Optimal Policies in Cities with Congestion and Agglomeration Externalities: Congestion Tolls, Labor Subsidies, and Place-Based Strategies

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

Kara M. Kockelman
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

Journal of Urban Economics 95C: 64-86, 2016.

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 policies and second-best pricing and place-based strategies using simulations. A first-best policy must combine both Pigouvian congestion tolling (PCT) and Pigouvian labor subsidies (PLS) instruments, or design an optimal toll (or subsidy) internalizing agglomeration externalities (or congestion externalities). We also examine second-best pricing policies if only one instrument is adopted. Congestion pricing alone policies (e.g., a partial PCT or a flat-rate toll) can improve social welfare only in heavy-congestion cities while their welfare gains could be trivial (e.g., below 10% of the welfare improvement achieved by first-best policies). In contrast, second-best labor subsidy alone policies are a more effective alternative to first-best policies. As to place-based policies, the firm cluster zoning (FCZ) regulation is more efficient than the urban growth boundary (UGB) policy. UGBs only have small effects on the agglomeration economy but could worsen land market distortion via the residential rent-escalation effects. These findings suggest that it is important to internalize firms' land use decisions and relax monocentricity assumptions, in order to appreciate the interplay of both urban externalities, since spatial adaptations to policy interventions can distort system efficiencies.

Key Words: Nonmonocentric Urban Economics, Agglomeration, Congestion, Optimal Policies, Land Use.

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 CO_2, 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. Many empirical studies have demonstrated that doubling job density results in a 4%-10% increase in productivity at the metropolitan level (Rosenthal and Strange, 2004; Combes et al., 2010) and even larger agglomeration benefits at smaller geographical scales (Arzaghi and Henderson, 2008; Rosenthal and Strange, 2008).

While urban economists have long recognized either negative congestion externalities (Pigou,1920; Mohring and Harwitz, 1962; Vickrey, 1963; 1969; Solow 1972; Henderson, 1974a; Kanemoto, 1977; Arnott, 1979; Small,1983; Pines and Sadka, 1985; Wheaton, 1998; Anas and Xu, 1999; Verhoef, 2005; Brueckner, 2007) or positive agglomeration externalities (Henderson, 1974b; Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002; Fujita and Thisse, 2002; Berliant et al., 2002; Duranton and Puga, 2004; Rossi-Hansberg, 2004; Borck and Wrede, 2009), few have considered their interactions *within a city*[1]. 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 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 (2007) developed a two-island model internalizing both negative congestion and positive production externalities. In the simplified model, residents living in an island and firms locate at

[1] Much literature has analyzed the interaction between agglomeration benefits and urban crowding costs in multicity systems (e.g., Henderson, 1974b; Fujita and Thisse, 2002; Eeckhout, 2004; Desmet and Henderson, 2014; Behrens and Robert-Nicoud, 2014), rather than within a city.

the other island, with a road of fixed capacity joining the two islands. He found that a Pigouvian congestion toll alone is not optimal policy, since it may harm agglomeration economies and productivity. Instead, optimal congestion tolls should be lower than the Pigouvian level when there is no policy in place to manage agglomeration externalities. Arnott believes that these findings are consistent, even though his model was extended to internalize time-varying congestion, heterogeneous individuals and firms, residential location and land use decisions, and multiple employment centers. These two studies identify the policy importance on incorporating multiple externalities. However, they either neglect the spatial distribution of externalities or assume an exogenously determined urban form (e.g., two islands), failing to fully analyze the interaction between externalities and urban form, which may significantly affect the optimal design of urban policies.

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 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, second-best pricing instruments, and second-best place-based policies. For the first-best interventions, we examine welfare gains and land use patterns in the social optimum and the challenges to designing first-best instruments. These topics are seldom discussed in cities with multiple externalities. We compare the welfare outcomes and the possible side effects of the second-best PCT-alone and Pigouvian labor subsidy (PLS) alone policies and a flat-rate congestion tolling (FRCT) scheme. A robustness analysis is conducted by changing the congestion and agglomeration parameters to investigate how these policies and their impacts vary with the scales of externalities.

We also examine the welfare effects and side effects of optimal urban growth boundaries (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). For comparison, we also discuss another coarse place-based policy by designating a cluster zone exclusively for firm/business use, i.e., firm cluster zoning (FCZ). Simulation results suggest that the optimal FCZ policy generates a larger welfare improvement than the optimal UGB policy. Such questions and comparisons relate closely to planning practice, 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; Sections 3 describes parameters and simulation settings; Sections 4-5 compare simulation results for welfare, externalities, and land use, evaluates first-best policy scenarios and provides robustness analysis; Sections 6-7 discuss the second-best pricing and place-based policies; Section 8 offers conclusions and suggestions for future work.

2 The Model

The model developed here mainly refers to Lucas and Rossi-Hansberg (LRH, 2002). While the LRH model has well established a nonmonocentic model with agglomeration externalities, our model extends it to consider traffic congestion and more importantly, contributes to the discussion on optimum versus equilibrium under congestion and agglomeration externalities. In addition, there are several differences with basic modeling settings. First, the model relaxes the constraint of the fixed city boundary in the LRH model, allowing for an endogenously determined boundary under an additional constraint where the city edge land rent equals a fixed agriculture land rent. The latter constraint is often used in monocentric models (e.g., Wheaton, 1998; Brueckner, 2007). This change can internalize city size, which may affect the spatial distribution of land use and the commute distance/cost. Second, while the LRH model measures commute time costs determined by travel time and wage, our model's measure is simplified to commute money costs determined only by distance (and traffic volume after considering congestion). In reality, the commute costs should consist of both costs of time and money. Third, our model is built in a closed-form city with a fixed population and all revenues (or subsidies) uniformly redistributed to residents (or firms), while the LRH model is built in an open-form city with a fixed utility and without revenue redistribution. These changes increase the complexity of computational simulation, but make this type of nonmonocentric model more flexible for optimal policy analysis.

The model 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 $x (0 \le x \le \bar{x})$, $\theta_f(x)$, $\theta_h(x)$ and θ_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 θ_t is exogenously given.

2.1 Household and Congestion Externality

Each household living in location x and working at location x_w 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 $u\big(c(x, x_w), q(x, x_w)\big)$. Its willingness to pay for land is rental rate $r_h(x)$. 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, $w(x_w)$, 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 $(0 < x \le \bar{x})$, *choose a job location* x_w $(0 < x_w \le \bar{x})$ *and evaluate functions* $c(x, x_w)$ *and* $q(x, x_w)$, *so as to maximize utility*

(1) $$u\big(c(x, x_w), q(x, x_w)\big)$$

subject to the budget constraint:

(2) $$c(x, x_w) + r_h(x) q(x, x_w) \le y(x, x_w) = w(x_w) + \bar{y} - T(x, x_w)$$

where

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

(4) $$T(x, x_w) = \int_x^{x_w} (t(s) + \tau(s)) ds$$

The budget constraint in Eq. (2) represents that the expenditure of goods and housing is no larger than the net income. Eq. (3) 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. (4) 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, φ (in dollars per mile). Thus,

$$t(x) = \begin{cases} -\varphi - \rho\left(\frac{-D(x)}{2\pi x\theta_t}\right)^{\sigma} & \text{if } D(x) < 0 \\ \varphi + \rho\left(\frac{D(x)}{2\pi x\theta_t}\right)^{\sigma} & \text{if } D(x) > 0 \\ \varphi \;\; or - \varphi & \text{if } D(x) = 0 \end{cases} \tag{5}$$

where ρ and σ $(\sigma \geq 1)$ are positive parameters designed to reflect road congestibility. 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).

Proposition 1: *Suppose* $c^*(x, x_w)$*and* $q^*(x, x_w)$ *are the solutions to Problem 1 and* $\bar{u}$ *is an equilibrium 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)* *Both the equilibrium consumption of goods and lot size are functions of the net income and the utility level, that is,* $c^*(x) = c^*(y(x), \bar{u})$ *and* $q^*(x) = q^*(y(x), \bar{u})$.
- *(c)* *The net income of households residing at x equals the wage income paid by firms at x plus redistributed revenues, that is,* $y(x) = w(x) + \bar{y}$.
- *(d)* *The condition that both the wage gradient and the net-income gradient equal the marginal travel cost should be satisfied when maximizing utilities, that is,* $y'(x) = w'(x) = t(x) + \tau(x)$. *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.*

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:

$$u\big(c(x), q(x)\big) = c(x)^{\alpha} q(x)^{1-\alpha}, \quad 0 < \alpha < 1 \tag{6}$$

then, the solutions to *Problem 1* are:

$$q^*(x) = \alpha^{-\alpha/(1-\alpha)} y(x)^{-\alpha/(1-\alpha)} \bar{u}^{1/(1-\alpha)} \tag{7}$$

$$c^*(x) = \alpha y(x) \tag{8}$$

and maximized bid-rents from households are:

(9) $$r_h^m(x) = (1-\alpha)\frac{y(x)}{q^*(x)} = (1-\alpha)\alpha^{\alpha/(1-\alpha)}\left(\frac{y(x)}{\bar{u}}\right)^{1/(1-\alpha)}$$

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, $y(x)$, as defined in the Proposition 1c. As $1/q^*(x)$ represents the optimal residential density at location x, from Eq. (9), the maximum bid-rent of households at location x, $r_h^m(x)$, is proportional to the optimal residential density and the net income.

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:

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

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

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

where $n(x)$ is labor density along ring x and κ is the production function's elasticity parameter. One can internalize agglomeration economies in the TFP, by assuming that the agglomeration externality $F(x)$ at location x determines the productivity:

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

Here, δ is the productivity scale parameter and γ 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[2]. Similar to LRH's setting, agglomeration 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 the city boundary[3]. Thus, the agglomeration externality at each location along the annulus at radius x is specified as

[2] One can set a more general formation of the agglomeration externality function, as in the following example:

$$F(x) = \int_0^{\bar{x}} b(r)d(r,x)dr$$

Here, $b(r)$ represents the density of firms or workers at location r. $d(r,x)$ is a distance-based decay function from location r to x. Two specifications of $d(r,x)$ are widely used. For example, in a linear city, $d(r,x)$ could be a linear form, $1-\phi|r-x|$ (e.g., Ogawa and Fujita, 1980; Duranton and Puga, 2014), or an inverse-exponential form, $e^{-\phi|r-x|}$ (e.g., Fujita and Ogawa, 1982). These two formations are actually equivalent when $\phi|r-x|$ is small enough. In our simulation experiments, we compared results using both externality specifications, and found negligible differences in land use and welfare outcomes. This finding also corresponds to those in the linear model (e.g., by comparing Ogawa and Fujita [1980] and Fujita and Ogawa [1982]). Thus, the following discussions only hinge on the inverse-exponential specification.

[3] LRH's model sets a fixed-boundary assumption, while our model estimates an endogenous $\bar{x}$ under the constraint of edge land rent. This change allows for endogenous city size.

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

where ζ is the production externality scale parameter, and is exogenously determined. ψ is the polar angle around the center (ranging from 0 to 2π), 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 $\bar{x}$ miles of the center (at a counter-clockwise angle of ψ from the first firm). Thus,

(14) $$l(x,r,\psi) = \sqrt{x^2 + r^2 - 2xr\cos(\psi)}$$

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):

(15) $$\text{Max } \pi(x) = \delta n(x)^{\kappa} F(x)^{\gamma} - n(x)\big(w(x) - s(x)\big) - r_f(x)$$

where $s(x)$ represents a potential labor subsidy for firms at location *x* to hire each worker. The per capita profit of firms at location *x*, $\pi(x)$, equals the corresponding production gains minus labor costs and land rents.

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

(16) $$n^*(x) = \left(\frac{\kappa\delta F(x)^{\gamma}}{w(x)-s(x)}\right)^{1/(1-\kappa)}$$

The optimal employment density increases with the locational productivity $A(x)$ (as shown in Eq. (12), or the agglomeration externality $F(x)$) and decreases with the labor cost. 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:

(17) $$r_f^m(x) = \frac{1-\kappa}{\kappa} n^*(x) = (1-\kappa)\delta^{1/(1-\kappa)} F(x)^{\gamma/(1-\kappa)} \left(\frac{\kappa}{w(x)-s(x)}\right)^{\kappa/(1-\kappa)}$$

Eq. (17) shows that the maximum bid-rent of firms at location *x*, $r_f^m(x)$, is only proportional to the optimal job density. The area with a higher job density is often of a higher land rent.

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:

(18) $$r(x) = max\{r_h^m(x), r_f^m(x), R_a\}$$

(19) $$r(\bar{x}) = R_a$$

Eq.(18) guarantees that the equilibrium land rent $r(x)$ is the maximum bid-rent provided by either households, firms, or the absent bidders for agricultural use. 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 $r_f^*(x)$ equals $r_h^*(x)$, 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 thus shown in Eq. (20):

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

(21) $$\theta_h^*(x) = 1-\theta_t-\theta_f^*(x)$$

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:

(22) $$y_{rent} = \int_0^{\bar{x}} 2\pi x\left\{\theta_f^*(x)\left(r_f^m(x)-R_a\right)+\theta_h^*(x)\left(r_h^m(x)-R_a\right)\right\}dx$$

Eq. (22) shows that y_{rent} includes the aggregate revenue of land rents at each residential location *x*, $r_h^m(x)-R_a$, and that of land rents at each firm location *x*, $r_f^m(x)-R_a$.

2.4 The Labor Market's Equilibrium Conditions

Under equilibrium, 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)[4]. Thus,

(23) $$D'(x)dx = 2\pi x\left(\frac{\theta_h^*(x)}{q^*(x)} - \theta_f^*(x)n^*(x)\right)dx$$

In Eq. (23), $2\pi x\frac{\theta_h^*(x)}{q^*(x)}dx$ represents an equilibrium number of workers living at the circular interval *dx* and $2\pi x\theta_f^*(x)n^*(x)dx$ represents an equilibrium level of job positions provided by firms at *dx*. 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:

(24) $$D(0) = 0 \quad and \quad D(\bar{x}) = 0$$

[4] The LRH paper explains *D*(*x*) (labeled as *H*(*x*)) as the stock (work hour) of unhoused workers at *x*. Since the LRH model measures commute costs using travel time and the total time for working and commuting is fixed, the changed stock of unhoused workers from *x* to *x*+*dx* (or *x*-*dx*) include two parts. The first part is the net number of unhoused workers in the interval *d*x. Another part is the lost work hours due to passing the interval. This part is not included in our model, since our model only considers the distance-based commute costs and the losses of work hours are not monetarized.

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*:

(25)
$$\int_0^{\bar{x}} 2\pi x \frac{\theta_h^*(x)}{q^*(x)} dx = \int_0^{\bar{x}} 2\pi x \theta_f^*(x) n^*(x) dx = N$$

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*: $c^*(x), q^*(x), r_h^m(x), y(x), t(x), \tau(x), D(x), w(x), n^*(x), r_f^m(x), s(x), F(x), r(x),$ $\theta_h^*(x), \theta_f^*(x)$, and 5 scalars: $\bar{x}, \bar{y}, y_{rent}, y_{toll}, y_{suby}$. Twenty equations are needed to resolve this model, including 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 policy instrument, $\tau(x)$ and y_{toll}, and the subsidy, $s(x)$ and y_{suby}, which vary across policy scenarios. Notice that analytical equilibrium results are very difficult to derive here, for a 20-equation system with several non-linear equations and differential equations. Thus, our analysis mainly relies on numerical simulations to compare the properties of the free-market, first-best and second-best equilibrium settings, by setting varying function values for $\{\tau(x), s(x), y_{toll}, y_{suby}\}$.

