MODELING CITIES WITH CONGESTION AND AGGLOMERATION EXTERNALITIES: THE EFFICIENCY OF CONGESTION TOLLS, LABOR SUBSIDIES, AND URBAN GROWTH BOUNDARIES

Wenjia Zhang  
Community and Regional Planning Program  
School of Architecture  
The University of Texas at Austin  
wenjiazhang@utexas.edu  

Kara M. Kockelman  
(corresponding author)  
E.P. Schoch Professor in Engineering  
Department of Civil, Architectural and Environmental Engineering  
The University of Texas at Austin  
6.9 E. Cockrell Jr. Hall  
Austin, TX 78712-1076  
kkockelm@mail.utexas.edu  
Phone: 512-471-0210

Under review for publication in the *Journal of Urban Economics*, September 2014

**ABSTRACT:** This paper develops a spatial general equilibrium model that accommodates both congestion and agglomeration externalities, while firms’ and households’ land-use decisions are endogenous across continuous space. Focusing on the interaction between externalities and land use patterns, we examine the efficiencies of first-best and second-best policy instruments using simulations. Pigouvian congestion tolling (PCT) alone can largely mitigate congestion but may erode agglomeration economies, while Pigouvian labor subsidies (PLS) alone correct labor market distortions, but raise congestion diseconomies. A first-best policy must combine both PCT and PLS instruments, or design an optimal toll (or subsidy) agglomeration externalities (or congestion externalities). The first-best and efficient second-best tolls or subsidies should lie below their Pigouvian levels. Urban growth boundaries (UGBs) can reduce congestion externalities and enhance agglomeration economies as first-best instruments do, while introducing land market distortions via rent-escalation and firm-centralization effects, even inducing a net welfare loss. These findings suggest that it is important to internalize firms’ land use decisions, and avoid monocentricity assumptions, in order to appreciate the interplay of both urban externalities, since spatial adaptations to policy interventions can distort system efficiencies.

**Key Words:** Urban Externalities, Congestion, Agglomeration, Pigouvian Congestion Tolling, Urban Growth Boundaries

1. **INTRODUCTION**

Cities are full of externalities. The external costs of traffic congestion and the external benefits of firm agglomeration are widely discussed in urban economics literature. Congestion, for example, delays other travelers, adds air pollution and greenhouse gases, and raises a community’s energy demands. Across the U.S.’s early 500 urban areas in 2011, congestion is estimated to generate
5.5 billion hours of travel delay every year, using 2.9 billion gallons of added fuel, and adding 56 billion pounds of CO2, tallying to over $120 billion in losses, or roughly $400 per capita per year (TTI, 2012). Firm agglomeration economies can largely explain the geographical centralization of firms, as well as the emergence and evolution of cities. Firms benefit from locating close to each another, via access to intermediate inputs and labor, easier job-worker matching, knowledge spillovers, and other sources (Fujita and Thisse, 2002; Rosenthal and Strange, 2004; Puga, 2010). Such agglomeration externalities rise with the density of economic activities and proximity to other firms. As a result, doubling job density or doubling city size is often associated with a 4%-10% or 3%-8% increase, respectively, in productivity (Rosenthal and Strange, 2004; Combes et al., 2010).

While urban economists have long recognized either negative congestion externalities (e.g., Solow 1972; Arnott, 1979; Pines and Sadka, 1985; Wheaton, 1998; Anas and Xu, 1999; Brueckner, 2007) or positive agglomeration externalities (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Berliant et al., 2002; Rossi-Hansberg, 2004; Borck and Wrede, 2009), few have considered their interactions. Incorporating both externalities in an urban economic analysis is important, since urban policies for coping with one externality in one distorted market may neglect the spillover effects of this policy on the other distorted market. For example, a Pigouvian congestion toll (PCT) strategy charges marginal external costs to travelers who impose such costs, and is regarded as a first-best instrument for correcting distortions from negative congestion externalities. In isolation, this strategy is not first-best for cities, because tolls affect labor costs, land use patterns, and rents, and thereby affect agglomeration economies and firm productivity. By better understanding the interactions between congestion and agglomeration, one can avoid policy distortions informed by partial equilibrium analyses with only one externality, and thereby design more appropriate “first-best” policies while evaluating the benefits and limitations of second-best tolling, labor subsidies, and land use policies.

Few researchers have endogenized multiple urban externalities, and most rely on aspatial settings. For instance, Parry and Bento (1999) explored the interaction of distorted labor and transportation markets and evaluated the welfare effects of a congestion tax in the presence of a labor tax. They found that the congestion tax could reduce labor supply if total toll revenues are equally redistributed to residents, and stimulate labor supply if revenues are used to subsidize labor, with the latter form of revenue recycling generating more welfare improvement. Arnott’s (2007) model internalized both negative congestion and positive agglomeration externalities, based on a straightforward assumption that lower travel costs increase non-market (e.g., social) interactions, enhancing agglomeration economies. He found that optimal congestion tolls should be lower than congestion’s (total) external costs when there is no policy in place to manage agglomeration externalities. These non-spatial studies identify the policy importance on incorporating multiple externalities; however, they cannot analyze the interaction between externalities and urban form, which may significantly affect the optimal design of urban policies (Verhoef and Nijkamp, 2004).

Externalities affect urban form, and urban form affects externalities. Some models rely on discrete spatial settings to track multiple externalities. For example, Anas and Kim (1996) presented a spatial computable general equilibrium (spatial CGE) model integrating congestion and agglomeration externalities for consumers in a linear city with discrete zones. Here,
consumers are assumed to make more shopping trips to larger shopping centers (i.e., those exhibiting retail-job agglomerations). Their simulation results suggest that congestion externalities disperse urban form, while shopping agglomeration favors more compact forms, with fewer and more job-rich centers. Anas (2012) also recently developed a valuable core-periphery model to explore social optima after first recognizing highway congestion’s external costs and transit’s external benefits, and then allowing for Marshallian agglomeration externalities. His comparative static analyses revealed that the optimal policy in a closed city with two or more externalities (or activities with economies of scale) should satisfy the general Henry George Theorem.

Other studies have internalized multiple spatial externalities by extending the traditional monocentric model. For example, Verhoef and Nijkamp (2004) modeled both agglomeration externalities (of firms) and pollution externalities (from commutes) under monocentric settings. They highlighted the importance of using a spatial equilibrium framework to understand urban externalities, since congestion pricing and labor subsidies are not perfect (opposite) substitutes in the presence of spatial interactions. Their simulations show how second-best tolls or subsidies are lower than the Pigouvian levels. Wheaton (2004) combined a congestion externality and center-agglomeration forces into a circular monocentric framework, suggesting that worse congestion is associated with more centralized firm agglomeration. However, such monocentric models often do not internalize land inputs/rents in any production function; they rely on simplified, aspatial measures of agglomeration and thus overlook interactions between agglomeration externalities and urban form.

This paper first develops and then applies a spatial general equilibrium model with endogenously determined congestion and agglomeration externalities in a continuous, non-monocentric city space. The agglomeration externality is a Marshallian production externality, and defined to be proportional to each site’s local jobs density and an integral of inverse-exponential distance-weighted job counts within a pre-existing cluster around the region’s center point. This assumption pivots off those in Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). Fujita and Ogawa (1982) were among the first to explore the economics of non-monocentric urban economies with production externalities, using a linear city form. Production externalities, or location potential (as defined in their paper), is reflected in firm productivity, which varies over space, thanks to clustering of economic activities. Lucas and Rossi-Hansberg (LRH, 2002) extend the Fujita-Ogawa model to a continuous, circular city setting. Rossi-Hansberg (2004) then applied the LRH model to evaluate labor subsidies and zoning restrictions, but without congestion externalities. Thus, our model is among the first to incorporate Fujita-Ogawa- and LRH-type agglomeration economies and congestion externalities in a continuous urban space, enabling more comprehensive policy assessments.

More importantly, relying on numerical simulations, we examine the efficiency of first-best interventions (using both PCT and PLS policies) and second-best instruments, like simply PCT, simply PLS, or an urban growth boundary (UGB). For the first-best interventions, we examine land use patterns in the social optimum and challenges to designing first-best instruments, since these topics are seldom discussed. For the PCT- or PLS-only settings, we emphasize welfare outcomes and the possible side effects associated with these second-best policies. Answers to these two questions are ambiguous in past studies, and have depended on the economic model used (e.g., spatial versus aspatial, and partial versus complete equilibrium) (Parry and Bento,
We also examine the welfare effects and side effects of optimal and imperfect UGBs. Some studies suggest that imposition of an UGB may be an effective second-best policy (since a UGB increases densities and reduces travel distances, much like optimal pricing will do) (Pines and Sadka, 1985), while others argue that UGBs have a lower, or even negative, welfare impact than PCT strategies (Anas and Rhee, 2006; Brueckner, 2007; Kono et al., 2012). Another debate concerning UGB regulation is whether such boundaries facilitate central-city revitalization via rising productivity and attraction for new development activities (Nelson et al., 2004). Such questions relate closely to planning practices and planning debates, and so merit exploration here.

