University of Toronto Working Paper

# Do We Need Eminent Domain? An Empirical Investigation of Urban Land Assembly

Leah Brooks
Research Division
Federal Reserve Board of Governors

Byron Lutz
Research Division
Federal Reserve Board of Governors

July 2011
Preliminary – Comments Welcomed

We are extremely grateful to staffers Bulmaro Borrero, Rex Hartline, and Alan Santos at the Los Angeles County Assessors office for providing data and guidance while we endlessly assembled the dataset for this paper. We are also appreciative to the Lincoln Institute for Land Policy, McGill University and the John Randolph Haynes and Dora Haynes Foundation, all of which contributed funds towards the construction of the dataset. We thank Gustavo Bobonis, Edward Coulson, Richard England, Lynn Fisher, Edward Glaeser, Robert Helsley, Robert Inman, Jeffrey Lin, Dan McMillen, Vasia Panousi, Kathy O'Regan, Jaren Pope, Raven Molloy, Hui Shan, David Wildasin and participants at the 2010 and 2011 International American Real Estate and Urban Economics Association conferences, 2010 Federal Reserve System Regional Analysis conference, 2010 Lincoln Institute conference (October), 2010 National Tax Association Annual Conference, 2011 ASSA meetings, 2011 Lincoln Institute Urban Economics and Public Finance Conference (May), the Philadelphia Federal Reserve, the Federal Trade Commission, the Urban Institute, 2011 Planning Law and Property Rights (Alberta), and the Federal Reserve Board of Governors. We are very grateful to the planners and developers that shared insights with us and greatly improved the paper; they are listed in the bibliography. This paper would not exist but for the superior research assistance of Brian McGuire.

**Do We Need Eminent Domain ? An Empirical Investigation of Urban Land Assembly**

Do urban land markets respond to changing economic conditions by producing land assembly – the legal joining together of two or more parcels of land? Failure to assemble land may produce urban blight, reduce agglomerative benefits, push growth to the city edge and away from the core; ultimately it may cause cities to forfeit economic growth. We develop a simple theoretical framework which provides a testable hypothesis: in the absence of market frictions, the price of land sold for assembly should not exceed the price of land sold for other uses. This hypothesis does not hold when frictions, such as holdouts and land use regulation, drive market behavior. We test this hypothesis using a novel dataset constructed by following each of the 2.2 million parcels in Los Angeles County over an eleven year period that allows us to observe all instances of assembly. Using panel data methods, including a repeat-sales approach, we find that to-be-assembled land trades at a 35 to 60 percent premium. Thus, we find that urban land markets are subject to significant frictions which prevent assemblies and produce sclerotic urban redevelopment. Additional empirical results suggest that private market frictions, such as holdouts, play an important role in blocking assembly.

Leah Brooks
Research Division
Federal Reserve Board of Governors
Stop # 93
20th and C Streets, NW
Washington, DC 20551-0001
USA
leah.brooks@frb.gov

Byron Lutz
Research Division
Federal Reserve Board of Governors
Stop # 83
20th and C Streets, NW
Washington, DC 20551-0001
USA
Byron.F.Lutz@frb.gov

Cities are composed of individual pieces of land called parcels. Just as atoms constrain the shape of matter, parcel size and shape constrain the built infrastructure of which cities are composed. Changes in technology and economic conditions cannot induce fundamental changes to this built infrastructure without changes to parcel boundaries. Thus, long-run urban development depends upon the ease of changing these boundaries.

In already built areas, changes to parcel boundaries usually require the assembly of multiple individually-owned parcels into one large singly-owned parcel – a process called land assembly. Land assembly occurs in two ways: private exchange in the urban land market and government exercise of eminent domain. In a recent and controversial decision, *Kelo v. City of New London*, 545 U.S. 469 (2005), the U.S. Supreme Court upheld the use of eminent domain for economic development, even when it involved transferring private property to a private corporation. In this paper, we ask whether there is empirical support for an economic argument for eminent domain. Specifically, does the market produce enough urban land assembly to put land to its highest value use?[1]

Economic historians give us many reasons to believe that the ability to assemble or concentrate land ownership is a crucial pre-requisite for economic growth. Rosenthal (1990) shows that fragmented powers of eminent domain in pre-Revolutionary France inhibited profitable irrigation projects. Bogart and Richardson (2009) contend that the British Parliament's willingness to "assemble" ownership interests in land following the Glorious Revolution of 1688 yielded dividends in economic growth. In a more recent period, Field (1992) argues that the land subdivision boom of the 1920s and the subsequent problems of land assembly it caused in the 1930s exacerbated the United States's exit from the Great Depression.[2]

[1]Few empirical studies have examined eminent domain. Munch (1976), Chang (2010), and Chen and Yeh (2010) are exceptions.

[2]Relatedly, differences in systems of land demarcation across the U.S. states (Libecap and Lueck, 2011) and across former British colonies (Libecap et al., 2010) yield divergent economic outcomes.

In the modern city, failure to assemble land can impede growth. Most fundamentally, it causes land to be misallocated: as cities grow, more capital intensive use of land is optimal (Henderson, 1977) and failure to assemble impedes adjustment in the capital-to-land ratio. The resulting lack of density may reduce agglomerative benefits (Jacobs, 1961). Similarly, the canonical model of urban redevelopment posits that land is redeveloped when the present value of capital is exceeded by the present value of redevelopment (Brueckner (1980b), Brueckner (1980a), Wheaton (1982) and Wheaton (1983); see Rosenthal (2008) for empirical work). If land assembly is required for capital adjustment, and if there is not enough assembly, capital decay alone may be insufficient to cause redevelopment. The failure of assembly in the urban core may also cause cities to expand at the edge (Miceli and Sirmans, 2007), yielding congestion and attendant environmental ills. Finally, many urban planners advocate "transit oriented development" – denser housing near transit hubs. An inability to assemble means that public transport investments may fail to generate denser housing. As a further consequence, the transit system then lacks the high ridership that yields economies of scale and makes transit viable.

Theorists have long been interested in the question of land assembly and suggest that the amount of land assembly is inefficiently low. Strange (1995) highlights inefficiencies due to asymmetric information and sellers who hold out for higher prices. Eckart (1985) also suggests that holdouts may prevent profitable assemblies. O'Flaherty (1994), highlights inefficiency due to the positive externalities received by neighboring parcels – potential members of the assembly – when an assembly occurs. Such spillovers make potential sellers less likely to participate in the assembly.[3]

[3] Other theoretical work on land assembly includes Asami (1988), Grossman et al. (2010), and Menezes and Pitchford (2004). Cadigan et al. (2011) conduct experimental research on the potential private market inefficiencies in land assembly. Kominers and Weyl (2010), Plassmann and Tideman (2008) and Tanaka (2007) suggest mechanisms for solving the holdout problem. A distinct literature explores whether the price per square foot of land increases in the size of the plot and is indirectly related to assembly (Coulson, 1989; Colwell and Munneke, 1999, 1997; Colwell and Sirmans, 1993; Brownstone and DeVany, 1991). A positive relationship between the per square foot price and the size of the lot is a necessary condition for assembly,

More broadly, theorists and legal scholars have considered how the problem of fragmented ownership – known as the "tragedy of the anticommons" – may impede efficient outcomes (Heller, 1998; Fennell, 2004). These include the resolution of sovereign debt consolidations (Pitchford and Wright, 2008), the efficient use of multiple related patented discoveries (Heller and Eisenberg, 1998), and the present case of land ownership.

Practitioners concur with theorists' dour assessment of land assembly efficiency. Nelson and Lang (2007) write that land assembly is the "single biggest obstacle to central city redevelopment." Shigley (2007) agrees, writing that Florida "has a number of very old cities, and some of them are crowded with dilapidated buildings on tiny lots. In addition, the state is crisscrossed with antiquated subdivisions drawn up during the first half of the 20th century that do not come close to meeting today's standards." Thus, problems with land assembly prevent land from being put to productive use. Similarly, Philadelphia has almost 60,000 vacant parcels and most are too small to be redeveloped. Land assembly is therefore crucial "to the stabilization and rebuilding of Philadelphia's neighborhoods" (Philadelphia Neighborhood Transformation Initiative; cited in Shoup (2008)).

Despite this theoretical and practical interest, there is extremely limited empirical evidence on land assembly, and none that directly addresses the central question of how successful the urban land market is at producing assemblies. Cunningham (2010) and Fu and Somerville (2002) document that the final seller in a land assembly receives a premium.[4]

Our limited empirical knowledge about land assembly is a major gap in our understanding of urban areas. Figure 1 presents the total number of assemblies in Los Angeles County divided by the total number of permits issued for residential construction by year from 1999 to 2008. We expect that this number actually underestimates the importance of land assembly. While one permit allows for construction of one unit, one assembly may (and

as we discuss in detail in section 1.

[4]Harding et al. (2003) provide evidence of more general deviations from the competitive equilibrium in the market for housing.

frequently does) result in more than one unit. For instance, three parcels of land may be assembled to construct a multi-story 30-unit condo building. Thus, while three parcels are assembled, permits are given for 30 units.[5] Even given this likelihood of underestimation, assemblies account for between 6 and 16 percent of residential permits per year.

Our paper makes three primary contributions. First, we directly address the question of central economic concern: does the urban land market produce the amount of land assembly consistent with putting land to its highest value use? Or, alternatively, do frictions such as holdouts and land use regulation significantly limit the supply of land for assembly? An answer of yes suggests there may be significant benefits to eminent domain (although these benefits would need to be weighed against the costs). Second, we use what is, to the best of our knowledge, the best existing dataset to study such a question. We have assembled a panel dataset that traces each of the 2.2 million parcels in Los Angeles County over an 11-year period. The dataset is based on annual cross-sections containing all parcels in the county and a database, provided by the county assessor, which identifies all instances of changes to parcel boundaries and links each parcel that ceases to exist with its descendant. Our dataset allows us to follow each individual piece of land in the county over this entire period. Third, we use novel, theoretically motivated tests to evaluate whether there is "too little" land assembly.

Our empirical test begins with a simple theoretical framework. The heart of the framework is the intuition that parcel definitions were decided long ago, and now no longer represent the optimal division of land. Prices of land are now such that larger pieces of land are worth more, per square foot, than smaller pieces of land. We assume free entry into the market for assembly and that, correspondingly, developers earn zero profits and all surplus from assembly goes to the landowner. This model yields the following testable assertion: In

[5]It is also possible that this number overstates the impact of assembly, since permits include only residential construction and our data include all land for assembly. However, the majority of assembly land (like all land) is used for residential purposes.

a market free of frictions, the price of land sold for assembly should not differ from the price of land that is sold not for assembly. Regardless of whether parcels are sold for assembly or not, any difference in the land price per square foot should be arbitraged away.[6]

To evaluate this contention we require a method to value land. We rely on the technique pioneered by Rosenthal and Helsley (1994) and refined by Dye and McMillen (2007) that identifies the price of land from parcels sold when the structure is subsequently torn down. Because the structure is worthless to the new owner, the sale value represents only the land value. We compare the price of properties sold as "teardowns" to properties sold for assembly, conditional on small neighborhood (tract or block group) fixed effects that net out the main component of land value: location. We even more rigorously test for a land assembly premium using a repeat sales approach, allowing us to net out time-invariant property features. Regardless of method, we find that land sold for assembly exceeds the price of land sold for teardown by roughly 35 to 60 percent. We interpret this as evidence that the market for land assembly is subject to frictions.

