Article

Restrictive Land Use Regulations and Economic Performance

Taner Osman

Abstract
There is emerging evidence that restrictive land use practices can misallocate economic activity across space and generate significant costs for regional and national economies. Work in this area has considered how land use regulations affect the supply of housing within, and the efficient migration of workers among, regions. According to some estimates, by restricting migration to the most productive regions in the United States, land use regulations generate an annual cost to the economy of up to US$1.95 trillion or 13.6 percent of gross domestic product. Yet this focus on labor migration provides an insufficient framework for understanding how land use regulations can shape economic performance. First, current approaches minimize the costs that could arise if firms and industries are misallocated across space. Second, current work overlooks the costs that might arise if land use restrictions misallocate economic activity within regions. Land use regulations can generate significant costs when their administration is designed to maximize the welfare of individual communities rather than the economies of regions and nations. This article will evaluate current research into, and provide a more complete assessment of, the welfare costs associated with restrictive land use practices.

Keywords
land use regulations, land use restrictions, economic performance

1 Department of Urban Planning, UCLA Luskin School of Public Affairs, Los Angeles, CA, USA

Corresponding Author:
Taner Osman, Department of Urban Planning, UCLA Luskin School of Public Affairs, 3250 Public Affairs Building, Los Angeles, CA 90095, USA.
Email: tanerosman@ucla.edu
we’re in a region that’s had extremely high job growth at a rate that is just not sustainable if we’re going to keep [Palo Alto] similar to what it’s been historically. Of course we know that the community is going to evolve. But we don’t want it to be a radical departure. We don’t want to turn into Manhattan . . . We’re looking to increase the rate of housing growth, but decrease the rate of job growth . . . we have to do away with this notion that Silicon Valley must capture every job available to it.

Patrick Burt, Mayor of Palo Alto, August 2016. (Brinklow 2016)

Introduction

Do land use regulations affect the performance of regional and national economies? For the most part, this question has been considered in relation to how land use constraints affect regional housing markets and the efficient migration of workers within nations (Glaeser 2006; Saks 2008; Hsieh and Moretti 2015, 2018; Ganong and Shoag 2017; Glaeser and Gyourko 2018). Considerably less attention has been given to understanding the impact of restrictive land use practices on the location of commercial and industrial activity (Cheshire and Hilber 2008). This is an important distinction because while land use regulations are used to restrict the development of new housing within communities, they are also targeted at constraining new commercial and industrial development (Fischel 2001; Manville and Osman 2017).

In 2016, motivated by local residents’ concerns about the level of job growth in their community, the city council of Palo Alto, CA—often referred to as the “capital” of Silicon Valley—voted to cap the annual growth of office development in the city to 1 percent per year (Brinklow 2016). In the face of high demand for office space in a given community, such constraints impose additional costs on businesses in the form of higher rents and land prices (Cheshire and Hilber 2008). Perhaps more importantly, if they “go against the market,” such restrictions can distort the efficient location of firms and sectors within and across communities, potentially generating significant losses in total output.

In the US context, economists have developed a framework in which land use constraints in the nation’s most economically vibrant regions impose high costs, not just on the people living in such places, but also on the nation as a whole (Hsieh and Moretti 2015, 2018; Glaeser and Gyourko 2018). According to this view, locally administered land use regulations restrict the population growth of the nation’s most productive regions and prevent workers from efficiently migrating among the nation’s cities. The subsequent misallocation of labor reduces output per worker, constraining the economic output of both specific metropolitan regions and, by extension, the nation (Cheshire, Nathan, and Overman 2014; Hsieh and Moretti 2015, 2018; Glaeser and Gyourko 2018). In the United States, there is evidence that such regulations suppress gross domestic product (GDP) by between 2 percent and 13.6 percent (Hsieh and Moretti 2015, 2018; Glaeser and Gyourko 2018).
While such approaches to understanding the relationship between land use regulations and economic performance hold intuitive appeal, partly because they suggest a relatively simple remedy for growing interregional inequality, they provide, at best, an incomplete measure of the true costs that land use regulations can generate. This is the case for two primary reasons. First, by focusing on the migration of workers within national economies, such approaches minimize the costs that arise if land use regulations misallocate firms and industries across space. The role that land use regulations play in shaping industrial geography is important since, in standard economic models and theories of economic geography, location is a critical component of the productivity and profitability of firms and industries (Marshall 1890; Alonso 1964; Krugman 1991; Rosenthal and Strange 2004; Cheshire, Nathan, and Overman 2014).

Second, the labor migration approach focuses on the efficient interregional movement of economic activity, not the costs to economic performance that might arise if land use regulations misallocate economic activity within regional economies. This is an important omission because, in the core models of urban economics, welfare is maximized when firms are efficiently located within cities. Furthermore, there is significant evidence that agglomeration economies attenuate within regional economies (Arzaghi and Henderson 2008; Graham 2007; Rosenthal and Strange 2003, 2011). Therefore, to the extent that land use regulations misallocate firms and sectoral activity within cities, costs to productivity should occur.

Any assessment of the impact of land use regulations on economic performance must consider that land use regulations, in principle, generate benefits in addition to the costs to which they can give rise (Cheshire and Sheppard 2002; Glaeser 2008; Cheshire, Nathan, and Overman 2014; Hsieh and Moretti 2015, 2018; Bunten 2017). Land use regulations that have the effect of misallocating workers and firms across space could be welfare enhancing if they generate net benefits for a region or nation. Yet the ability to make such an assessment is clouded by the highly fragmented administration of local land use regulations, which is especially pronounced within the United States. This fragmentation is particularly marked within the nation’s metropolitan regions (Fischel 2001).