The paper is organized as follows: Section 2 describes the new model’s assumptions, equilibrium conditions, and general equilibrium outcomes; Section 3 compares simulation results for welfare, externalities, and land use, and evaluates first-best and second-best policy scenarios; Section 4 offers conclusions and suggestions for future work.

2. THE MODEL

The model developed here assumes a continuous symmetric circular region of radius $\bar{x}$. The symmetry assumption implies that workers travel only towards or away from the center, along radial street networks. Two homogeneous agent types, households and firms, exist and can reside at the same location inside the region. For any location $x (0 \leq x \leq \bar{x})$, $\theta_f(x)$, $\theta_h(x)$ and $\theta_t$ represent the fractions of land area used by firms, households, and transportation infrastructure. $\theta_f(x)$ and $\theta_h(x)$ are endogenously determined, while $\theta_t$ is exogenously given.

2.1 Household and Congestion Externality

Each household living in location $x$ and working at location $x_w (0 \leq x_w \leq \bar{x})$ consumes a quantity of goods $c(x, x_w)$ (with price $p = 1$) and enjoys a residential lot size $q(x, x_w)$, resulting in utility level $u(c(x, x_w), q(x, x_w))$. Its willingness to pay for land is rental rate $r_h(x)$, in dollars per acre, for example. Each household has one worker, earning net income $y(x, x_w)$. This net income is comprised of three components: wage income paid by firms at location $x_w$, or $w(x_w)$, minus commuting costs $T(x, x_w)$, plus the return of aggregate rent and toll revenues, $\bar{y}$. Thus, the optimization problem of each household is as follows:

**Problem 1.** For each household living at location $x (0 \leq x \leq \bar{x})$, choose a job location $x_w (0 \leq x_w \leq \bar{x})$ and evaluate functions $c(x, x_w)$ and $q(x, x_w)$, so as to maximize utility $u(c(x, x_w), q(x, x_w))$

subject to the budget constraint:

1. $c(x, x_w) + r_h(x)q(x, x_w) \leq y(x, x_w) = w(x_w) + \bar{y} - T(x, x_w)$

where

2. $\bar{y} = \frac{1}{N} (y_{rent} + y_{toll} - y_{suby})$

3. $T(x, x_w) = \int_x^{x_w} (t(s) + \tau(s)) ds$
Eq. (2) guarantees that aggregate revenues from land rents \( y_{rent} \) and tolls \( y_{toll} \), net of the labor subsidy \( y_{suby} \), are uniformly distributed to households, consistent with a closed-form city of (given) population \( N \). This setting allows one to more equitably compare the welfare effects of different policy scenarios.

Eq. (3) shows that \( T(x, x_w) \) is an accumulation of marginal travel costs, from \( x \) to \( x_w \). Here, \( t(x) \) represents the average travel cost per mile at location \( x \), with a negative sign representing inward travel and a positive sign representing outward travel. \( \tau(x) \) represents a potential congestion toll on drivers passing location \( x \). Consistent with prior works (e.g., Wheaton, 1998, 2004; and Brueckner, 2007), \( t(x) \) is proportional to a power function of the traffic volume crossing the ring at \( x \), \( D(x) \), relative to the road supply or width at \( x \) – plus the free-flow travel-cost component, \( \varphi \) (in dollars per mile). Thus,

\[
\begin{align*}
\tau(x) = \begin{cases} 
- \varphi - \rho \left( -\frac{D(x)}{2\pi x \theta_c} \right)^\sigma & \text{if } D(x) < 0 \\
\varphi + \rho \left( \frac{D(x)}{2\pi x \theta_c} \right)^\sigma & \text{if } D(x) > 0 \\
\varphi & \text{or } D(x) = 0 
\end{cases}
\end{align*}
\]

where \( \rho \) and \( \sigma (\sigma \geq 1) \) are positive parameters designed to reflect network congestibility (very much like the standard Bureau of Public Roads [BPR 1964] formulation for travel times). As with travel costs, traffic volumes, \( D(x) \), are negative when flow is inward at location \( x \), and positive when flows are outward. When \( D(x) = 0 \), no traffic crosses location \( x \), and the marginal travel cost equals the free-flow cost (which can be either positive or positive).

According to Eq. (4), one can compute the marginal congestion externality at each \( x \), \( \tau_{mce}(x) \):

\[
\tau_{mce}(x) = \frac{\partial \tau(D(x))}{\partial D(x)} D(x) = \begin{cases} 
\rho \sigma \left( \frac{|D(x)|}{2\pi x \theta_c} \right)^\sigma, & \text{if } D(x) \geq 0 \\
- \rho \sigma \left( \frac{|D(x)|}{2\pi x \theta_c} \right)^\sigma, & \text{if } D(x) < 0 
\end{cases}
\]

Here, the derivative of \( t(x) \) of \( D(x) \) represents the added marginal travel cost on each individual driver across \( x \) when one new driver is added, while \( \tau_{mce}(x) \) represents all additional travel costs (imposed on other drivers), as caused by the added driver.

**Proposition 1:** Suppose \( c^*(x, x_w) \)and \( q^*(x, x_w) \) are the solutions to Problem 1 and \( \bar{u} \) is the maximized utility level; then, the following are true:

(a) For those households living in location \( x \), regardless of where they work, they earn an identical net income, \( y(x) \), so that: \( y(x, x_w) \equiv y(x), \forall x_w > 0 \); and they consume the same amount of goods and lot size, \( c^*(x) \) and \( q^*(x) \), so that: \( c^*(x, x_w) \equiv c^*(x) \) and \( q^*(x, x_w) \equiv q^*(x), \forall x_w > 0 \).

(b) \( q^*(x) = q^*(y(x), \bar{u}) \) and \( c^*(x) = c^*(y(x), \bar{u}) \) satisfy the equations \( c(x) + q(x) u_q / u_c = y(x) \) and \( u(c(x), q(x)) = \bar{u} \);

(c) \( y(x) = w(x) + \bar{y} \); and

(d) \( y'(x) = w'(x) = t(x) + \tau(x) \).

**Proof:** See A1 in the Appendix.
From Proposition 1a, household attributes at location $x$, including $c(x, x_w)$, $q(x, x_w)$, and $y(x, x_w)$, can be written simply as $c(x), q(x)$, and $y(x)$ in the rest of this article. From Proposition 1b, if one assumes a Cobb-Douglas utility function, as follows:

$$u(c(x), q(x)) = c(x)\alpha q(x)^{1-\alpha}, \quad 0 < \alpha < 1$$

then, the solutions to Problem 1 are:

$$q^*(x) = \alpha^{-\alpha/(1-\alpha)}y(x)^{-\alpha/(1-\alpha)}\tilde{u}^{1/(1-\alpha)}$$

$$c^*(x) = \alpha y(x)$$

and maximized bid-rents from households are:

$$r^*_m(x) = (1 - \alpha)\alpha^{\alpha/(1-\alpha)}\left(\frac{y(x)}{\tilde{u}}\right)^{1/(1-\alpha)}$$

Equations (7) to (9) show that optimal lot size and good consumption and maximum bid-rent at location $x$ are determined by household’s net income, $y(x)$, which relates to wages earned and commuting costs, as shown in Eq. (1). Proposition 1c demonstrates that the net income of households residing at $x$ equals the wage income paid by firms at $x$ plus redistributed revenues. From Proposition 1d, the condition that both the wage gradient and the net-income gradient equal the marginal travel cost should be satisfied when maximizing utilities. This condition supports the intuition that no worker can achieve a higher net income (net of commute costs, plus labor subsidies or toll revenue redistributions) by changing his or her job location.

### 2.2 Firms and Agglomeration Externalities

Each firm is a price taker in input and output markets. If a competitive firm located at $x$ operates under constant returns to scale, its total production $P(x)$ depends on the amounts of labor $L(x)$ and land area $H(x)$ used, and its total factor productivity (TFP) $A(x)$, such that:

$$P(x) = A(x)L(x)^\kappa H(x)^{1-\kappa} \quad (0 < \kappa < 1)$$

The production per unit of land, $p(x)$, is therefore as follows:

$$p(x) = \frac{P(x)}{H(x)} = A(x)n(x)^\kappa$$

where $n(x)$ is labor density along ring $x$ and $\kappa$ is the production function’s elasticity parameter.