The frictions we document could result from imperfections in both the public or private markets. Public market failures are due to the regulation of land by local government, such as zoning restrictions, development fees, and building codes. Private market failures stem from bargaining problems between the developer of the assembled land and the land sellers. We perform two additional tests that provide suggestive evidence about the sources of the frictions. The theoretical framework suggests that, because land prices increase more than linearly with lot size, small parcels receive a larger price increment from selling a parcel into assembly relative to their current use than do large parcels. Because small parcels therefore have a lower opportunity cost to assembly, it should be optimal to assemble them. On the other hand, theoretical models of land assembly such as Strange (1995) and Eckart (1985)

[6]Clearly, assembly is also a function of the value and vintage of the capital on a given lot. We address this issue both theoretically and empirically below.

emphasize that small parcel owners are prone to creating private market frictions by asking excessive prices. Our first additional test therefore evaluates the contention that in a well-functioning market developers prefer to assemble smaller parcels. Our second additional test examines the relationship between parcel size and the sales price into assembly. Evidence that small parcels command a significant premium relative to large parcels would support the theoretical contention that small parcel owners are prone to creating frictions.

These additional tests provide compelling evidence that private market imperfections are substantial. We find that developers prefer to assemble larger parcels, suggesting that private market frictions such as holdouts are substantial.[7] In addition, we find that, within a given assembly, smaller parcels sell into the assembly for a substantially higher price per square foot than larger properties. Again, this suggests an important role for holdouts. Although these two findings together suggest that private market failures are substantial, we do not interpret them as evidence against an equally substantial role of public market failures.

# 1 Theoretical Framework

We begin our analysis with a simple theoretical framework to generate testable predictions about land assembly. We first consider the case where no frictions prevent land prices and quantities from adjusting to market conditions. We then consider cases where frictions inhibit the market.

Assume that land was developed at time $t-j$ ($j > 0$) into identical parcels of size $p$. The parcel size optimally reflected market conditions at time $t-j$. We use this framework as a rough approximation for a neighborhood where parcels are defined at the time of development. As time passes, those initial parcel definitions may become sub-optimal. In this vein, we assume that market conditions have evolved such that larger assembled parcels size $2p$

[7] It is also possible that developers prefer to assembly fewer parcels because this yields fewer interactions with the government; in this case both public and private market frictions may be at work.

command a premium relative to smaller parcels. In other words, the relationship between land value and parcel size has become convex, such that

$$V(2p) > 2V(p), \tag{1}$$

where $V(x)$ is the market price of a parcel of size $x$. Consistent with our empirical approach, which focuses on the value of land and not capital, we assume that any capital placed on the land at time $t - j$ has depreciated to zero by time t. The convexity of land prices is presented graphically in figure 2.

Land values tend to become convex when the optimal capital to land ratio increases. Convexity arises because the density implied by high capital to land ratios requires large lots. Builders may require large lots for technical reasons. For example, a footprint of a given size may be required to construct a building of a given height. Similarly, builders may require large lots in order for buildings to be of sufficient size to absorb fixed costs such as elevators.

The optimal capital to land ratio in a metropolitan area may increase for any number of reasons. Population growth will tend to increase the optimal capital to land ratio (Henderson, 1977). Similarly, increased commute times in an urban area may push the optimal ratio up in the urban core. The optimal ratio may also rise when land use shifts geographically with time. For instance, land initially developed into small single family lots may eventually become more valuable for commercial purposes. Commercial uses typically require a higher capital to land ratio, and thus larger lot sizes. Finally, technological shocks may also change this ratio. For example, Willis (1995) describes how the invention of fluorescent lighting changed the shape of office buildings and thus increased the optimal lot size for such buildings.

The convexity in the land value function implies that assembly generates a surplus relative

to the land in its current use. The surplus value, $s$, is defined as

$$s = V(2p) - \delta - 2V(p), \tag{2}$$

where $\delta$ is the costs of assembly and captures factors such as conversion costs (e.g., demolition, grading to-be assembled parcels with different slopes, etc.) and "good-institution" transactions costs. Crucially, $\delta$ only includes costs which would reasonably arise in a well-functioning land market free of frictions. We use the term "good-institution" to sharply distinguish transactions costs consistent with a well-functioning land market from 'bad-institution" transactions costs better viewed as *frictions*. "Bad-institution" transactions costs might include zoning and holdouts. For example, $\delta$ includes the cost of changing title to a property, but not delays to change in title caused by protesting neighbors.

Convexity in the land price function is a necessary condition for land assembly to occur (see Colwell and Munneke (1999) and Shoup (2008)). The cost of assembly, $\delta$, can only be covered when the value of the assembled parcel, $V(2p)$, exceeds the value of the unassembled parcels, $2V(p)$. The value of the unassembled parcels, $2V(p)$, can be interpreted as the opportunity cost of assembly; it represents the economic returns to the unassembled land foregone in exchange for realizing the higher return to the assembled land.

In a frictionless world, arbitrage ensures that all surplus is realized and that assemblies continue until the market price of land has adjusted such that any surplus is eliminated. Specifically, as assemblies occur the supply of lots sized $2p$ expands and the price of these lots falls. Assembly ceases when the return to assembled and unassembled lots has equalized: $V(2p) = \delta + 2V(p)$ and $s = 0$. It is also possible that the market reaches a corner solution such that all parcels have been assembled. As we do not see this solution in practice, we do not further analyze this case.

If frictions are present, all surplus available from assembly may not be arbitraged away.

For instance, holdouts may ask excessive prices for their parcel and thereby make projects infeasible for the developer (Eckart, 1985; Strange, 1995). Similarly, strategic delay on the part of individual landowners may cause assemblies to fail (Menezes and Pitchford, 2004; Miceli and Segerson, 2007; Miceli and Sirmans, 2007). The public goods aspect of land assembly – the fact that assembly may increase the value of neighboring parcels not participating in the assembly – may also block arbitrage opportunities (O'Flaherty, 1994). Finally, land use regulations may systematically block arbitrage opportunities. For example, regulation could bar a large building that would optimally occupy an assembled site.

Given this framework, we now lay out a strategy for assessing whether or not land assembly is inhibited by frictions. Our test estimates the magnitude of the surplus $s$ accruing to successful assemblies in order to obtain a rough sense of the magnitude of the assembly market frictions. A large estimate of $s$ is consistent with substantial frictions in the market for assembly. This approach is similar in spirit to the work of Glaeser et al. (2005) and Glaeser and Gyourko (2002) on the regulatory tax. Glaeser et. al. reason that in the absence of regulation the extensive value of land – the value of land with a house on it – will equal the intensive value of land – the value of a marginal increase in the area of a lot. If the extensive value exceeds the intensive value, landowners should optimally choose to subdivide their land and sell a portion of it. Our approach applies what is, in essence, the reverse of this logic: in the absence of market imperfections, if land is worth more combined than divided, owners will choose to combine it.

Inferring the existence of frictions from the presence of surplus requires two assumptions. The first assumption is free entry into the market for development (or assembly). Developers earn zero profits and the owners of the initial parcels size $p$ therefore capture any surplus $s$ available from assembly. The value of an assembled parcel is $V(2p) + K$, where $K$ is the amount of capital placed on the newly assembled parcel. If the developer earns zero profits, this post-assembly value must equal his costs. The developer's costs are capital, $K$, assembly

costs, $\delta$, and the purchase price of the unassembled land, $p_u$. Thus, $V(2p) + K = K + \delta + p_u$ which yields $p_u = V(2p) - \delta$. We can therefore estimate surplus as the difference between the sales price of to-be-assembled parcels, $V(2p) - \delta$, and the sales price of not assembled parcels, $2V(p)$: $V(2p) - \delta - 2V(p)$.[8]

The second assumption is that the frictions in the urban land market operate purely as a supply constraint on assembly. This occurs if regulation prohibits assembly or if landowners cause assemblies to fail by asking prices that drive the developer profits below zero. These supply restraints prevent arbitrage from entirely eliminating the surplus to assembly.

If the free entry and supply constraint assumptions fail to hold we will likely *understate* frictions in the market for assembly. First, if there are barriers to entering the market for development, developers may capture a portion of the assembly surplus. The portion of the surplus accruing to the developer is reflected in the post-assembly sales price of the newly assembled parcel, not in the pre-assembly price we use to infer surplus. As a result, our estimate of $s$ would be biased downward. Second, although the frictions described above almost certainly act as supply constraints on assembly, there may be other types of frictions. For instance, frictions such as regulatory costs (e.g. the time spent getting approval for a project) and strategic delay may increase the developer's costs. Such an increase will reduce our measured value of surplus. A given assembly may have no measurable surplus under our methodology, but a much larger surplus in the absence of the friction-induced increase in developer costs. Thus, even a finding of no price premium for to-be-assembled parcels does not rule out the possibility that frictions influence the market for assembly.

There are two potential objections to our test that merit discussion. Our empirical procedure recovers the average surplus to assembly. A first objection is that a strict claim of inefficiency requires that surplus fail to equal zero for the marginal assembly parcel. In our simple framework this is not a concern because there is no heterogeneity in costs and the

[8] We remain deliberately agnostic on how sellers of input parcels split the surplus from assembly.

average and marginal surplus are equal. Of course, in reality there is surely heterogeneity in the costs of assembly (keeping in mind that "good-institution" transactions costs are not driven by market frictions). Assume that there are contiguous pairs of parcels with unique assembly costs, $\delta_i$, that there are no frictions in the land market, and that the market is in equilibrium (and hence no assembly takes place because all positive surplus assemblies have already occurred). Then a shock increases $V(2p)$. Order the surplus of each pair, $s_i = V(2p) - \delta_i - 2V(p)$, so that $s_1 > s_2 > s_3$.... Unless all parcels assemble, there must be an $s_{i^*}$ such that $s_{i^*-1} > s_{i^*} = 0 > s_{i^*+1}$. The index $i^*$ implicitly defines the number of assemblies produced by the shock. The marginal parcel $i^*$ earns zero surplus, and all inframarginal parcels earn a surplus. Hence, with heterogeneous assembly costs, it is possible to generate positive average surplus even in the absence of frictions. [9]

This type of heterogeneity in assembly costs, however, is unlikely to explain our findings for two reasons. First, we find that average surplus equals 50 percent of the market value of a parcel in its unassembled state. Assembly costs – demolition, land grading, "good-institution" transaction costs, etc. – are negligible relative to the value of some of the most expensive urban real estate in the U.S.[10] Yet for heterogeneous assembly costs to explain our estimates, assembly costs would have to be extremely large relative to the value of the land in its unassembled state. Second, it is very unlikely that the *distribution* of assembly costs is disperse enough to generate an average surplus of 50 percent and a marginal surplus of 0.[11]

The second potential objection to our test is that it may take time for the market to complete enough assemblies to drive surplus to zero. During such a transition, surplus

[9] We thank Robert Helsley for discussion on this point.

[10] Rosenthal and Helsley (1994) document that demolition costs are only around $1\frac{1}{2}$ percent of the value of the underlying land for single family homes in Vancouver (which, like LA, has relatively expensive land).

[11] Of course there are other forms of heterogeneity. For instance, there is likely variation across locations in the difference between the value of a parcel in its assembled state, $V(2p)$, and unassembled state, $V(p)$, and hence in the surplus to assembly. However, in the absence of frictions arbitrage should drive the assembly surplus to zero in all locations regardless of the initial surplus available.

would be positive, even in the absence of frictions. However, the market for construction is quite deep in Los Angeles County. Many developers participate and even large construction projects can typically be completed quickly. The Census Bureau reports that a small sized multi-family building worth less than \$3 million is completed in less than a year on average. Even very large buildings worth in excess of \$10 million and containing over 100 units take less than two years to complete.[12] Thus, a transition period should be relatively short and it is unlikely to explain a 50 percent surplus to assembly. While assemblies often take a significant amount of time to complete, the delay is typically caused by the very frictions, such as holdouts, whose impact we are attempting to quantify, not by the normal adjustment process of a well functioning land market.