Table 1 summarizes these four functions, $\tau(x)$, $s(x)$, y_{toll}, and y_{suby}, across six spatial equilibria. In the free-market equilibrium, neither a toll nor a subsidy is imposed, so $\tau(x) = 0, s(x) = 0, y_{toll} = 0,$ and $y_{suby} = 0$. 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. Six types of intervention are considered here: the simultaneous application of two first-best instruments, second-best PCT scenarios, second-best flat-rate congestion tolling (FRCT) scenarios, second-best PLS scenarios, second-best UGB scenarios, and second-best zoning regulation scenarios

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

Policy Interventions	**Equations**
Free-Market Case	$\tau(x) = 0; s(x) = 0; y_{\text{toll}} = 0; y_{\text{suby}} = 0$
First-Best Case	$\tau(x) = \tau_{\text{pct}}(x);\ s(x) = s_{\text{pls}}(x);\ y_{\text{toll}} = \int_0^{\bar{x}} \tau(x) D(x) dx\,;$ $y_{\text{suby}} = \int_0^{\bar{x}} 2\pi x \theta_f^*(x) n^*(x) s(x) dx.$
Second-Best Pigouvian Congestion Toll (PCT) Scenarios	$\tau(x) = \varsigma_{\text{pct}} \tau_{\text{pct}}(x), \varsigma_{\text{pct}} \in [0,1]; s(x) = 0; y_{\text{toll}} = \int_0^{\bar{x}} \tau(x) D(x) dx\,; y_{\text{suby}} =$ 0. For $\varsigma_{\text{ts}} = 0$, the scenario equals to the free-market case. For $\varsigma_{\text{ts}} = 1$, the policy is defined as a 100% PCT-alone policy.
Second-Best Flat-Rate Congestion Tolling (FRCT) Scenarios	$\tau(x) = \varsigma_{\text{frct}}$, ς_{frct} is a flat rate of congestion toll, e.g., \$0.2/mile. $s(x) = 0; y_{\text{toll}} = \int_0^{\bar{x}} \tau(x) D(x) dx\,; y_{\text{suby}} = 0.$
Second-Best Pigouvian Labor Subsidy (PLS) Scenarios	$\tau(x) = 0; s(x) = \varsigma_{\text{pls}} s_{\text{pls}}(x); y_{\text{toll}} = 0;\ y_{\text{suby}} = \int_0^{\bar{x}} 2\pi x \theta_f^*(x) n^*(x) s(x) dx.$ $\varsigma_{\text{pls}} \in [0,1]$. For $\varsigma_{\text{pls}} = 0$, the scenario equals to the free-market case. For $\varsigma_{\text{pls}} = 1$, the policy is defined as a 100% PLS-alone policy.

Urban Growth Boundary (UGB) Scenarios	$\tau(x)=0; s(x)=0; y_{\text{toll}}=0; y_{\text{suby}}=0;$ $\bar{x}=\bar{x}_{\text{ugb}}$ in place of Eq. (19). $\bar{x}_{\text{ugb}}$ is exogenously given.
Firm Cluster Zoning (FCZ) Scenarios	$\tau(x)=0; s(x)=0; y_{toll}=0; y_{\text{suby}}=0;\ \theta_f^*(x)=\begin{cases}1-\theta_t, x\in[x_0,x_1]\\ 0, \quad\quad\quad \text{others}\end{cases}$ in place of Eq. (20). x_0 and x_1 are exogenously given.

Notes: Under different policy interventions, some instrument values (e.g., y_{toll} and y_{suby}) have the same equation expression but different quantities given that the underlying equilibrium will be different (e.g., the equilibrium $\theta_f^*(x)$ and $n^*(x)$ are different at each location in the first-best and the PLS-alone case).

Proposition 2: *First-best instruments to correct congestion and agglomeration externalities satisfy either one of following conditions:*

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

(26) $$\tau_{pct}(x)=t'\left(D(x)\right)D(x)=\begin{cases}\rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^{\sigma}, & if\ \ D(x)\geq 0\\ -\rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^{\sigma}, & if\ \ D(x)\leq 0\end{cases}$$

(27) $$s_{pls}(x)=\begin{cases}\frac{\partial\left(\int_0^{\bar{x}}2\pi r\theta_f^*(r)p(r)dr\right)}{\partial g^*(x)}-\frac{\partial p(n^*(x))}{\partial n^*(x)}, & \text{if}\ \ \theta_f(x)>0\\ 0, & \text{if}\ \ \theta_f(x)=0\end{cases}$$
$$=\begin{cases}\gamma\delta\zeta\int_0^{\bar{x}}\int_0^{2\pi}r\theta_f^*(r)n^*(r)F(r)^{\gamma-1}e^{-\zeta l(r,x,\psi)}\,d\psi dr, & \text{if}\ \ \theta_f^*(x)>0\\ 0, & \text{if}\ \ \theta_f^*(x)=0\end{cases}$$

where $g^*(x)=2\pi x\theta_f^*(x)n^*(x)dx$, representing the number of workers in the interval dx (from the locations x+dx to x or x to x-dx).

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

(28) $$\tau_{fb}(x)=\begin{cases}\tau_{pct}(x) & \text{if}\ \ \theta_f^*(x)=0\\ \tau_{pct}(x)-\frac{\partial s_{pls}(x)}{\partial x}, & \text{if}\ \ \theta_f^*(x)=0\end{cases}$$

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_i and works at x, $s_{fb}(x_i,x)$ will be as follows:

(29) $$s_{fb}(x_i,x)=s_{pls}(x_i,x)-\int_{x_i}^{x}\tau_{pct}(r)\,dr$$

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

Proof. *See A2 in the Appendix.*

In the socially optimal city, market's failures from both congestion and agglomeration externalities need to be corrected by first-best instruments. As noted in Proposition 2, the social optimum can be achieved via three types of first-best instruments. The city can simultaneously impose PCT and PLS, both of which equal corresponding marginal externalities, as shown in Eqs. (26) and (27). The marginal congestion externality at each *x* equals $\tau_{pct}(x)$, i.e., $t'\big(D(x)\big)D(x)$. Intuitively, the derivative of $t(x)$ with respect to $D(x)$ times $D(x)$ represents the added marginal travel cost on all individuals traveling across location *x* when another new driver is added, while $\tau_{pct}(x)$ represents total added travel costs, as caused by this same added driver. The marginal external benefit by hiring an additional worker at location *x* equals $s_{pls}(x)$, calculated by the marginal social benefit (i.e., $\frac{\partial\left(\int_0^{\bar{x}} 2\pi r\theta_f^*(r)p(r)dr\right)}{\partial g^*(x)}$) minus the marginal private benefit (i.e., $\frac{\partial p(n^*(x))}{\partial n^*(x)}$). Notice that $\int_0^{\bar{x}} 2\pi r\theta_f^*(r)p(r)dr$ (and $p(r) = \delta n^*(r)^{\kappa}F(r)^{\gamma}$) is the aggregate product and the agglomeration externality at any location *r* $F(r)$ is a function of the number of workers in the interval *dx*, $g^*(x)$.

The city can also impose first-best tolls by internalizing external benefits of agglomeration into PCT levels. Proposition 2b suggests that the first-best tolls largely vary with location. They should be set at corresponding Pigouvian levels in residential areas but not within firm clusters. After considering the impact on agglomeration economies, the optimal tolls could be positive or negative (e.g., an incentive or subsidy), depending on the margin of agglomeration benefits at each location, $s'_{pls}(x)$. In addition, 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 total optimal subsidy will then lie below the total agglomeration benefit. But Proposition 2c suggests that such an optimal labor subsidy will be very complicated, since it varies not only with firms' locations but also with each worker's residence.

This study also compares the second-best pricing and place-based policies. Second-best pricing instruments in practice have various forms of imposition[5]. This article concentrates on a PCT-alone policy, by which each traveler passing location *x* is levied a fixed share ς_{pct} of PCT, i.e, $\tau(x) = \varsigma_{\text{pct}}\tau_{\text{pct}}(x)$, and a PLS-alone policy, by which each firm at location x is subsidized a fixed share ς_{pls} of PLS, i.e., $s(x) = \varsigma_{\text{pls}}s_{\text{pls}}(x)$ (Table 1). The scenarios thus change ς_{pct} (or ς_{pls}) from 0 to 1 to find the second-best PCT (or PLS) policy. When $\varsigma_{\text{pct}} = 1$ (or $\varsigma_{\text{pls}} = 1$), the total amount of congestion externalities (or agglomeration externalities) is fully corrected. Although this type of 100% PCT-alone (or PLS-alone) instrument is rarely found in reality, it

[5] In practice, transportation-side pricing schemes include increasing vehicle registration fees, imposing higher fuel taxes, pricing road use such as building high-occupancy toll lanes, zone-based or area-wide pricing, and eliminating free parking or parking subsidies (USDOT, 2009). Among them, congestion pricing and parking pricing are two major topics widely discussed in urban economic studies. It is worth to note that parking pricing could be an alternative second-best pricing policies for reducing traffic congestion, especially in the downtown area or employment centers. Related work refers to Arnott et al. (1991), Arnott and Rowse (1999), Anderson and de Palma (2004), Shoup (2005), Arnott and Inci (2006), and Inci (2015). In contrast, labor subsidies for correcting agglomeration externalities appear less found in our living cities. But the investment on public transit infrastructure and service at job centers (e.g., the CBD) and subcenters could be regarded as a form of subsidies to firm/job agglomeration (Anas, 2012).

deserves a thorough discussion. It is important to show researchers and policy makers the extreme consequence of overlooking the interaction of multiple externalities when pricing policies are designed. For comparison, we also introduce a flat-rate congestion toll (FRCT) by imposing a fixed toll on each commute miles[6].

This paper particularly focuses on two types of land use regulation policy: urban growth boundary (UGB) and firm cluster zoning (FCZ). The UGB policy is a land-use regulation without any pricing adjustments, where the fixed-land-rent assumption at the city edge is replaced by fixing a city boundary, $\bar{x}_{\text{ugb}}$. The FCZ policy imposes an idealistic exclusionary zoning regulation by designating one or more cluster areas only for firm use and all remaining areas for residential use. While UGB policies have been applied in several metropolitan areas (Nelson et al., 2004), FCZ policies are less debated. Nevertheless, many cities have implemented place-based policies similar to the FCZ, such as industrial parks and high-tech development zones (see a review of Neumark and Simpson, 2014).

3 Simulation Settings

This paper simulates an abstract circular, close-form city, where the number of households (or workers) N is fixed at 600,000 and the edge 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 the 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 shows the parameter values of the base scenario[7]. Parameters of Cobb-Douglas utility and production functions rely on LRH's (2002) assumptions, where $\alpha = 0.90$ and $\kappa = 0.95$. The agglomeration parameters γ and ζ are set at 0.06 and 2, which are well in line with the empirical estimates ranging from 0.04 to 0.10 (Combes et al., 2010). The constant part of total factor productivity, δ, 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' share of land is assumed to be 30%. The intercept parameter φ in Equation (16)'s average travel cost function represents an average cost of free-flow travel, and is set at \$20 dollar per mile per year. This figure is generated from the calculation that marginal free-flow travel cost is about \$0.04 per mile when each worker works about 250 days a year. ρ and σ reflect road congestibility, and are set as 0.00001 and 1.5, respectively. For simplification, we can rescale the parameter ρ to ρ_0 with $\rho = \rho_0 \times 10^{-6}$. 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

[6] We also tested a flat-rate labor subsidy (FRLS) by providing a fixed subsidy for firms to hire each worker. However, this FRLS strategy has no impact on the city efficiency and welfare in our modeling simulation. In theory, under the closed-form setting, the flat-rate subsidy to a worker is fully paid by the worker herself.

[7] While calibrating a realistic city using empirical data under the model framework developed here is possible and important in the future, it is not a major focus of this paper. Some calibration examples can refer to several studies relying on monocentric models (e.g., De Lara et al., 2013; Rappaport, 2014) and non-monocentric models (e.g., Brinkman, 2013).

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.

Table 2 Parameter value assumptions in the base scenario

N	R_a	α	κ	γ	ζ	δ	θ_t	λ	φ	ρ_0	σ
600,000	$4M/sq.mi	0.9	0.95	0.06	2	30,000	0.3	0.5	20	10	1.5

The following sections also discuss optimal land use policies in cities with varying congestion and/or agglomeration parameters (e.g., ρ_0 and γ) for robustness analysis. The parameter settings here also guarantee that all numerical simulations reported in following sections have equilibrium solutions in cities of decreasing returns to the city population. While cities of increasing returns could exist in theory (e.g., Henderson, 1974b), they may be rarely found as stable configurations in the reality. Cities below the optimal size may be either in the growing transition by attracting workers or in the dying process by losing population (Duranton and Puga, 2013). It is more common to see cities over the efficient size, in which an increase in population leads to a decrease in utility. In addition, the spatial equilibria are solved using a nested fixed-point algorithm, as described in Appendix A3.

4 First-Best Policies in the Base Scenario

This section examines the welfare and land use effects of first-best policies, comparing them to those in the free-market equilibrium. These policies are first investigated in the base scenario with parameters in Table 2 and then in cities with varying agglomeration scales (by changing γ from 0.04 to 0.08) and congestion levels (by changing ρ_0 from 1 to 30), for robustness analysis in the next section.

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. Here, we use the combination of PCT and PLS to simulate the optimum. Table 3 shows major characteristics of free-market and first-best equilibria in the base scenario. Under the social optimum, the city needs to impose an average toll of $584 per commuter per year while delivering an annual average labor subsidy of $2,049 per job position (Table 3). This result does not imply that a combined, equivalent subsidy of $1,465 (i.e., $2,049-$584) on each worker will achieve the first-best optimum: spatial variations in tolls and labor subsidies need to be considered.

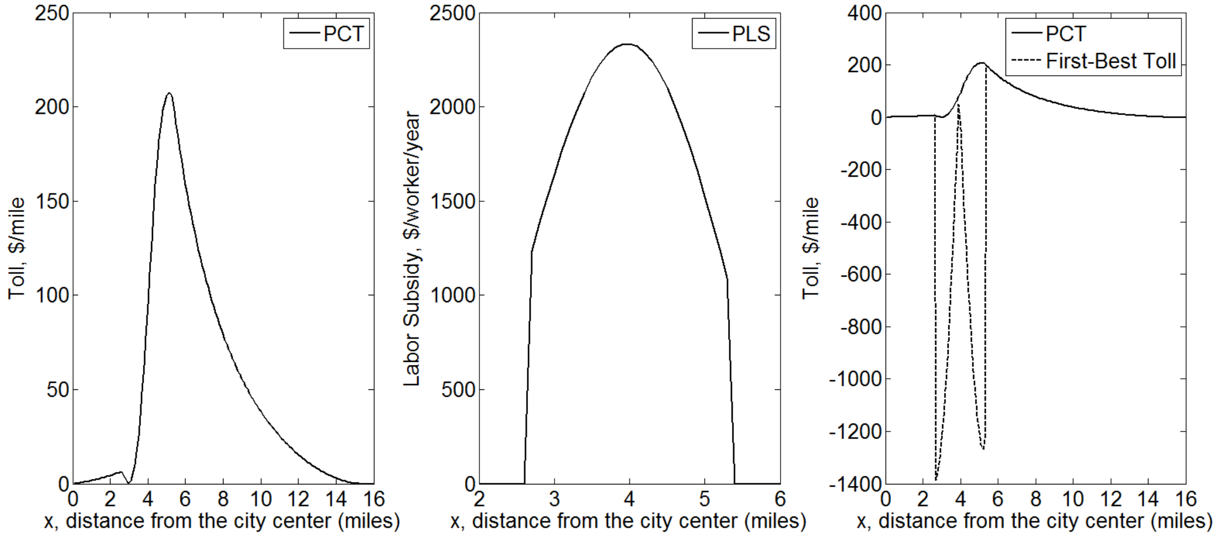


Figure 1 Levels of toll (a) and subsidy (b) under the first-best instrument combining both PCT and PLS and levels of tolls (compared to PCT-only) under the first-best tolling instrument after internalizing agglomeration externalities (c)

Figure 1a-b shows the corresponding toll and 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 (from 2.8 to 5.4 miles in radius), subsidies increase from about \$1,200 to \$2,350 per year at the locations of peak labor density and then fall to about \$1,100 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. Social optimum can be achieved by levying an optimal toll after internalizing agglomeration externalities. Figure 1c shows that the first-best toll equals the PCT in the residential areas, but varies quite a bit within the annulus of jobs, consistent with Proposition 2. The optimal toll levels across locations in the firm cluster area lie below the PCT and even become negative (thereby incentivizing such travel). These findings extend Arnott's (2007) analytical discussion, underscoring the importance of enabling spatial variation in policy interventions, in order to optimally address urban externalities.