Consistent with many theoretical and empirical studies (e.g., Verhoef and Nijkamp, 2004; and Combes et al., 2010), we internalize agglomeration economies in the TFP, by assuming that the agglomeration externality $F(x)$ at location $x$ determines the productivity:

$$A(x) = \delta F(x)^\gamma \quad (\delta > 0, 0 < \gamma < 1)$$

Here, $\delta$ is the productivity scale parameter and $\gamma$ is the elasticity of productivity with respect to agglomeration externalities at location $x$. Fujita and Ogawa (1982) provided a measure of agglomeration economies for firms based on location potential in a linear city setting: they used job densities and distances to other firms or workers. Lucas and Rossi-Hansberg (2002) extended this measurement to circular space. Similar to LRH’s setting, agglomeration

---

1 The only difference is that LRH’s model considers production externalities from all firms in the entire city (inversely weighted by distance), and assumes a fixed city boundary. Our model assumes production externalities
externalities are defined here to be proportional to the local employment density (at location \(x\)) and the integral of an inverse-exponential distance-weighted job count within a pre-existing cluster around the region’s center point, up to an (exogenously set) boundary distance of \(\bar{r}\). This implies that a firm’s productivity is affected by the presence of firms within a pre-existing urban core, \([0, \bar{r}]\), and any external effect of firms outside the urban core is negligible. This assumption is typically used in the monocentric model or the core-periphery model (e.g., Glaeser and Kahn, 2001; Wheaton, 2004; Anas, 2012), in which only firms in an abstract center exhibit agglomeration effects. Thus, the agglomeration externality at each location along the annulus at radius \(x\) is specified as

\[
F(x) = \zeta \int_0^\bar{r} \int_0^{2\pi} r \theta_f(r) n(r) e^{-\zeta l(x,r,\psi)} d\psi dr
\]

where \(\zeta\) is the production externality scale parameter, and is exogenously determined. \(\psi\) is the polar angle around the center (ranging from 0 to \(2\pi\)), and \(l(x,r,\psi)\) is the straight-line distance between a firm at a specific location along annulus \(x\) and each firm lying within \(\bar{r}\) miles of the center (at a counter-clockwise angle of \(\psi\) from the first firm). Thus,

\[
l(x, r, \psi) = \sqrt{x^2 + r^2 - 2 rx \cos(\psi)}
\]

The firms then maximize the profit function with respect to employment density \(n(x)\), with firm output price set at 1 (without loss of generality):

\[
\text{Max } \pi(n(x)) = \delta n(x)^{\kappa} F(x)^{y} - n(x) \left( w(x) - s(x) \right) - r_f(x)
\]

where \(s(x)\) represents a potential labor subsidy for firms at location \(x\) to hire each worker.

From the first-order condition of profit maximization with respect to \(n(x)\), one can obtain optimal employment density at location \(x\) as follows:

\[
n^*(x) = \left( \frac{\kappa \delta F(x)^y}{w(x) - s(x)} \right)^{1/(1-\kappa)}
\]

Given perfectly competitive input and output markets, all firms make zero (excess) profit, with land rents rising to their maximum values to ensure this, as follows:

\[
r^m_f(x) = (1 - \kappa) \delta^{1/(1-\kappa)} F(x)^{y/(1-\kappa)} \left( \frac{\kappa}{w(x) - s(x)} \right)^{\kappa/(1-\kappa)}
\]

### 2.3 The Land Market’s Equilibrium Conditions

Since both firms and households can exist in the same location, a competitive market requires they bid for the land via their willingness to pay (or maximum bid rents). Given the maximized bid-rents from the partial equilibrium of households and firms at each location \(x\) (as shown in Eqs. (9) and (17)), the land market equilibrium requires that land rents, \(r(x)\), satisfy the following two equations:

\[
r(x) = \max\{r^m_h(x), r^m_f(x), R_a\}
\]

\[
r(\bar{x}) = R_a
\]

come only from firms within a pre-set area, and the city’s boundary/limit is endogenously determined.
Eq. (19) defines the *edge* land rent \( r(\bar{x}) \), which equals the agricultural land rent (or opportunity rent) \( R_a \). If both \( r_f^m(x) \) and \( r_h^m(x) \) are less than \( R_a \), the equilibrium land use share for firms \( \theta_f(x) \) and the equilibrium land use share for household \( \theta_h(x) \) will equal zero. If \( r_f^m(x) \) equals \( r_h^m(x) \), a mixed land use pattern will emerge at location \( x \), and the equilibrium number of jobs at that location will equal the number of households (or residing workers) at that location (LRH, 2002). Given that both \( r_f^m(x) \) and \( r_h^m(x) \) will exceed \( R_a \) (except at the developed region’s edge), \( \theta_f(x) \) and \( \theta_h(x) \) at each location \( x \) are as follows:

\[
\theta_f^*(x) = \begin{cases} 
1 - \theta_t & \text{if } r_f^m(x) > r_h^m(x) \\
\frac{n^*(x)q^*(x)}{n^*(x)q^*(x)+q^*(x)}(1 - \theta_t) & \text{if } r_f^m(x) = r_h^m(x) \\
0 & \text{if } r_f^m(x) < r_h^m(x)
\end{cases}
\]

\[
\theta_h^*(x) = 1 - \theta_t - \theta_f^*(x)
\]

Eq. (21) represents the land market clearing so that all available land or properties are assigned to either firms/jobs, households, or transport infrastructure. Moreover, total city/region land rents (net of the base rent, \( R_a \)), \( y_{rent} \), in a spatial equilibrium will satisfy the following equation:

\[
y_{rent} = \int_0^\infty 2\pi x\{\theta_f^*(x)(r_f^m(x) - R_a) + \theta_h^*(x)(r_h^m(x) - R_a)\}dx
\]

### 2.4 The Labor Market’s Equilibrium Conditions

Under equilibrium, the labor market meets two conditions. First, the commute demand generated in the interval \( dx \) from \( x \) to \( x+dx \) (or absorbed in \( dx \) from \( x+dx \) to \( x \) ), \( D'(x)dx \) (or \(-D'(x)dx\)), will equal the number of workers who need to work outside the interval (or the job vacancies in \( dx \)). Thus,

\[
D'(x) \leq 2\pi x\left(\frac{\theta_h^*(x)}{q^*(x)} - \theta_f^*(x)n^*(x)\right)
\]

A spatial equilibrium requires that travel demand at the city edge, \( D(\bar{x}) \), and in the city center point, \( D(0) \), equals zero (since there are no jobs or workers beyond this boundary, to attract or generate such trips). Thus, the two boundary conditions for commute demand are:

\[
D(0) = 0 \text{ and } D(\bar{x}) = 0
\]

These two boundary constraints also guarantee the second condition for labor market clearing: the total number of workers will equal the number of households, \( N \):

\[
\int_0^\infty 2\pi x\frac{\theta_h^*(x)}{q^*(x)}dx = \int_0^\infty 2\pi x\theta_f^*(x)n^*(x)dx = N
\]

### 2.5 Spatial General Equilibrium

One can combine households’ and firms’ partial equilibria with equilibrium conditions for labor and land markets, thereby creating a spatial general equilibrium model for the region. Given \( \bar{u} \) and other parameters, this model has 20 unknowns, including 15 functions of \( x \):
\( c^*(x), q^*(x), r^m_h(x), y(x), t(x), \tau(x), D(x), w(x), n^*(x), r^m_f(x), s(x), F(x), r(x), \theta^*_h(x), \theta^*_f(x) \), and 5 scalars: \( \bar{x}, \bar{y}, y_{\text{rent}}, y_{\text{toll}}, y_{\text{suby}} \). The 20 equations needed to resolve this model include 16 equations described above (Eqs. (2) and (4), Proposition (c) and (d), Eqs. (7)-(9), (13), and (16)-(23)) plus 4 other equations that define the tolling instrument, \( \tau(x) \) and \( y_{\text{toll}} \), and the subsidy, \( s(x) \) and \( y_{\text{suby}} \), which vary across policy scenarios. Table 1 summarizes these four functions, \( \tau(x) \), \( s(x) \), \( y_{\text{toll}} \), and \( y_{\text{suby}} \), across five spatial equilibria. In the free-market equilibrium, neither a toll nor a subsidy is imposed, so \( \tau(x) = 0, s(x) = 0, y_{\text{toll}} = 0, \) and \( y_{\text{suby}} = 0 \).