While such fragmentation is theorized to enhance efficiency in the provision of local public goods (Tiebout 1956; Oates 1972; Teaford 1979; Fischel 2001), in other areas, such as when communities zone for low-density housing, affecting the extent of urban sprawl, fragmentation is considered to be inefficient (Fischel 2001; Glaeser and Kahn 2010; Glaeser 2011). Importantly for the purposes of the current article, the administrative landscape of land use regulations has not been designed to maximize regional or national economic output but to enhance the welfare of residents in local communities (Teaford 1979; Fischel 2001; Hilber and Robert-Nicoud 2013; Glaeser and Gyourko 2018). This creates a problem since, within a metropolitan region, land use decisions might maximize the welfare of an individual community but create a loss of welfare for the wider region and nation.
This article reviews and assesses the theoretical and empirical work into the relationship between land use regulations and economic performance. In particular, this article considers the role that land use regulations play in shaping the efficient allocation of economic activities across space. The first section of this article evaluates the relationship between the fragmentation of land use decision-making and economic performance. This article will next explore existing research into the economic costs of land use regulations, considering both direct and indirect costs. This article concludes that existing approaches to understanding the relationship between land use regulations and economic performance are incomplete because they do not fully account for the costs that arise when firms and sectors are misallocated across space nor the costs that arise when economic activity is misallocated within, not just across, regional economies.

The Administration of Land Use Regulations, Local Government Fragmentation, and Economic Performance

Local zoning regulations are the most direct way in which local governments regulate how land is used. Zoning codes determine the extent, type, and the physical character of development within communities. In addition to zoning stipulations, other factors can influence the location of development across communities such as building permit fees, the length of development approval processes, and mitigation provisions to which developers must conform (Glaeser, Gyourko, and Saks 2005; Quigley, Raphael, and Rosenthal 2007; Fulton and Shigley 2012; Cheshire, Nathan, and Overman 2014; Duranton and Puga 2015; Glaeser and Gyourko 2018). Taken together, land use regulations can shape the allocation of firms and workers both within and across metropolitan regions.

In the United States, the administration of land use regulations is highly fragmented.\(^1\) According to the US Census of Governments, there were 90,056 local governments in the country in 2012. This landscape includes 38,910 general-purpose governments—3,031 counties, 19,519 municipalities, and 16,360 townships—which typically exercise control over land use regulations (Hogue 2013). Such fragmentation is especially pronounced within metropolitan regions as a result of homeowner driven municipal incorporation (Fischel 2001). In 2016, the economies of 382 metropolitan regions accounted for 90 percent of US GDP (US Bureau of Economic Analysis 2017a), yet these urbanized areas accounted for less than 3 percent of the nation’s total land area (Bigelow and Borchers 2012). Despite the importance of metropolitan regions to the output of the US economy, local government fragmentation means that, in many policy areas, the governance of such regions—including land use regulation—is diffuse and uncoordinated.

The regulation of land is a profoundly political process. Land use decisions can pit “pro-growth” constituencies—such as developers, business coalitions, labor unions, and city councils—against homeowners, who tend to have relatively more conservative views about land development in their communities (Logan and
In theories of the politics of land use decision-making, homeowners are considered to be an especially influential force (Fischel 2001; Glaeser, Gyourko, and Saks 2005; Hilber and Robert-Nicoud 2013). Fischel’s (2001) homevoter hypothesis provides an important insight into the motivation of homeowners regarding new development. According to Fischel, due to the perceived negative externalities associated with land development, homeowners work together to restrict new development within their communities in an effort to enhance the value of their properties. It is widely held that restrictive land use restrictions have increased over time (Glaeser, Gyourko, and Saks 2005; Saks 2008; Gyourko and Molloy 2014; Glaeser and Gyourko 2018). Furthermore, there is emerging evidence that the influence of homeowners on the planning process, especially in the most economically vibrant regions, entails economic costs for regions and nations (Cheshire, Nathan, and Overman 2014; Hsieh and Moretti 2015, 2018; Ganong and Shoag 2017; Glaeser and Gyourko 2018).

Fragmentation and Welfare

Tiebout’s (1956) theory of local government expenditure provides the primary justification for local government fragmentation, and since his work, the efficiency of local government fragmentation has been widely considered and debated (Tiebout 1956; Oates 1972; Fischel 2001). In relation to the provision of public goods, the consumption of which are restricted to the residents of particular communities, fragmentation is considered to increase welfare by better tailoring public goods to the diverse preferences of the population (Tiebout 1956; Oates 1972; Fischel 2001). Decentralized policy making is not considered to be welfare enhancing when a particular policy action spills over into neighboring communities (Oates 1972). Such a scenario can lead to the undersupply of public goods when the benefits of a given policy cannot be contained within the boundaries of a given jurisdiction. Likewise, governments are inclined to oversupply public goods for which the benefits are localized but the costs are diffuse. In terms of economic performance, the efficiency of government fragmentation is commonly considered in relation to interjurisdictional competition for business activity (Bartik 1991; Donahue 1997; Cheshire and Gordon 1998). Many believe that such competition creates a “race to the bottom,” where communities outbid one another for mobile capital, inefficiently reducing public services and damaging the local environment (Oates and Schwab 1988; Donahue 1997; Cheshire and Gordon 1998; Fischel 2001).