Given that we show a non-zero surplus, we propose two additional tests to examine the source of frictions in the market for land assembly. For the first additional test, assume that at time $t$ parcels of size $\frac{p}{2}$ and $p$ exist. Assembly technology allows for generating parcels of size $2p$ through any combination of parcels yielding an area of $2p$. The convexity of the land value function suggests that

$$V(p) > 2V(\frac{p}{2}). \tag{3}$$

The assembly surplus is therefore higher for assemblies involving only parcels of size $\frac{p}{2}$ because these smaller parcels have a lower opportunity cost than the larger parcels. (Returning to equation 2 and concentrating on the negative final term which represents the opportunity cost, we see that $2V(p) > 4V(\frac{p}{2})$). As a result, in a market free of frictions, small parcels should be more likely to be assembled than larger parcels.[13]

However, small parcels may also increase the likelihood of private market frictions. The-

[12] http://www.census.gov/const/C30/length.html

[13] Assembly costs, $\delta$, may be a function of the number of parcels included in an assembly (e.g. real estate transaction fees would be expected to be a function of the number of parcels). In such a case, we are implicitly assuming that as the number of input parcels increases, holding the size of the end assembly fixed, the value of the assembled parcel increases faster than the increase in the assembly costs (roughly $\frac{\partial \delta(n)}{\partial n} < \partial \frac{V(n)}{\partial n}$). This is a reasonable assumption given that assembly costs are small relative to the value of land.

oretical work argues that owners of small parcels are more likely to ask excessive prices (Eckart 1985, Strange 1995). Similarly, the greater number of parcel owners involved in an assembly, the greater the odds of strategic delay (Miceli and Sirmans 2007). Both these factors may cause assemblies with positive surplus to fail. Our first additional test therefore examines the influence that parcel size has on the probability of assembly. A finding that larger parcels are more likely to be assembled suggests that private market frictions inhibit assembly.[14]

Our second additional test examines the relative sales prices of to-be-assembled parcels by size. Evidence that small parcels command a significant per-square foot price premium over large ones (despite a lower opportunity cost of selling into assembly) would support the theoretical predictions that small parcels owners are unusually likely to ask excessive prices and engage in strategic delay. These two additional tests shed light on the source of the frictions, as both focus on frictions likely caused by private market failures.

# 2 Empirical Approach

## 2.1 Primary Test: Price per square foot of land should equalize

Our primary test compares the value of assembled and unassembled land. We model this comparison using two empirical strategies: a comparison of land values, where identification comes from within-neighborhood differences; and a repeat sales approach which controls for parcel-specific attributes.

To estimate surplus, $s$, we must recover the land value of both assembled and unassembled parcels. In our first specification, we focus on recovering, $V(p)$, the land value of an unassembled parcel, by using the technique pioneered by Rosenthal and Helsley (1994) and

[14]We view this as very suggestive evidence of private market failures, but acknowledge that there are also public market frictions (zoning changes, regulatory hurdles) that could be dampened by assembling fewer parcels.

refined by Dye and McMillen (2007). This technique recovers the value of land using the sales of homes to be torn down. The value of such a teardown sale reflects only the value of the underlying land, since the capital is discarded.

We apply a similar logic to value the land used in assemblies. Most assemblies discard the existing capital and place new capital on the assembled site to take advantage of the larger building area. Intuitively, if the capital on the initial parcels were retained, there would be no gain from assembly. As a result, we assume that to-be-assembled parcels are teardowns and also use their sales price as a measure of the value of the land. We therefore use the pre-assembly sale price to recover the value of assembled land, $V(2p) - \delta$, and estimate any incremental value that to-be-assembled parcels command relative to teardown parcels. The larger any incremental value, the larger the surplus earned by landowners selling into an assembly and the greater the frictions in the market for assembly.

One possible objection to the strategy of comparing to-be-assembled parcels to unassembled parcels is that many plots are constrained from assembly. For instance, physical barriers such as steep slopes may prevent assembly. Public capital, such as roads, may separate parcels and prevent assembly. A parcel ready for redevelopment may be next to a parcel with new or valuable capital, making redevelopment as part of an assembly economically infeasible. These factors are reasonably viewed as materially different from factors such as zoning and holdouts which may also prevent assembly. It would be unreasonable to label the failure to assemble two parcels separated by a road as due to market "frictions." However, arbitrage opportunities should cause the price of assembled and unassembled teardown parcels to converge as long as a corner solution is not reached in which no feasible assemblies exist. In other words, as long as there are available assembly opportunities, arbitrage should drive the assembly surplus to zero. At least in Los Angeles it seems clear that ample assembly opportunities remain and that a corner solution has not been reached (Landis and Hood, 2005).

A related potential objection concerns reservation prices. A particular individual may have a reservation price that exceeds the market price for his parcel if, for instance, he has a strong emotional attachment to his home. If a sufficient number of owners have reservation prices that are so far above market prices that assemblies fail to occur, and thus the market fails to clear, we may misinterpret a positive surplus estimate as due to frictions when it correctly reflects high reservation prices. [15] We believe that this is not likely to be an impediment to our estimates for two reasons. First, while individuals with elevated reservation prices undoubtedly exist, the test requires only that there are assembly opportunities, not that all parcels can be assembled. Second, commercial properties are the most likely to sell into assembly in our data. Commercial owners are plausibly less likely than home owners to have reservation prices significantly above market prices.

Our test relies on the co-existence, in geographically comparable areas, of both teardown and assembly parcels. Figure 3 shows census tracts in Los Angeles County, and marks areas that ever (1999 to 2010) have both at least one assembly and one teardown; areas that ever have only teardowns; areas that have only assemblies; and areas that have neither. Rough three-quarters of all tracts have an assembly over the period of study; this figure is 80 percent for tracts with teardowns. Roughly 65 percent of all tracts have at least one assembly and at least one teardown over the period of study.[16] The map shows that these redevelopment areas are located widely across the county.

The empirical model for the primary test is

$$\begin{aligned} \log(\frac{\textit{real sale price}}{\textit{lot square footage}})_{i,g,t} &= \alpha_{10} + \alpha_{11} \textit{assembly}_i + \textit{year*quarter}_t \\ &+ \textit{neighborhood}_g + \alpha_{12} \textit{amenities}_i + \varepsilon_{i,g,t}, \end{aligned} \tag{4}$$

[15] The reservation prices would need to extremely elevated. They would need to exceed not only the market price of the parcel in its current use, but also the surplus available from selling into assembly.

[16] If we restrict analysis to tracts with qualifying *sales* of assemblies or teardowns, 16 percent of tracts have assembly sales, 61 percent have teardown sales; 12 percent of tracts have both.

where $\frac{real\ sale\ price}{lot\ square\ footage}_{i,g,t}$ is the per-square foot price of land for parcel $i$ in neighborhood $g$ at time $t$ and $assembly_i$ equals one for an assembly and zero for a teardown. The estimation sample includes only teardown and assembly sales. Specifically, we choose the sample in the following way. Each assembly has a start and end year (which may be the same, but are frequently not). We choose all parcels that are "input" parcels to assembly. Of these, we keep sales that are no more than three years before the beginning of the assembly, no more than three years after the beginning of the assembly, and before the end of the assembly. Of these sales, we keep only the last sale on each parcel. We use this window to keep only sales "close" to the time of assembly, which should be sales that reflect the value of land.

With this sample, $\alpha_{11}$ captures the surplus to assembly relative to redevelopment within the existing parcel boundaries. If $\alpha_{11} = 0$, there is no surplus to assembly, consistent with an absence of market frictions.

Assembly may be correlated with unobservable determinants of price, $\varepsilon_{i,g,t}$, for at least two reasons. First, rising land values often dictate increasing the capital-to-land ratio and assembly may be required to increase this ratio. Thus, land with a particularly high value may be more likely to assemble than less valuable land (e.g. there may be a reverse causality problem). Second, some of the very frictions we are attempting to quantify, such as zoning, may make high value land less likely to assemble. For instance, more stringent zoning in Malibu may make assembly less likely there than in Watts. Because land is more valuable in Malibu than in Watts for reasons having nothing to do with the likelihood of assembly, this may introduce bias.

We tackle the above endogeneity concerns by observing that the value of land, virtually by definition, is a function of location. We therefore include a very fine set of geographic fixed-effects, either census tract indicators or census block group indicators, $neighborhood_g$. The comparison of land price between assemblies and teardowns is therefore made only within very small areas, either a census tract or census block group. There are 2,054 census

tracts and 6,346 block groups in Los Angeles County. The median tract contains 985 parcels, and the median block group 290.

Of course, some elements of location vary even within small geographic areas. For instance, access to a highway may differ within a neighborhood. We therefore control for distance to the nearest major highway, and major road, the parcel-specific elevation, and the shortest distance to urban and commuter rail with the $amenities_i$ vector. Finally, to control for market-wide evolution in price over time we include a full set of indicators for each quarter in our sample, $year*quarter_t$ (i.e., indicators for 1999Q1, 1999Q2, etc, are included).

Our approach departs from the teardown literature in one important regard. The teardown literature controls for selection into redevelopment using a standard Heckman two-stage procedure. Accordingly, coefficient estimates yield marginal effects for the untruncated, latent dependent variable. As the goal of the teardown literature is to recover the value of all land, whether redeveloped or not, this is the correct approach. Because our aim differs, we estimate equation (4) using OLS. Two points bear emphasis. First, the OLS coefficients recover marginal effects for the truncated, observed dependent variable. These effects are precisely those required for the primary test. Specifically, the OLS estimate of $\alpha_{11}$ answers the question: among redeveloped parcels (i.e. in the observed portion of the distribution), is there excess value to assembly? The arbitrage argument underlying the test is only valid for parcels actually undergoing redevelopment. Arbitrage should not eliminate the surplus to assembly for parcels whose condition makes redevelopment suboptimal (e.g. those containing capital too valuable for redevelopment). Second, although the conditional expectation function is non-linear, OLS provides a well-defined minimum mean squared error linear approximation.[17]

This first empirical model is appealing because it plausibly identifies land value. The

[17] Angrist and Pischke (2009, page 102-3) and Cameron and Travedi (2005, page 542) discuss using OLS to fit the conditional expectation function of a left-truncated distribution.

second method we use, a repeat-sales approach, has alternate strengths. We estimate

$$\begin{aligned}\log(\frac{\textit{real sale price}}{\textit{lot square footage}})_{i,t} &= \beta_0 + \beta_1 1\{\text{redevelop}\} * 1\{\text{ever assembly}\}_{i,t} \\ &+ \beta_2 1\{\text{redevelop}\} + \beta_3 K_{i,t} + \beta_4 K_{i,t} * 1\{\text{redevelop}\} \\ &+ \textit{parcel}_i + \textit{year} * \textit{quarter}_t + \varepsilon_{i,t}, \end{aligned} \quad (5)$$

where $K_{i,t}$ is a measure of capital on parcel $i$ at time $t$ and $1\{\text{redevelop}\}$ equals one for the final sale immediately before either an assembly or teardown. Here, we restrict the sample to parcels from the single difference test that have at least one additional sale before "redevelopment" sale used in the previous estimation. Thus, $1\{\text{redevelop}\} * 1\{\text{ever assembly}\}_{i,t}$ equals one if this sale is the final sale before assembly, and $\beta_1$ is the coefficient of interest.

Many of the things that required controls in the previous estimation – the size of the parcel, time-invariant neighborhood features such as greenery and freeway access, and even intra-neighborhood time-invariant characteristics such as elevation and proximity to commercial strips – are now subsumed in the parcel fixed effect. This repeat sales estimation can be viewed as a double difference: relative to the same parcel over time, and relative to teardowns, is there a premium for being the last sale before an assembly?