Welfare improvement is visible under the first-best instruments. The utility level increases from 5242 to 5257, so it appears to be just 0.3% 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 \$106 per year (as a compensating variation[8]). When congestion externalities are internalized, the average commute costs rise from \$0.4 to \$0.7 per mile per day, leading to a decrease in travel demand and commute distance (which falls by 22%). The PLS allows firms to hire workers by lower labor cost (the average

[8] Given that utility levels are u_0 in the free-market case and u under a specific policy scheme, the average compensating variation (CV) is computed as $\mathrm{CV} = \frac{1}{N}\int_0^{\bar{x}} 2\pi \frac{\theta_h(x)}{q(x)}(u-u_0)\frac{dy(u,q(x))}{du}dx$. Here, $\frac{u}{c}\frac{dy}{du}$ represents the elasticity of net income with respect to utility at location x, and $(u-u_0)\frac{dy}{du}$ represents the effective income change that comes with changing utility from u to u_0.

labor cost drops by 0.4%, Table 3) and raise the average wage level in the city by about 6% (Figure 2a). The TFPs in most job locations significantly improve (Figure 2b), with the average TFP rising 0.4%. These findings suggest that after correcting congestion and agglomeration externalities, first-best instruments could simultaneously increase travel costs and wage levels and enhance citywide productivity.

Table 3 Simulated results of policy scenarios

	Free Market	**First Best**	**Change (%)**
Utility Level, $\bar{u}$	5241.86	5256.99	0.29
Avg. CV (relative to the FM case, \$/household/year)		106.31	
City Boundary, $\bar{x}$ (miles)	15.6	15.3	-1.92
Tolls, $\bar{y}_{toll}$ (\$/household/year)	0	584	
Subsidy, $\bar{y}_{\mathrm{suby}}$ (\$/worker/year)	0	2049	
Rent Revenues Returned, $\bar{y}_{rent}$ (\$/household /year)	1512	1670	10.48
Average Commute Distance (miles/day/worker)	8.16	6.37	-21.94
Average Commute Cost (\$/worker/year)	811	1102	35.95
Average TFP (compared to the constant δ)	1.81	1.82	0.39
Average Labor Cost (\$/worker/year)	32570	32435	-0.41
Average Wage Income (\$/worker/year)	32570	34484	5.88
Average Labor Density (workers/square mile)	10699	12323	15.18
Average Residential Density (households/square mile)	1270	1306	2.80
Average Rent for Firms (times R_a)	4.59	5.26	14.6
Average Rent for Housing (times R_a)	1.05	1.09	3.81

Land use patterns are also affected, 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 job density rising by 15% (in that ring, versus the original jobs zone, Table 3). 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 under the closed-form setting, 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 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.

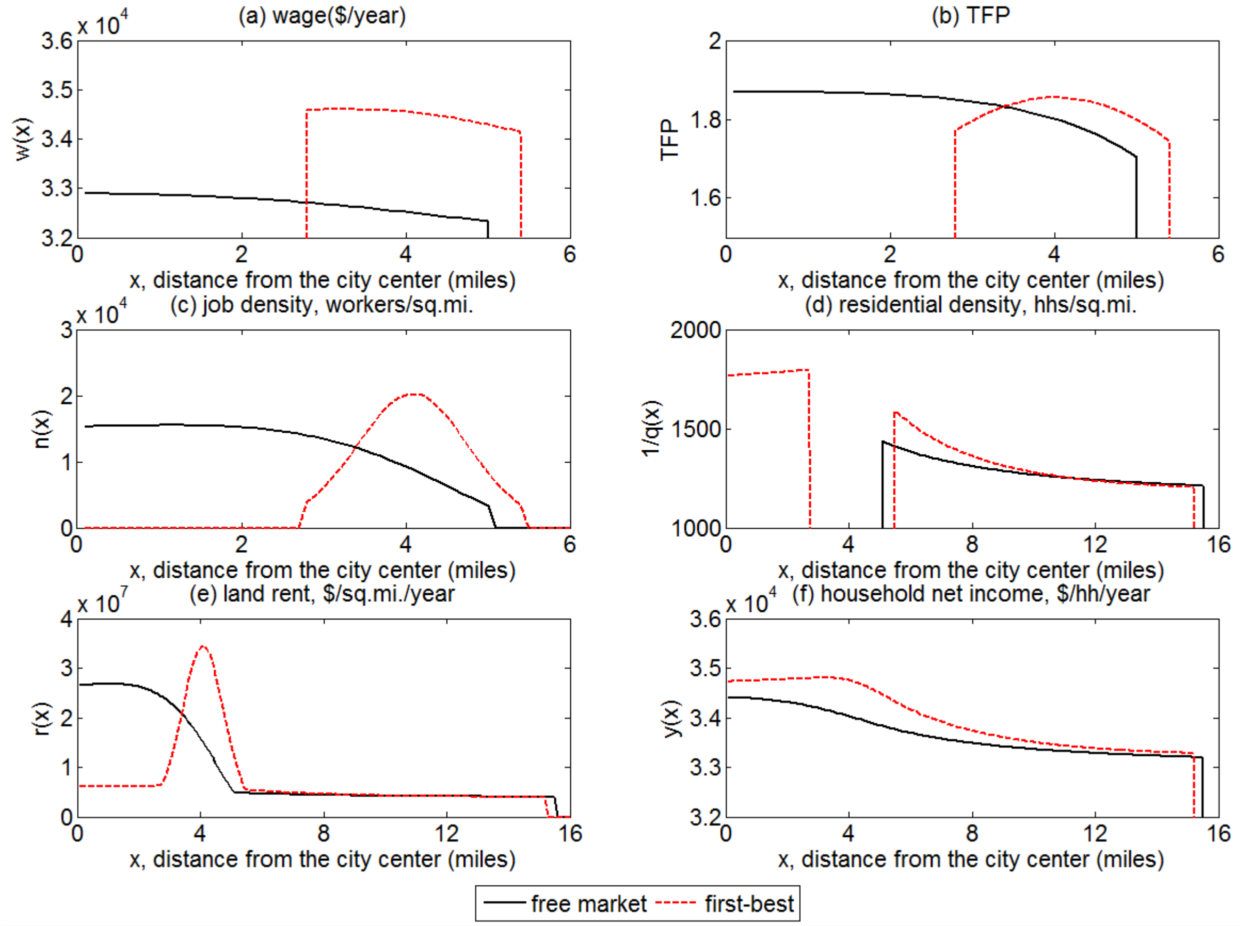


Figure 2 Spatial distribution of wages (w), TFP ($A/\boldsymbol{\delta}$), job (n) and residential densities ($1/q$), land rents (r) and net income (y) in the first-best optimum versus the free-market equilibrium in the base scenario (ρ_0=10, γ=0.06).

First-best instruments also centralize households, resulting in a shrinking city boundary (from 15.6 to 15.3 miles) and 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 about 4.6 and 1.1 times the opportunity rent (i.e., the rent at the city edge, R_a), and about 15% and 4% higher than those in the free-market equilibrium (Table 3). Given that all congestion tolls and rent revenues (net of labor subsidies) are uniformly returned to each household, net incomes rise in all locations (Figure 2f), with average net income rising by 0.7%. 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 is about five times 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).

5 Optimal Policies Varying with Congestion and Agglomeration Parameters

This section conducts a robustness analysis on the optimal policies in cities with varying agglomeration and congestion levels. First, we investigate the impact of the levels of congestion and agglomeration on welfare and urban form in the free-market scenarios. Second, we focus on the corresponding effects of first-best policies.

5.1 Free-Market Reactions to the Changes in Congestion and Agglomeration Levels

In the free-market scenarios, an increase in congestion (e.g., ρ_0 rises) or a decrease in agglomeration (e.g., γ falls) will lower social welfare (as the CV values shown in Table 4) and transform the urban spatial structure from a "FH" to a "HFH" configuration (Figure 3). Here, we define "FH" as a typical monocentric urban structure, with jobs and firms at the center and households outside the employment center, and "HFH" as a non-monocentric structure, in which housing occupies the center and the edge areas and firms agglomerate at the middle annulus[9]. These findings are generally consistent with those theoretical findings of Fujita and Ogawa (1982), Berliant et al. (2002), and Lucas and Rossi-Hansberg (2002), although these models do not consider congestion and wealth redistribution. Intuitively, the congestion cost (or travel cost) is a major centrifugal force of jobs/firms while the agglomeration benefit is a major centripetal force. The city configuration is thus determined by the interaction between congestion and agglomeration.

Simulations also support the notion that job decentralization could be driven by worsening congestion levels in free-market regions (Figure 3a and 3c) and in turn, help relieve traffic congestion, as suggested by commute times and costs in Giuliano and Small's (1991) and Crane and Chatman's (2004) empirical work. As ρ_0 increases from 1 to 30, the firm clusters in the free-market equilibrium decentralize from the locational interval (0, 4.1] to [6.5,9.3], halving the average commute distance from 8.8 to 4.2 miles. In addition, higher congestion levels may make the city area less compact. This finding differs from, or not easily detected in, a monocentric model. In the monocentric models, an increase in congestion levels can only affect households' behaviors and make them live closer to the city center, leading to a compact city size (Wheaton, 1998; Brueckner, 2007; Kono et al., 2012). However, when firms' spatial decisions are internalized, increasing congestion may not only encourage job-housing proximity but also decentralize firms, leading to an expansive city size. The city boundaries increase from 15.6 to 16 miles when ρ_0 rises from 10 to 30.

Congestion also affects the city's agglomeration, productivity, employment, and land rents. Heavy congestion significantly constrains firm agglomeration: the job density falls by 62% as ρ_0 rises from 1 to 30 (Table 4 and Figure 3). The decreasing job densities indicate that the marginal cost of agglomeration (due to a rise in congestion) has exceeded its marginal benefits. Less agglomeration leads to a decrease in the average TFP and wage levels by about 5% and 2%, respectively, from the light to heavy congestion cases. These correspond to several empirical studies, which report that traffic congestion can harm the city economy through slowing employment growth (Hymel, 2009) and reducing marginal agglomeration benefits (Graham,

[9] It is also worth to note that mixed land use could be an equilibrium solution but never *Pareto*-optimal (See detailed discussion in Appendix A4).The family of mixed urban forms is thus not this paper's focus, since we only compare *Pareto*-optimal equilibria in the policy scenarios.

2007). In addition, congestion exerts a variant impact on the bid-rents of firms and households. When the congestion level increases, firms are more willing to avoid overcrowding and their bid-rents decrease. Instead, households are more willing to pay higher rents for living closer to the firm cluster, in exchange for shorter commute and lower travel costs. The average bid-rent of households and the residential density thus increase with the congestibility parameter (Table 4).

In contrast, as the agglomeration parameter γ increases from 0.04 to 0.08, firms become increasingly centralized and more willing to locate closer to other firms for earning external benefits. A higher γ is associated with a smaller firm cluster area but a larger city size (Figure 3e and 3g). The boundaries are enlarged from 15.3 miles at $\gamma = 0.04$ to 16.8 miles at $\gamma = 0.08$ (Table 4). The agglomeration-enhancing effect also largely raises job densities (by 388%), locational productivity (by 62%), firms' bid-rents (by 650%), and wage levels (by 54%). The increase in workers' wage income thus allows them to live in larger house and pay for farther commute, leading to lower residential densities and a larger city size (Figure 3). Also, higher level of agglomeration could exacerbate congestion and raise commute costs. As γ increases from 0.04 to 0.08, the average commute distance lifts from 3.7 to 9.9 miles, an increase of 167%, while the average commute costs sharply raise by around ten times.

5.2 Free-Market *versus* First-Best

First-best policies appear to be more effective in cities with lighter congestion (i.e., smaller ρ) or higher agglomeration levels (i.e., larger γ), although this tendency is nonlinear. In the light-congestion case (ρ_0=1), the welfare gained from first-best policies are $295 per year per worker, accounting for 0.8% of the average wage income. As ρ_0 increases from 1 to 10, the welfare gains fall from $295 to $106. They respectively account for 0.8% and 0.3% of the average annual wage income. However, when ρ_0 rises from 10 to 30, the welfare gains first jump up to $162 and then drop to $132. The jump-up effect is linked with an abrupt change of urban configurations (from "FH" to "HFH" as ρ_0 rises from 10 to 15).

This finding contradicts the partial equilibrium findings relying on traditional monocentric models, which suggest that welfare gains of first-best policies are larger in cities with larger congestibility levels (e.g., Brueckner, 2007). This is because the traditional monocentric model only recognizes the negative congestion externalities but overlooks the external benefits from crowding (that causes congestion) due to the agglomeration effect. After positive agglomeration externalities are considered, the net benefit of first-best instruments equals the marginal external benefit minus the marginal external cost of crowding. First-best instruments are thus more efficient when the marginal external cost is lower (e.g., in cities with a smaller ρ_0). Similarly, first-best policies can produce more net benefits when the marginal agglomeration benefit is larger. As γ increases from 0.04 to 0.08, the CV values gained by first-best policies rise from $56 to $188, accounting for 0.2% and 0.4% of the corresponding annual wage income. Therefore, these findings demonstrate the importance to internalize firm agglomeration and it may be inappropriate for policy makers to apply optimal policies found in partial equilibrium models to improve market efficiency in cities with multiple externalities.

Compared to the free-market equilibria, first-best policies lead to more compact urban forms, regardless of the congestibility and agglomeration scales. Figure 3 compares land use differences of free-market and first-best equilibria. The city area reduces by 4% to 8% after implementing first-best policies (Table 4). Meanwhile, first-best policies significantly raise the average commute cost and wage income, along with a decrease in commute distance and an increase in citywide productivity, respectively. These are consistent with findings in the base scenario and demonstrate the nature of first-best instruments, which are used to correct the commute and labor price to reflect the full cost of driving and the full benefit of employment.

