S \frac{\partial^2 v}{\partial N_S^2}}\right)(\beta - \alpha)$$

where

(A16) $$\alpha = 1 + N_S \frac{\partial^2 v}{\partial N_S^2} \bigg/ \frac{\partial v}{\partial N_S}$$

and

(A17) $$\beta = N_S \frac{\partial^2 v}{\partial \sigma \partial N_S} \bigg/ \frac{\partial v}{\partial \sigma}$$

The term inside the large brackets in (A15) is negative: the numerator is negative by $\partial v / \partial \sigma < 0$ and $\partial v / \partial N_S > 0$, and the denominator is positive by strict convexity of $N_S v(N_S, \sigma)$ in N_S. Thus, the sign of this expression takes the sign of $(\beta - \alpha)$.♣

Proof of Proposition 3

Part (b) follows from (19) in the text. Since $v(N_S, \sigma)$ is increasing in N_S, part (a) follows directly from part (b).♣

Proof of Proposition 4

There are two effects to consider for both *G* and *x*: a wealth effect and a price effect. First consider the wealth effect. Let Y^* denote regional disposable income at the first-best solution:

(A18) $$Y^* = m - (1 - N_S^*)l_C - N_S^*[l_S + v(N_S^*, \sigma)]$$

Substituting for $v(N_S^*, \sigma)$ from (8), this reduces to

(A19) $$Y^* = m - l_C + N_S^* \tau(N_S^*, \sigma)]$$

Now let y^* denote the corresponding regional income *per capita*. Since $N = 1$ by our normalization, $y^* = Y^*$. In comparison, disposable income *per capita* for city residents in the non-cooperative equilibrium is

(A20) $$\hat{y}_C = m - l_C$$

From (A19) and (A20), it is clear that $\hat{y}_C < y^*$. Next consider the price effect. The *marginal price* of *G* for each city resident in the first-best solution is $g^* \equiv \partial t_C^* / \partial G = 1$; see equation (20) in the text. In contrast, the marginal price of *G* for each city resident in the non-cooperative equilibrium is $\hat{g} \equiv 1/(1 - N_S) > 1$. Thus, $\hat{g} > g^*$.

(a) If *G* is normal (and hence *not* Giffen) then the price and wealth effects on *G* are both negative. It follows that $\hat{G} < G^*$.

(b) If *G* and *x* are complements, and *x* is normal, then the price and wealth effects on *x* are also both negative. It follows that $\hat{x} < x^*$. ♣

Proof of Proposition 6

We begin with a standard result from consumer theory, stated here as Lemma 1.

Lemma 1. Suppose there are only two goods, z_1 and z_2, and let ρ_1 denote the own-price elasticity of demand for z_1. Then z_1 and z_2 are

(a) complements if $|\rho_1| < 1$

(b) substitutes if $|\rho_1| > 1$

(c) cross-price neutral if $|\rho_1| = 1$

Proof. In general, from the consumer budget constraint we have

(A21) $$\sum_{i=1}^{n} p_i z_i(p, m) = m$$

where p_i denotes the price of good *i*, *m* denotes income, and $z_i(p,m)$ is the Marshallian demand for good *i*. Differentiating both sides of (A21) with respect to p_j yields the Cournot aggregation condition:

(A22) $$z_j + \sum_{i=1}^{n} p_i \frac{\partial z_i}{\partial p_j} = 0$$

In the case of just two goods, this reduces to

(A23) $$z_1 + p_1 \frac{\partial z_1}{\partial p_1} = -p_2 \frac{\partial z_2}{\partial p_1}$$

Dividing both sides by z_1 yields

(A24) $$1 + \rho_1 = -p_2 \frac{\partial z_2}{\partial p_1}$$

where ρ_1 is the own-price elasticity of z_1. Thus, $\partial z_2 / \partial p_1$ takes the sign of $-(1+\rho_1)$.♠ From (24) in the text we know that $dv/d\sigma$ takes the sign of $1+\varepsilon$, where ε is the own-price elasticity of *G*. The result then follows directly from Lemma 1.♣

REFERENCES

Alonso, W. (1964), *Location and Land Use*. Cambridge: Harvard University Press.