However, this estimation does not clearly isolate land value. Abstracting from the neighborhood and amenity controls, equation (4) estimates the single difference $P_{A2} - P_{T2}$, where $P$ is price, 2 stands for the redevelopment sale (i.e. the last sale before assembly or teardown), and $A$ and $T$ denote an assembly and teardown, respectively. Equation (5) estimates the double-difference $(P_{A2} - P_{A1}) - (P_{T2} - P_{T1})$, where 1 denotes sales which occur before the redevelopment sale. Decomposing price, P, into its land, L, and capital, K, components, the double difference becomes

$$[L_{A2} - (L_{A1} + K_{A1})] - [L_{T2} - (L_{T1} + K_{T1})] \quad (6)$$

As before, the redevelopment sales, denoted by 2, reflect only the value of land. Rearranging terms, the estimate becomes

$$(L_{A2} - L_{A1}) - (L_{T2} - L_{T1}) + (K_{T1} - K_{A1}) \tag{7}$$

Our object of interest in the equation above is the the first two terms in parentheses: the difference in land prices across the pairs of assembly and teardown sales. Note that the regression specification nets out any time-invariant element of sale price associated with a given parcel and any land premium common to redevelopment. However, the double difference in equation (7) includes capital terms. $\beta_1$ an unbiased estimator of our object of interest when $K_{T1} - K_{A1} = 0$, which is when the capital on the assembled and teardown parcels, at the time of sales prior to redevelopment (in time period one), are equal. We underestimate $\beta_1$ if if $K_{A1} > K_{T1}$, and overestimate when $K_{T1} > K_{A1}$. We account for this by including a set of controls in equation (5) for capital, $K_{i,t}$, and capital on the final sale ($K_{i,t} * 1\{\text{redevelop}\}$).

## 2.2 First Additional Test: Smaller parcels more likely to assemble

For the first additional test – in a well functioning market, smaller parcels should be more likely to assemble – we use the full cross-section of *all* county parcels from the first year of the sample (1999). Specifically, we estimate

$$\begin{aligned} assembly_{i,g} &= \gamma_0 + \gamma_1 \log(lot\ square\ footage_i) + neighborhood_g \\ &+ \gamma_2 amenities_i + K\ vintage_i {*} K\ quantity_i + \gamma_3 \frac{K\ value_i}{land_i} + \varepsilon_{i,g}, \end{aligned} \tag{8}$$

where $assembly_{i,g}$ equals one if the parcel is involved in an assembly over the following 11 years of the sample and zero otherwise, and $lot\ square\ footage_i$ is the size of the parcel. $\beta_{21}$

captures the marginal effect of lot size on the probability of assembly and is the coefficient of interest.

$\gamma_1$ is properly identified when lot size is uncorrelated with unobserved determinants of assembly, $\varepsilon_{i,g}$. There are at least three possible reasons to be concerned that this may not be the case. First, the likelihood of assembly may vary by location, and location may be correlated with parcel size. For example, parcels near the urban center may be ripe for redevelopment and assembly due to increased commute times. These parcels may also be smaller, but it is their lower commute time, not their size, driving the assembly decision. As in the first test, we net out such neighborhood-specific factors with tract or block group fixed effects, so that all our comparisons are made within a given, limited geographic area.

The second threat to the credibility of the identifying assumption is the possibility that large plots tend to have more valuable capital per square foot of land than smaller parcels. All else equal, the more valuable the capital on a parcel, the less likely the capital is scrapped to allow for redevelopment and the less likely is assembly. Larger lots may very well have more valuable capital. For instance, if small lots tend to contain two-story single family homes, whereas large lots tend to contain multi-story buildings, larger lots will have systematically more capital. We therefore control for the presence of capital in a flexible, relatively non-parametric manner by including a full set of interaction terms between indicator variables for the vintage of the existing capital, $K\ vintage_i$, and indicator variables for the size or quantity of the capital per square foot of land, $K\ quantity_i$. This approach fails to account for both differences in depreciation across properties and differences in the initial quality of the capital given its size. We therefore also control for the ratio of the value of improvements (i.e. capital) to the value of land, $\frac{K\ value_i}{land_i}$. A third reason lot size may be correlated with the error term is because not all lots are "assemble-able," and "assemble-ability" may be correlated with lot size. To be assembled, a parcel needs a contiguous neighbor, one not separated by an existing road or other physical barrier. The most extreme example of an

"un-assemble-able" parcel would be one that takes up an entire city block. Such a parcel cannot be assembled given the existing structure of roads. In general, larger parcels should be more likely to be unable to assemble for such reasons. However, this possibility will tend to bias our estimation toward *failing* to reject the hypothesis that smaller parcels are more likely to be assembled (and hence make it less likely we will conclude the market for assembly is being inhibited by frictions).

## 2.3 Second Additional Test: Larger parcels sell for more per square foot

We implement the second additional test – in a well functioning market, small parcels should not sell into assembly at a premium – with a panel specification

$$\begin{aligned} \log\left(\frac{\textit{real sale price}}{\textit{lot square footage}}\right)_{i,g,t} &= \theta_0 + \theta_1 \log(\textit{lot square feet}_i) + \textit{year*quarter}_t \\ &+ \textit{assembly group}_g + \theta_2 \textit{amenities}_i + \varepsilon_{i,g,t}, \end{aligned} \tag{9}$$

where $\textit{assembly group}_g$ is an indicator variable for a set of contiguous parcels which are assembled together. (We provide a more precise definition below.) $\theta_1$ captures the marginal effect of lot size on sales price and is the coefficient of interest. The sample is limited to the last sale of parcels involved in an assembly and the inclusion of the $\textit{assembly group}_g$ term ensures the influence of lot size on sales price is measured solely *within* groups of parcels assembled together. The identifying assumption required for equation (2) – that lot size is uncorrelated with the unobserved determinants of price – is therefore extremely plausible. To-be-assembled parcels are sold only for their land value. Within an assembly group the parcels are contiguous to each other and thus have the same locational, or land, value.

## 3 Land Assembly Definition and Institutions

Before estimation, we discuss the institutions of land assembly and our data. This section presents our empirical definition of land assembly and then discusses the institutions for assembly in Los Angeles County.

To define land assembly in our dataset, we first define a "parcel change group." A parcel is part of a parcel change group if it or its ancestor or descendant parcel(s) ever changes. However, parcels combine and disaggregate in a number of different ways, so membership in a parcel group does not alone indicate assembly. The list below presents three examples of possible parcel changes.

A. $3 \rightarrow 1 \rightarrow 7$

B. $1 \rightarrow 5$

C. $8 \rightarrow 1 \rightarrow 2$

We define a change group to have land assembly if any part of the parcel change group goes from $n > 1$ properties to one. In the example above, this includes cases A and C, but excludes case B. Case C involves land assembly (8 parcel into 1 parcel) followed by disassembly (1 parcel into 2 parcels). We use this definition, despite the fact that it encompasses episodes of net disassembly, because in most cases the redevelopment cannot occur without the assembly step. For instance, a developer may wish to convert multiple detached single family home parcels into denser townhouse style homes. The lots first need to be combined before they can be properly subdivided into the smaller townhouse lots.

In addition, this definition is a good match with our data. Parcel identification numbers change only when the physical boundaries of a parcel change. We observe all changes in parcel numbers and thus measure all assemblies as defined above.

When is the type of land assembly we have defined required? Regulations for land assembly are the province of cities, or the county for unincorporated areas. In general,

cities do not require developers to assemble parcels, even when a new structure spans more than one parcel.[18] However, there are two circumstances under which formal assembly is required.[19] The first of two exceptions is when the new land use will be condominiums. Each unit in a condominium must have a separate parcel, as each unit may have a unique legal owner. Therefore, any land combined for condos must be assembled.

The second exception from the laissez-faire policy is a function of the use of the property. Suppose that a city's zoning requires two parking spaces for each multi-family unit, and that the developer has purchased two parcels upon which to build a new multi-family development. If the developer builds the parking on one parcel and the structure on the other, he is required to legally assemble the parcels. Cities make such a requirement to ensure that all future sales keep parcels in compliance with zoning regulations. The developer can avoid the requirement to legally assemble by building a structure that spans both parcels.

Outside of these two exceptions, a developer may purchase adjoining land with the intent of building new structures but not go through the formal process of legal assembly.[20] It is this legal assembly which we observe in the data. For the purposes of our estimation, this type of underreporting likely biases our estimates of the prevelance of frictions downward, by mistakenly putting some assembled parcels in the teardown comparison group. Interviews with practitioners, however, suggest that legal assembly is very likely for projects requiring financing, since assembly lessens the financier's cost in the case of default (interview citations appear after the bibliography). In addition, selling an assembled parcel, rather than multiple unassembled parcels, reduces paperwork and uncertainty in future transactions. While there are substantial benefits to legal assembly, the costs, in terms of administrative burden and

[18]Information in this section comes primarily from an interview with Wolfgang Krause, Chief Planner, City of Glendale, May 2010. We plan to improve this section through additional interviews.

[19]To the best of our knowledge, these two circumstances do not vary across regulatory jurisdictions in the county.

[20]In such cases, with multiple owners, owners usually write legally binding easement agreements across properties.

fees are exteremely low.

Importantly, the legal land assembly process does not trigger a reassessment under California's Proposition 13. Proposition 13 limits the increase of a property's assessed value to two percent per year, and the assessed value raises to market value at sale. A developer may face a increase assessment due to property purchase, but does not face an increased assessment due to the legal act of assembly.[21] Thus, our interviews suggests that the combination of substantial benefits and low costs results in developers choosing legal assembly over informal assembly in most cases.

# 4 Data

Our project relies on multiple sources of data. We summarize the data here, and refer interested readers to our lengthy Data Appendix for full details on all data inputs and data construction details. The three key components of our data are the annual property-level data for Los Angeles County, sales data for properties, and census neighborhood measures.

Our annual property data consist of three key parts: (i) ten annual cross-sectional observations of the 2.2 million parcels in the County of Los Angeles, (ii) a dataset listing all parcels that change, and the number of the parcel(s) to which they change, and (iii) electronic maps with geographic information on all properties.

The annual cross-sections are the heart of the dataset. In each year from 1999 to 2010 (except for 2003[22]) we observe attributes about each individual piece of property in the 88 cities and the large unincorporated area of Los Angeles County. We observe too many attributes to list here, but briefly the data include attributes about the property itself (e.g., size and location); attributes about the building on the property (e.g., building size); and attributes about the legal regime that governs the property (i.e., the use and zoning rules

[21] This information comes from the Special Investigations Section, Los Angeles County Assessor.
[22] We have been unable to obtain this cross-section.

for each property). Thus, this part of the dataset includes somewhat more than 24 million observations with many descriptive variables.

The second part of the data is a file that allows us to take the 11 cross-sections and make them into a true panel by linking property identification numbers over time. Though most properties retain a constant identification number throughout the sample, some properties split or merge. Our dataset of all property identification number changes allows us to follow each initial piece of land to its current, perhaps aggregated or disaggregated, form. While this task is conceptually simple, in practice it has been exceedingly difficult, and the bulk of our data assembly has been devoted to making sure that we have built these linkages correctly.

The third and final part of the annual property data is electronic maps of all parcels. These maps, which we have from 2006 onward, allow us to pinpoint the exact location of each individual property and calculate distances from one property to another, or from a given property to key urban amenities, such as light rail stops or freeways. These maps also allow us to assign each property to a unique census block group.[23]

We combine this panel of all properties in the county with all property transactions by property identifier. Specifically, we observe the last three sales on each property as of 2006, and sales in the last two years from 2009 and 2010. This leaves a small gap of sales in 2006. We limit the sample of transactions to include only arms' length transactions and make other small adjustments as defined in the Data Appendix.

We measure neighborhood economic and demographic factors with data from the 1990 and 2000 Decennial Censuses at the block group level. To use the 1990 block group data, we use GIS mapping to make a correspondence from 1990 to 2000 census block groups.

[23]On average, populated block groups in Los Angeles contain approximately 1,400 people.

# 5 Results

We now turn to describing the results of each of our tests in turn.