First-best policies also affect the spatial distribution of firms and jobs. In the light-congestion (ρ_0=1 and 5) or high-agglomeration case (γ=0.07 and 0.08), optimal policies centralize jobs and firms in a smaller firm cluster area in the city center (Figure 3a and 3e), causing a significant increase in job densities (Table 4). Job centralization may simultaneously reinforce the agglomeration effect and worsen congestion. But the benefit from increased agglomeration is larger than the loss due to worsened congestion, leading to a positive net benefit. In contrast, in the base scenario case (Figure 2c), optimal policies also enhance agglomeration by increasing job densities, but meanwhile, lead to a significant job decentralization. In this case, the decentralization strategy may not raise the agglomeration benefit as high as the centralization strategy while it could largely reduce travel and congestion costs. In the high-congestion (ρ_0=15 and 30) or low-agglomeration (γ=0.04 and 0.05) cases, firms in the free-market equilibria are over-decentralization and first-best policies can adjust it to be a more centralized firm cluster (Figure 3c and 3g). Accordingly, first-best policies will cluster firms in smaller areas with higher job densities and cause either job centralization or decentralization.

Similar to the base scenario, first-best policies also affect the spatial distribution of housing and land rents. Imposition of socially optimal policies can largely raise the residential densities at locations near the firm cluster (Figure 3). Percentage changes of the average residential density after imposing the optimal policies increase from 0.6% to 5% when ρ_0 rises (Table 4). These densification effects are similar to monocentric studies (e.g., Pines and Sadka, 1985; Wheaton, 1998; Kono et al., 2012). Optimal policies also cause a relatively small increase in residential densities and housing rents, compared to that in job densities and land rents for firms (Table 4).

Table 4 Policy scenario results under varying congestion and agglomeration levels

	Road Congestibility Parameter ρ_0					**Agglomeration Parameter** γ				
	30	**15**	**10**	**5**	**1**	**0.08**	**0.07**	**0.06**	**0.05**	**0.04**
Free-Market Equilibrium Outputs										
Urban Form	HFH	HFH	FH	FH	FH	FH	FH	FH	HFH	HFH
Firm Cluster Interval (miles)	[6.5,9.3]	[4,7.4]	(0,5]	(0,4.5]	(0,4.1]	(0,3.3]	(0,3.9]	(0,5]	[4.6,8.1]	[6.7,9.9]
City Boundary (miles)	16.0	15.9	15.6	15.6	15.6	16.8	16.0	15.6	15.6	15.3
CV (relative to the base case, $/hh./year)	-360	-202	0	481	979	9525	4354	0	-3578	-6606
Commute Distance (miles/day/worker)	4.24	5.79	8.10	8.52	8.83	9.91	9.06	8.10	5.10	3.71
Commute Cost ($/hh./year)	356	455	811	576	273	1540	1187	811	270	142
TFP (compared to the constant δ)	1.74	1.76	1.81	1.83	1.84	2.33	2.06	1.81	1.59	1.44
Wage/Labor Cost ($/worker/year)	32197	32345	32570	32766	32852	41245	36577	32570	29410	26805
Labor Density (workers/square mile)	5955	6838	10699	13181	15844	24317	17490	10699	5968	4981
Residential Density (hhs./square mile)	1298	1282	1271	1225	1207	1007	1135	1271	1381	1522
Rent for Firms (times R_a)	2.52	2.91	4.59	5.68	6.85	13.20	8.42	4.59	2.31	1.76
Rent for Housing (times R_a)	1.07	1.06	1.05	1.04	1.03	1.05	1.05	1.05	1.05	1.03
Free-Market versus First-Best										
CV (relative to free-market, $/hh./year)	132	162	106	141	295	188	110	106	100	56
Congestion Toll ($/year/hh.)	342	444	584	917	263	2604	1906	584	234	109
Labor Subsidy ($/year/hh.)	2028	2037	2049	2081	2087	3515	2716	2049	1542	1125
Percentage Change from the Free-Market to the First-Best Equilibria										
City Area (%)	-7.94	-8.00	-3.80	-6.73	-3.96	-4.94	-5.19	-3.80	-8.16	-6.87
Commute Distance (%)	-10.23	-11.86	-21.94	-1.64	-0.18	-1.99	-2.31	-21.94	-8.55	-2.19
Commute Cost (%)	81.65	85.14	35.95	195.34	128.37	194.46	182.49	35.95	79.14	79.23
TFP (%)	1.43	1.96	0.39	3.73	4.41	4.36	3.56	0.39	1.39	0.85
Wage Income (%)	6.02	6.04	5.88	6.91	6.96	9.73	8.19	5.88	4.87	3.86
Labor Density (%)	38.10	52.17	15.18	96.02	127.78	72.62	67.74	15.18	40.68	28.41
Residential Density (%)	3.04	2.42	2.80	2.40	0.47	3.26	2.71	2.80	2.67	1.66
Rent for Firms (%)	37.72	51.77	14.60	97.13	129.16	74.71	69.03	14.60	40.16	27.98
Rent for Housing (%)	4.59	4.82	3.81	2.57	0.63	4.48	4.57	3.81	3.10	2.22

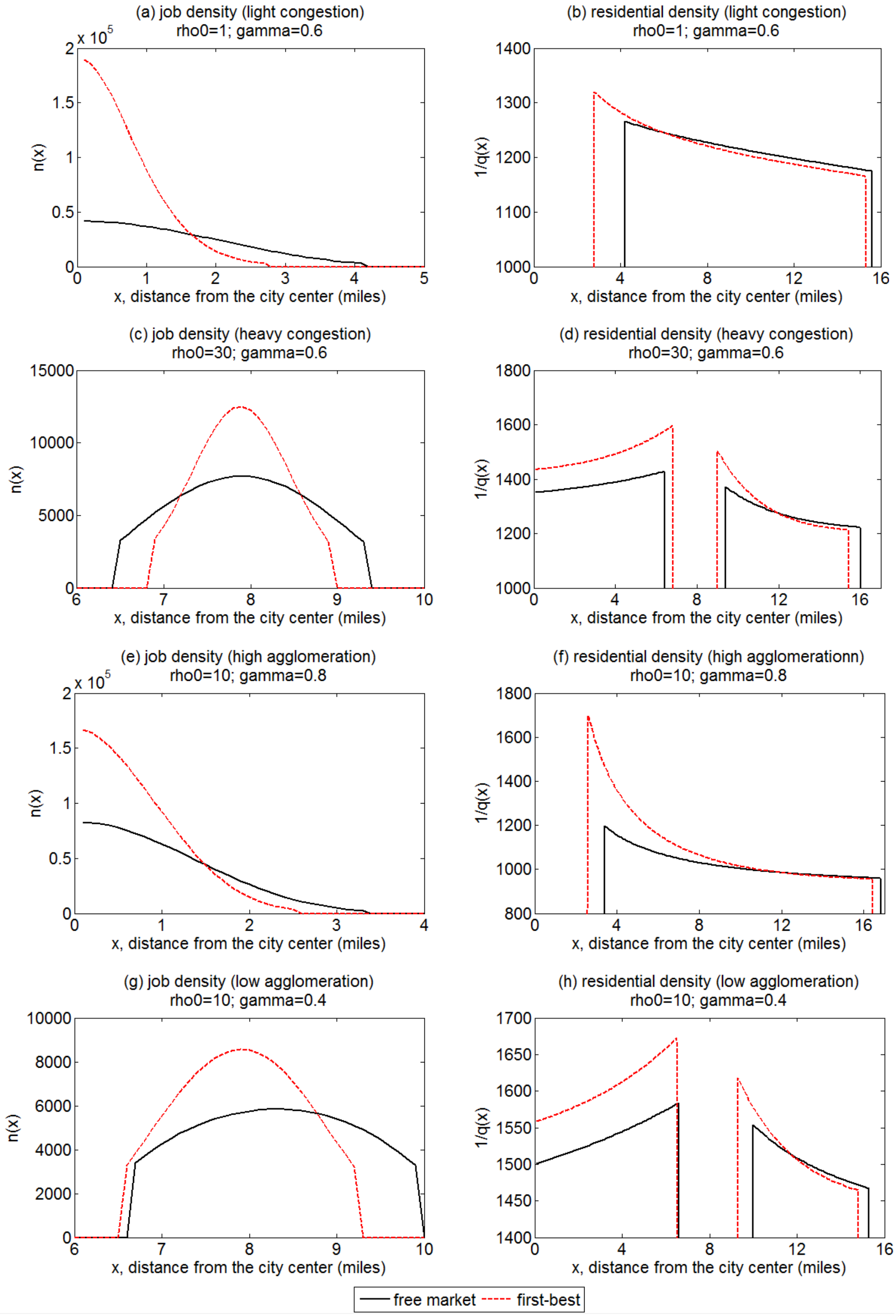
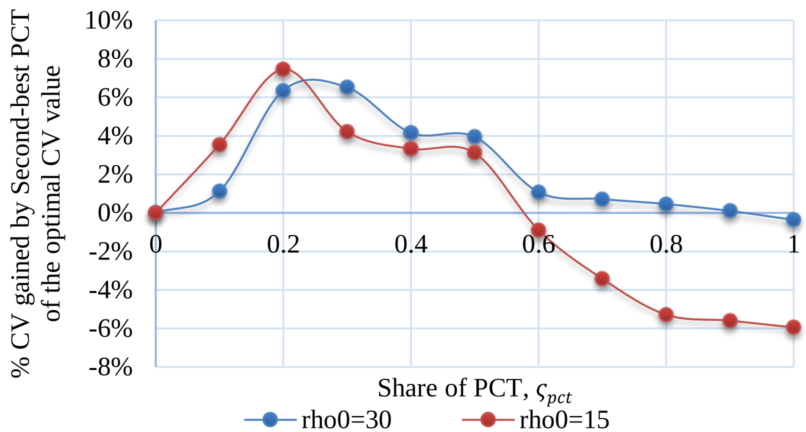
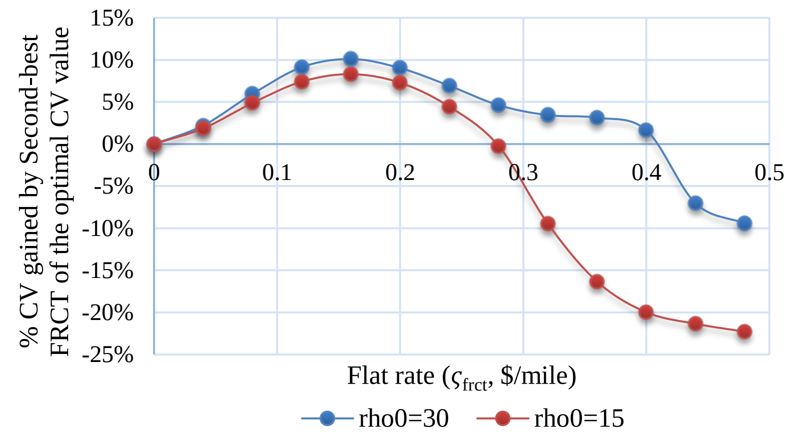


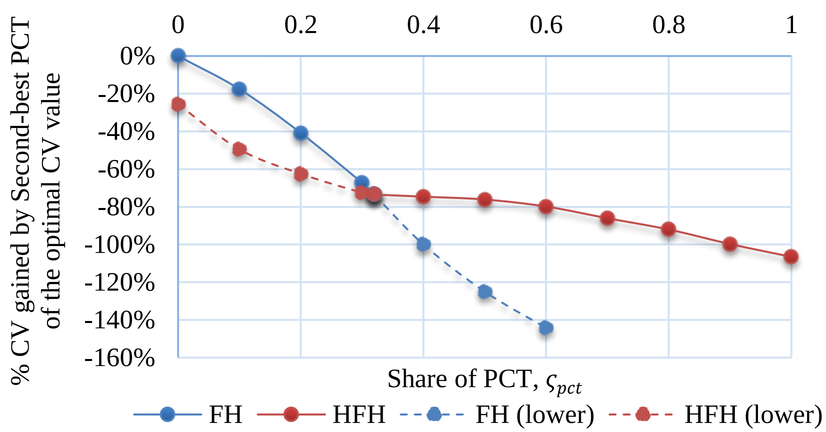
Figure 3 Spatial distribution of job density (n) and residential densities ($1/q$) in the first-best optimum versus the free-market equilibrium in different levels of congestion and agglomeration.



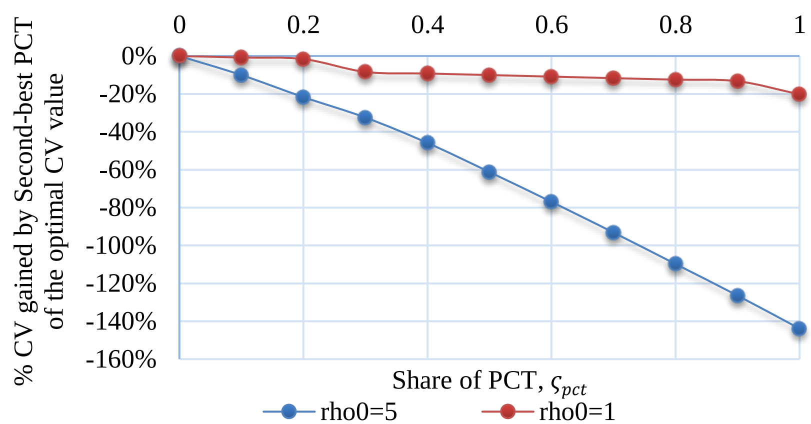
(a) PCT scenarios in relatively heavy-congestion cases



(b) FRCT scenarios in relatively heavy-congestion cases



(c) PCT scenarios in the base case (rho0=10)



(d) PCT scenarios in relatively light-congestion cases

Figure 4 Welfare effects of second-best PCT and FRCT scenarios

6 Second-Best Pricing Policies

While first-best interventions presumably are the best choice for a city authority wishing to pursue welfare improvements, they may be associated with major construction and operations costs (for variable toll collection, for example) that are not internalized in theoretical models. In addition, a combination of PCT and PLS may require much coordination between transportation agencies and departments of labor, which presents added transaction costs and political difficulties. In practice, policymakers tend to neglect the impact of anti-congestion policies on the agglomeration economy (Banister and Berechman, 2003; Litman, 2006) and rarely connect transport pricing with labor subsidies. When a city has two externalities, the efficiency of the optimal policy is no longer determined by one externality but the interaction of two. It is thus important to understand in which situation a congestion tolling-only or a subsidy-only policy is second best, how much welfare improvement could be gained by second-best tolling or subsidies, and how severe being aware of just one externality.