**Table 1. Policy instrument values \( \{\tau(x), s(x), y_{\text{toll}}, y_{\text{suby}}\} \) for urban equilibria under five policy interventions**

<table>
<thead>
<tr>
<th>Policy Interventions</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-Market</td>
<td>( \tau(x) = 0; s(x) = 0; y_{\text{toll}} = 0; y_{\text{suby}} = 0 )</td>
</tr>
</tbody>
</table>
| First-Best            | \( \tau(x) = \pm \rho \sigma \left( \frac{|D(x)|}{2\pi \sigma} \right)^\sigma; \)  
\( s(x) = \int_0^\bar{x} \int_0^{2\pi} r \theta^*_f(v) f(n^*(v)) g \left( F' (v) \right) e^{-\zeta_l(x,r,\psi)} d\psi dv; \)  
\( y_{\text{toll}} = \int_0^\bar{x} \tau(x) D(x) dx; \)  
\( y_{\text{suby}} = \int_0^\bar{x} 2\pi x \theta^*_f(x)n^*(x)s(x)dx \) |
| PCT-Alone             | \( \tau(x) = \pm \rho \sigma \left( \frac{|D(x)|}{2\pi \sigma} \right)^\sigma; s(x) = 0; y_{\text{toll}} = \int_0^\bar{x} \tau(x) D(x) dx; y_{\text{suby}} = 0 \) |
| PLS-Alone             | \( \tau(x) = 0; s(x) = \int_0^\bar{x} \int_0^{2\pi} r \theta^*_f(v) f(n^*(v)) g \left( F' (v) \right) e^{-\zeta_l(x,r,\psi)} d\psi dv; \)  
\( y_{\text{toll}} = 0; y_{\text{suby}} = \int_0^\bar{x} 2\pi x \theta^*_f(x)n^*(x)s(x)dx \) |
| UGB                   | \( \tau(x) = 0; s(x) = 0; y_{\text{toll}} = 0; y_{\text{suby}} = 0; \)  
\( \bar{x} = \bar{x}_{\text{ugb}} \) is given as \( R_a \) is endogenous. |

**Proposition 2:** First-best instruments to correct congestion and agglomeration externalities, as defined in Equations (5) and (13), satisfy various conditions, as follows:

(a) A first-best combination of the *Pigouvian Congestion Toll* \( \tau_{\text{pct}}(x) \) at each location \( x \) and the *Pigouvian Labor Subsidy* \( s_{\text{pls}}(x) \) on every unit of labor supplied at each firm location \( x \) can be defined as follows:

\[
\tau_{\text{pct}}(x) = \begin{cases} 
\rho \sigma \left( \frac{|D(x)|}{2\pi \sigma} \right)^\sigma, & \text{if } D(x) \geq 0 \\
-\sigma \left( \frac{|D(x)|}{2\pi \sigma} \right)^\sigma, & \text{if } D(x) < 0 
\end{cases}
\]

(26)

\[
s_{\text{pls}}(x) = \begin{cases} 
\int_0^\bar{x} \int_0^{2\pi} r \theta^*_f(v) f(n^*(v)) g \left( F' (v) \right) e^{-\zeta_l(x,r,\psi)} d\psi dv, & \text{if } \theta^*_f(x) > 0 \\
0, & \text{if } \theta^*_f(x) = 0 
\end{cases}
\]

(27)

(b) *First-best road tolling* for each mile driven at each location \( x \), \( \tau_{\text{fb}}(x) \), is as follows:

\[
\tau_{\text{fb}}(x) = \tau_{\text{pct}}(x) - \frac{\partial s_{\text{pls}}(x)}{\partial x}
\]

(28)

and the revenue generated by optimal tolls equals the aggregate congestion externality costs minus the aggregate agglomeration externality benefits.
(c) *First-best labor subsidy* on every worker who lives at $x_l$ and works at $x$, $s_{fb}(x_l, x)$ will be as follows:

\[
s_{fb}(x_l, x) = s_{pls}(x_l, x) - f^x_{x_l} \tau_{pct}(r) \, dr
\]

and the aggregate optimal subsidy equals the aggregate agglomeration externality benefits minus the aggregate congestion externality costs.

**Proof.** See A2 in the Appendix.

Given the simultaneous existence of two externalities in the model, a free-market equilibrium is inefficient; thoughtful policy intervention is needed to cope with market inefficiency. As noted earlier, four types of intervention are considered here: the simultaneous application of two first-best instruments, application of just PCT, application of just PLS, and application of an UGB. As noted in Proposition 2, the social optimum can be achieved via three types of first-best instrument. The city can simultaneously impose PCT and PLS, both of which equal corresponding marginal externalities, as shown in Eqs. (26) and (27). In addition, the city can impose first-best tolls by internalizing external benefits of agglomeration into PCT levels (Eq. 28). As one would expect, model solutions find that the aggregate optimal toll should lie below the aggregate congestion externality cost. This finding is consistent with Arnott’s (2007) result for a relatively straightforward, non-spatial model, where the optimal toll is lower than congestion externality cost and even negative, if the agglomeration externality cannot be subsided. Similarly, when congestion tolls are not feasible (e.g., they may not be politically acceptable), the city can supply first-best subsidies to firms, and the average optimal subsidy will then lie below the average agglomeration benefit.

Both the PCT-alone and PLS-alone policy instruments are second-best, rather than first-best, strategies, since they only endogenize one externality. Similarly, the UGB policy is a second-best land-use regulation without any pricing adjustments, where the fixed-land-rent assumption at the city edge is replaced by fixing a city boundary, $x_{ubb}$. Analytical equilibrium results are very difficult to derive here, for a 20-equation system with several non-linear equations and differential equations. In the following section, we rely on numerical results, to compare the properties of the free-market, first-best and second-best equilibrium settings, by setting function values for \{$\tau(x), s(x), y_{toll}, y_{suby}$\} according to Table 1.

### 3. NUMERICAL SIMULATIONS

#### 3.1 Parameter Values and Algorithm

Table 2 shows parameter values calibrated using U.S. data and the existing literature (e.g., LRH, 2002; Wheaton, 1998; Brueckner, 2007). For example, the Cobb-Douglas utility function’s parameter $\alpha$ reflects a household’s expenditure shares on goods and services, versus rents (relative to a household’s or worker’s net income). Here, we rely on LRH’s (2002) data-based value of 0.9, with $1-\alpha = 0.1$ as the share of net income going to housing rent. Each household’s utility level is set at 3500 utils in the free-market equilibrium, implying an annual expenditures of about $25,000 (per household) on the compound good’s consumption, after paying for housing and transportation; and average residential lot size is around 2000 square feet per household. Agricultural land rent, $R_a$, is set to $4,000,000 per square mile per year. This comes from the assumption that farmland at the edge of a city sells for about $50,000 per acre, with
amortization of such costs over 40 years at a discount rate of 5% resulting in rural land rents over $4,000,000 per square mile per year.

**Table 2 Parameter value assumptions**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.9</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>3,500</td>
</tr>
<tr>
<td>$R_a$</td>
<td>$4M/sq.mi$</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.95</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\delta$</td>
<td>30,000</td>
</tr>
<tr>
<td>$\xi$</td>
<td>5</td>
</tr>
<tr>
<td>$\theta_l$</td>
<td>0.3</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>2 miles</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>$200$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.00</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Key parameters for firm behaviors also rely on LRH’s assumptions, where $\kappa = 0.95$ and $\gamma = 0.04$, which is well in line with the empirical estimates ranging from 0.04 to 0.10 (Combes et al., 2010). Total factor productivity, $\delta$, is set at 30,000, by calibrating Eq. (16) under the assumption that per-capita money income is $30,000 (per year) and the city center holds over 100 persons per acre, on average. Following Wheaton’s (1998) study, roadways’ or transportation’s share of land is assumed to be 30%. The agglomeration-limit boundary, $\bar{r}$, is set to 2 miles. This is a somewhat arbitrary setting and has an impact on urban form, since it constrains job decentralization. Ceteris paribus, larger $\bar{r}$ values result in greater jobs and population, larger city limits/wider boundaries, and greater job decentralization. However, the general variations in the city’s land use and welfare under different policy interventions are less connected to $\bar{r}$’s pre-set value.

The intercept parameter $\varphi$ in Equation (16)’s average travel cost function represents an average cost of free-flow travel, and is set at $200 dollar per mile per year. Following Brueckner (2007), this figure is generated from the calculation that marginal free-flow travel cost is about $0.40 per mile when each worker works about 250 days a year. $\rho$ and $\sigma$ reflect link congestibility, and are set as 0.00001 and 1.5, respectively. In a highly congested location, for example, if there are 50,000 travelers passing a point $x = 1$ mile from the region’s center, the marginal congestion cost at $x = 1$ will be $0.17 per vehicle-mile, accounting for about 30% of total marginal costs. In a lightly congested location, say 5,000 travelers per day at a distance $x = 10$ miles away, the marginal congestion cost will account for only 0.4% of total marginal social costs at that point in the network.

In order to iteratively solve for location-specific values, the circular city can be divided into discrete, narrow rings, each of width $\Delta x = 0.01$ mile. Each location $x$ can then be labeled as $x_i = i\Delta x$ (with $i = 1,2,\ldots,l$), with $x_1$ representing the city center and $x_l$ representing the city’s boundary, $\bar{x}$. The spatial equilibria were solved using MATLAB, and by referring to LRH’s fixed-point algorithm (2002). There are two steps in a simulation of a free-market equilibrium. The first step computes the equilibrium land use and labor distribution levels at each location $x_i$, given an initial production externality function $F_0(x_i)$, an initial utility level $u_0$, and initial lump-sum redistributed revenue (including rent, toll and subsidy), $\bar{y}_0$. The second step calculates a new production externality function, $F^1(x_i)$, and redistributed revenue, $\bar{y}_1$, based on Equations (13) and (2). This setup will converge when $F(x_i)$ and $\bar{y}$ reach a fixed point, after several iterations. A third step is needed for those simulations of equilibria under PCT or PLS interventions, by increasing or decreasing the pre-assumed utility value and

---

2 The algorithm’s convergence is proven in LRH’s model when congestion effects are absent. Although the lemmas that convergence can be reached using the algorithm introduced in this paper are not rigorously proven to apply, this paper’s results are all convergent in practice.
repeating the first two steps until the derived population equals the simulated population in the free-market equilibrium.