Cheshire and Gordon (1998) develop a comprehensive framework for assessing the welfare effects of interjurisdictional competition for business activity. In the European Union, the authors find most interjurisdictional competition to be wasteful, and this is especially the case the wider the geographic range over which welfare effects are considered. For example, for diversionary economic development policies, which are targeted at attracting mobile capital, the benefit
to a particular jurisdiction—such as increased employment—could be offset by the loss of activity from other communities or the opportunity cost associated with providing unnecessary subsidies to firms. When such policies reduce aggregate welfare, the authors call for superregional action to constrain competition among jurisdictions. Yet little attention has been paid to the welfare effects of local government fragmentation in relation to the efficient location of activities across space.

Fischel (2001) challenges the idea that interjurisdictional competition for economic activity will lead to a “race to the bottom.” Instead, Fischel describes a “race to the top” in which local governments “err on the side of caution in admitting industry” because of the perceived negative externalities associated with industrial development. Yet this perspective ignores the possibility that, if communities in economically vibrant regions, or parts of regional economies, “err on the side of caution” in relation to new commercial and industrial development, this can misallocate firms and industries across space, creating losses in efficiency and productivity.

In relation to the location of business activities, the fragmentation of land use decision-making within regions creates a number of challenges. The costs of new commercial and industrial development can be relatively localized—such as increased levels of traffic locally—but the benefits of new development can be more diffuse, such as job creation for the broader community. Under this scenario, there is likely to be a suboptimal supply of development in certain parts of regions, from an economic efficiency perspective. In other words, while locally administered land use regulations might maximize the welfare of individual communities, they can lead to an outcome where economic activity is inefficiently allocated across space and generate losses of welfare for regional and national economies.

There is little direct testing of the relationship between local government fragmentation and economic performance. Ahrend et al. (2014) is a rare exception in this regard. For cities in five Organization for Economic Cooperation and Development (OECD) countries—Germany, Mexico, Spain, United Kingdom (UK), and the United States—the authors find that, as the number of municipalities in a metropolitan region doubles, productivity declines by between 3 percent and 4 percent. The authors believe that, as local government fragmentation increases within a region, complexity arises in the coordination and implementation of growth-enhancing infrastructure, such as regional transportation projects. The authors also contend that fragmentation may make it more difficult for businesses to navigate across the landscape of local regulations. For cities in the European Union, Cheshire and Magrini (2009) also provide evidence that economic growth is lower in regions where decision-making is more fragmented. Yet more research is required to support these findings but also to test the mechanisms that might lead greater levels of fragmentation to inhibit regional economic performance.
The Direct Cost of Land Use Regulations

In relation to the economy, land use regulations generate both direct and indirect costs and benefits. The direct effects of land use regulations can be broadly divided into two categories. First, land use regulations can cap construction activity below market clearing levels, which would directly lower economic activity within local and regional economies. Second, there are a number of transaction and administration costs to which land use regulations can give rise (Cheshire, Nathan, and Overman 2014). I will consider the indirect costs of land use regulations in the following section.

Forgone Construction Activity

Suppressed construction activity is the clearest direct cost of restrictive land use regulations. If land use regulations within a region restrict the supply of new development below the level that meets the demand for housing and commercial space, regional output will be suppressed in the construction sector. While the construction industry is cyclical in nature, since World War II it has, on average, directly accounted for between 4 percent and 5 percent of national economic output in the United States (US Bureau of Economic Analysis 2017b). Beyond the direct output from new construction, the industry also creates income for other sectors of regional economies. Through the multiplier effect, Woetzel et al. (2016) estimate that every dollar of output from the construction industry generates an additional US$2.15 in total economic output in local economies—from the wages spent by construction workers, but also transactions between the construction industry and downstream and upstream suppliers. From the local and regional perspective, it is possible that restrictions on new development can be welfare enhancing, if they prevent the generation of urban costs greater than the benefits to which new development gives rise. Yet a number of studies provide evidence that constraints on land use development cap the size of metropolitan regions before such a point is reached (Hsieh and Moretti 2015, 2018; Au and Henderson 2006; Glaeser and Gyourko 2018).

Furthermore, as a nation grows, foregone construction activity in one community may generate construction in some other community. Therefore, any gains in welfare from restricting development in one place should be weighed against any welfare effects of substitute development in another location (Glaeser and Gyourko 2018). There is little evidence that the negative externalities associated with new development vary across locations (Hilber and Vermeulen 2016; Gyourko and Glaeser 2018). In some cases, forgone construction in a community might create greater negative externalities in an alternative location—such as when suppressed construction in temperate parts of California pushes development to other parts of the state, or other regions all together, where greater energy consumption is required to heat and cool homes (Glaeser and Kahn 2010).
Impact Fees, Uncertainty, and Unpredictability

A second category of direct costs generated by land use regulations includes the time it takes to approve permit applications, the level of uncertainty that can arise from the development approval process, building permit and mitigation fees which developers must pay, and the cost to public agencies from creating, administering, and enforcing land use regulations (Glaeser, Gyourko, and Saks 2005; Evans 2008; Cheshire, Nathan, and Overman 2014).

In principle, the time taken to approve a development can impose costs on individual developers and the economy. From a developer’s perspective, any delay in construction due to a lengthy planning approval process defers the realization of profit from land development. From the perspective of the broader economy, a lengthy planning approval process can delay the development or redevelopment of land to a more productive use and defer any positive externalities that a proposed development might generate (Keogh and Evans 1992; Evans 2008). Likewise, the more uncertainty that is embedded in the planning process, the higher an investor’s targeted rate of return, meaning that, all else being equal, less development will occur in communities with stringent and unpredictable land use regulations.