Anas, A. and Xu, R. (1999), Congestion, Land Use and Job Dispersion, *Journal of Urban Economics*, 45, 451–472.

Arnott, R. and Small, K. (1994), The Economics of Traffic Congestion, *American Scientist*, 82(5), 446-455.

Barrett, S. (1994), Self-enforcing International Environmental Agreements, *Oxford Economic Papers*, 46, 878–894.

Baum-Snow, N. (2007a), Suburbanization and Transportation in the Monocentric Model, *Journal of Urban Economics*, 62, 405–423.

Baum-Snow, N. (2007a), Did Highways Cause Suburbanization?, *The Quarterly Journal of Economics*, 122(2), 775-805.

DeBartolome, C. and Ross, S. (2003), Equilibria with Local Governments and Commuting: Income Sorting vs. Income Mixing, *Journal of Urban Economics*, 54(1), 1-20.

Duranton, G. and Turner, M. (2011), The Fundamental Law of Road Congestion: Evidence From US Cities, *American Economic Review, 101(6), 2616-2652.*

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

Glaeser, E.L. and Khan, M.E. (2004), Sprawl and Urban Growth, in Henderson, J. and Thisse, J-F. (eds.), *Handbook of Urban and Regional Economics, Volume 4*. Elsevier: Amsterdam.

Mills, E.S. (1967), An Aggregative Model of Resource Allocation in a Metropolitan Area, *American Economic Review*, May, 57, 197-210.

Mogridge, M. (1997), The Self-Defeating Nature of Urban Road Capacity Policy, *Transport Policy*, 4(1), 5-23.

Muth, R. (1969), *Cities and Housing*. Chicago: University of Chicago Press

Nechyba, T.J. and Walsh, R.P. (2004), Urban Sprawl, *The Journal of Economic Perspectives*, 18(4), 177-200.

Taylor, C. (1972), Parasitic Suburbs, *Society*, September/October, 35-41.

Wu, J. (2006), Environmental amenities, urban sprawl, and community characteristics, *Journal of Environmental Economics and Management*, 52, 527–547.