## 5.1 Primary Test: Surplus Value of Assembly

We motivate our comparison of assembled to teardown parcels by illustrating their characteristics. As a reminder, the sample is sales for assembly and teardown parcels sold before and close to the date of assembly or teardown.[24] We define a property as a teardown if the structure's age changes in our panel. Specifically, we require that the replacement structure be newer than the old structure, that the new structure is built after 1998, and that the old structure was built before 1990.

Columns (1) and (2) of Table 1 present summary statistics for our sample of assembly and teardown parcels, respectively, while column (3) displays the mean difference between these types of parcels. Assemblies and teardowns differ on many observable dimensions, a potential problem for our primary test given its requirement that assemblies and teardowns be comparable on all unobserved determinants of price other than assembly status. While we could control for these covariates in equation (4), the systematic differences raise the possibility that assembly and teardown parcels differ on unobserved dimensions.

Column (4), however, presents regression-adjusted mean differences conditional on census tract fixed effects.[25] In this and all following regressions, we report standard errors clustered at the tract level. With the comparison between the two parcel types restricted to within small neighborhoods, differences are very small and imprecise. The notable exception is the probability of being a single family parcel. Overall, we view the results of column (4) as supportive of our identification strategy for the single-difference estimation which relies

[24]Although we observe an exact date for the sale, we observe only the year for teardown or assembly.

[25]The demographic characteristics, such as the poverty rate, are from the census and are measured at the block group level. The census tracts in the sample on Table 1 have an average of 2.5 block groups within them.

on geographic fixed effects to control for unobservable differences between the two types of parcels. The right half of the table repeats this exercise for neighborhood demographic variables which are expressed as the level difference between 1990 and 2000 values. The same pattern of no difference conditional on tract-level fixed effects holds.

Given the results in Table 1, we take care below to address the difference in single family status. Most specifications include a control for use type. We also present specifications that restrict the sample to only single family parcels or non-residential parcels to make teardowns and assemblies as comparable as possible.

Table 2 presents the results for our primary test (absent frictions, the price of teardowns and assembly parcels should equate), implemented by estimating equation (4) on teardown and assembly sales. The coefficient estimate in column (1) indicates that being in an assembly is associated with an almost 50 percent sales price premium relative to being sold for redevelopment without a change in parcel boundaries. The extremely large magnitude of the assembly surplus suggests substantial frictions in the market for assembly.

Columns (2) - (5) present alternative specifications. Column (2) addresses within neighborhood variation by adding neighborhood demographic controls at the block group level and controls for local amenities. Column (3) further addresses intra-neighborhood variation by controlling for very finely grained neighborhood effects. Instead of tract fixed effects, we use block group fixed effects. Column (4) controls for the evolution in price specific to each tract by using tract fixed effects and tract-level linear trend terms. Column (5) includes a set of municipality-year fixed effects to control for evolution in price specific to individual municipalities. Regardless of specification, the magnitude of the surplus estimate changes only slightly.

The remaining columns use different subsets of the sample. Column (6) includes assemblies only if we observe that the existing capital is torn down and replaced with new capital following assembly. It is possible that assemblies may occur with the aim of redeveloping at

some point in the future, not immediately. If so, the sales price of the to-be-assembled parcels reflects the return to any existing capital over the period before redevelopment (McMillen and O'Sullivan, 2011). Such a scenario would bias our surplus estimates upward. In this case, estimates in Columns (1) to (4) should be interpreted as upper bounds.

However, we can identify assemblies that lead to immediate new construction. Due to the nature of our data, we under-identify these parcels. For a given assembly, pre-assembly, we observe attributes for one structure per parcel, though a parcel may, in fact, contain more than one structure. For each parcel change group before assembly, we find the minimum and maximum of the age of the structure on each parcel: $\max(age_{c,before})$ and $\min(age_{c,before})$ ($c$ indicates a parcel change group). We observe similar ages post-assembly: $\max(age_{c,after})$ and $\min(age_{c,,after})$. We define parcels as being assembly teardowns only if $\max(age_{c,before}) < \min(age_{c,after})$. The magnitude of the surplus estimate is similar to the unrestricted sample.

To address the concern that assembled parcels are systematically less likely to be single family than are teardowns, Columns (7) and (8) restrict the sample to single family parcels only and report estimates similar to that produced by the full sample. Columns (9) and (10) restrict the sample to non-residential properties only. The results are again similar to those produced by the full sample.

Table 3 presents the repeat sales version of the comparison between assembled and teardown parcels. The data requirements of the approach yield a somewhat thinner sample than was used on Table 2 (see the observation rows). The first column presents the results from a specification with parcel fixed effects, year by quarter fixed effects, use controls, and a redevelopment premium. This specification reports that, relative to the previous assembly sale, and relative to changes in teardown prices, the final sale before an assembly receives a 54 percent price premium per square foot.

Column (2) adds four controls for capital: structure square feet per unit of lot size, assessed improvement value divided by assessed land value, assessed improvement value per

unit of lot size and the year the structure on the property was built. The estimate is little changed. While capital should influence sales price before the redevelopment (final) sale, it should not influence the price of the redevelopment sale. Column (3) therefore allows the coefficient on the capital variables to differ for the redevelopment sale. While the estimate falls somewhat, it remains a very sizeable 35 percent.

A drawback of our data is that, while we use sales from 1985 onward, we only observe parcel characteristics from 1999 onward. Columns (1) and (3) use the 1999 parcel characteristics for all pre-1999 sales. We address whether this biases our results by restricting ourselves to only sales from 1999 onward in column (4). This cuts the sample of unique assembly parcels by almost 2/3, and the sample of unique teardown parcels by about 4/5. Regardless, we still find an assembly premium of nearly 40 percent per square foot.

Broadly, we construe the results from our primary test from Tables 2 and 3 as robust support for the existence of an economically significant premium paid to assemble land.

## 5.2 Additional Test One: The Influence of Lot Size on the Probability of Assembly

Our first additional test states that in a market free of frictions, developers should prefer to assemble larger parcels. Table 4 presents the results for this test, implemented by estimating equation (8) on the 1999 cross-section of all parcels.[26] The coefficient in column (1) indicates that a 10 percent increase in the size of a parcel increases the probability of ever being assembled by 0.1 percent. This is a large effect, equal to 10 percent of the sample mean probability of ever assembling (see the bottom row of the table). The estimate also indicates that the elasticity of the probability of assembly with respect to lot size is roughly 1.

Columns (2) - (4) present specification permutations which result in very little change in the lot size coefficient. In column (2), the estimates are robust to conditioning on the

[26]Summary statistics for this sample are in Appendix Table 1.

extremely finely grained block group fixed-effect. To better purge intra-neighborhood variation, Columns (3) and (4) add parcel- and neighborhood-specific covariates. Parcel-specific covariates are dummies for use type (four categories), the improvement to land ratio (from 1999 reported assessed values for each) and the distance to key amenities as above. The improvement to land ratio is measured with error, as Proposition 13 caps increases in assessed values to 2 percent per year until sale. Neighborhood demographics are as in the primary test. As neighborhood demographics are observed at the block group level, we omit them in all specifications with block group fixed effects.

Across specifications, and despite the fact that the model argues that the surplus to assembly is largest for small parcels, large parcels are more likely to assemble. This evidence is consistent with the theory that holdouts and strategic delay impede the market for assembly.

The $K\ vintage_i * K\ quantity_i$ terms (full interactions of decile dummies for structure size per lot square foot and structure age) are not included in columns (1) - (4). We are interested in results conditional on these variables, because they are a relatively non-parametric method of capturing the quantity and quality of capital per parcel. Including these terms, which are not available for all parcels, truncates the sample. For comparison purposes, columns (5) and (6) replicate columns (3) and (4) using the truncated sample without adding the $K\ vintage_i * K\ quantity_i$ terms. Although around one-third of the assemblies are lost from the sample – see the bottom row – the lot size coefficient is essentially unchanged. Columns (7) and (8) add in the $K\ vintage_i * K\ quantity_i$ terms and again the coefficient is little altered.

Table 5 presents robustness checks based on columns (7) and (8) of Table 4. To examine whether effects differ by initial use, consistent with the concern raised in Table 1, columns (1) and (2) restrict the sample to only single family parcels, while columns (3) and (4) restrict the sample to only non-residential parcels. The marginal effect of lot size on the likelihood of assembly is remarkably consistent across these use categories. The remaining columns exclude the least dense block groups in order to ensure our results are driven by

outcomes in a dense urban setting. The parcels excluded are generally located in outlying areas of the county where land subdivision and new development, as opposed to assembly and redevelopment, are likely more prevalent. Again, the results are little changed.

### 5.3 Additional Test Two: Influence of Lot Size on Price of To-Be-Assembled Parcels

We now move to our second additional test: in a market free of frictions, small parcels should not sell into assembly at a premium. Table 6 presents the results for this test, implemented by estimating equation (9) on the same sample as the single-difference primary test, excluding teardowns. The result in column (1) indicate that a 10 percent increase in parcel size reduces the sales price of a to-be-assembled parcel by roughly 6 percent. Alternatively, the results suggest an elasticity of price with respect to lot size of around 0.6. Column (2) permits the coefficient on parcel size to vary with use type of the parcel. The evidence suggests that both residential and non-residential parcels have a similar elasticity. The substantial premium to small parcels supports the hypothesis that owners of small parcels of land tend to hold out and demand higher than average prices for their land – behavior which works to reduce the number of successful assemblies.

## 6 Conclusion

In sum, we provide robust confirmation of the hypothesis that the market for land assembly is inhibited by frictions. Our evidence on the premium paid to small parcels suggests that at least part of these frictions are due to private market failures; this does not rule an important role for governmental regulation of land. These findings form an economic rationale for the benefits of eminent domain. To understand the net effect of eminent domain, we need additional research on its costs.

If the market for private land assembly operates as poorly as we suggest in this work, it is natural to look to government for remediation. The most recent large-scale government action to assemble land is known as "urban renewal," a process in the United States, Canada and in some parts of Europe in the 1960s and 70s. Urban renewal used the government's power of eminent domain to assemble small parcels of land, predominantly in the urban core. While readers may associate renewal with tall towers of public housing, this was not its predominant output. Urban renewal generated substantial high and middle income housing, as well as non-residential construction. Copley Place in Boston, and the "new downtown" atop Bunker Hill in Los Angeles are both examples of urban renewal. Urban renewal ended amid charges of developer cronyism and racism, has largely been judged harshly by historians (Cord, 1974) and undoubtedly has significant costs. It may also yield significant benefits, though, given the frictions in the private market for land assembly documented in this paper.

Alternatively, there may be a middle ground between the private exchange of land and direct government intervention. Under the practice of "land readjustment", redevelopment districts are formed and land assembly occurs if a majority of land owners in the district vote for it. The owners are given a stake in the new development as compensation. The practice has been successfully used in a number of countries, including Germany, South Korea, and Israel (Hong, 2007). Land readjustment may provide a mechanism for overcoming the private market failures inherent in land assembly, while mitigating the costs associated with direct government intervention.

Another possible middle ground approach is graduated density zoning which awards the right to build at greater density to larger lots. The "density bonus" increases the value of large lots and encourages assembly. Although it does not eliminate holdouts, it may generate a fear of being left out and therefore reduce the prevalence of holdouts (Shoup, 2008).

More broadly, our evidence argues that "tragedies of the anti-commons," in which problems of fragmented ownership impede socially optimal outcomes, are of economically signifi-

cant magnitude. Specifically, for urban areas this presents a challenge to the canonical model of urban redevelopment. These dynamic urban models typically assume that land is redeveloped when the present value of capital is exceeded by the present value of redevelopment and that, similarly, when population growth renders existing population densities suboptimal, redevelopment occurs (e.g. Brueckner (1980b), Brueckner (1980a), Wheaton (1982) and Wheaton (1983)). The market frictions we document imply that urban redevelopment depends not just on capital, but crucially on the legal boundaries of land.