The costs associated with a lengthy planning application process must be offset against the benefits that the approval process yields (Evans 2008; Cheshire, Nathan, and Overman 2014). Reducing planning approval time will only yield a benefit if there is no reduction in the ability of planning agencies to reach effective decisions (Keogh and Evans 1992; Evans 2008). While it is difficult to assess the extent to which planning approval processes take longer than is necessary, there are notorious examples, from around the globe, of major construction projects that become embroiled in years of deliberation and litigation before final planning approval is reached. For example, government approval of London Heathrow Airport’s Terminal 5 took eight years, costing in excess of £100 million in legal and public service costs (Barker 2006). Boston’s notorious “Big Dig” was delayed due to a number of financial and political factors, including the seven years it took to receive federal environmental clearance (Altshuler and Luberoff 2004). To the extent that such delays take longer than the time needed to reach effective decisions, they generate costs for local, regional, and, potentially, national economies, as described above (Altshuler and Luberoff 2004; Evans 2008).

Mitigation and development fees can also generate costs if such fees do not approximate the marginal costs associated with new development. In theory, mitigation fees should approximate a Pigouvian-style tax on new development, which compensates for the social costs new development generates. All else being equal, if mitigation fees exceed the impact of new development, they will lead to the undersupply of new development. Therefore, inefficient fees can misallocate economic activity among communities and can also inefficiently allocate money between the private and public sector.
Whether impact fees approximate the marginal cost of new development is the subject of debate (Bartik 1991; Burge and Ihlanfeldt 2009; Jones 2015), although there is evidence that impact fees exceed the marginal cost of new construction (Bartik et al. 1987; Fox and Neel 1987; Burge and Ihlanfeldt 2009). There are few studies that examine the relationship between mitigation fees and economic development. There is some evidence that higher impact fees positively affect economic development outcomes—such as employment growth—perhaps because the fees generate productivity enhancing infrastructure (Nelson and Moody 2003; Jeong and Feiock 2006). By contrast, in a study of commercial impact fees in the State of Florida, it was found that impact fees suppressed employment growth because they exceeded the costs generated by new development (Burge and Ihlanfeldt 2009).

In sum, the direct costs of restrictive land use regulations emerge from suppressed levels of construction activity in local economies, inefficiencies that are introduced into urban land markets, and the nonmarginal construction fees that are charged to developers. More research is required to tease out the relationship between these direct costs and economic performance. However, there is some evidence that the benefits of restrictive land use regulations do not justify the costs they generate (Cheshire and Sheppard 2002; Glaeser, Gyourko, and Saks 2005).

The Indirect Cost of Land Use Regulations

I will consider five primary ways in which land use restrictions can indirectly influence economic performance. Land use regulations can influence the size of metropolitan regions, the efficient spatial allocation of firms and industries, urban structure, the performance of specific sectors of the economy, and the price of housing and commercial properties.

The Size of Cities and the Allocation of Workers within Nations

*The efficiency of large cities.* Agglomeration economies mean that, as a city grows in size, factors of production, such as labor and capital, become more productive. Yet at a certain size, the benefits of agglomeration can be offset by diseconomies of crowding—such as traffic congestion—potentially causing the marginal output of factors of production to decline and the population of a given city to fall (Henderson 1974; Au and Henderson 2006; Duranton and Puga 2013; Combes, Duranton, and Gobillon 2012).

Economies of agglomeration include both urbanization and localization economies. Urbanization economies are the general benefits that emerge from the clustering of people and firms in space, which are not specific to a particular industry. These economies include sharing the high cost of fixed infrastructure over a large number of users. Urbanization economies may also refer to Jacobsian economies, whereby greater levels of diversity in large cities engender the cross-fertilization of ideas, generating innovation and the creation of new economic activities (Jacobs
Localization economies are the benefits that firms of particular trade industries gain from locating in close proximity to one another and will be discussed further below.

There is significant evidence that workers are more productive and earn higher wages in large cities compared to workers of similar characteristics in smaller cities and rural locations (Combes, Duranton, and Gobillon 2008; Glaeser 2011; Gibbons, Overman, and Pelkonen 2014; Ahrend et al. 2014; De La Roca and Puga 2017). In the United States, for example, workers in metropolitan regions with populations of greater than one million people earn 30 percent more than workers in rural locations (Glaeser 2011). In Spain, when a city doubles in size, worker productivity increases by roughly 5–6 percent, which is reflected in higher wages (De La Roca and Puga 2017), while in the UK, Gibbons, Overman, and Pelkonen (2014) find that workers in the city at the seventy-fifth percentile of the size distribution of cities earn 3.8 percent more than workers in the city at the twenty-fifth percentile of the distribution. For cities in five OECD countries, Ahrend et al. (2014) find that, as a city’s population doubles in size, productivity increases by between 2 percent and 5 percent.

Despite the apparent economic payoff of large cities, governments throughout the world, both local and national, have undertaken explicit efforts to restrict the size of large cities. These measures include controlling migration patterns, employing urban growth boundaries, and other restrictive land use practices, such as density limits and height controls (Evans 2008; Koster, van Ommeren, and Rietveld 2013; Cheshire, Nathan, and Overman 2014; Albouy et al. 2016). Despite these actions, there is little evidence in support of the claim that large cities are inefficient and that their growth should be constrained. In fact, there is evidence that concerns about the inefficiencies of large cities may be misplaced (Au and Henderson 2006; Combes, Duranton, and Gobillon 2012; Albouy et al. 2016).