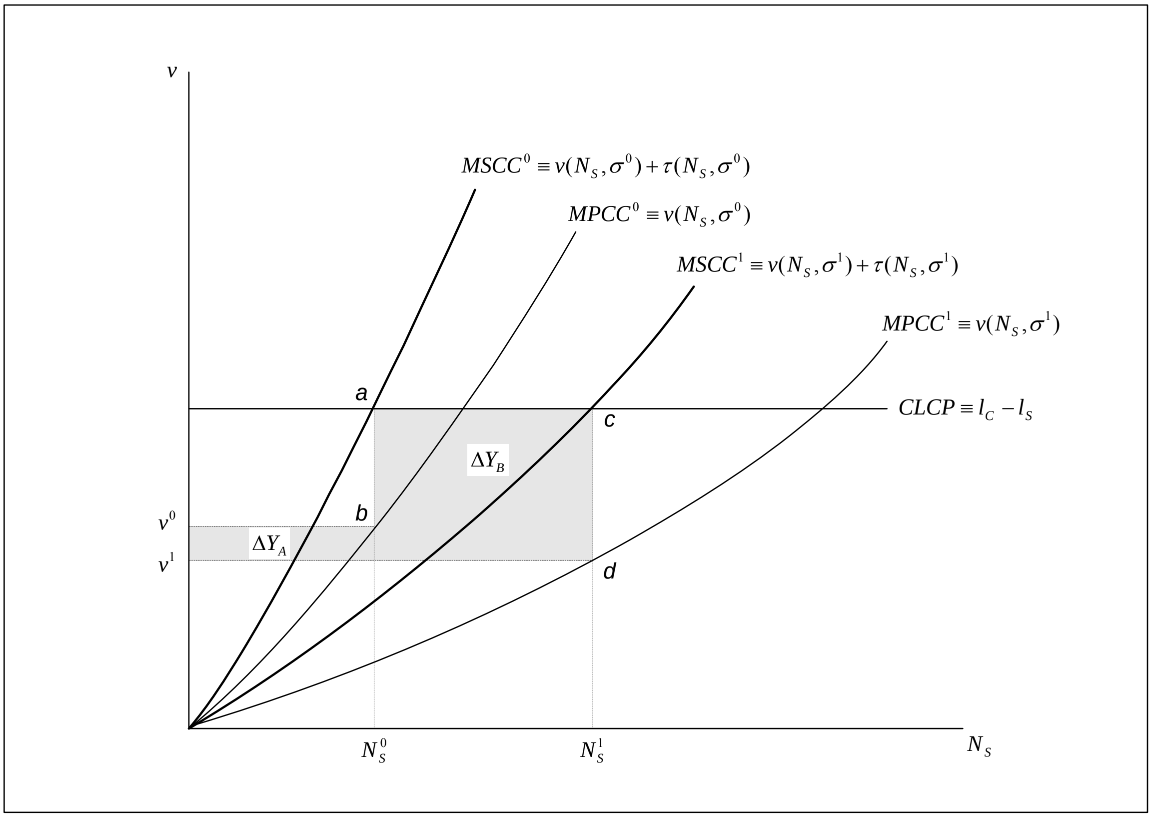
v
$MSCC^0 \equiv v(N_S, \sigma^0) + \tau(N_S, \sigma^0)$
$MPCC^0 \equiv v(N_S, \sigma^0)$
$MSCC^1 \equiv v(N_S, \sigma^1) + \tau(N_S, \sigma^1)$
$MPCC^1 \equiv v(N_S, \sigma^1)$
a
c
$CLCP \equiv l_C - l_S$
ΔY_B
b
v^0
ΔY_A
v^1
d
N_S^0
N_S^1
N_S