3.2 Policy Scenarios
In this section, we examine the welfare and land use effects of four policy instruments, comparing to those in the free-market equilibrium. The policies include both the first-best and second-best instruments discussed above. These policy scenarios are compared in a closed-form city with a fixed population, about 1 million (the exact number is 1,054,230, which is derived as the population of the base equilibrium). Table 3 shows major characteristics of urban equilibria under different policy schemes. In the free-market equilibrium, the city’s boundary is 13.25 miles away from the city center when the edge land rent equals the agricultural land rent, $R_a$. The free-market equilibrium generates about $3.6 billion in annual congestion diseconomies, which is about four times the agglomeration economies ($0.89 billion in benefits). The average commute distance is about 6.49 miles per day, while the average worker’s commute cost is about $3,600 per year.

Table 3 Simulated results of policy scenarios

<table>
<thead>
<tr>
<th></th>
<th>Free Market</th>
<th>First Best</th>
<th>PCT-Alone</th>
<th>PLS-Alone</th>
<th>UGB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility level, $\bar{u}$</td>
<td>3,500</td>
<td>3,546</td>
<td>3,520</td>
<td>3,503</td>
<td>3,503</td>
</tr>
<tr>
<td>City boundary, $\bar{x}$ (miles)</td>
<td>13.25</td>
<td>12.32</td>
<td>12.41</td>
<td>13.18</td>
<td>12.26</td>
</tr>
<tr>
<td>Edge rent, $(\bar{x})$ (million $/sq.mi.)$</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5.1</td>
</tr>
<tr>
<td>Average tolls, $y_{toll}$ ($)/year</td>
<td>0</td>
<td>2466</td>
<td>2357</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Average subsidy, $y_{suby}$ ($)/year</td>
<td>0</td>
<td>854</td>
<td>0</td>
<td>939</td>
<td>0</td>
</tr>
<tr>
<td>Rent revenues returned, $y_{rent}$ ($)/year</td>
<td>2,191</td>
<td>2,497</td>
<td>2,437</td>
<td>2,221</td>
<td>2,459</td>
</tr>
<tr>
<td>Congestion Diseconomies ($) millions</td>
<td>3,655</td>
<td>2,600</td>
<td>2,485</td>
<td>3,891</td>
<td>3,544</td>
</tr>
<tr>
<td>Agglomeration Economies ($) millions</td>
<td>891</td>
<td>900</td>
<td>769</td>
<td>990</td>
<td>896</td>
</tr>
<tr>
<td>Average Commute Distance (miles/day)</td>
<td>6.49</td>
<td>5.02</td>
<td>5.11</td>
<td>6.46</td>
<td>5.97</td>
</tr>
<tr>
<td>Average Commute Cost ($)/year</td>
<td>3,609</td>
<td>2,647</td>
<td>2,592</td>
<td>3,752</td>
<td>3,436</td>
</tr>
<tr>
<td>Average Labor Wage ($)/year</td>
<td>24,907</td>
<td>25,641</td>
<td>24,519</td>
<td>26,007</td>
<td>24,910</td>
</tr>
<tr>
<td>Average TFP (compared to the constant)</td>
<td>1.51</td>
<td>1.53</td>
<td>1.50</td>
<td>1.54</td>
<td>1.52</td>
</tr>
<tr>
<td>Average Labor Density (workers/sq. mile)</td>
<td>47,785</td>
<td>66,954</td>
<td>53,872</td>
<td>56,365</td>
<td>49,261</td>
</tr>
<tr>
<td>Average Residential Density (hhs/sq. mile)</td>
<td>1,991</td>
<td>2,286</td>
<td>2,271</td>
<td>2,000</td>
<td>2,363</td>
</tr>
<tr>
<td>Average Rent for Firms (times $R_a$)</td>
<td>22.37</td>
<td>31.20</td>
<td>24.83</td>
<td>26.56</td>
<td>23.07</td>
</tr>
<tr>
<td>Average Rent for Housing(times $R_a$)</td>
<td>1.67</td>
<td>2.01</td>
<td>1.98</td>
<td>1.68</td>
<td>2.02</td>
</tr>
<tr>
<td>Avg. Net Income ($) /year</td>
<td>27,098</td>
<td>29,750</td>
<td>29,314</td>
<td>27,290</td>
<td>27,369</td>
</tr>
</tbody>
</table>

3.2.1 First-Best Instruments
According to Proposition 2, there are three first-best interventions – a combination of PCT and PLS, a first-best congestion toll (that varies by road location), and a first-best labor subsidy (that varies by firm or job location) – and these first-best instruments can each produce the same social optimum (according to Proposition 2). Here, we use the combination of PCT and PLS to simulate the optimum. Results show that under the social optimum, the city need to impose an
average toll of $2,466 (per year) per worker/commuter while delivering an annual average labor subsidy of $854 (per job) (Table 3).

This result does not imply that a combined, equivalent tax of $1,612 (i.e., $2,466-$854) on each worker will achieve the first-best optimum: spatial variations in tolls and labor subsidies need to be considered. Figure 1 shows the corresponding toll and/or subsidy levels across locations in the social optimum. Under this combination instrument, as job densities (or travel flows) increase, the amount of optimal labor subsidy (or optimal tolling) rises. Within the firm cluster area increases from 1.1 mile to 2.5 mile in radius, subsidies increase from $727 to $1,172 per year at the locations of peak labor density, and then fall to about $80 per year at the other edge of the firm cluster. Congestion tolls peak at the two ends of the firm cluster area, since these two places accumulate of the highest levels of outward and inward commute flows, generating the largest marginal negative externalities. These findings underscore the importance of enabling spatial variation in policy interventions, in order to optimally address urban externalities.

![Figure 1](image.png)

**Figure 1** Levels of toll and subsidy under the first-best instrument combining both PCT and PLS

Welfare improvement is also visible under the first-best instruments. The utility level increases to 3545.61, so it appears to be just 1.3% higher than that of the free-market equilibrium (Table 3). But utils are only ordinal in nature; the average worker’s willingness to pay to live in this optimally managed city, versus the free-market setting, is about 348 per year (as an equivalent variation). This welfare gain comes from lowered congestion diseconomies (which fall by 29%) and higher agglomeration economies (which rise by 1.1%). Such shifts are also evident in workers’ travel patterns; the average commute distance and cost fall by 23% and 27%, respectively. Wage levels also rise, in most locations (Figure 2a), with the average wage increasing 2.9% (Table 3). Similarly, the TFPs in most job locations significantly improve (Figure 2b), with average TFP rising 1.3%. These findings suggest that first-best instruments simultaneously reduce congestion and enhance agglomeration benefits.

Land use patterns are also affected, of course, as shown in Figure 2. The first-best instrument causes firms to decentralize, away from the city center, and agglomerate in a smaller cluster, as an annulus, with average labor density rising by 40% (in that ring, versus the original jobs zone).
This is a combined consequence of the imposition of PCT and PLS. First, the PLS encourages firms at locations of higher productivity to hire more workers, thereby reinforcing agglomeration externalities of their locations. Since labor supply is assumed fixed, firms at locations with lower productivity will lose labor and thus productivity. These shifts stimulate firms to locate closer to each other, clustering in a smaller area, raising job densities, wages, and total agglomeration economies (Figure 2c). Second, the PCT increases the per-mile commuting costs, thereby encouraging firms and workers to co-locate closer together, to reduce travel costs. While road tolls are paid by workers, firms need to provide an attractive wage that internalizes much of the toll to remain competitive. Firm decentralization (and some inward migration of households) can bring them closer to their workers while reducing inward traffic flows.

**Figure 2** Spatial distribution of wages ($w$), TFP ($A$), job ($n$) and residential densities ($1/q$), household bid-rents ($r$) and net income ($y$) in the first-best optimum (dashed) versus the free-market equilibrium (solid).

First-best instruments also centralize households, resulting in higher residential densities over most areas of the city, especially at locations closer to the firm cluster (Figure 2d). If comparing
Figure 2d and 2e, we find that higher residential densities raise household bid-rents. The average land rents for firms and houses in the socially optimal setting are 31.2 and 2.01 times the opportunity rent (i.e., the rent at the city edge, $R_o$), and 39% and 21% higher than those in the free-market equilibrium (Table 3). Given that all congestion tolls and rent revenues (net of labor subsidies) are equally/uniformly returned to each household, net incomes rise in all locations (Figure 2f), with average net income rising by 9.8% (Table 3). Notice that utility values rise with net income levels and fall with residential rents, everything else equation (as evident in Eq. 8). Even though housing’s rent growth doubles the net income growth, households still experience higher utility, since the elasticity of utility with respect to residential rent is much lower than that with respect to net income (0.1 versus 1).