For example, for Chinese cities, Au and Henderson (2006) find that when a city grows from 25 percent below its optimal size to the optimum, net output per worker grows by 2.9 percent. Yet when a city grows to a size 25 percent greater than the optimal, a 2.3 percent loss in net worker productivity occurs. If a city size is capped at 50 percent below the optimum, this results in a loss of 14 percent in net output per worker, but growing to a level 50 percent larger than this point amounts to a loss in net worker productivity of around 8 percent. For French cities, Combes, Duranton, and Gobillon (2012) also find that the costs of large cities are overstated. They find that the elasticity of wages with respect to city population almost entirely offsets the elasticity of urban costs with respect to population. The authors conclude, therefore, that cities operate at close to constant returns in the aggregate.

The migration of workers among regions. According to regional science and urban economics (RSUE), there are two primary reasons why a given city might not reach its optimal size, from the perspective of total output. First, government actions, such as restrictions on land use development, can limit the supply of housing and,
consequently, restrict population growth in a given location. Second, crowding costs and amenity differences among cities affect their desirability to households, shaping migration away from certain cities before total factor productivity is maximized (Rosen 1979; Roback 1982; Glaeser 2008; Hsieh and Moretti 2015, 2018).

After a sustained period of convergence, real income differences among cities in the United States have grown since the 1980s (Hsieh and Moretti 2015, 2018; Diamond 2016; Ganong and Shoag 2017). According to RSUE, regional differences in income levels occur when workers are unable to migrate among regions and equalize income. Hsieh and Moretti (2015, 2018) and Ganong and Shoag (2017) provide evidence that migration among US cities has been shaped by restrictive land use regulations, not the undesirability of certain cities to workers. In their study of 220 metropolitan regions in the United States, Hsieh and Moretti (2015, 2018) provide evidence that relaxing land use regulations in the nation’s most productive cities (to the level of the median city) would reallocate workers within the country, increasing the nation’s annual per capita income by almost US$7,000. This would lead GDP to increase by between US$1.4 and US$1.95 trillion or by as much as US$13.6 percent. Employing more conservative assumptions, Glaeser and Gyourko (2018) estimate that output in the United States would increase by around 2 percent by relaxing land use controls in the most productive regional economies.

While loosening land use regulations in economically vibrant regions might lead to an increase in GDP—according to the RSUE—this outcome might not lead to an increase in welfare. For example, if land use regulations lower total output, this might be offset by the benefits that such land use practices yield. Some studies reveal that the costs associated with local land use regulations outweigh any benefits that land use regulations may generate (Cheshire and Sheppard 2002; Glaeser, Gyourko, and Saks 2005). While Bucten (2017) shows that relaxing land use regulations could increase GDP by 6 percent but decrease welfare by 5.9 percent. Generally, the relationship between land use regulations and welfare can be shaped by a number of factors, including the degree to which regulations restrict development and their effectiveness in limiting negative externalities that may be associated with land use development (Cheshire and Sheppard 2002; Hilber and Vermeulen 2016).

The assumption that land use regulations are a critical determinant of interregional migration patterns and that the labor supply response to house prices is infinitely elastic is disputed. Storper (2018), for example, asserts that the demand side of the economy shapes migration patterns, city growth, and interregional inequality, not housing supply constraints. While RSUE provides a useful framework for understanding the relationship between land use regulations and economic performance, the models are lacking in two key areas. First, they minimize the importance of industrial geography to regional economic performance. In the RSUE models, for example, industry is represented in reduced form. In this respect, the models do not fully consider the welfare effects of land use regulations as they affect the endogenous actions of firms and industries, as represented in the new economic geography (NEG). Second, the framework does not
consider costs to economic performance when land use regulations misallocate economic activity within regions.

**Land Use Regulations and the Spatial Allocation of Firms**

Compared to the analysis of how land use regulations affect housing markets and the distribution of workers among cities, there is little research into the welfare effects of land use regulations as they affect the location of firms and sectors and shape patterns of commercial and industrial development. This is surprising since, in both standard economic models and other theories of regional development, location is a critical component of both the efficiency of industries and the productivity of firms (Marshall 1890; Alonso 1964; Krugman 1991; Rosenthal and Strange 2004; Cheshire, Nathan, and Overman 2014). If land use regulations “go against the market” and prevent firms and industries from locating where productivity is maximized, we should expect significant costs to occur in the form of reduced output, for both regional and national economies.

Beyond urbanization effects, firms in trade industries cluster together to benefit from localization economies. These benefits include input sharing, labor market matching, and knowledge spillovers (learning; Marshall 1890; Krugman 1991; Duranton and Puga 2004; Rosenthal and Strange 2004). In the formal models of localization, such as the NEG, localization economies are assumed to be region-wide in scope (Krugman 1991). However, recent empirical work has shed new light on the geographic scope of localization economies (Rosenthal and Strange 2003, 2011; Graham 2007; Arzaghi and Henderson 2008). There is mounting evidence that localization economies attenuate within regions and that some parts of regions enhance the productivity of firms and sectors more than others. I will first consider the effects that might emerge if land use regulations misallocate industrial activity among regions, before considering the effects if regulations misallocate industrial activity within regions.

**The NEG and the allocation of firms among regions.** The NEG formally models the process by which external economies of scale in production, in the presence of transportation costs, lead cost-minimizing, rational producers in trade industries to congregate in space (Krugman 1991). In the basic form of the model, two industries—an immobile agricultural sector and a mobile manufacturing industry—are efficiently allocated between two regions, and production maximized, when one region (the core) is home to all of the firms in the manufacturing industry (Krugman 1991). Under this core-periphery scenario, manufacturing output is maximized through Marshallian increasing returns to scale that are realized from the concentration of manufacturing production in one region.