# References


Angrist, Joshua and Pischke, Jorn-Steffen, 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Asami, Y., 1988. "A game-theoretic approach to the division of profits from economic land development." *Regional Science and Urban Economics* 18: 233–246.

Bogart, Dan and Richardson, Gary, 2009. "Making property productive: reorganizing rights to real and equitable estates in Britain, 1660-1830." *European Review of Economic History* 13: 3–30.

Brownstone, David and DeVany, Arthur, 1991. "Zoning, Returns to Scale, and the Value of Undeveloped Land." *Review of Economics and Statistics* 73(4): 699–704.

Brueckner, Jan K., 1980a. "Residential Succession and Land-Use Dynamics in a Vintage Model of Urban Housing." *Regional Science and Urban Economics* 10: 225–240.

Brueckner, Jan K., 1980b. "A Vintage Model of Urban Growth." *Journal of Urban Economics* 8: 389–402.

Cadigan, John, Schmitt, Pamela, Shupp, Robert, and Swope, Kurtis, 2011. "The Holdout Problem and Urban Sprawl: Experimental Evidence." *Journal of Urban Economics* 69(1): 72–81.

Cameron, A. Colin and Travedi, Pravin K., 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Chang, Yun-chien, 2010. "An Empirical Study of Compensation Paid in Eminent Domain Settlements: New York City, 1990 - 2002." *Journal of Legal Studies* 39(201).

Chen, Daniel L. and Yeh, Susan, 2010. "The Economics Impacts of Eminent Domain." Mimeo.

Colwell, Peter and Sirmans, C. F., 1993. "A Comment on Zoning, Returns to Scale and the Value of Undeveloped Land." *Review of Economics and Statistics* 75(4): 783–786.

Colwell, Peter F. and Munneke, Henry J., 1997. "The structure of urban land prices." *Journal of Urban Economics* 14: 321–336.

Colwell, Peter F. and Munneke, Henry J., 1999. "Land Prices and Land Assembly in the CBD." *Journal of Real Estate Finance and Economics* 18(2): 163–180.

Cord, Steven, 1974. "Urban Renewal: Boon or Boondoggle?" *American Journal of Economics and Sociology* 33(2): 184–186.

Coulson, Edward, 1989. "The Empirical Content of the Linearity-as-Repackaging Hypothesis." *Journal of Urban Economics* 25: 295–309.

Cunningham, Christopher, 2010. "Estimating the holdout problem in land assembly." Federal Reserve Bank of Atlanta Working Paper.

Dye, Richard F. and McMillen, Daniel P., 2007. "Teardowns and land values in the Chicago metropolitan area." *Journal of Urban Economics* 61: 45–63.

Eckart, Wolfgang, 1985. "On The land assembly problem." *Journal of Urban Economics* 18: 364–378.

Fennell, Lee Anne, 2004. "Common Interest Tragedies." *Northwestern University Law Review* 98(3): 907–990.

Field, Alexander James, 1992. "Uncontrolled Land Development and the duration of the Depression in the United States." *Journal of Economic History* 52(4): 785–805.

Fu, Daniel P. McMillen, Yuming and Somerville, Tsur, 2002. "Land Assembly: Measuring Holdup." University of British Columbia Centre for Urban Economics Working Paper 02-08.

Glaeser, Edward and Gyourko, Joseph, 2002. "The Impact of Zoning on Housing Affordability." NBER Working Paper No. 8835.

Glaeser, Edward, Gyourko, Joseph, and Saks, Raven, 2005. "Why is Manhattan so Expensive? Regulation and the Rise in Housing Prices." *Journal of Law and Economics* XLVIII: 331–369.

Grossman, Zachary, Pincus, Jonathan, and Shapiro, Perry, 2010. "A Second-Best Mechanism for Land Assembly."

Harding, John P., Rosenthal, Stuart, and Sirmans, C. F., 2003. "Estimating Bargaining Power in the Market for Existing Homes." *Review of Economics and Statistics* 85(1): 178–188.

Heller, Michael A., 1998. "The Tragedy of the Anticommons: Property in the Transition from Marx to Markets." *Harvard Law Review* 111(3): 621–688.

Heller, Michael A. and Eisenberg, Rebecca S., 1998. "Can Patents Deter Innovation? The Anticommons in Biomedical Research." *Science* 280: 698–701.

Henderson, J. Vernon, 1977. *Economic Theory and the Cities.* New York: Academic Press.

Hong, Yu-Hung, 2007. "Assembling Land for Urban Redevelopment: Issues and Opportunities." In Yu-Hung Hong and Barrie Needham, (Eds.) "Analyzing Land Readjustment: Economics, Law and Collective Action," Cambridge, MA: Lincoln Institute of Land Policy.

Jacobs, Jane, 1961. *The Death and Life of Great American Cities.* New York: Vintage Books.

Kominers, Scott Duke and Weyl, E. Glen, 2010. "Concordance Among Holdouts." April.

Landis, John D. and Hood, Heather, 2005. "Ther Future of Infill Housing in California: Opportunities, Potential, Feasibility and Demand." Tech. rep., Institute of Urban and Regional Development, University of California, Berkeley.

Libecap, Gary D. and Lueck, Dean, 2011. "The Demarcation of Land and the Role of Coordinating Property Institutions." *Journal of Public Economics* forthcoming.

Libecap, Gary D., Lueck, Dean, and O'Grady, Trevor, 2010. "Large Scale Institutional Changes: Land Demarcation Within the British Empire." NBER Working Paper No. 15820.

McMillen, Dan and O'Sullivan, Arthur, 2011. "Option Value and the Price of Teardown Properties." Mimeo.

Menezes, Flavio and Pitchford, Rohan, 2004. "The land assembly problem revisited." *Regional Science and Urban Economics* 34: 155–162.

Miceli, Thomas and Segerson, Kathleen, 2007. "A bargaining model of holdouts and takings." *American Law and Economics Review* 9(1): 160–174.

Miceli, Thomas J and Sirmans, C. F., 2007. "The holdout problem, urban sprawl, and eminent Domain." *Journal of Housing Economics* 16: 209–219.

Munch, Patricia, 1976. "An Economic Analysis of Eminent Domain." *Journal of Political Economy* 84(3): 473–497.

Nelson, Arthur and Lang, Robert, 2007. "The Next 100 million." *Planning* 73(1): 4–6.

O'Flaherty, Brendan, 1994. "Land Assembly and Urban Renewal." *Regional Science and Urban Economics* 24: 284–300.

Pitchford, Rohan and Wright, Mark L. J., 2008. "Holdouts in Soverign Debt Restructuring: A Theory of Negotiation in a Weak Contractual Environment."

Plassmann, Florenz and Tideman, T. Nicolaus, 2008. "Efficient Urban Renewal Without Takings: Two Solutions to the Land Assembly Problem."

Rosenthal, Jean-Laurent, 1990. "The Development of Irrigation in Provence, 1700-1860: The French Revolution and Economic Growth." *Journal of Economic History* 50(3): 615–638.

Rosenthal, Stuart, 2008. "Old Homes, Externalities, and Poor Neighborhoods. A Model of Urban Decline and Renewal." *Journal of Urban Economics* 63(3): 816–840.

Rosenthal, Stuart and Helsley, Robert W., 1994. "Redevelopment and the Urban Land Price Gradient." *Journal of Urban Economics* 35: 182–200.

Shigley, Paul, 2007. "Round 3 for eminent domain." *Planning* 73(3): 11–15.

Shoup, Donald, 2008. "Graduated Density Zoning." *Journal of Planning Education and Research* 28(2): 161–179.

Strange, William C., 1995. "Information, Holdouts and Land Assembly." *Journal of Urban Economics* 38: 317–332.

Tanaka, Tomomi, 2007. "Resource Allocation with Spatial Externalities: Experiments on Land Consolidation." *The B.E. Journal of Economic Analysis and Policy* 7(1): article no. 7.

Wheaton, William, 1982. "Urban Spatial Development with Durable but Replaceable Capital." *Journal of Urban Economics* 12: 53–67.

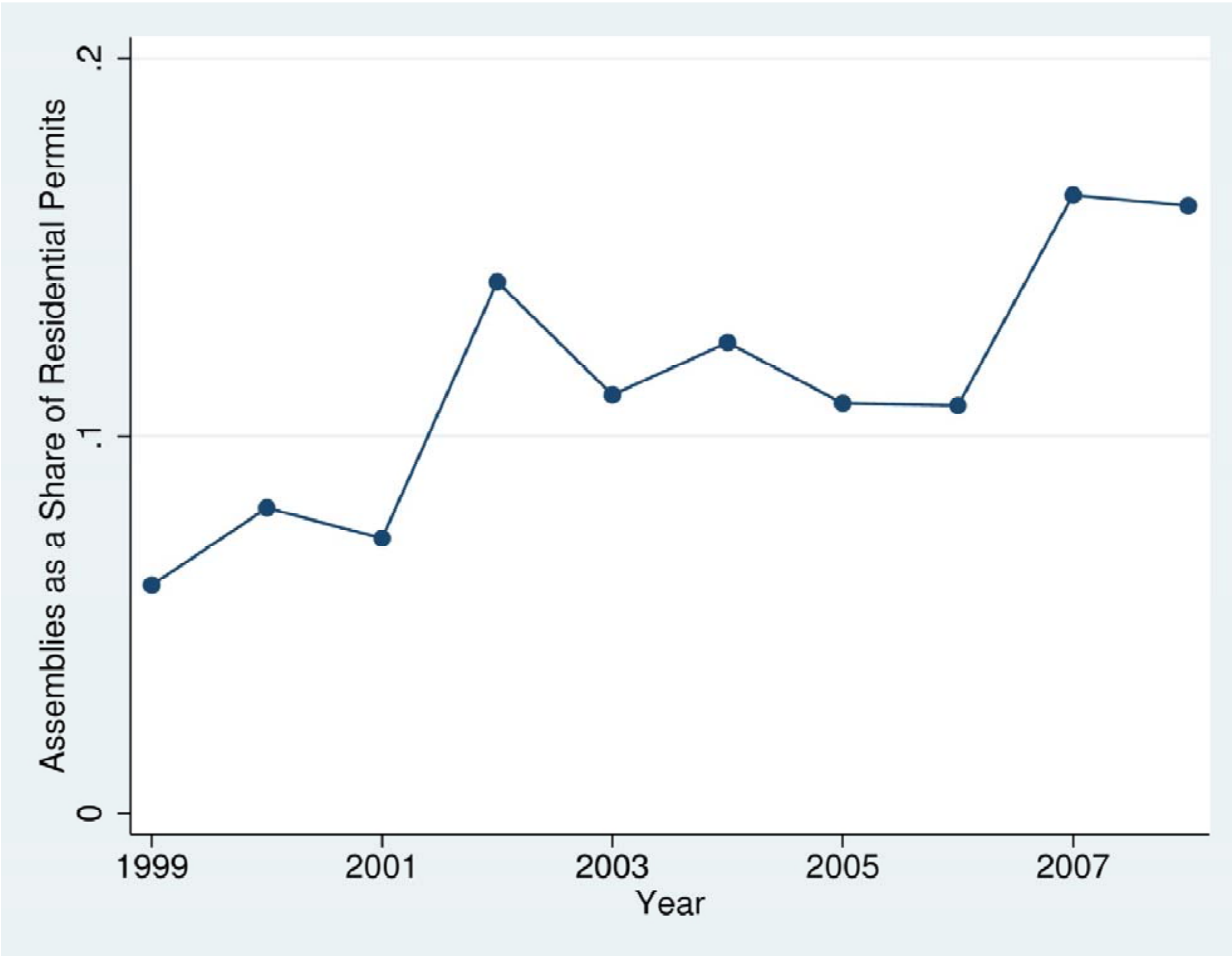
Wheaton, William, 1983. "Theories of Urban Growth and Metropolitan Spatial Development." In J.V. Henderson, (Ed.) "Research in Urban Economics, Vol. 3," Greenwich, Conn.: JAI Press.