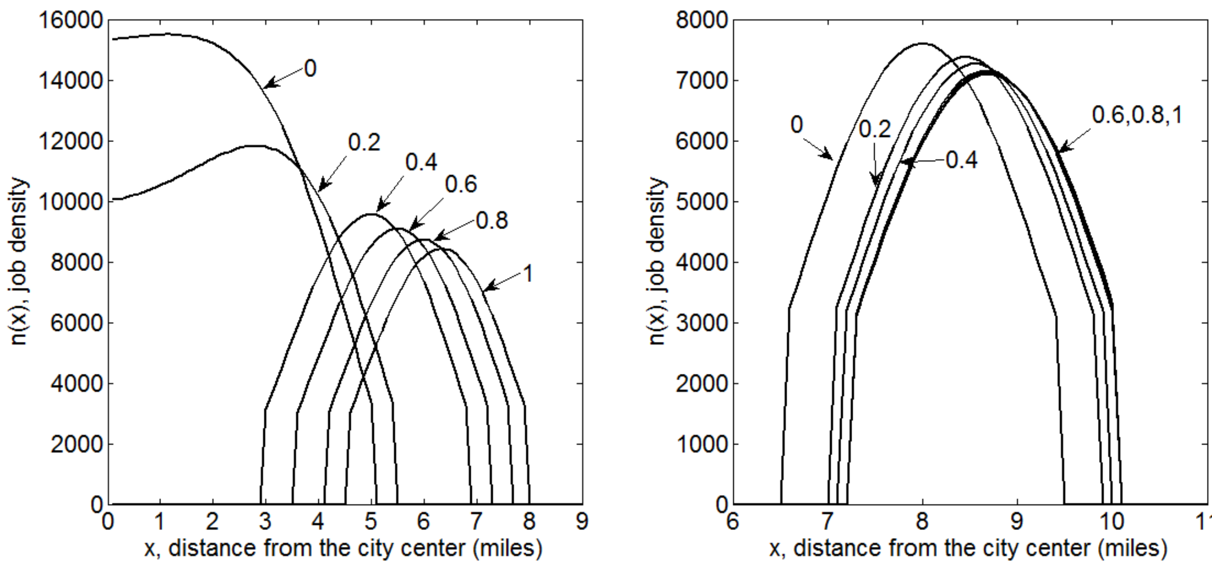
6.1 Second-Best Congestion Tolls

This section focuses on second-best congestion tolling, including a PCT-alone and an FRCT scenario. Firstly, for seeking more efficient second-best PCT-alone policies, we tracked the change in utility under ten additional tolling schemes, which impose a fixed share (ranging from 0.1 to 1) of the PCT level on each mile driven. Figure 4(a)-(c) present the percent of CV gains relative to that in the first-best optimum. The PCT-alone policies can improve social welfare only in heavy-congestion cities (e.g., ρ_0=15 and 30). For example, for ρ_0=15, the second-best utility gains peak at about 8% of the first-best utility gains (compared to the free-market equilibrium), when the toll level is set as about 20% of the PCT level. As the share increases, the welfare gain declines and even becomes negative. These findings suggest that an efficient second-best toll level, if exists, should lie below the PCT level, as agglomeration economies are internalized. For comparison, Figure 4(b) shows the percent of CV gains by imposing a flat rate of toll on each mile driven (ranging from 0 to \$0.5 per mile) in congested cities. The FRCT policies can generate a peak welfare gain when the flat rate is set at around \$0.16 per mile, and the peak welfare gain accounts for about 8-10% of optimal CV gains. In terms of the welfare effects, the second-best FRCT scheme appears similar to the second-best PCT-alone policies.

In relatively light-congestion cities, congestion-pricing policies alone may be not second best since they could produce a welfare loss than the free-market allocation (Figure 3c and 3d)[10]. Those strategies only correcting congestion externalities without considering their impacts on agglomeration may be ineffective. Figure 4(c) depicts the welfare effects of the share of PCT level ς_{pct} in the base scenario and implies the firm-decentralization process as the toll level increases. When the toll level is low (i.e., $\varsigma_{\text{pct}} \leq 0.3$), both the monocentric (i.e., “FH”) and nonmonocentric (i.e., “HFH”) configurations could be an equilibrium solution but the monocentric equilibrium has a larger level of welfare than the nonmonocentric one. As ς_{pct} increases (e.g., $\varsigma_{\text{pct}} \geq 0.4$), the rising commuting costs make the monocentric form less efficient than the nonmonocentric allocation, and thus the city becomes a “HFH” configuration (Figure 5a). The change of city configuration helps to mitigate the welfare-reduction impact of PCT-

[10] We also examined the second-best FRCT policies in the relatively light-congestion city and found that the corresponding welfare effects are similar to those of PCT-alone policies.

alone policies. Even that, the annual welfare loss by imposing a 100% PCT is \$110 per household. If the PCT-alone policies do not change the city configuration (e.g., for ρ_0=1 and 5), their negative impact on welfare appears smaller in cities with lower congestibility levels, as shown in Figure 4(d).



(a) ρ_0=10 (the base case) (b) ρ_0=30

Figure 5 Decentralization and deagglomeration impact of second-best PCT scenarios (ς_{pct} ranges from 0 to 1)

An implementation of a 100% PCT-alone instrument (i.e., ς_{pct}=1) is proven as socially optimal if a city only has congestion externalities in traditional monocentric analysis (Solow 1972; Pines and Sadka, 1985; Wheaton, 1998). Our simulations, however, suggest that the 100% PCT-alone policy is very likely to worsen, rather than improve, the citywide welfare even when the congestion condition is severe. These welfare-reduction effects mainly result from the side effect of PCT on the agglomeration economy and productivity. For example, in the 100% PCT case, the tolling policy can cause a 3-8% decrease in the average TFP and 7-16% decrease in the city wage (varying with ρ), although the congestion externalities are fully corrected.

Intuitively, we can understand the impact of PCT on agglomeration in a spatial interaction process (Figure 5). 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. 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 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. Figure 5 depicts the spatial process of job decentralization and deagglomeration along with an increase in ς_{pct}.

6.2 Second-Best Labor Subsidies

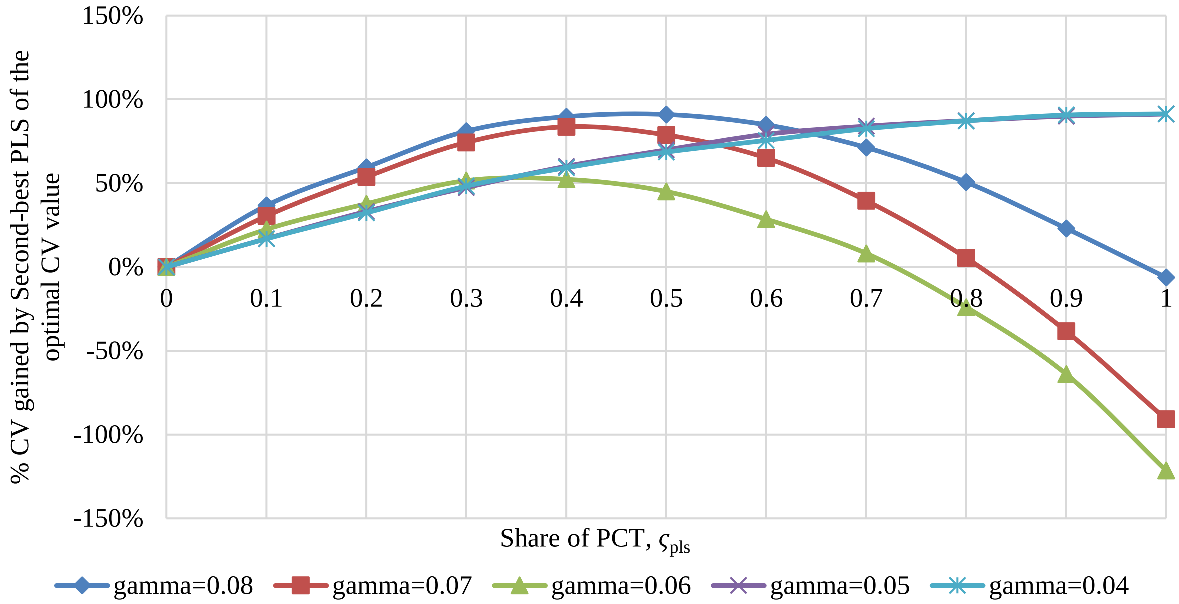
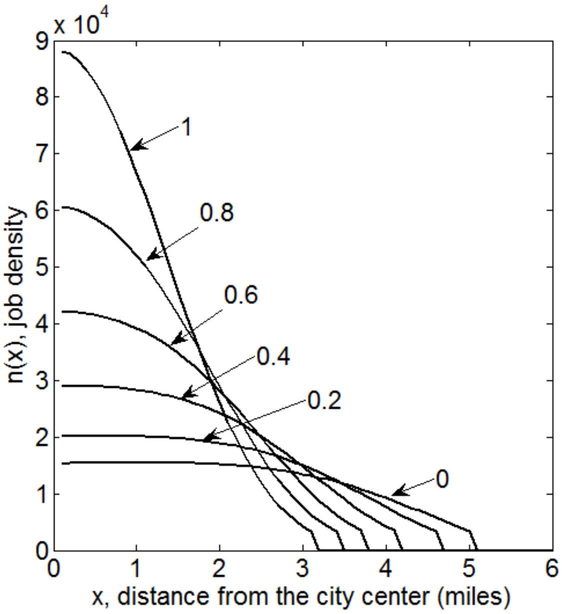


Figure 6 Welfare effects of second-best PLS scenarios

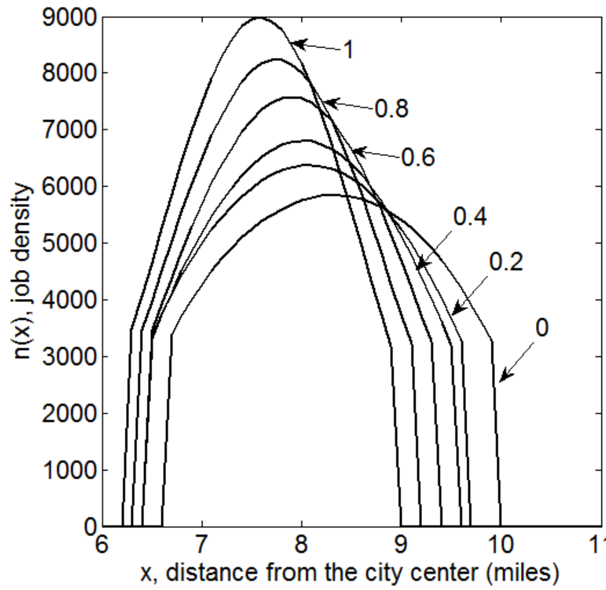
Figure 6 shows the welfare effects of second-best PLS schemes with different levels of agglomeration scale. Compared to second-best PCT policies, the PLS-alone policies are a more efficient alternative to first-best instruments. The welfare gains from second-best PCT policies may account for about 52% to 91% of that from first-best instruments. In cities with low-agglomeration scales (γ=0.04 and 0.05), the welfare improvement is larger as the share of the PLS level increase. The 100% PCT-alone scheme can generate 90% of the CV gains of the first-best level. In contrast, in cities with large-agglomeration scales (γ=0.06 to 0.08), the policies with the labor subsidy level setting below the PLS level may generate more welfare gains than those at the exact Pigouvian level (i.e., ς_{pls}=1). For example, in the base scenario case (γ=0.06), the utility gains relative to the first-best level peak at 52%, when the labor subsidy is set at about 30% of the PLS level. The welfare improvement increases and reaches up to 91% of the first-best level as γ rises. In these cases, the 100% PLS-alone policy could cause a welfare loss, although it fully corrects the market distortion due to the existence of agglomeration externalities. These findings remind policy makers to take into consideration that only correcting for one externality in cities with multiple externalities may achieve low, or even negative, welfare gains. An optimal policy should internalize multiple externalities.

Spatially, the PLS-alone policy produces a more compact firm cluster than the free-market equilibrium (Figure 7). Without the PCT's congestion correction, the PLS-alone intervention could encourage firms to locate closer to each other. After levying a PLS policy, firms at a location with relative low productivity (often at the edges of firm clusters) will move to a location with higher productivity. This tendency agglomerate firms in a smaller area, and job densities increase near the centerpoint and drop at the edge of the firm cluster, as shown in Figure 7. The traffic volumes will thus rise within the firm cluster, triggering a rise in congestion.

Simulations show that the PLS-alone policy links with a 3-5% increase in average productivity, a 4-6% decrease in average labor cost, and a 6-9% increase in average commute distance.



(a) γ=0.06 (the base case)



(b) γ=0.04

Figure 7 Centralization and agglomeration impact of second-best PLS scenarios (ς_{pls} ranges from 0 to 1)

7 Second-Best Place-Based Policies

This section turns to evaluate the second-best place-based policies, including urban growth boundary (UGBs) and firm cluster zoning (FCZ) regulation. For each scenario, we focus on the optimal UGB or FCZ that maximizes the citywide utility level. The search process of an optimal place-based policy is introduced in Appendix A5. Table 5 shows simulated impacts of optimal UGBs and FCZ on welfare and land use using different parameters of congestion and agglomeration levels.

7.1 Urban Growth Boundaries

According to Table 5, optimal UGB policies can improve citywide welfare, but the welfare gains are small. The UGB CV values are about \$9 to \$14 per household per year, ranging from 3% to 22% of the first-best optimum. The UGB policies appear more effective (compared to the optimum level) in the cities with lower agglomeration scales or higher congestion levels. Under the base scenario, the CV of the UGB policy relative to the free-market case is above 9% of the first-best CV level. For γ= 0.04, the UGB's CV value accounts for 22% of the corresponding first-best level, although it is only about \$12 per household per year, lower than 0.05% of the average annual wage income.

The UGB equilibrium could cause worse land market distortion than the free-market equilibrium. Figure 8 compares the spatial patterns of job and residential densities and land rents for firm and residential use in the UGB, first-best, and free-market equilibria under varying agglomeration parameters. The UGB policies could largely raise residential densities and escalate residential rents over the optimum levels at most locations. The average residential rents under the optimal UGB policies are above 50% larger than first-best instruments (Table 5). For example, the average residential rent in the UGB equilibrium is three and eight times that of the corresponding first-best optimum in the city with heavy congestion (ρ_0=30) and light congestion (ρ_0=1), respectively. In addition, UGBs can slightly agglomerate firms, leading to a trivial increase in productivities. As shown in Figure 8, however, the increases in job densities and firms' rents caused by the optimal UGB policies are much smaller than those by first-best instruments (accounting for about 7-25%). Thus, restrictive UGBs cannot effectively reduce the distortion of firm's land market and excessively raise residential densities and rents. These consequences may explain why the optimal UGB regulation gains a relatively low welfare improvement.

7.2 Firm Cluster Zoning

The optimal FCZ policies are more effective than the UGB policies. The annual CV values gained by the FCZ policies range from \$45 to \$166 per household. Taking the base scenario as an example, the CV value of the optimal FCZ equilibrium relative to the free-market case is 62% of the first-best welfare gain, about seven times the UGB level. The lower the agglomeration parameter γ or the larger the congestibility parameter ρ_0, the larger the welfare improvement the FCZ policies can achieve. For γ=0.04, the FCZ welfare gain is up to 80% of the first-best level and about four times the UGB level. For ρ_0=30, the FCZ welfare gain is about 70% of the first-best level and eight times the UGB level. These findings show that the FCZ regulation is a better place-based policy than UGB for correcting both agglomeration and congestion externalities. However, such an effective policy appears less discussed in literature. The major reason is probably related to the fact that many urban economic studies remain heavily reliant on monocentric models, failing to explore firms' spatial reactions to optimal policies.

Differing from the UGB policy, the FCZ policy can raise the locational productivity and correct the distortions in the firms' land market and cause fewer distortions in the housing market. The average job densities and commercial rents in the FCZ equilibria are very close to the corresponding optimum levels, regardless of ρ_0 and γ. According to Figure 9, the spatial distribution of job densities and commercial rents in the FCZ equilibrium are similar to the first-best optimum, and the distributions of residential densities and rents are similar to the free-market equilibrium. While FCZ policies have more sensitive impact on firm's spatial decision and land market, they will not cause an excessive escalation in residential rents as UGB policies do.