FIGURE 1

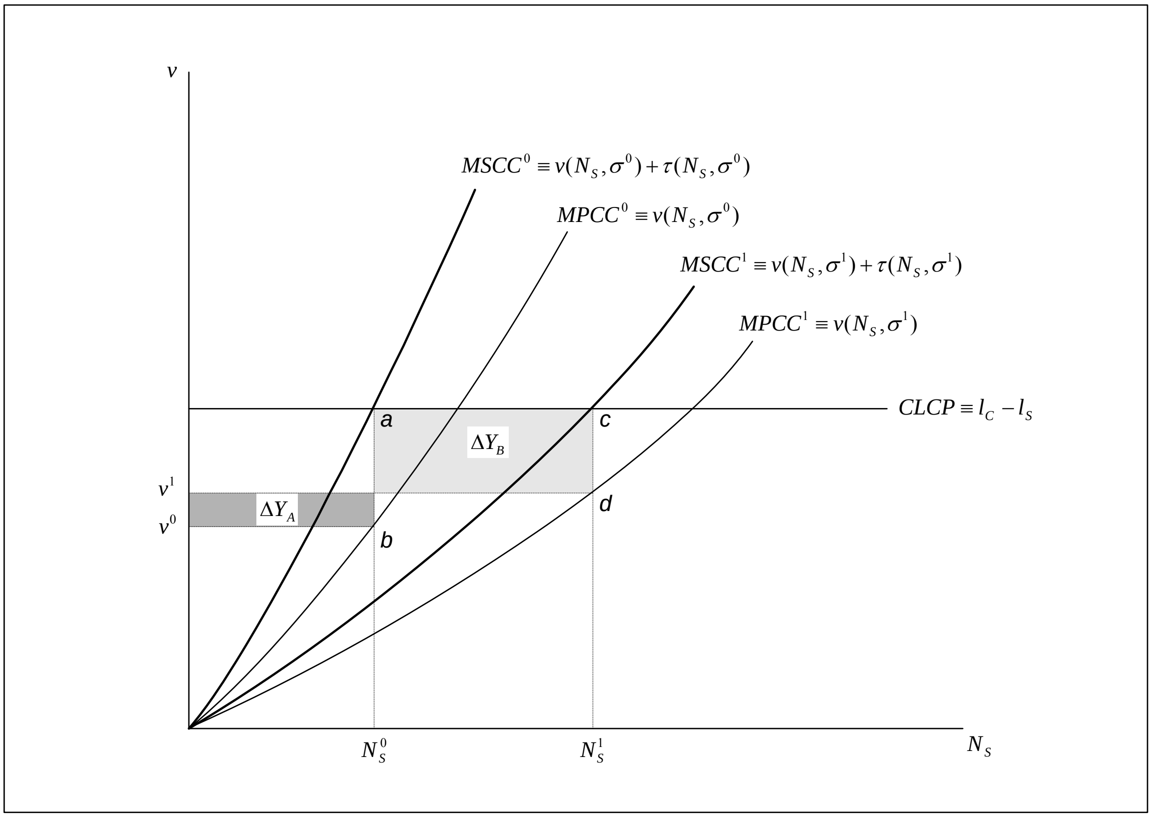
v
$MSCC^0 \equiv v(N_S,\sigma^0)+\tau(N_S,\sigma^0)$
$MPCC^0 \equiv v(N_S,\sigma^0)$
$MSCC^1 \equiv v(N_S,\sigma^1)+\tau(N_S,\sigma^1)$
$MPCC^1 \equiv v(N_S,\sigma^1)$
$CLCP \equiv l_C - l_S$
a
c
ΔY_B
v^1
ΔY_A
d
v^0
b
N_S^0
N_S^1
N_S