Willis, Carol, 1995. *Form Follows Finance.* New York: Princeton Architectural Press.

**Interviews**

Krause, Wolfgang. Principal Planner, City of Glendale. By phone May 18, 2010.
Garcia, Jose. Planner, City of Maywood. By phone July 2, 1010.
Ashworth, Paul, Director of Planning and Development, City of Santa Fe Springs. In person, November 30, 2010.
Morrell, Wayne M., Principal Planner, City of Santa Fe Springs. In person, November 30, 2010.
Fuentes, Al, Redevelopment Consultant, City of Santa Fe Springs. In person, November 30, 2010.
Pastucha, Simon, Planner, Los Angeles Urban Design Studio. In person, December 1, 2010.
Olson, Todd, President, Community Development, The Olson Company. In person, December 2, 2010.

**Figure 1: Ratio of Assemblies to Permits**




Sources: Permit data from Census Bureau, assemblies calculated from authors' dataset. See Data Appendix for details.

Notes: While the Census Bureau counts one permit for each new unit, one assembly (in our terms) may result in more than one unit. This means that our measure in the chart understates the importance of assembly. However, the Census Bureau counts residential permits only, while our measure of assemblies includes assemblies that result in non-residential construction. This leads to this measure overstating the importance of assembly. Because we do not have a cross-section of properties from 2003, we all assemblies in 2002 and 2003 are attributed to 2002; for the purposes of this chart, we split the assemblies in 2002 evenly between 2002 and 2003. Though our assembly information continues through 2010, this chart ends in 2008 when our permit data ends.

**Figure 2: The Price of Land is Convex in Lot Size**

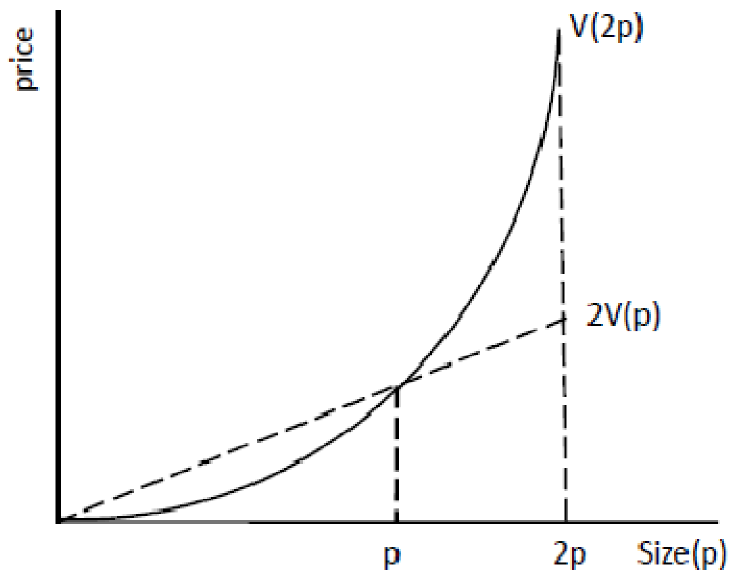
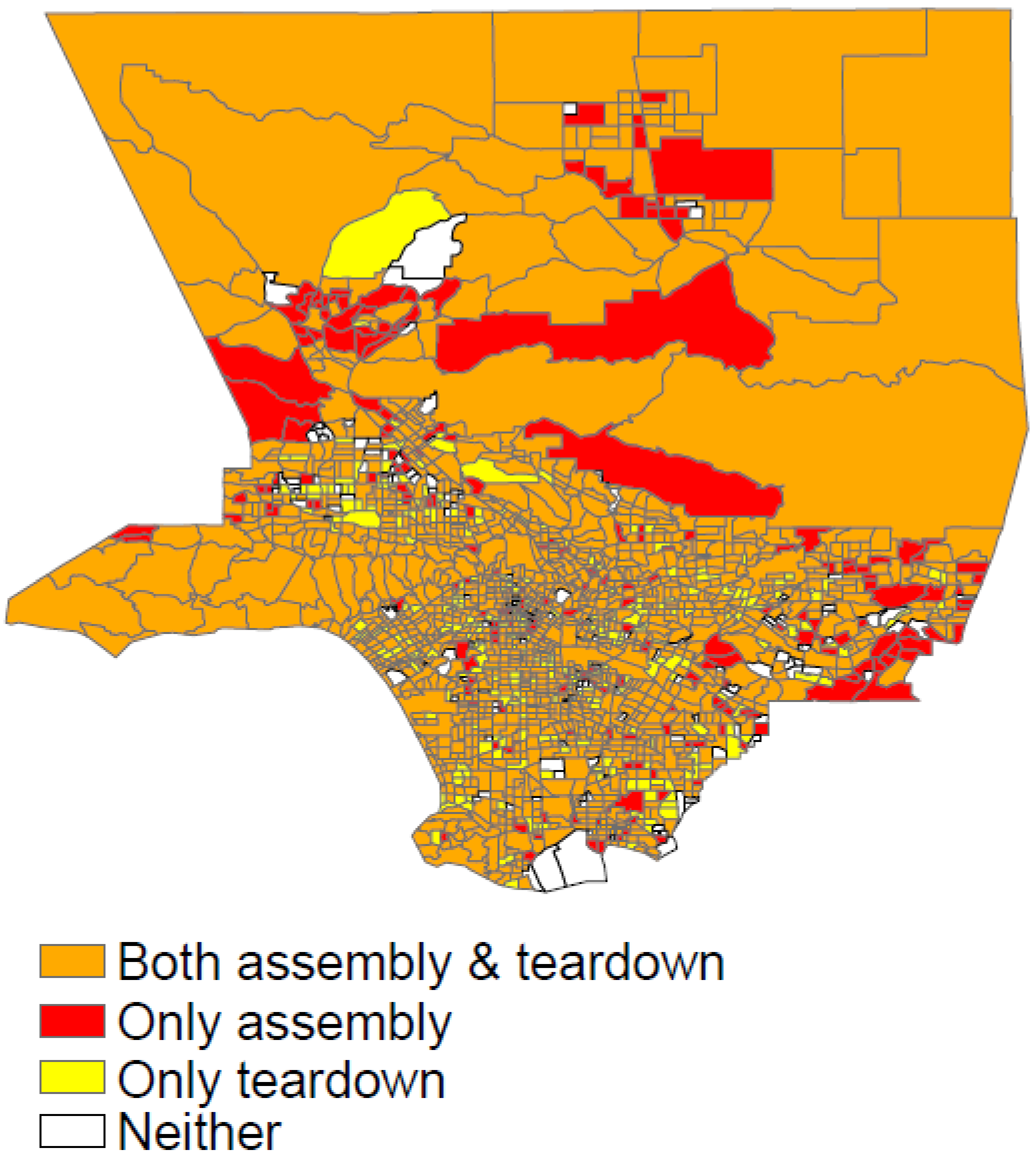


**Figure 3: Map of Los Angeles County Census Tracts by Redevelopment Status**




Notes: Each polygon in this map is a census tract, which is defined to have roughly between 3,000 and 4,000 people.

## Table 1: Summary Statistics for Price Analysis

| | Levels | | | | Change from 1990 to 2000 | | | |
|---|---|---|---|---|---|---|---|---|
| | Assembled | Teardown | Difference | Difference conditional on tract FE | Assembled | Teardown | Difference | Difference conditional on tract FE |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Price per square foot | 215.921 | 115.950 | 99.971 | n/a | n/a | n/a | n/a | n/a |
| | [404.041] | [186.378] | (7.855)*** | | | | | |
| Single family | 0.193 | 0.721 | -0.528 | -0.273 | n/a | n/a | n/a | n/a |
| | [0.394] | [0.449] | (0.011)*** | (0.032)*** | | | | |
| Poverty rate | 0.095 | 0.120 | -0.025 | 0.003 | -0.008 | 0.015 | -0.023 | -0.011 |
| | [0.111] | [0.122] | (0.003)*** | (0.006) | [0.103] | [0.068] | (0.002)*** | (0.008) |
| Neighborhood share | 0.044 | 0.050 | -0.006 | 0.002 | -0.003 | -0.006 | 0.004 | 0.002 |
| Black | [0.083] | [0.112] | (0.003)** | (0.003) | [0.055] | [0.060] | (0.002)*** | (0.004) |
| Neighborhood share | 0.217 | 0.247 | -0.030 | 0.006 | 0.025 | 0.030 | -0.005 | -0.001 |
| Hispanic | [0.251] | [0.272] | (0.007)** | (0.008) | [0.093] | [0.099] | (0.003)** | (0.005) |
| Median Household | 72303 | 71730 | 573 | 503 | 20471 | 16967 | 3504 | 1138 |
| Income | [50687] | [43838] | (1222) | (1267) | [50068] | [20086] | (952)*** | (976) |
| Observations | 2642 | 3362 | 6004 | 6004 | 2398 | 3351 | 5749 | 5749 |

Sources: See Data Appendix for complete information.
Notes: Columns (1), (2), (5) and (6) present means. Columns (3) and (7) present equality of means test for the means presented in columns (1) and (2), and (5) and (6), respectively. Columns (4) and (8) present coefficient estimates from a regression of the variable in the row on an indicator variable for assembly and a set of census tract fixed-effects. [] denotes standard deviation and () denotes standard error. Significance levels are denoted by *** for significant at the 1% level, ** for significant at the 5% level, and * for significant at the 10% level.

## Table 2: Excess Value of Land in Assembly

| | Full Sample | | | | | | Single Family | | Non-Residential | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 1 if Parcel is in an assembly | 0.491*** | 0.460*** | 0.565*** | 0.568*** | 0.452*** | 0.476* | 0.516*** | 0.516*** | 0.609** | 0.557* |
| | (0.105) | (0.096) | (0.121) | (0.119) | (0.115) | (0.205) | (0.108) | (0.107) | (0.232) | (0.230) |
| Observations | 6,004 | 5,749 | 6,004 | 6,004 | 5,749 | 3,575 | 2,933 | 2,911 | 2,256 | 2,156 |
| Assemblies | 2,642 | 2,642 | 2,642 | 2,642 | 2,642 | 232 | 509 | 509 | 1,522 | 1,522 |
| Teardowns | 3,362 | 3,362 | 3,362 | 3,362 | 3,362 | 3,362 | 2,424 | 2,424 | 734 | 734 |
| Geographic Fixed Effects | | | | | | | | | | |
| Tract | X | X | | X | X | X | X | X | X | X |
| Block Group | | | X | | | | | | | |
| Additional Covariates | | | | | | | | | | |
| Year-Quarter of Sale | X | X | X | X | X | X | X | X | X | X |
| Use Classifications | X | X | X | X | X | X | X | X | X | X |
| Cubic in Lot Size | X | X | X | X | X | X | X | X | X | X |
| Neighborhood Demographics | | X | | | X | X | | X | | X |
| Distance to Key Amenities | | X | | | X | X | | X | | X |
| Linear Tract Trends | | | | X | | | | | | |
| City * Year Fixed Effects | | | | | X | | | | | |
| Assembly Teardowns | | | | | | X | | | | |

Sources: See Data Appendix for complete information.
Notes: The dependent variable in these regressions is log(real sales price per square foot). () denotes standard error. Significance levels are denoted by *** for significant at the 0.1% level, ** for significant at the 1% level, * for significant at the 5% level, and + for significant at the 10% level. Neighborhood demographics, obtained from the census, include the following variables in both 2000 level form and as changes between 1990 and 2000 levels: poverty rate, neighborhood share black, neighborhood share Hispanic, share of housing units vacant and share of housing units owner-occupied. Use classifications include indicator variables for single family, non-condo multi-family, condo, vacant and other. Distance to key amenities include measures for the shortest distance from each parcel to three amenities: a highway entrance or exit, a metrolink stop (commuter rail), and a metrorail stop (subway or light rail).