Though first-best interventions presumably are the best choice for a city authority wishing to pursue welfare improvements, they may be associated major construction and operations costs (for variable toll collection, for example) that are generally not internalized in theoretical models. And a combination of PCT and PLS may require much coordination between transportation agencies and departments of labor (Verhoef and Nijkamp, 2005), which presents added transaction costs and political difficulties. A first-best tolling-only may reduce the need for coordination, but optimal toll levels for each location are difficult to set, especially in the presence of all sorts of non-work travel, in realistic networks. Figure 3a shows optimal toll levels across locations after internalizing agglomeration externalities. The first-best toll equals the PCT in the residential areas, but varies quite a bit within the annulus of jobs. Optimal tolls at some locations, where the spatial derivative of marginal agglomeration economies exceeds marginal congestion diseconomies, may lie below the PCT or even become negative (thereby incentivizing such travel). Optimal tolls could also lie above the PCT levels in some locations, where the spatial derivative of agglomeration economies $s_{pls}(x)$ is negative. This seems impossible from an aspatial or monocentric perspective, which generally presume that the marginal agglomeration externality is constant or decreasing at locations away from the CBD, so that the optimal tolls should never exceed the PCT levels. But it is in fact possible when the model is made more flexible. Such findings suggest that it is meaningful to internalize firms’ spatial behaviors in order to anticipate spatial externalities and related remedies.

Figure 3 Two alternative first-best instruments
Similarly, a first-best labor subsidy may avoid require cross-agency cooperation, but may deliver a confusing subsidy system, one which varies by workplace (and therefore a worker’s home location, since spatial [zone to zone] worker-to-job matches are essentially one to one) (Figure 3b). In this simulation case, when congestion externalities dominate (exceeding agglomeration externalities), most labor subsidies may be negative, implying a labor tax (rather than subsidy) on firms. In addition, comparing to these first-best pricing strategies, land use regulation can be an alternative for social optimum. Several existing studies have shown how regulating residential density is equivalent to levying first-best congestion tolls when only congestion externalities are internalized (e.g., Pines and Sadka, 1985; Wheaton, 1998). After internalizing agglomeration economies, our results suggest that regulating only residential density is not enough to achieve the first-best optimum. It appears that an optimal land use strategy should probably allow for both job decentralization and residential densification. While density could be adjusted by zoning regulations, job or firm decentralization seems rather difficult to regulate by city authorities. On account of these difficulties in implementing first-best interventions, the following sections focus on three types of second-best policies.

### 3.2.2 Second-Best Congestion Tolls
Figure 4 Spatial distribution of wages ($w$), TFP ($A$), job ($n$) and residential densities ($1/q$), household bid-rents ($r$) and net income ($y$) in the PCT-alone equilibrium (dashed) versus the free-market equilibrium (solid).

The PCT-alone policy imposes a congestion toll that equals the marginal congestion externality, but does not correct the agglomeration externality. Simulation shows that each worker driver needs to pay an average toll of $2,357 per year (Table 3). The utility rises to 3520, 0.57% higher than that in the free-market case and about 44% of the utility gain by the first-best instruments. This lower utility gain may result from the negative side effect of PCT on the agglomeration economy. Compared to the free-market equilibrium, the PCT-alone equilibrium leads to a 32% decrease in congestion diseconomies and a 14% decline in the agglomeration benefits. While the average commute distance and costs decrease by 21% and 28%, the average wage and productivity drop by 1.6% and 1.1% (Table 3).

Without the incentive of a PLS to guarantee labor supply, the PCT-alone policy incentivizes firms and workers to locate closer to each other, to reduce commuting costs and better match labor supply and demand. For example, a PCT levied in location $x$ will reduce the level of commute volume passing $x$ to a socially optimal level, making some workers relocate to avoid paying the toll at $x$. Some workers will change their workplace to the location outside $x$, while some workers will move inside to live near the city centerpoint for outward commuting. These demand-side adjustments will decentralize firms to relatively low-productivity locations, since the lower-productivity locations are closer to the edge of the firm cluster and thus households. Compared to the free-market equilibrium, the PCT-alone policy causes a decrease in wage and TFP in most locations (Figure 4a-b), increase residential densities in areas near the firm cluster’s edges (Figure 4d), and thus lift firms’ and households’ bid-rents in the areas with increasing densities (Figure 4e). In average, the residential and labor densities rise by 14% and 13%, and the residential and firm’s rents increase by 18% and 10% (Table 3).

Figure 5 Utility gains relative to the first-best level (left) and the percent changes in aggregate congestion diseconomies and agglomeration economies (right) under 10 tolling schemes.

For seeking more efficient “second-best” tolling policies, we tracked the change in utility and externalities under ten additional tolling schemes, which impose a fixed share (ranging from 0 to 0.9) of the PCT level on each mile driven. Figure 5 presents the percent of utility gains relative to
that in the first-best optimum and the percent change of two aggregate externalities relative to the free-market case. The second-best utility gains peak at about 47% of the first-best utility gains (compared to the based equilibrium), when the toll level is set as about 72% of the PCT level. As the share increases, the congestion and agglomeration externalities decline. These findings suggest that an efficient second-best toll level should lie below the PCT level, as agglomeration economies are internalized.

3.2.3 Second-Best Labor Subsidies
The PLS-alone policy offers labor subsidies to firms cluster in the amount of agglomeration’s marginal externality benefits. In the PLS-alone equilibrium, the city needs to deliver an average subsidy of $939 per job per year, so that agglomeration benefits can be redistributed back to firms. The utility level improves by 0.09% than that of the free-market equilibrium, about 6.9% and 16% of utility gains by the first-best and the PCT-alone equilibria (Table 3). In contrast to the PCT-alone instrument, the PLS-alone policy increases agglomeration and congestion externalities by 11% and 6.5%, respectively. The growing agglomeration economies links with a 4.4% and 2.0% increase in average wage and productivity, while the decreasing congestion diseconomies lead to a 0.5% decrease in average commute distance, but a 4.0% increase in average commute costs (Table 3).
Figure 6 Spatial distribution of wages ($w$), TFP ($A$), job ($n$) and residential densities ($1/q$), household bid-rents ($r$) and net income ($y$) in the PLS-alone equilibrium (dashed) versus the free-market equilibrium (solid)

Without the PCT’s congestion correction, the PLS-alone intervention results in more agglomeration and clustering of firms than one sees under first-best instruments, triggering a rise in congestion diseconomies. According to Figure 6a, firms agglomerate in a smaller area with higher levels of wage than those in the base equilibrium. The locational productivities rise near the center of the firm cluster area and slightly fall near the city centerpoint (Figure 6b); accordingly, job densities increase at the locations with higher productivities and drop at the edges of the firm cluster with lower productivities (Figure 6c). However, PLS appears to have trivial impact on household’s spatial decision, as shown in the spatial patterns of residential densities, bid-rents, and net income (Figure 6d-f). As the firm cluster narrows down, some residents can live closer to the city center, incurring a slightly compact city.

Similar to those second-best tolling schemes, the policies with the labor subsidy level setting below the PLS level may generate more welfare gains than those at the exact Pigouvian level. Figure 7 shows that the utility gains relative to the first-best level peak at 8.7%, when the labor subsidy is set at about 68% of the PLS level. As the subsidy levels increase, both the aggregate congestion externality cost and agglomeration externality benefit rise.

Figure 7 Utility gains relative to the first-best level (left) and the percent changes in aggregate congestion diseconomies and agglomeration economies (right) under 10 subsidy schemes.

3.2.4 UGB Policies

This section simulates the welfare and land use effects of restricted UGB regulations, with the corresponding city boundaries being set at 80%-100% of that in the free-market equilibrium (i.e., 13.25 mile). Results show that the optimal second-best UGB is 12.26 miles, leading to a 14.4% reduction in the total land supply for housing and firms, compared to the base case (Table 3). The UGB regulations deliver lower welfare improvement than the second-best tolling and subsidy interventions. The optimal utility gain is only 0.085% of the first-best optimum, 15% of the PCT-only equilibrium, and 95% of the PLS-only equilibrium (Table 3). Differing from the PCT-alone and PLS-alone equilibria, the UGB equilibrium can simultaneously decrease congestion diseconomies by 3% (a 8% and 5% decrease in commuting distance and costs) and
increase agglomeration economies by 0.58% (a slightly 0.01% and 0.32% increase in wage and TFP) (Table 3). These effects are similar to those in the first-best optimum, though the changes in aggregate externalities are much smaller.