Implicitly, the NEG assumes that manufacturing producers will be able to fully realize the benefits from external economies of scale in production; namely, that all
of the manufacturing producers that seek to will be able to locate in the core region. Yet what if the effect of land use regulations is to cap the number of manufacturing firms that can locate in the core region at a level before increasing returns to scale are fully realized? Manufacturers would be inefficiently divided between two regions, reducing the productivity of firms, and, therefore, total output. Under this scenario, economic inefficiencies arise, not because of the level of housing supply, but the supply of industrial space among regions.

Expanding beyond a two-city model, land use constraints could lead to a scenario where firms in a given trade industry are distributed across a system of cities, where firms in a series of smaller clusters produce less output than would be the case if the firms could agglomerate together in one, or a small number of, regions. Furthermore, land use constraints could also limit the scope for co-agglomeration benefits, which are the benefits that pairs of industries gain from colocation (Ellison, Glaeser, and Kerr 2010). Again, restraints that limit the efficient allocation of firms and industries across space could be welfare enhancing, if they generate net benefits in the aggregate. Yet, at present, the effect of land use regulations is primarily considered in relation to the supply of housing, not industrial space.

The allocation of firms within regions. The relationship between land use regulations and industrial geography can also be considered in relation to the within-city location of firms and sectors. In models of urban economics, the allocation of firms and workers within cities is critical to total output. Furthermore, a relatively recent body of research has shown that localization economies are not constant within regions and that they attenuate over markedly short distances (Rosenthal and Strange 2003, 2011; Graham 2007; Arzaghi and Henderson 2008). This would suggest that it is also possible for land use regulations to generate costs by misallocating firms and sectors within regional economies. I will consider theories relating to the efficient allocation of commercial and industrial activity within regions, before considering the empirical literature related to the attenuation of agglomeration economies.

Urban form and economic performance
As discussed above, the concentration of economic activities in space generates positive externalities in the form of agglomeration economies, rendering cities more productive than other locations. While larger cities are associated with higher levels of productivity, there is also evidence that the aggregate density of cities (jobs divided by land area) enhances the returns to factors of production. For example, in the US context, Glaeser and Kahn (2004) find that as metropolitan-level density increases by 10 percent, wages increase by around 1.3 percent. For French cities, Combes, Duranton, and Gobillon (2008, 2012) find a density-wage elasticity of 4.9 percent, and for Swedish cities, Andersson, Klaesson, and Larsson (2016) find a citywide density-wage elasticity of around 1 percent. There is also evidence that metropolitan-wide density facilitates innovation and information spillovers. Carlino,
Chatterjee, and Hunt (2006) find that, as the level of metropolitan area job density doubles, the rate of patents per capita increases by 20 percent. Despite these citywide findings, there is little evidence of how the physical arrangement of economic activities within cities affects total factor productivity. In other words, whether some types of urban structure are more conducive to maximizing returns to factors of production than others. The relationship between economic performance and urban structure is important since urban structure is directly influenced by land use regulations.

In the core models of urban economics, general equilibrium can be reached in both monocentric and polycentric cities (Alonso 1964; Fujita and Ogawa 1982; Lucas and Rossi-Hansberg 2002; Duranton and Puga 2015). Theory, therefore, is somewhat ambiguous as to the form of urban structure that maximizes welfare. However, certain modeling extensions and approaches have provided a theoretical basis for understanding the importance of within-city density. For example, Ogawa and Fujita (1980) assume that transaction costs (such as communication and information exchange) are critical to firm productivity and that the cost of a transaction between any two businesses increases as the distance between the two businesses grows. Similarly, Duranton and Puga (2015) model agglomeration economies that are rooted in communication spillovers between workers and specify that efficient communication depends on how far apart jobs are located within a city (a similar approach can be found in Lucas and Rossi-Hansburg 2002).

In these models, firms benefit from density (proximity to other businesses) but, as they cluster within a city, land prices and congestion increase, meaning that firms must compensate their workers for longer commutes (Ogawa and Fujita 1980; Duranton and Puga 2015). The costs of agglomeration will lead some firms to decentralize within cities. This process can lead to the emergence of secondary clusters of activity (subcenters) within a city (Anas, Arnott, and Small 1998; Duranton and Puga 2015). Subcenters are considered to be smaller versions of central business districts (CBDs)—namely, areas of relatively high density and land values, where localized agglomeration economies are present.

If a firm relocates from a CBD to another part of a city, it will reduce its access to agglomeration benefits, but it can lower its land costs and the wages it pays to workers. What emerges is a “system of cities” within a city where firms locate with respect to the benefits they gain from, and the spatial decay of, agglomeration economies, their intensity of land use in production, and the minimization of commuting costs. General equilibrium will be reached when each firm is optimally located within a city with respect to these factors. In this modeling approach, city-level productivity is the aggregate of a number of localized density effects within a city.

Within this framework, it is conceivable for politically motivated land use regulations to constrain the intensity of land use in certain parts of a city, limiting the scope of local agglomeration economies. While such regulations may reduce negative externalities in some parts of cities, it is also conceivable that they could render
firm location suboptimal and generate losses in citywide welfare. While there is a significant body of work examining the welfare effects of land use regulations on within-city residential land use patterns—such as the effects of urban sprawl on carbon emissions, the extent of social interaction and public health (see Duranton and Puga 2015)—there is little work in this area in relation to within-city commercial and industrial land use patterns (Gyourko and Molloy 2014; Duranton and Puga 2015).