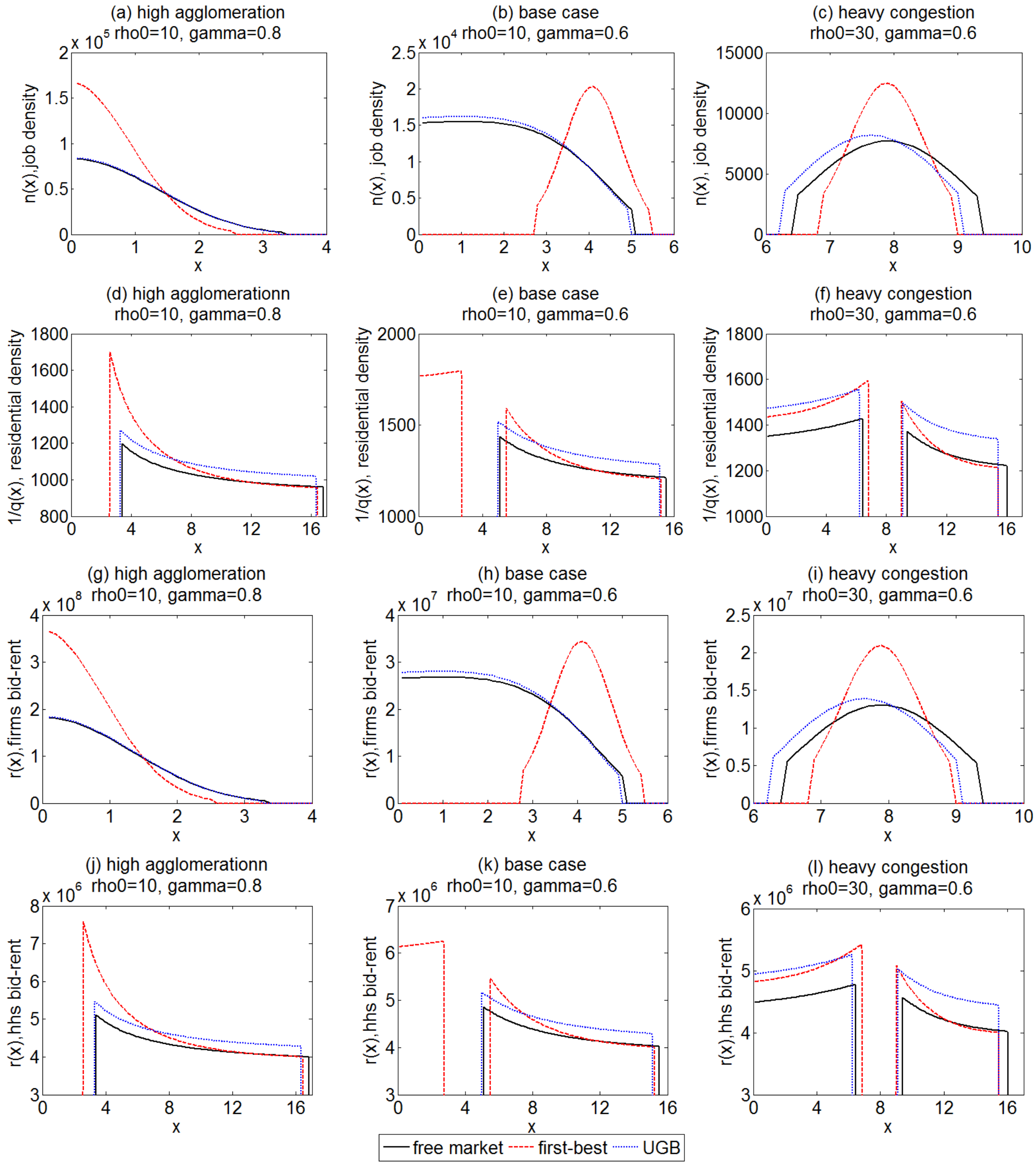


Figure 8 Spatial distributions of job and residential densities and land rents for firm and residential use in the UGB, first-best, and free-market equilibria varying between the FH(i.e., monocentric, Left) and HFH (i.e., nonmonocentric, Right) urban forms

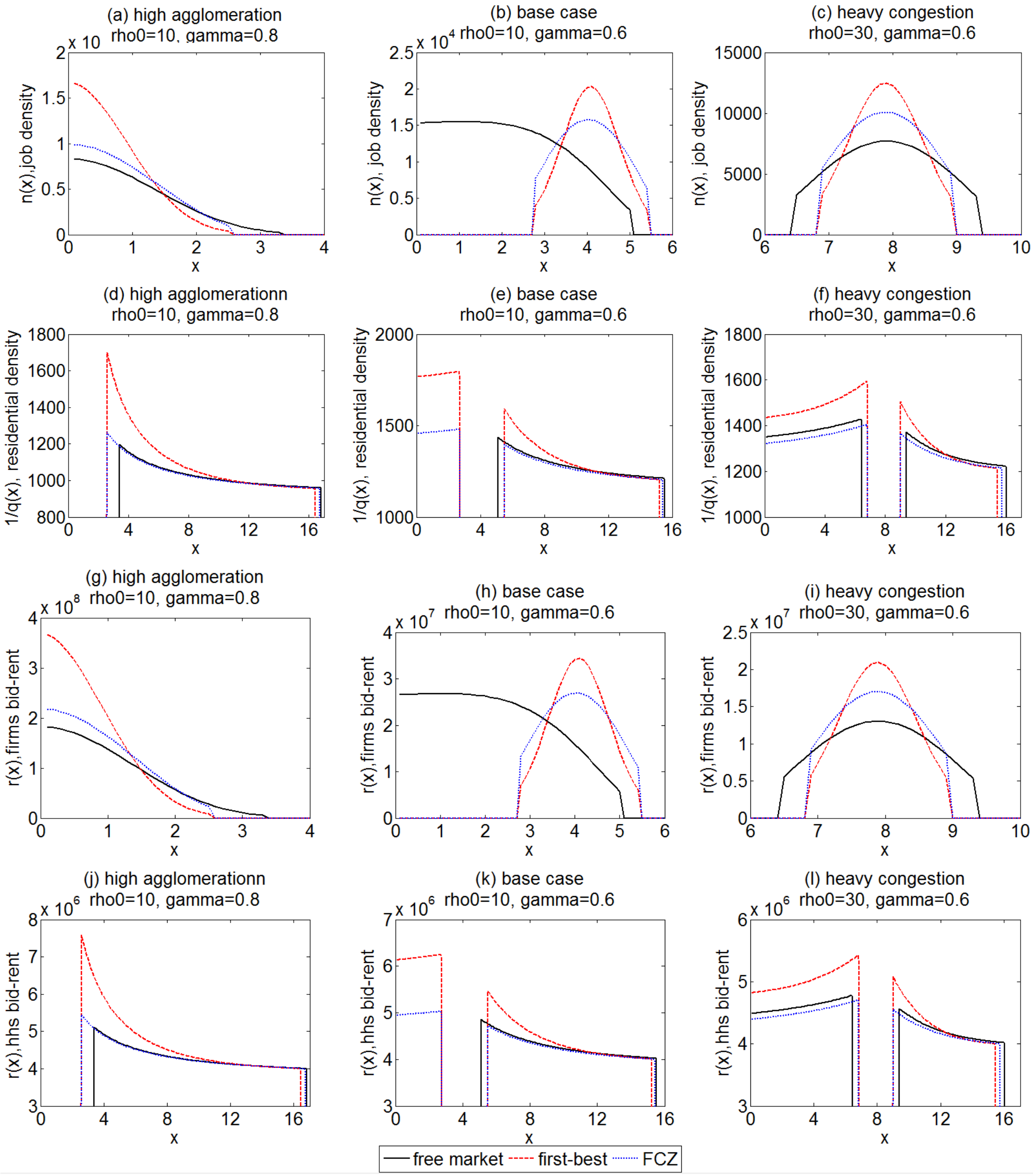


Figure 9 Spatial distributions of job and residential densities and land rents for firm and residential use in the FCZ, first-best, and free-market equilibria varying with the FH (i.e., monocentric, Left) and HFH (i.e., nonmonocentric, Right) urban forms

8 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 pricing and land use policies, such as urban growth boundary (UGB) and firm cluster zoning (FCZ) 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, and internalize spatial interactions 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. The exact PCT-alone or PLS-alone policies could be the optimal policies in the partial equilibrium model that internalizing congestion or agglomeration externalities only. But in many realistic cities with both externalities, the PCT- or PLS- alone policies could lead to significant land market distortions and 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 first-best toll (or labor subsidy) lies below its related marginal externality cost (or benefit), as also found in Arnott (2007) and Thissen et al.'s (2011) empirical analysis in 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 in theory. 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 Borck and Wrede (2009).

We also examine the second-best pricing policy if only one instrument is adopted. The congestion pricing alone policies (a partial PCT or a flat-rate toll) can improve social welfare only in heavy-congestion cities while their welfare gains could be trivial (e.g., below 10% of the welfare improvement achieved by first-best policies). The second-best toll should be set below the PCT level (e.g., 20-30% of the Pigouvian level). The congestion toll-only policies may cause significant welfare loss especially in cities with low congestibility. The inefficiency of second-best tolling policies primarily results from the over-decentralization and deagglomeration effects of congestion pricing on firms and jobs, causing lower productivity and wage level. In contrast, the second-best labor subsidy alone policy is a more effective alternative to first-best policies. PLS-alone policies can enhance the agglomeration economy and raise productivity as first-best policies do, although in some cases, they may generate an overcrowding employment cluster and thus worsen traffic congestion.

The UGB regulations may partially correct distortions in both transport and labor markets, but may worsen land market distortion via the residential rent-escalation effects, leading to trivial utility gains. 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). Of course, real cities are much more complex than the models allowed here use. Human health, ecological conservation, social interaction, and other variables are at play and may counteract some or much of the rent escalation losses that tend to come with tight UGBs. In contrast, FCZ policies, enforced by regulating a zone's land use exclusively for firm/business use, are more effective than UGB policies for reducing congestion and enhancing agglomeration. They can generate welfare improvement closer to the first-best levels by effectively regulating firm's locations and do not result in excessive escalation of housing rents, avoiding the housing affordability issue raised by UGB policies. Planning practice should pay more attention to such an effective land use policy, and urban economics models should allow for land use decision scenarios related to firms.

Multiple opportunities exist to make these models more realistic. Since urban spatial structures regularly depend on the specification of agglomeration externalities (the function $F(x)$), future investigations should seek to compare results from different specification assumptions and endogenize the generation of such agglomeration externalities. For example, several studies have modeled locational agglomeration externalities as a consequence of distance-decay knowledge spillovers and firms investment decisions on innovation (e.g., Berliant et al., 2002; Desmet and Rossi-Hansberg, 2014). Also, 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, such as a vehicle-miles-traveled (VMT) tax and a cordon toll (Zhang and Kockelman, 2016) and the investment on public transit, 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.

Acknowledgements

The authors thank the editor Gilles Duranton and two anonymous reviewers for their valuable comments, Alex Anas, Jan Brueckner, and Esteban Rossi-Hansberg for their inspired questions and comments in an early version of this article, and Scott Schauer-West and Annette Perrone for their editing and administrative support.

References

Anas, A., 2012. The optimal pricing, finance and supply of urban transportation in general equilibrium: A theoretical exposition. *Economics of Transportation*, 1: 64–76.

Anas, A. and Kim, I. 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *Journal of Urban Economics*, 40, 232–256.

Anas, A. and Rhee, H. 2006. Curbing excess sprawl with congestion tolls and urban boundaries. *Regional Science and Urban Economics*, 36, 510–541.

Anas, A. and Xu, R. 1999. Congestion, land use and job dispersion. *Journal of Urban Economics*, 45, 451–473.

Anderson, S.P., de Palma, A., 2004. The economics of pricing parking. *J. Urban Econ.*55, 1–20.

Arnott, R. J. 1979. Unpriced transport congestion. *Journal of Economic Theory*, 21(2), 294-316.

Arnott, R. 2007. Congestion tolling with agglomeration externalities. *Journal of Urban Economics*, 62 (2), 187–203.

Arnott, R., de Palma,A., Lindsey,C.R.,1991.A temporal and spatial equilibrium analysis of commuter parking. *J. Public Econ.* 45,301–335.

Arnott, R., Inci,E.,2006.An integrated model of downtown parking and traffic congestion. *J. Urban Econ.* 60,418–442.

Arnott, R., Rowse, J.,1999. Modeling parking. *J. Urban Econ.* 45,97–124.

Arzaghi, M. and Henderson, J. V. 2008. Networking off madison avenue. *The Review of Economic Studies*, 75(4), 1011-1038.

Banister, D. and Berechman, J., 2003. *Transport investment and economic development*. Routledge., J., 2003.

Behrens, K. and Robert-Nicoud, F. (2014). Agglomeration theory with heterogeneous agents. in G. Duranton, J.V. Henderson, and W. Strange (eds.) *Handbook of Regional and Urban Economics*, Volume 5. Amsterdam: Elsevier.

Berliant, M., Peng, S., and Wang, P., 2002. Production externalities and urban configuration. *Journal of Economic Theory,* 104, 275–303.

Borck, R., & Wrede, M. (2009). Subsidies for intracity and intercity commuting. *Journal of Urban Economics*, 66(1), 25-32.

Boucekkine, R., C. Camacho, and Zou, B. 2009. Bridging the Gap Between Growth Theory and the New Economic Geography: The Spatial Ramsey Model, *Macroeconomic Dynamics*, 13, 20–45.

Brinkman, J. C. 2013. Congestion, Agglomeration, and the Structure of Cities. Federal Reserve Bank of Philadelphia Working Paper Number 13-25. May 23.

Brueckner, J. 2007. Urban growth boundaries: an effective second-best remedy for unpriced traffic congestion? *Journal of Housing Economics*, 16, 263-273.

Cheshire, P. and Sheppard, S. 2005.The introduction of price signals into land use planning decision-making: a proposal[online]. London: LSE Research Online.Available at: http://eprints.lse.ac.uk/archive/00000568.

Cho, S. H., Poudyal, N., and Lambert, D. M. 2008. Estimating spatially varying effects of urban growth boundaries on land development and land value. *Land Use Policy*, 25 (3), 320-329.

Cox, W. 2010. Property values 11 times higher across Portland's urban growth boundary. http://www.newgeography.com/content/001808-property-values-11-times-higher-across-portlands-urban-growth-boundary, 2013, July.

Combes, P. P., Duranton, G., Gobillon, L., & Roux, S. 2010. Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics* (pp. 15-66). University of Chicago Press.

De Lara, M., De Palma, A., Kilani, M. and Piperno, S., 2013. Congestion pricing and long term urban form: Application to Paris region. *Regional Science and Urban Economics*, 43(2), pp.282-295.

Dong, X. and Ross, S.L., 2015. Accuracy and efficiency in simulating equilibrium land-use patterns for self-organizing cities. *Journal of Economic Geography*, 15 (4): 707-722.

Downs, A. (Ed.). 2004. *Growth management and affordable housing: Do they conflict*. Brookings Institution Press.

Desmet, K. and E. Rossi-Hansberg. 2010. On spatial dynamics, *Journal of Regional Science*, 50, 43-63.

Desmet, K. and Henderson, J. V., 2014. The Geography of Development within Countries. CEPR Discussion Paper No. DP10150. Available at SSRN: http://ssrn.com/abstract=2501633

Desmet, K. and Rossi-Hansberg, E., 2014. Spatial development. *The American Economic Review*, 104(4), pp.1211-1243.

Duranton, G. and Puga, D., 2004. Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics*, 4, 2063-2117.