FIGURE 2

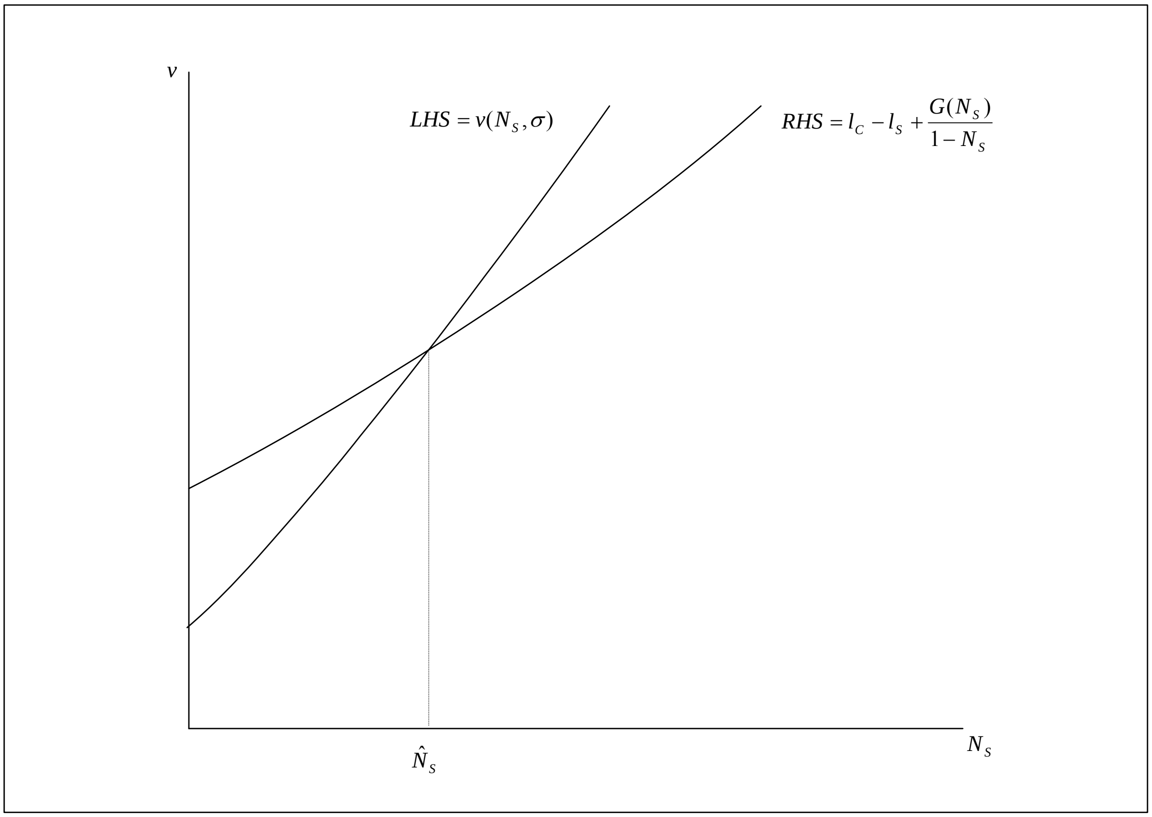

FIGURE 3

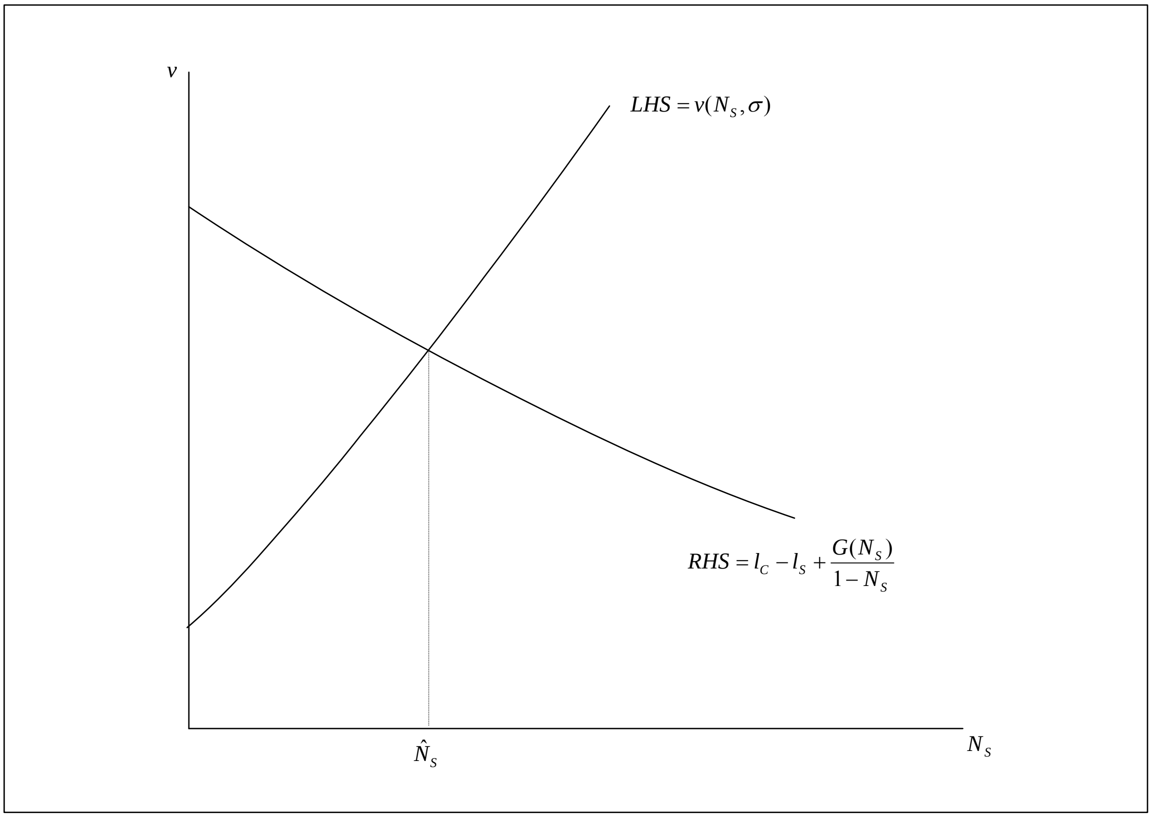

FIGURE 4

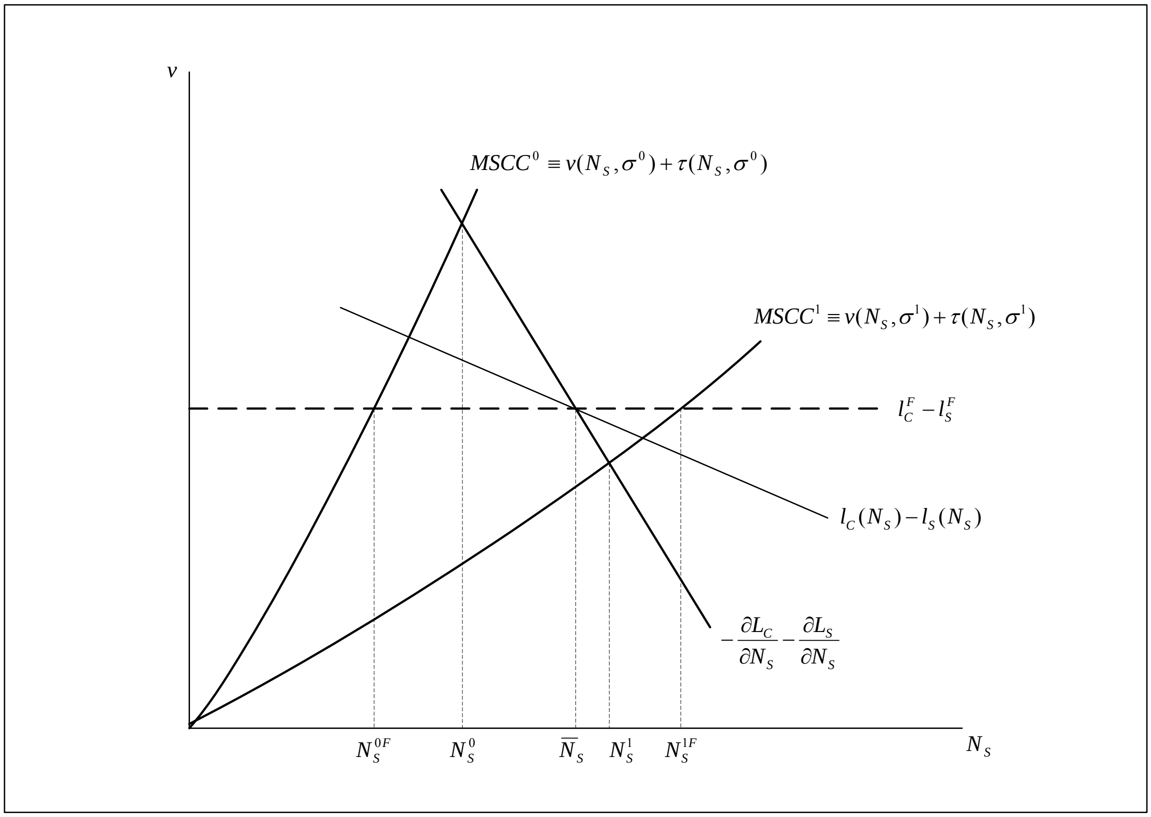
v
$MSCC^0 \equiv v(N_S, \sigma^0) + \tau(N_S, \sigma^0)$
$MSCC^1 \equiv v(N_S, \sigma^1) + \tau(N_S, \sigma^1)$
$l_C^F - l_S^F$
$l_C(N_S) - l_S(N_S)$
$-\frac{\partial L_C}{\partial N_S} - \frac{\partial L_S}{\partial N_S}$
N_S^{0F}
N_S^0
$\overline{N}_S$
N_S^1
N_S^{1F}
N_S

FIGURE 5