The relationship between within-city employment density and productivity is commonly considered in relation to whether jobs are concentrated within a CBD or decentralized. Glaeser and Kahn (2004) provide limited evidence that decentralized cities are more productive than cities in which a high share of jobs are located in CBDs. Yet the authors do not consider the nature of decentralization within cities, such as whether, outside of the CBD, employment is decentralized into subcenters, or whether employment is more diffuse and random in nature. In fact, there is significant empirical work into the presence of subcenters within cities (Anas, Arnott, and Small 1998; McMillen 2001; McMillen and Smith 2003). While, in US cities, employment has decentralized from CBDs over time, there is evidence that a small number of subcenters account for as much as 48 percent of metropolitan area employment (Cervero and Wu 1997). When employment is decentralized from a CBD, it can take many forms, and more research is required to understand under what circumstances decentralized employment patterns affect productivity.

While there is limited research into the importance of within-city employment density effects, there are some interesting exceptions. Andersson, Klaesson, and Larsson (2016) find that the employment-density wage elasticity is twice as high at the neighborhood, compared to the city-level. They conclude that city-level density effects are the outcome of neighborhood-level density. This is a local version of Ciccone and Hall’s (1996) finding that a doubling of employment density in US counties results in a 6 percent increase in average labor productivity at the state level. There is also evidence that “new economy” activities thrive in denser parts of cities (Spencer 2015; Duvivier, Polese, and Apparicio 2018), suggesting that within-city density is important to the development of information-based, new economy activities. In the Netherlands, Koster, van Ommeren, and Rietveld (2013) provide evidence of vertical agglomeration economies, while Liu, Rosenthal, and Strange (2018) also find evidence of vertical agglomeration economies in the US context.

Overall, there is significant evidence that the arrangement of firms and employment within cities is important to citywide productivity, but there is little consideration of where land use regulations fit into this story. Just as land use regulations can produce inefficient outcomes with respect to the arrangement of residential activities within cities, it is plausible that they can inefficiently allocate commercial activity within cities and generate losses in welfare.

**Attenuation.** Research has found that the benefits of localization can attenuate at a distance of as little as a quarter of a mile and be exhausted by a distance of ten miles
(Rosenthal and Strange 2003, 2011; Arzaghi and Henderson 2008; Melo, Graham, and Noland 2009). The presence of within-building agglomeration effects provides further evidence that agglomeration economies attenuate sharply with distance (Liu, Rosenthal, and Strange 2018). There is evidence that firms of traded-services industries benefit more from agglomeration than firms in manufacturing sectors of the economy (Graham 2007; Combes, Duranton, and Gobillon 2008; Cheshire, Nathan, and Overman 2014). Yet this crude sectoral dichotomy probably misrepresents the underlying causes that lead a given industry to cluster in space. Ultimately, the extent to which a firm benefits from localization economies depends on the underlying nature of its core activities (Audretsch and Feldman 1996; Duranton and Puga 2005). For example, some parts of sectors include activities that are comprised of nonroutine, cognitive tasks, which cannot be easily embedded into simple rules. Such activities are not simply automated and replaced by machines and, therefore, are bound to locations where specific workers are found (Glaeser 2011; Storper et al. 2015).

It is considered that such activities thrive in urban environments, where face-to-face contact and proximity remains a critical medium of communication (Storper and Venables 2004; Glaeser 2011). For example, even in a world in which the cost of long-distance communication approaches zero, it is theorized that face-to-face contact remains crucial to communicating noncodified information, the transmission of tacit knowledge and learning, building trust and relationships between customers and clients, and establishing thick social networks (Duranton and Puga 2004; Storper and Venables 2004).

Within-city localization, therefore, is better understood as the result of functional rather than sectoral specialization (Duranton and Puga 2005). In other words, business functions—regardless of sector—that are reliant on the positive externalities associated with proximity and density will locate in certain parts of cities, and such proximity benefits likely attenuate over relatively short distances. This explains why we see evidence of attenuation in sectors as diverse as advertising (Arzaghi and Henderson 2008; Graham 2007); software, food products, fabricated metal, and machinery (Rosenthal and Strange 2003); biotech and life sciences (Aharonson, Baum, and Feldman 2007; Kolymphiris and Kalaitzandonakes 2013); and finance and insurance and the manufacture of motor vehicles (Graham 2007). In Graham’s study of twenty-seven industries in the UK, he finds no consistent pattern of attenuation by sector type. For example, in the thirteen industries for which attenuation does exist, seven are in the manufacturing sector, while six are in the services sector of the economy.

The relationship between land use regulations and attenuation might be better understood if there is a clearer understanding of the mechanisms of attenuation. For example, is attenuation the result of distance decay effects related to networks and knowledge spillovers, or the efficiency of labor market matching? There is little formal testing of the mechanism of attenuation, with an exception in this regard including Rosenthal and Strange (2001), who find evidence that knowledge spillover...
effects are more localized than the other mechanisms of agglomeration. This finding is consistent with the broader literature on the geographic scope of patent citations, where there is significant evidence of the clustering of citation activity at fine-grained geographic scales (Carlino et al. 2012; Kerr and Kominers 2015). Evidence of attenuation adds further importance to the need to understand the relationship between land use regulations and the physical arrangement of firms and employment within cities.

The Impact of Land Use Regulations on Specific Sectors

Beyond influencing the extent of localization economies, there is evidence that land use regulations can affect the productivity of particular sectors in other ways. In England, Cheshire, Hilber, and Kaplanis (2014) find evidence that land use constraints have limited both the size and potential location of supermarkets compared to the regulations in place in other nations of the UK. Since larger supermarkets are, on average, more productive than smaller stores, and store location affects access for deliveries and customers, constraints that affect the size and location of stores directly affect the sector’s productivity. The authors estimate that, over time, land use constraints in England have reduced supermarket output by 32 percent. In the hotel industry, Lewis et al. (1998) have found evidence that land use constraints in the UK affect the price of constructing and refurbishing hotels, which has meant that the breakeven occupancy rate for hotels in the UK is two times greater than in the United States—80 percent, compared to 40 percent.