Duranton, Gilles and Puga, Diego, The Growth of Cities (August 2013). CEPR Discussion Paper No. DP9590. Available at SSRN: http://ssrn.com/abstract=2309234

Duranton, G., Puga, D. 2014. Urban land use. Discussion Paper No. 10282, Centre for Economic Policy Research, London, UK.

Eeckhout, Jan. 2004. Gibrat's law for (All) cities. American Economic Review 94(5):1429–1451.

Fujita, M. and Ogawa, H. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12, 161–196.

Fujita, M. and Thisse, J. 2002. *Economics of Agglomeration*. Cambridge Univ. Press.

Glaeser, E. and Kahn, M. 2001, Decentralized employment and the transformation of the American city, NBER Working Paper 8117.

Graham, D.J., 2007. Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, 62(1), pp.103-120.

Henderson, J. V., 1974a. Road congestion: A reconsideration of pricing theory, *Journal of Urban Economics*, 1(3), 346-365.

Henderson, J V, 1974b. The Sizes and Types of Cities, *American Economic Review*, 64(4), 640-456

Hymel, K. 2009. Does traffic congestion reduce employment growth?. *Journal of Urban Economics*, 65(2), 127–135.

Inci, E., 2015. A review of the economics of parking. *Economics of Transportation*, 4(1), pp.50-63.
Kanemoto, Y. 1977. Cost–benefit analysis and the second best land use for transportation. *Journal of Urban Economics* 4: 483–503.
Kono, T., Joshi, K.K., Kato, T., and Yokoi, T. 2012. Optimal regulation on building size and city boundary: An effective second-best remedy for traffic congestion externality. *Regional Science and Urban Economics*, 42, 619–630.
Litman, T. 2006. What's it worth? Life cycle and benefit/cost analysis for evaluating economic value, Victoria Transport Policy Institute (2006) Available at: www.vtpi.org/worth.pdf
Lucas, R.E. Jr., Rossi-Hansberg, E., 2002. On the internal structure of cities. *Econometrica*, 70 (4), 1445–1476.
Mohring, H., Harwitz, M., 1962. *Highway Benefits: An Analytical Framework*. Northwestern University Press, Evanston, Illinois.
Nelson, Arthur C., Raymond J. Burby, Edward Feser, Casey J. Dawkins, et al. 2004, Urban containment and central-city revitalization. *Journal of the American Planning Association* 70: 411-425.
Neumark, D. and Simpson, H., 2014. Place-based policies (No. w20049). National Bureau of Economic Research.
Ogawa, H. and Fujita, M., 1980. Equilibrium land use patterns in a nonmonocentric city. *Journal of regional science*, 20(4), pp.455-475.
Parry, I. W., & Bento, A. (2001). Revenue recycling and the welfare effects of road pricing. *The Scandinavian Journal of Economics*, 103 (4), 645-671.
Pigou, A.C., 1920. *The Economics of Welfare*. Macmillan, London.
Pines, D. and Sadka, E. 1985. Zoning, first-best, second-best and third-best criteria for allocating land for roads. *Journal of Urban Economics*, 17, 167-183.
Puga, D. 2010. The magnitude and causes of agglomeration economies, *Journal of Regional Science*, 50 (1), 203-219.
Rappaport, J., 2014. Monocentric city redux. Federal Reserve Bank of Kansas City Working Paper, (14-09).
Rossi-Hansberg, E. 2004. Optimal urban land use and zoning. *Review of Economic Dynamics*, 7, 69–106.
Rosenthal, S. and Strange, W.C. 2004. Evidence on the nature and sources of agglomeration economies. In Henderson, V. and Thisse, J-F (eds.), *Handbook of Regional and Urban Economics*, V4:2119-2174.
Rosenthal, S. S., and Strange, W. C. 2008. The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2), 373-389.
Shoup, D.C., 2005. *The High Cost of Free Parking*. American Planning Association, Chicago.
Small, K. 1983. The incidence of congestion tolls on urban highways. *Journal of Urban Economics* 13: 90–111.
Solow, R.M. 1972. Congestion, Density and the Use of Land in Transportation. *The Swedish Journal of Economics*, 74, 161-173.
Staley, S., Edgens, J. G., & Mildner, G. C. 1999. *A line in the land: Urban-growth boundaries, smart growth, and housing affordability.* Los Angeles: Reason Public Policy Institute.
Texas A&M Transportation Institute (TTI). 2012. *Congestion data for your city*. http://mobility.tamu.edu/ums/congestion-data/

Thissen, M., Limtanakool, N. and Hilbers, H., 2011. Road pricing and agglomeration economies: a new methodology to estimate indirect effects applied to the Netherlands. *The Annals of Regional Science*, 47(3), pp.543-567.

U.S. Department of Transportation (USDOT). 2009. *Transit and Congestion Pricing: A Primer.* At http://ops.fhwa.dot.gov/publications/fhwahop09015/ fhwahop09015.pdf

Verhoef, E.T., 2005. Second-best congestion pricing schemes in the monocentric city. *Journal of Urban Economics* 58: 367–388.

Vickrey, W.S., 1963. Pricing in urban and suburban transport. *American Economic Review* 53, 452–465.

Vickrey, W.S., 1969. Congestion theory and transport investment. *American Economics Review*, 59, 251–260.

Wheaton, W.C. 1998. Land use and density in cities with congestion. *Journal of Urban Economics* 43, 258-272.

Wheaton, W.C. 2004. Commuting, congestion, and employment dispersal in cities with mixed land use. *Journal of Urban Economics*, 55, 417–438.

Zhang, W. and Kockelman, K.M., 2016. Congestion pricing effects on firm and household location choices in monocentric and polycentric cities. *Regional Science and Urban Economics*, 58, pp.1-12.

Appendix

A1: Proof of Proposition 1

(a) Since utility maximization and expenditure minimization are 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, $c^*(x, x_w) = c^*\big(y(x, x_w)\big) \equiv c^*\big(y(x)\big) = c^*(x)$, and $q^*(x, x_w) = q^*\big(r_h(x), y(x, x_w)\big) \equiv c^*\big(r_h(x), y(x)\big) = c^*(x)$.

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

(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 is 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 $n(x), q(x), c(x), \theta_f(x), F(x)$ at each location x $(0 \le x \le \bar{x})$ so as to maximize

$$u\big(c(x), q(x)\big)$$

subject to

(A1) $$\int_0^{\bar{x}} \left\{2\pi x \left[\theta_f(x)\delta n(x)^\kappa F(x)^\gamma - \frac{\theta_h(x)}{q(x)} c(x) - (1 - \theta_t)R_A\right] - t(x)D(x)\right\} dx \ge 0$$

(A2) $$\theta_h(x) + \theta_f(x) + \theta_t = 1$$

(A3) $$F(x) = \zeta \int_0^{\bar{x}} \int_0^{2\pi} r\theta_f(r)n(r)e^{-\zeta l(x,r,\psi)} d\psi dr$$

(A4) $$|t(x)| = \varphi + \rho\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma$$

(A5) $$D'(x) = 2\pi x\left(\frac{\theta_h(x)}{q(x)} - \theta_f(x)n(x)\right)$$

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

(A6) $$D(0) = 0 \ \ and \ \ D(\bar{x}) = 0$$

(A7) $$r(\bar{x}) = R_A$$

(A8) $$\int_0^{\bar{x}} 2\pi x \frac{\theta_h(x)}{q(x)} dx = N$$

Equations (A1)-(A8) are present in the body text of this paper, with the exception of constraint (A1), 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:

$$H\big(x;n,F,q,c,\theta_f,t,D,\beta_1,\beta_2,\beta_3\big)$$
$$= \lambda(x)u\big(c(x),q(x)\big) + 2\pi x\left[\theta_f(x)\delta n(x)^\kappa F(x)^\gamma - \frac{1-\theta_t-\theta_f(x)}{q(x)}c(x) - (1-\theta_t)R_A\right]$$
$$- t(x)D(x) + \left(\beta_1(x)F(x) - \beta_1(x)\zeta\int_0^{\bar{x}}\int_0^{2\pi} r\theta_f(r)n(r)e^{-\zeta l(x,r,\psi)}d\psi dr\right)$$
$$+ \beta_2(x)\left(t(x) - \varphi - \rho\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma\right) + \ \beta_3(x)2\pi x\left(\frac{1-\theta_t-\theta_f(x)}{q(x)} - \theta_f(x)n(x)\right)$$

From the conditions of the maximum principle, some of the first-order conditions are derived as:

(A9) $\frac{\partial H}{\partial n} = \frac{\partial H(x)}{\partial n(x)} + \int_0^{\bar{x}}\frac{\partial H(r)}{\partial n(x)}dr = 2\pi x\theta_f(x)[\delta\kappa n(x)^{\kappa-1}F(x)^\gamma - \beta_3(x)] - x\theta_f(x)\zeta\int_0^{\bar{x}}\int_0^{2\pi}\beta_1(r)e^{-\zeta l(r,x,\psi)}\,d\psi dr = 0$

(A10) $\frac{\partial H}{\partial F} = 2\pi x\theta_f(x)\gamma\delta n(x)^\kappa F(x)^{\gamma-1} + \beta_1(x) = 0$

(A11) $\frac{\partial H}{\partial D} = -\beta_3{}'(x) \ \rightarrow\ \beta_3{}'(x) = t(x) + \rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma$

From (A9) and (A10), one can obtain the following relationship in the firm cluster:

(A12) $\delta\kappa n(x)^{\kappa-1}F(x)^\gamma = \beta_3(x) - \gamma\delta\zeta\int_0^{\bar{x}}\int_0^{2\pi} r\theta_f(r)n(r)^\kappa F(r)^{\gamma-1}e^{-\zeta l(r,x,\psi)}\,d\psi dr$

When firms' profits are maximized, from Eq. (16), one can derive the following:

(A13) $\delta\kappa n(x)^{\kappa-1}F(x)^\gamma = w(x) - s(x)$

In a socially optimal city, both conditions (A12) and (A13) should be satisfied. Thus,

(A14) $\beta_3(x) = w(x) - s(x) + \gamma\delta\zeta\int_0^{\bar{x}}\int_0^{2\pi} r\theta_f(r)n(r)^\kappa F(r)^{\gamma-1}e^{-\zeta l(r,x,\psi)}\,d\psi dr$

Comparing the first-order condition (A11) and Eq.(A14), one can derive the following equations:

(A15) $w'(x) = \left(s(x) - \gamma\delta\zeta\int_0^{\bar{r}}\int_0^{2\pi} r\theta_f(r)n(r)^\kappa F(r)^{\gamma-1}e^{-\zeta l(x,r,\psi)}\,d\psi dr\right)' + t(x) + \rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma$

When household's utility is maximized, from Proposition 1d and (A15), one can obtain the following relationship:

(A16) $\tau(x) - \rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma = \left(s(x) - \gamma\delta\zeta\int_0^{\bar{r}}\int_0^{2\pi} r\theta_f(r)n(r)^\kappa F(r)^{\gamma-1}e^{-\zeta l(x,r,\psi)}\,d\psi dr\right)'$

In order to fulfill Eq. (A16) for each location *x*, we have three strategies:

(a) A combination of two instruments:

(A17) $$\begin{cases}\tau(x) = \tau_{pct}(x) = \rho\sigma\left(\frac{|D(x)|}{2\pi x\theta_t}\right)^\sigma \\ s(x) = s_{pls}(x) = \gamma\delta\zeta\int_0^{\bar{r}}\int_0^{2\pi} r\theta_f(r)n(r)^\kappa F(r)^{\gamma-1}e^{-\zeta l(x,r,\psi)}\,d\psi dr,\ \ if\ \ \theta_f(x) > 0\end{cases}$$

(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:

(A18) $\int_0^{\bar{x}}\tau(x)D(x)dx = \int_0^{\bar{x}}\left(\tau_{pct}(x) - s_{pls}{}'(x)\right)D(x)dx = \int_0^{\bar{x}}\tau_{pct}(x)D(x)dx - \int_0^{\bar{x}}2\pi x\theta_f(x)n(x)s_{pls}(x)dx$

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:

$$\text{(A19)}\qquad \int_0^{\bar{x}}2\pi x\theta_f(x)n(x)s(x)dx=-\int_0^{\bar{x}}s(x)D'(x)dx=\int_0^{\bar{x}}s'(x)D(x)dx-s(x)D(x)\Big|_0^{\bar{x}}=$$
$$\int_0^{\bar{x}}\left({s_{pls}}'(x)-\tau_{pct}(x)\right)D(x)dx=\int_0^{\bar{x}}2\pi x\theta_f(x)n(x)s_{pls}(x)dx-\int_0^{\bar{x}}\tau_{pct}(x)D(x)dx$$

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

A3: A Nested Fixed-Point Algorithm

In order to iteratively solve for location-specific values, the circular city is divided into discrete, narrow rings, each of width Δx (e.g., $\Delta x=0.1$mileused in this article). Each location x can then be labeled as $x_i=i\Delta x$ (with $i=1,2,\ldots,I$), with x_1 representing the city center and x_I representing the city's boundary, $\bar{x}$. According to the boundary condition in Eq. (24), both location's commute traffic demand, $D(x_1)$ and $D(x_I)$, equal zero.

The spatial equilibria were solved using a nested fixed-point algorithm (three loops) coded in MATLAB. The inner part of algorithm refers to LRH's (2002) algorithm for finding the fixed points of the agglomeration function $F(x)$. The middle loop of algorithm is applied to find the fixed points of the redistributed revenue $\bar{y}$. Notice that the boundary conditions in our simulation differ from those in LRH's models. While LRH's simulation assumes a fixed utility level and city boundary, our simulation assumes a fixed population and edge land rent. Finally, the outer part of our algorithm is used to find the fixed points of the land share function $\theta_f(x)$.

LRH(2002) provided a strict proof of the existence of a set of equilibrium solutions under a certain assumption on the specification of utility and production functions (e.g., when these two functions are Cobb-Douglas form). Rossi-Hansberg (2004) provided a proof of a set of optimal solutions in his extension of LRH model to correct for agglomeration externalities. The substantial difference of our model is the inclusiveness of congestion externalities and wealth redistribution (rents, tolls, and subsidies). Instead of providing complicated and elusive analytical proof, the model in our paper is solved computationally, so if an equilibrium can be computed, it exists. This is true for all models of this genre such as Fujita-Ogawa (1982), Anas-Kim (1996), and Brueckner (2007) etc. Our simulation results suggest that there exists a set of equilibrium/optimal solutions to Problem A if the parameters are appropriately selected.

In addition, in order to check the existence of multiple equilibria, simulations in this paper use several different initial functions of $\theta_f(x)$, $F(x)$, and $\bar{y}$. Simulations show that given $\theta_f(x)$ and a fixed utility level $\bar{u}$, the equilibrium solution, if exists, is unique. We thus define the optimal $\theta_f^*(x)$ when it maximizes the utility. All simulated results reported in this article are thus *Pareto*-optimal. The detailed algorithms are described below.

Step 1: Given an initial land share function $\theta_f^0(x)$, there exist a set of equilibrium functions $\{F^*,w^*,q^*,n^*,D^*,t^*,\tau^*\}$ and equilibrium values $\{y_{rent}^*,y_{toll}^*,y_{suby}^*\}$ that solve Problem A.

Step 1.0: Designate initial values to the function $\theta_f^0(x)$.