The spatial patterns of wage, TFP, and labor densities demonstrate that the optimal UGB has no significant impact on firms’ spatial behaviors (Figure 8a-c). The UGB can slightly centralize firms, leading to a trivial increase in productivities. On the other hand, the optimal UGB makes housing supply scarce, thus raising residential densities and escalating residential rents (Figure 8d-e). The average residential density and bid-rent are 19% and 21% higher than those in the base equilibrium (Table 3). Moreover, though the UGB causes a similar escalation in residential rent as first-best instruments do, it brings much lower improvement to net income (1% versus 9.8%, Figure 8f). This explains why even the optimal UGB regulation gains a relatively low welfare improvement.

**Figure 8** Spatial distribution of wages ($w$), TFP ($A$), job ($n$) and residential densities ($1/q$), household bid-rents ($r$) and net income ($y$) in the UGB equilibrium (dashed) versus the free-market equilibrium (solid)
Figure 9 presents the utility gains relative to the first-best level and the percent changes of externalities as the urban boundaries range from 80% to 100% of the free-market level. A too-restrictive UGB regulation could cause an overall welfare loss, due to rent-escalation effects and over-centralization of jobs, though it can largely reduce congestion diseconomies and enhance agglomeration economies. For example, in the 20%-off UGB equilibrium, the average productivity rises only by 0.8% and the average firm’s rent rise by 8% (due to over-centralization), leading to an 8% decline in the average wage. As the average wage decreases and the average residential rents largely increase by 67% (due to land supply constraint), the affordability of housing, transportation, and goods for households would drop, causing welfare loss.

![Utility gains relative to the first-best level (left) and the percent changes in aggregate congestion diseconomies and agglomeration economies (right) under six UGB regulations.](image)

3.2.5 Sensitivity Analysis

The simulations described above are all based on a city where the congestion diseconomies dominate (i.e., they are much larger than the agglomeration benefits). We conducted a sensitivity analysis by raising the agglomeration elasticity $\gamma$ from 0.04 to 0.08 and then decreasing the congestion parameter $\sigma$ from 1.5 to 1, in order to obtain a city with agglomeration economies that dominate. Similar welfare and land use effects are found in the equilibria with different combinations of new parameters. When agglomeration economies dominate in a city, PLS-alone policies can improve welfare better than PCT-alone policies, and the welfare gains by the second-best PLS policies are closer to the optimal level than those by the second-best PCT policies.

We also resolve the urban equilibria under different agglomeration limit $\bar{r}$ from 1 to 5 miles. As $\bar{r}$ increases, the firm cluster in the free-market equilibrium becomes increasingly decentralized, while the city transforms from monocentric to non-monocentric structures (forming an annulus, instead of a CBD cluster). This is because the city center has no pre-set/historical productivity advantage, so firms can agglomerate away from the old city center until reaching a location limit, trading off lower densities and rents. If the city center already offers agglomeration benefits (as in most existing regions), one can obtain duo-centric or multi-centric urban forms as well. For a more detailed discussion, readers can refer to Zhang and Kockelman (2014). However, the change in $\bar{r}$ will not fundamentally alter the comparative welfare and land use effects. Moreover,
sensitivity analyses (via various following parameter value changes) also deliver no significant changes in interpretation of results; parameters varied here include the elasticity parameter for goods consumption (or $\alpha$ in the utility function, from 0.9 to 0.6), the utility level $\tilde{u}$ (from 3500 to 5000), and the productivity scale parameter (from 30,000 to 60,000).

4. CONCLUSION AND DISCUSSION

This paper develops and then applies a new spatial general equilibrium model in order to explore the welfare and land use effects of first-best instruments and second-best instruments, such as PCT-alone policies, PLS-alone policies, and UGB regulations, in cities with both congestion and agglomeration externalities. This new model differs from many existing studies (e.g., Fujita and Ogwa, 1982; Anas and Kim, 1996; Lucas and Rossi-Hansberg, 2002; Wheaton, 2004; Verhoef and Nijkamp, 2004; Arnott, 2007; Anas, 2012) by recognizing both congestion externalities and agglomeration externalities on production, while allowing endogenous land use decisions by households and firms over continuous space.

Our findings serve as a fresh contribution to two important debates surrounding multiple urban externalities. The first debate focuses on the modeling framework applied in analyzing interactions between externalities. Both our analytical and simulation results support previous studies’ results, supporting the notion that it is important to use general equilibrium frameworks, rather than non-spatial or partial equilibrium models (see, e.g., Arnott, 2007), and internalize spatial interactions (e.g., Verhoef and Nijkamp, 2004) when analyzing urban externalities. Our model further suggests that it is critical to endogenize firms’ land use decisions (e.g., decentralization and agglomeration), which are always neglected in the traditional monocentric model. Firms’ spatial adaptations to some policy interventions (e.g., PLS-only policies) could improve welfare, even though these interventions are expected to increase total external costs. In contrast, even though some policies (e.g., very restrictive UGB regulations) can largely reduce congestion externalities and raise agglomeration economies as first-best instruments do, they may still cause welfare loss. Only by considering the land use decisions of both firms and households can one quantify such policy impacts. This work does not imply that aspatial, partial equilibrium, or monocentric models should be not used for policy analysis, but that decision makers should recognize the potential distortions when using such models in cities full of distinctive externalities.

The second debate concerns the efficiency and design of different urban policies. First-best instruments may maximize social welfare but are difficult to implement in practice, especially when recognizing spatial variations. The aggregate first-best toll (or labor subsidy) lies below its related aggregate externality cost (or benefit), as also found in Arnott (2007) and Thissen et al.’s (2011) empirical analysis for the Netherlands. However, the specific optimal tolls levied on drivers can be both positive and negative, varying over space, while the subsidies to firms for hiring workers are even more complex to design, since they depend on both worker residence and workplace. While both first-best tolling and subsidy policies are equivalent in theory, some may suggest that it is easier to subsidize firms than charge drivers, because the public prefers to earn the subsidy rather than pay the tolls and subsidizing a few firms may be much easier than tolling the masses. However, our findings challenge this belief, since the aggregate optimal subsidy will equal the aggregate optimal toll. If the optimal toll is a true negative tax, firms need to pay a labor tax, rather than receive a positive subsidy, when hiring/paying a worker. Also,
when agglomeration economies are larger than congestion diseconomies, commuting subsidies can replace labor subsidies, similar to findings in Wrede (2009) and Borck and Wrede (2009).

Second-best instruments may be less challenging to implement than first-best instruments, but offer less utility gain, since they often come with side effects. For example, the PCT-alone policies can largely mitigate congestion diseconomies, but concurrently erode productivity and agglomeration economies, since congestion tolls cause some firms to shift labor supply to less productive locations, decentralizing jobs. In contrast, the PLS-alone policies can correct labor market distortions and thereby raise agglomeration economies, causing firms and households to centralize and agglomerate, resulting in greater traffic congestion. Interactions between these two externalities suggest that an efficient second-best toll or subsidy should be set at the naïve Pigouvian level, as most partial equilibrium models and many travel demand modelers do.

The UGB regulations may partially correct distortions in both transport and labor markets, but distort the land market via the rent-escalation effects and over-centralization of jobs, leading to even lower utility gains, as compared to the other two second-best policies investigated here. Such UGB distortions in land markets appear present in regions like Portland, Oregon and Knoxville, Tennessee, where housing rents/prices inside the UGBs rise faster than properties in areas without UGBs (Staley and Mildner, 1999; Cho et al., 2008). London, England and Auckland, New Zealand also have reported major rent escalations due to relatively low housing or land supply for new development (Cheshire and Sheppard, 2005; Cox, 2010). Home affordability remains a key topic for debate under growth management discussions (Downs, 2004; Nelson et al., 2004). Moreover, development activities in 95 relatively “contained” U.S. metro areas (those with topographical limitations, greenbelts, and/or UGBs) are more agglomerated near their central cities than those in uncontained areas (Nelson et al., 2004). These effects may facilitate the central-city revitalization, while causing land market distortions across the region. Of course, real cities are much more complex than the models allowed here use, and many more variables are at play and may counteract some or much of the rent escalation losses that tend to come with tight UGBs.

Multiple opportunities exist to make these models more realistic. Allowing for travel mode and trip scheduling flexibility is important in appreciating congestion toll effects. Moreover, a model that enables a gradual, dynamic city evolution is important to explore. The one-shot, static equilibrium typical of papers in urban economics is never achieved in practice. In reality, most cities already exist, and populations regularly expand, in the midst of great uncertainty and imperfect information, along with speculation and other complex -- but very realistic -- human behaviors. Several recent studies have explored this topic (e.g., Boucekkine et al., 2009; Desmet and Rossi-Hansberg, 2010). Finally, allowing for more diverse, and realistic policies, like flat-rate tolls and only on freeways, or cordon area congestion pricing (Zhang and Kockelman, 2014), would be meaningful, since PCTs and PLSs are not common in practice. Nevertheless, the tool developed here extends urban economic modeling while illuminating multiple impacts of several important policies and a wide range of behavioral assumptions, relating to human settlement in the past, present and future.
ACKNOWLEDGMENTS
The authors thank Alex Anas, Jan Brueckner, and Esteban Rossi-Hansberg for their inspired comments and questions, and Annette Perrone for her administrative support.