In sum, there is a broad body of theory and empirical evidence underscoring the importance of industrial geography, both among and within regions. Yet there is little work into the importance of land use regulations in facilitating or inhibiting the efficient arrangement of such activity.

Land Use Regulations and Property Prices

There is a significant amount of research devoted to understanding the relationship between land use regulations and house prices (see, e.g., Quigley and Rosenthal 2005; Glaeser and Gyourko 2003; Cheshire, Nathan, and Overman 2014; Gyourko and Molloy 2014; Duranton and Puga 2015; Hilber and Vermeulen 2016; Glaeser and Gyourko 2018). In a well-functioning housing market, the price of a given home would be equal to the cost of construction and the minimum profit that it requires a developer to construct the home (what Glaeser and Goyourko [2018] refer to as the minimum profitable production costs [MPPC]).

For much of the United States, homes are valued at or below MPPC. Yet for around a quarter of the metropolitan regions studied by Glaeser and Gyourko (2018), house prices exceed MPPC by more than 25 percent, and in roughly 40 percent of these cases, house prices were more than double MPPC. In these areas, restrictive land use regulations have decreased the elasticity of the supply curve for housing,
meaning house prices are driven by regulatory supply constraints and not by market fundamentals, amounting to a “zoning tax” (Glaeser, Gyourko, and Saks 2005; Quigley and Raphael 2005; Saiz 2010). Similar results have been found in the UK, where Hilber and Vermeulen (2016) find that house prices in the South East of England would be 25 percent lower if local governments had pursued zoning policies that were as unrestricted as those implemented in the North East of the country.

Beyond the costs that might arise if restrictive land use regulations misallocate economic activity across space, is there an independent cost of high house prices? The homeowner hypothesis, described above, identifies that homeowners restrict development to preserve and enhance the value of their primary asset, their homes. There is a welfare gain for a homeowner when zoning constraints raise home prices (Glaeser 2014). However, this gain must be weighed against a loss in welfare for renters, the potential loss of wealth for workers who are priced out of property markets, and the general costs associated with housing unaffordability, such as increased levels of sprawl, as low-paid workers search for cheaper housing in remote parts of regions. Rising house prices also represent a transfer of wealth from buyers to sellers, the welfare effects of which are unclear (Glaeser and Gyourko 2018).

Land use constraints can also create a “regulatory tax” on commercial property markets. Such constraints can amount to an 800 percent “tax” on new construction in economically vibrant cities with tight land use regulations, such as London (Cheshire and Hilber 2008). The “regulatory tax” represents an additional cost to doing business in a given region, the welfare effects of which depend on assumptions that are made about the production function of a given firm. If rents increase due to regulatory constraints, and space is not perfectly substitutable in production, higher costs should reduce output and employment. Again, these costs should be weighed against the benefits that restrictive zoning practices might yield.

**Conclusion**

There are a number of ways in which land use regulations can affect economic performance. The most promising research in this area considers how land use regulations can misallocate economic activity among cities, generating losses in regional and national output. Yet this work focuses on the efficient migration of workers among cities and minimizes the importance of industrial geography to economic output. This is problematic for at least two reasons. First, by assuming that the labor supply response to house prices is infinitely elastic, the RSUE approach minimizes the effect of the demand side of the economy in shaping migration patterns. Consider the case of Palo Alto, CA, mentioned at the outset of this article. The city’s policies seek to restrict land development for new commercial and industrial activity, while increasing the supply of housing in the community (Brinklow 2016). By restricting localization economies within the community and, by extension, the region, the city’s zoning policies will generate a cost independent
of the ability of workers to migrate among cities. Second, by minimizing the relationship between land use regulations and industrial geography, current research fails to account for the cost to economic performance from misallocating activity at the intraregional scale.

The production function of a given firm is shaped by geography, and the presence of agglomeration economies is critical to firm output. This article has detailed evidence demonstrating how the arrangement of firms and sectors within cities is important to firm and sector productivity. For firms in certain industries, localization economies cannot be perfectly substituted across space—either among or within regions. For example, if a new tech firm is zoned out of the parts of Silicon Valley that would optimize the firm’s productivity, the firm’s location in some other place would be suboptimal. This outcome would also reduce the extent of localization economies in Silicon Valley. More broadly, further evidence is required better to understand the relationship between urban form and economic performance.

This picture is made more complicated by local government fragmentation and the tensions between local and regional welfare. More research is required to tease out the relationship between local government fragmentation and the economic performance of regional economies. Land use regulations are justified on the grounds that they can limit the impact of the negative externalities that are associated with certain forms of development. Yet the perceived benefits to individual communities from restricting land use development can come at a great cost to the regions and nations within which communities are located. To understand fully the welfare implications of land use regulations, more research is required to understand how such regulations affect the efficient allocation of firms and sectors across space. Finally, if, as the evidence suggests, land use constraints in one community can reduce aggregate welfare, there is a strong case for intervention from higher levels of government to minimize these effects.

**Declaration of Conflicting Interests**
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The UCLA Ziman Center for Real Estate provided some funding for this research.

**ORCID iD**
Taner Osman https://orcid.org/0000-0002-8572-1290

**Notes**
1. For the illustration of key arguments, this article primarily references the administrative landscape of land use regulation in the United States. However, this article’s reasoning
applies to other planning contexts where land use decision-making is highly fragmented and independent.

2. The indirect costs of artificially restricting economic activity within and across regions will be discussed below.

References