Step 1.1: Given a set of initial values, F^0,$y^0_{rent}, y^0_{toll}, y^0_{suby}$, one can find a unique wage at the city center $w^*(x_1)$ and a unique utility level u^* that satisfies the first-order conditions and the Maximum Principle conditions of Problem A.

Step 1.1.0: Define the initial values of F^0,y^0_{rent}, y^0_{toll} and y^0_{suby}. Our simulations set y^0_{rent}, y^0_{toll} and y^0_{suby} as 2000, 0, and 0. The initial values of $F^0(x_i)$ vary with the setting of $\theta^0_f(x_i)$. For example, $F^0(x_i) = \theta^0_f(x_i) \times 10^6$.

Step 1.1.1: Given an initial utility u_0, select an initial wage at x_1, $w_0(x_1)$, calculate $q_0(x_1)$ and $n_0(x_1)$ by Eqs. (7) and (16), then $D'_0(x_1)$ using Eq. (23). Given $D_0(x_1)$ is known, calculate $D_0(x_2) = D_0(x_1) + D'_0(x_1)\Delta x$. Given $D_0(x_2)$, calculate $t_0(x_2)$ by Eq. (5) and $\tau_0(x_2)$ under different policy scenarios as defined in Table 1. Given $t_0(x_2)$, $\tau_0(x_2)$, and $w_0(x_1)$, calculate $w_0(x_2)$=$w_0(x_1)$+$(t_0(x_2) + \tau_0(x_2))\Delta x$. Repeat the previous calculation, one can derive a set of paths $\{w_0(x), q_0(x), n_0(x), D_0(x), t_0(x), \tau_0(x)\}$, $\forall x_1 \leq x \leq x_I$. These iterative calculations stop at x_I, that satisfies:

$$D_0(x_{I-1}) \leq 0 \ \text{ and } \ D_0(x_I) \geq 0$$

Step 1.1.2: Calculate the edge household bid-rent $r_h(x_I)$. If the boundary condition satisfies

$$\begin{cases} |r_h(x_I) - R_a| < \epsilon_1, \text{if the instrument is not UGB policies} \\ x_I = x_{ugb}, \qquad\qquad \text{if the instrument is UGB policies'} \end{cases}$$

return $w^*(x_1) = w_0(x_1)$ and go to Step 1.1.3. Instead, repeat Step 1.1 to find a continuous series of central wage $w_0(x_1)$, $w_1(x_1)$, …, $w_{n_w}(x_1)$ until finding the $w^*(x_1)$.

Step 1.1.3: Based on $w^*(x_1)$, calculate a set of equilibrium function $\{w^*, q^*, n^*, D^*, t^*, \tau^*\}$. If the city population reaches the given number, i.e., satisfying:

$$\left|\sum_{i=1}^{I} 2\pi x \theta^0_f(x_i) n^*(x_i)\Delta x - N\right| < \epsilon_2$$

return $u^* = u_0$ and go to Step 1.2. Else, adjust the value of u^0 and repeat the Step 1.1.1and 1.1.2 to find a continuous series of u^0_0, u^0_1, …, $u^0_{n_u}$until the population condition is satisfied

Step 1.2: Based on u^* and $\{w^*, q^*, n^*, D^*, t^*, \tau^*\}$, compute land rent as follows:

$$r(x) = \begin{cases} r_f(x), & \text{if } \theta^0_f(x) > 0 \text{ and } r_f(x) > R_a \\ r_h(x), & \text{if } \theta^0_f(x) = 0 \text{ and } r_h(x) > R_a \\ R_a, & \text{if } \ r_h(x) \leq R_a \text{and } r_f(x) \leq R_a \end{cases}$$

Calculate y_{rent} and $F(x)$ using Eq.(22) and Eq. (13). The method calculating the integral in F(x) follows LRH's (2002), by using an approximation over a radial coordinate system. Dong and Ross (2015) suggested that the approximation of the production externality function F(x) over a rectangular grid system is more precise than a radial coordinate system. Our simulation experience suggests that the two coordinate systems could generate similar approximation of F(x) if the interval of angle (or grid) is small enough. While both approximation approaches could result in inaccuracy, we believe the imprecision generated by radial coordinate approximation is tolerable here. Later, we calculate y_{toll} and y_{suby} according to the definition in different policy scenarios (Table 1). If the following conditions are satisfied:

$$|y_{rent} - y^0_{rent}| < \epsilon_3$$
$$|y_{toll} - y^0_{toll}| < \epsilon_4$$

$$\left|y_{suby} - y_{suby}^0\right| < \epsilon_5$$
$$\max_{\forall x_i}\left|F(x_i) - F^0(x_i)\right| < \epsilon_6$$

return $y_{rent}^* = y_{rent}$, $y_{toll}^* = y_{toll}$, $y_{suby}^* = y_{toll}$, $F^* = F$ and go to Step 2. Else, replace y_{rent}^0, y_{toll}^0, y_{suby}^0, and F^0 with y_{rent}, y_{toll}, y_{toll}, and F, and go back to Step 1.1.

Step 2: Based on the equilibrium functions $\{F^*, w^*, q^*, n^*, D^*, t^*, \tau^*\}$ and equilibrium values $\{y_{rent}^*, y_{toll}^*, y_{suby}^*\}$, calculate a new land use share function $\theta_f(x)$ using Eqs. (20) and (21). If $\theta_f(x) = \theta_f^0(x)$, the simulation ends. Else, set $\theta_f^0(x) = \theta_f(x)$ and go back to Step 1.

A4: A Discussion on Mixed Urban Configuration

The existence of mixed-use equilibrium has been discussed in several studies (e.g., Ogawa and Fujita, 1982; Lucas and Rossi-Hansberg, 2002; Rossi-Hansberg, 2004; Duranton and Puga, 2014). These require urban models that endogenize both firms' and households' location decisions and their interactions, which are difficult to examine through traditional monocentric models. Our theoretical and simulation analyses suggest that the partially or completely mixed land use pattern could be an equilibrium solution when the congestion level is high, or the agglomeration scale is low, as found in those existing literature (e.g., Ogawa and Fujita, 1982; Lucas and Rossi-Hansberg, 2002; Duranton and Puga, 2014). However, our findings also show that mixed-use equilibrium allocation is never *Pareto*-optimal. If a non-mixed use equilibrium exists, it is always more efficient than the mixed-use allocation.

Here, the question of whether mixed land use patterns is Pareto-optimal is discussed in three situations. The first is a free market where both congestion and agglomeration externalities are not internalized. The second one is that the society recognizes both externalities but do correct them by introducing policy instruments. The third one is the social optimum, where the externalities are internalized and fully corrected.

In the free-market case, the constraints (A3) and (A4) in Problem A are relaxed. Suppose firms exist at location *x*, i.e., $\theta_f(x) > 0$, the solutions to Problem A satisfy a condition on $n^*(x)$:

(A20) $\quad \delta\kappa n^*(x)^{\kappa-1}F(x)^{\gamma} - \beta_3(x) = 0$,

and the solutions to the firms' profits maximization problem require the optimal $n^*(x)$ satisfies:

(A21) $\quad \delta\kappa n^*(x)^{\kappa-1}F(x)^{\gamma} = w(x)$

Thus, the optimal $\beta_3^*(x)$ in the free-market equilibrium should equal $w(x)$, i.e.,

(A22) $\quad \beta_3^*(x) = w(x),\ if\ \theta_f(x) > 0$

If households co-exist at location x, i.e., $\theta_h(x) > 0$, from the first-order conditions on $c(x)$ and $q(x)$ of Problem A, one can derive that the optimal $c^*(x)$ and $q^*(x)$ satisfy the following condition:

(A23) $\quad \frac{c^*(x) - \beta_3^*(x)}{q^*(x)} = \frac{\partial u}{\partial q} \Big/ \frac{\partial u}{\partial c}$

By comparing the condition (A23) and the conditions of utility maximization, i.e., Eq.(7) and (8), one can derive:

(A24) $\quad \beta_3^*(x) = y(x) = w(x) + \bar{y},\ if\ \theta_h(x) > 0$

Combining Eqs. (A22) and (A24):

$$\text{(A25)} \qquad \beta_3^*(x) = \begin{cases} w(x) + \bar{y}, & if \ \ \theta_h(x) > 0 \\ w(x), & if \ \ \theta_f(x) > 0 \end{cases}$$

Thus, if $\bar{y} \neq 0$, there exist no mixed land use at any location x. If the governmental income, including rent and toll revenues net of subsidy expenditures, is redistributed back to residents, a mixed urban form would never be Pareto-optimal. However, if the governmental income is assumed to be owned by an absent landlord and/or city authority (i.e., $\bar{y} = 0$), a mixed land use pattern could be an optimal solution. This is why a completely or partially mixed urban configuration could be a Pareto-optimal solution to the models of Ogawa and Fujita (1982) and Lucas and Rossi-Hansberg (2002).

Under the second situation, Problem A includes the constraints (A3) and (A4) and sets $\tau(x) = 0$ and $s(x) = 0$. Similar to the free-market case, one can compute the optimal $\beta_3^*(x)$ as follows:

$$\text{(A26)} \qquad \beta_3^*(x) = \begin{cases} w(x) + \bar{y}, & \text{if} \quad \theta_h(x) > 0 \\ w(x) + \zeta\gamma\delta \int_0^{\bar{x}} \int_0^{2\pi} r\theta_f(r) n(r)^{\kappa} F(r)^{\gamma-1} e^{-\zeta l(x,r,\psi)}\, d\psi dr, & \text{if} \quad \theta_f(x) > 0 \end{cases}$$

Obviously, there is no mixed land use at any location x even the governmental income equals zero. Thus, if externalities are realized in the city market but no policy instruments are adopted, the optimal urban configuration has no mixed land use areas. This finding is consistent with Theorem 1 in Rossi-Hansberg (2004), although his research only internalizes agglomeration externalities.

Under the third situation, Problem A includes the constraints (A3) and (A4) and both $\tau(x)$ and $s(x)$ are set at their optimal levels (equaling their corresponding marginal externalities). The optimal $\beta_3^*(x)$ equals that in the free-market case, as follows:

$$\text{(A27)} \qquad \beta_3^*(x) = \begin{cases} w(x) + \bar{y}, & \text{if} \quad \theta_h(x) > 0 \\ w(x), & \text{if} \quad \theta_f(x) > 0 \end{cases}$$

Thus, similar to the free-market case, the socially optimal land use patterns would have no mixed areas, if the amount of wealth redistribution is not zero.

A5: A Search for the Optimal UGB and FCZ Regulation

There is no analytical solution to the optimal UGB and FCZ setting. To find the optimal UGBs in simulations, we applied an enumeration algorithm to search an optimal location for setting UGBs in the interval $[\bar{x}_{min}, \bar{x}_{fm}]$. Here, $\bar{x}_{fm}$ is the equilibrium boundary in the free-market case, and $\bar{x}_{min}$ is the minimum boundary location or the most restrictive UGB set in the simulation. Here, we select $\bar{x}_{min}$ as one mile away from the free-market boundary, i.e., $\bar{x}_{min} = \bar{x}_{fm} - 1$mile. Figure A1 shows the change of welfare gains (% CV value of that gained in the first-best optimum) by setting different restrictive UGBs under different parameter settings. A simulation finding is that the optimal UGBs are often near the first-best city boundaries.

We used a similar algorithm to search the optimal area for FCZ regulation, i.e., the interval $[x_0, x_1]$. We started from the firm cluster boundaries of the first-best case, i.e., x_0^* and x_1^* and searched the best combination of x_0 and x_1 that maximizing the utility level. Here, $x_0 \in [x_0^* - r_0, x_0^* + r_0]$ and $x_0 \in [x_1^* - r_1, x_1^* + r_1]$. r_0 and r_1 represents the search distance away from x_0^* and x_1^*. Figure A2 visualizes the search result in the base scenario case. The vertical axis of the matrix represents x_0 and the horizontal axis represents x_1 while the colors of each cell represent the utility level when setting the firm cluster within $[x_0, x_1]$. Simulations suggest that the optimal FCZ boundaries should be set at the firm cluster boundaries of the first-best case.

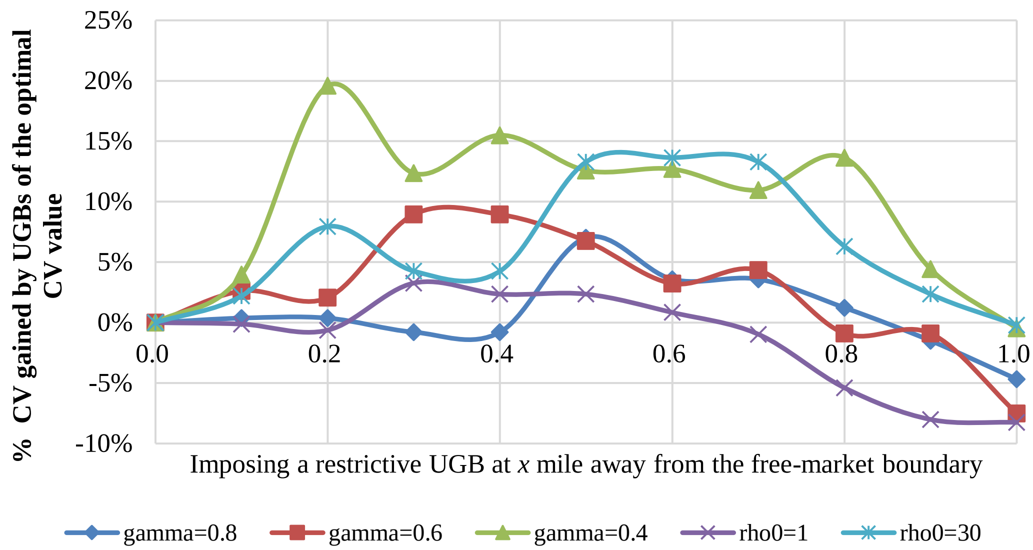


Figure A1 A search for optimal UGBs

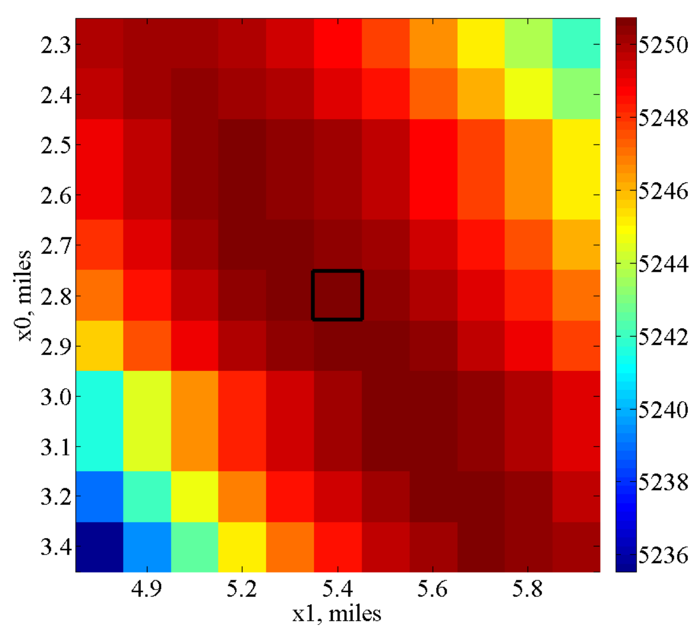


Figure A2 A search for optimal FCZ regulation in the base scenario case
(Colors represent the utility levels and the maximum utility is gained when the firm cluster locates at [2.8, 5.4], the same as the firm cluster area of the first-best optimum)