REFERENCES
Glaeser, E. and Kahn, M. 2001, Decentralized employment and the transformation of the

**APPENDIX**

**A1: Proof of Proposition 1**

(a) Since utility maximization and expenditure minimization are fully equivalent, the minimum expenditure at the equilibrium utility $\bar{u}$ equals the net income $y(x, x_w)$, i.e., $y(x, x_w) = e(r_h(x), \bar{u})$. Since $r_h(x)$ is only relevant to location $x$, one has $y(x, x_w) \equiv y(x)$. Under utility maximization, $c^*(x, x_w) = c^*(y(x, x_w)) = c^*(y(x)) = c^*(x)$, and $q^*(x, x_w) = q^*(r_h(x), y(x, x_w)) \equiv c^*(r_h(x), y(x)) = c^*(x)$.

(b) From the first-order conditions of this utility maximization problem, one can derive the following: $c(x) + q(x)u_q/u_e = y(x)$. In combination with $u(c(x), q(x)) = \bar{u}$, one calculates that $q^*(x) = q^*(y(x), \bar{u})$ and $c^*(x) = c^*(y(x), \bar{u})$. 

25
(c) Since \( t(x, x) = 0 \), \( y(x) = y(x, x) = w(x) = t(x, x) = w(x) \).

(d) Since \( w(x) \equiv w(x_w) - t(x, x_w), \forall x_w > 0 \), \( w(x_w) - w(x) = \int_x^{x_w}[t(s) + \tau(x)]ds \).

Thus, \( w'(x) = t(x) + \tau(x) \). From (c), \( y'(x) = t(x) + \tau(x) \).

**A2: Proof of Proposition 2**

The solutions to the social optimum are achieved by determining each of six factors, \( \{n(x), q(x), c(x), \theta_f(x), F(x), t(x)\} \), at each location \( x \) so as to maximize the households’ utility level under constraints (A1)-(A5), as defined in Problem A.

**Problem A.** Choose functions \( n(x), q(x), c(x), \theta_f(x), F(x) \) so as to maximize

\[
U(c(x), q(x))
\]

subject to

\[
\begin{align*}
(A1) & \quad \int_0^\pi \left\{ 2\pi x \left( \theta_f(x) \delta n(x)^2 F(x)^r - \frac{\theta_h(x)}{q(x)} c(x) - (1 - \theta_t)R_a \right) - t(x)D(x) \right\} dx \geq 0 \\
(A2) & \quad \theta_h(x) + \theta_f(x) + \theta_t = 1 \\
(A3) & \quad F(x) = \xi \int_0^\pi \int_0^{2\pi} r \theta_f(r)n(r)e^{-\xi(x,r,\psi)}d\psi dr \\
(A4) & \quad t(x) = \varphi - \rho \left( \frac{D(x)}{2\pi x \theta_t} \right) \sigma \\
(A5) & \quad D'(x) \leq 2\pi x \left( \frac{\theta_h(x)}{q(x)} - \theta_f(x)n(x) \right)
\end{align*}
\]

for all \( x \in [0, \bar{x}] \), with boundary conditions:

\[
\begin{align*}
(A6) & \quad D(0) = 0 \text{ and } D(\bar{x}) = 0 \\
(A7) & \quad \int_0^\pi 2\pi x \frac{\theta_h(x)}{q(x)} dx = N
\end{align*}
\]

Equations (A1)-(A11) are present in the body text of this paper, with the exception of constraint (A3), which guarantees a non-negative net social surplus. Given that aggregate land rents (net of the opportunity costs) are equally returned to each household (in this closed system), the net surplus equals the total value of production, minus general consumption, minus and opportunity costs of land, and minus workers’ commute costs.

The Hamiltonian function of the Problem A is given by:

\[
H(n, F, q, c, \theta_f, \beta_1, \beta_2, \beta_3, \beta_4) =
\]

\[
u(c(x), q(x))/\lambda(x) + 2\pi x \left[ \theta_f(x) \delta n(x)^2 F(x)^r - \frac{\theta_h(x)}{q(x)} c(x) - (1 - \theta_t)R_a \right] -
\]

\[
t(x)D(x) + \left( \beta_1(x)F(x) = \xi \int_0^\pi \int_0^{2\pi} \beta_1(r) \theta_f(r)n(r)e^{-\xi(x,r,\psi)}d\psi dr \right) + \beta_2(x) \left( t(x) - \varphi - \rho \left( \frac{D(x)}{2\pi x \theta_t} \right) \sigma \right) \right) + \beta_3(x) 2\pi x \left( \frac{1-\theta_t - \theta_f(x)}{q(x)} - \theta_f(x)n(x) \right)
\]

From the Maximum Principle, some of the first-order conditions are derived as:

\[
(A8) \frac{\partial H}{\partial n} = \frac{\partial H}{\partial n(x)} + \frac{\partial H}{\partial n(r)} = 2\pi x \theta_f(x) \left[ \delta \kappa n(x)^{r-1}F(x)^r - \beta_3(x) \right] - \xi \int_0^\pi \int_0^{2\pi} \beta_1(r) \theta_f(r)e^{-\xi(x,r,\psi)}d\psi dr = 0
\]

\[
(A9) \frac{\partial H}{\partial F} = 2\pi x \theta_f(x) \gamma \delta n(x)^2 F(x)^r - 1 - \beta_1(x) = 0
\]

\[
(A10) \frac{\partial H}{\partial D} = -\beta_3'(x) \rightarrow \beta_3'(x) = t(x) + \rho \sigma \left( \frac{D(x)}{2\pi x \theta_t} \right) \sigma
\]

26
In a socially optimal city, both conditions (A11) and (A12) should be satisfied. Thus,

\[
\beta_3(x) = w(x) - s(x)
\]

Comparing the first-order condition (A10) and Eq.(A13), one can derive the following equations:

\[
\tau(x) - \rho \sigma \left( \frac{|D(x)|}{2\pi \theta_r} \right)^{\sigma} = \left( s(x) - \zeta \delta \int_0^\infty r \theta_f(r)n(r)F(r) e^{-\zeta l(x,r)} dr \right) + \tau(x) + \rho \sigma \left( \frac{|D(x)|}{2\pi \theta_r} \right)^{\sigma}
\]

When household’s utility is maximized, from Proposition 1d and (A14), one can obtain the following relationship:

\[
\tau(x) - \rho \sigma \left( \frac{|D(x)|}{2\pi \theta_r} \right)^{\sigma} = \left( s(x) - \zeta \delta \int_0^\infty r \theta_f(r)n(r)F(r) e^{-\zeta l(x,r)} dr \right) + \tau(x) + \rho \sigma \left( \frac{|D(x)|}{2\pi \theta_r} \right)^{\sigma}
\]

In order to fulfill Eq. (A15) for each location \( x \), we have three strategies:

(a) A combination of two instruments:

\[
\begin{align*}
\tau(x) &= \tau_{pct}(x) = \rho \sigma \left( \frac{|D(x)|}{2\pi \theta_r} \right)^{\sigma} \\
(s(x) &= s_{pls}(x) = \zeta \delta \int_0^\infty r \theta_f(r)n(r)F(r) e^{-\zeta l(x,r)} dr, \text{ if } \theta_f(x) > 0)
\end{align*}
\]

(b) When \( s(x) = 0 \), \( \tau(x) = \tau_{pct}(x) - s_{pls}′(x) \), which represents the first-best toll at location \( x \). Given Eq. (A5), the total toll revenues thus equal:

\[
\int_0^\infty \tau(x)D(x)dx = \int_0^\infty \left( \tau_{pct}(x) - s_{pls}′(x) \right)D(x)dx = \int_0^\infty \tau_{pct}(x)D(x)dx - \int_0^\infty 2\pi \theta_f(r)n(x)s_{pls}(x)dx
\]

Therefore, revenues provided by optimal tolling across the region equal the total congestion externality costs of the work commute traffic (or total revenues from the PCT policy) minus total agglomeration externality benefits (or total payments under the PLS policy).

(c) When \( \tau(x) = 0 \), \( s′(x) = s_{pls}′(x) - \tau_{pct}(x) \). Thus, \( s(x_i, x) = s_{pls}(x_i, x) = \int_{x_i}^x \tau_{pct}(x)dx \), which represents the first-best subsidy to workers living at \( x_i \) but working at \( x \). Given Eq. (A5) and the fact that \( \theta_h(x) = 0 \), the total first-best subsidies equals the following:

\[
\int_0^\infty 2\pi \theta_f(r)n(x)s(x)dx = -\int_0^\infty s(x)D′(x)dx = \int_0^\infty s′(x)D(x)dx - s(x)D(x)
\]

Thus, total optimal subsidy to workers equals the overall benefits of agglomeration to the region’s firms minus total external congestion costs.