## Table 3: Premium to Assembly in Difference Framework

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| 1{Redevelopment}*1{Assembly} | 0.536* | 0.533* | 0.348* | 0.369* |
| | (0.260) | (0.262) | (0.149) | (0.170) |
| Observations | 5,324 | 5,324 | 5,324 | 1,690 |
| No. Unique Assembled Parcels | 583 | 583 | 583 | 220 |
| No. Unique Teardown Parcels | 1,725 | 1,725 | 1,725 | 359 |
| Repeat Sales Sample | | | | |
| Assemblies | x | x | x | x |
| Teardowns | x | x | x | x |
| Non-Assembly, Non-Teardowns | | | | |
| Only Sales 1999 onward | | | | x |
| Additional Covariates | | | | |
| Parcel Fixed Effect | x | x | x | x |
| Year*Quarter | x | x | x | x |
| Use | x | x | x | x |
| 1{Redevelopment} | x | x | x | x |
| Capital | | x | x | x |
| Capital * 1{Redevelopment} | | | x | x |

Notes: Significance levels are denoted by ** for significant at the 1% level, * for significant at the 5% level, and + for significant at the 10% level. Capital (3) includes the assessed improvement value per units of lot size, the assessed improvement value divided by the assessed land value, and the structure square footage per units of lot size. Capital (4) includes all variables in Capital (3) and also includes the structure's year built. Use classifications include indicator variables for single family, non-condo multi-family, condo, vacant and other. Last sale premium denotes inclusion of a dummy variable for being the final sale before assembly or teardown.

## Table 4: Impact of Lot Size on Likelihood of Assembly

| | Biggest Possible Sample | | Biggest Possible Sample | | Sample with full information | | Sample with full information | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Ln(Lot Size) | 0.0123 | 0.013 | 0.0109 | 0.0119 | 0.0108 | 0.0116 | 0.0097 | 0.0104 |
| | (0.0012)*** | (0.0012)*** | (0.0016)*** | (0.0016)*** | (0.0013)*** | (0.0013)*** | (0.0012)*** | (0.0012)*** |
| Geographic Fixed Effects | | | | | | | | |
| Tract | x | | x | | x | | x | |
| Block Group | | x | | x | | x | | x |
| Additional Covariates | | | | | | | | |
| Use Classification | | | x | x | x | x | x | x |
| Improvement/Land Ratio | | | x | x | x | x | x | x |
| Neighborhood Demographics | | | x | | x | | x | |
| Distance to key amenities | | | x | x | x | x | x | x |
| Full interaction of year built and structure/lot size decile dummies | | | | | | | x | x |
| R-squared | 0.082 | 0.100 | 0.110 | 0.127 | 0.082 | 0.101 | 0.086 | 0.105 |
| Observations | 2,155,932 | 2,155,932 | 2,155,932 | 2,155,932 | 1,993,303 | 1,993,303 | 1,993,300 | 1,993,303 |
| Share Ever Assembled | 0.012 | 0.012 | 0.012 | 0.012 | 0.008 | 0.008 | 0.008 | 0.008 |

Sources: See Data Appendix for complete information.
Notes: The dependent variable is 1 if the parcel is every engaged in assembly from 1999 to 2010, and zero otherwise; results are estimated via OLS. The sample is all parcels that exist in 1999, and excludes public land as denoted by the parcel number. Use classification is as in Table 2. Improvement/land ratio is calculated with assessed values as of 1999. Census variables and amenities are as listed in the notes for Table 2. The “full interaction” is the interaction of indicators for decile of structure age with deciles of structure square feet divided by lot square feet.

## Table 5: Robustness Results for Impact of Lot Size on Likelihood of Assembly

| | By Use Type | | | | Parcel is in a Block Group with Density is Greater Than | | | |
|---|---|---|---|---|---|---|---|---|
| | Single-Family Residential Only | | Non-Residential Only | | 1st Percentile of Block Groups | | 10th Percentile of Block Groups | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Ln(Lot Size) | 0.0101 | 0.0109 | 0.014 | 0.0138 | 0.0097 | 0.0104 | 0.0077 | 0.0082 |
| | (0.0016)*** | (0.0016)*** | (0.0019)*** | (0.0016)*** | (0.0012)*** | (0.0012)*** | (0.0009)*** | (0.0009)*** |
| Geographic Fixed Effects | | | | | | | | |
| Tract | x | | x | | x | | x | |
| Block Group | | X | | x | | x | | x |
| Additional Covariates | x | X | x | x | x | x | x | x |
| R-squared | 0.057 | 0.078 | 0.203 | 0.265 | 0.086 | 0.105 | 0.079 | 0.093 |
| Observations | 1,389,272 | 1,389,272 | 148,411 | 148,414 | 1,993,300 | 1,993,303 | 1,759,088 | 1,759,091 |
| Share Ever Assembled | 0.005 | 0.005 | 0.052 | 0.052 | 0.008 | 0.008 | 0.006 | 0.006 |

Sources: See Data Appendix for complete information.
Notes: The dependent variable is 1 if the parcel is every engaged in assembly from 1999 to 2010, and zero otherwise; results are estimated via OLS. The sample is all parcels that exist in 1999, and excludes public land as denoted by the parcel number. All covariates are as in final column of the previous table. Results estimated when using the biggest possible sample and omitting controls for structure year built and structure square feet per lot square feet are very similar.

**Table 6: Elasticity of Sales Prices with Respect to Lot Size**

| | (1) | (2) |
|---|---|---|
| Ln(Lot Size) | -0.598*** | -0.595*** |
| | (0.154) | (0.155) |
| Ln(Lot Size) * 1{Non-residential property} | | -0.006 |
| | | (0.008) |
| Parcel Group Fixed Effects | x | x |
| Year * Quarter | x | x |
| Neighborhood Demographics | x | x |
| Distance to Key Amenities | x | x |
| Use Type Fixed Effects | x | |
| Observations | 2,642 | 2,642 |
| Number of Unique Parcel Groups | 947 | 947 |

Sources: See Data Appendix for complete information.
Notes: The dependent variable in these regressions is log(real sales price per square foot). () denotes standard error. Significance levels are denoted by *** for significant at the 0.1% level, ** for significant at the 1% level, * for significant at the 5% level, and + for significant at the 10% level. A "parcel group" is the set of parcels in which any ancestors or descendants experience assembly.

## Appendix Table 1: Summary Statistics for 1999 Cross-section

| | Maximal Sample | | Sample with full covariates | |
|---|---|---|---|---|
| | Ever Assembled | Never Assembled | Ever Assembled | Never Assembled |
| | (1) | (2) | (3) | (4) |
| Log(Lot Size) | 9.728 | 8.912 | 9.667 | 8.783 |
| | [1.569] | [1.093] | [1.393] | [0.848] |
| Improvement/Land Assessed Value | 36.878 | 3.937 | 59.933 | 4.285 |
| | [2156.9] | [864.9] | [2751] | [902.7] |
| 1 if Single Family | 0.257 | 0.643 | 0.400 | 0.699 |
| | [0.437] | [0.479] | [0.49] | [0.459] |
| Structure Square Feet in 100,000s | 0.080 | 0.027 | 0.121 | 0.029 |
| | [0.421] | [0.149] | [0.514] | [0.155] |
| Year Structure Was Built | 1960.0 | 1955.2 | 1960.4 | 1955.2 |
| | [28.881] | [21.613] | [28.565] | [21.612] |
| Census 2000 Block Group Covariates | | | | |
| Poverty Rate | 0.128 | 0.135 | 0.133 | 0.134 |
| | [0.124] | [0.113] | [0.129] | [0.113] |
| Share Black | 0.068 | 0.086 | 0.072 | 0.088 |
| | [0.119] | [0.162] | [0.122] | [0.166] |
| Share Hispanic | 0.292 | 0.353 | 0.317 | 0.359 |
| | [0.271] | [0.284] | [0.28] | [0.286] |
| Share Housing Units Vacant | 0.050 | 0.043 | 0.046 | 0.038 |
| | [0.062] | [0.053] | [0.053] | [0.037] |
| Share Housing Units | 0.619 | 0.606 | 0.605 | 0.600 |
| Owner-Occupied | [0.296] | [0.262] | [0.308] | [0.264] |
| Distances to Amenities | | | | |
| Nearest Highway Entrance | 1.741 | 1.637 | 1.342 | 1.272 |
| or Exit | [2.907] | [2.683] | [1.771] | [1.573] |
| Nearest Subway or Light Rail | 10.362 | 8.172 | 8.743 | 7.040 |
| | [10.717] | [9.316] | [9.462] | [7.506] |
| Nearest Commuter Rail | 5.919 | 6.195 | 5.552 | 5.929 |
| | [4.891] | [4.982] | [4.414] | [4.703] |
| Observations | 26,907 | 2,157,067 | 16,538 | 1,979,876 |

Sources: See Data Appendix.

**Data Appendix**

This appendix describes how we created the Los Angeles County parcel dataset. Section 1 describes the input datasets. Section 2 describes how we join these datasets, and reports statistics on the uncleaned data. Section 3 describes how we clean the joined dataset, and Section 4 reports statistics on the quality of the final dataset.

# 1 Data Sources

The basic unit of analysis is the parcel, which is an individual property as legally defined by the Los Angeles County Assessor and Recorder. In any year, there are roughly 2.3 million parcels in the County of Los Angeles. We rely on a number of different sources for information about parcels.

## 1.1 Parcel-Level Data

For detailed property information on parcels, we rely on data from three separate vendors: DataQuick, Applied Geodetics, and the Los Angeles County Assessor directly.

### 1.1.1 DataQuick: 1999 - 2002

Dataquick is a property information vendor. It purchases property information from the Los Angeles County Assessor to sell to real estate professionals. We rely on Dataquick data for 1999-2002, and data are reported as of January of each year.

As far as we can ascertain, the Dataquick data are a slightly modified version of the "Secured Basic File" that the Assessor prepares. Dataquick modifies the data from the original Los Angeles County format, and we re-modify it to be consistent with the two following datasets. We discuss modifications at length in Section 3. This data vendor is abbreviated in tables as DQ.

### 1.1.2 Applied Geodetics: 2004 - 2006

Applied Geodetics is a mapping firm in Los Angeles County. Applied Geodetics sold us data for 2004 (April), 2004 (February) and 2005 (May). To the best of our knowledge, these data are the unmodified Assessor's "Secured Basic File," which is the most complete record of property attributes available to the public from the Assessor. This data vendor is abbreviated in tables as AG.

### 1.1.3 Los Angeles County Assessor, Local Roll: 2007 - 2009

From 2007 through 2009, we have purchased data directly from the Assessor. Due to financial constraints we purchased the "Local Roll" database (roughly $400) instead of the "Secured Basic File" (roughly $13,000). The Local Roll has fewer parcel attributes than the Secured Basic File, and comes out annually in July. This data source is abbreviated in tables as LR.

### 1.1.4 Los Angeles County Assessor, Secured Basic File: 2010

In 2010, we purchased the Secured Basic File, which is the County's most complete publicly available dataset about properties. These data are from July 2010, and this data source is abbreviated in tables as SB.

## 1.2 Sales Data

### 1.2.1 Last Three Sales, 1980 to 2006

In 2006, Brooks purchased a file from the County Assessor that contains information on the last three transactions for each property in the county. For each transaction, we observe transaction type, sale amount (if applicable), and date of transaction.

### 1.2.2 Sales Within Two Years: 2008, 2009, and 2010 Files

In 2008, 2009, and 2010, we purchased additional lists of sales data from the County Assessor. These contain information on all transactions in the prior two years. For each transaction, we observe transaction type, sale amount (if applicable), and date of transaction.

These files leave a small gap from May through December 2006 which we have not been able to obtain.

## 1.3 Parcel Change Database

At our request, the Assessor made a special file that includes all parcel changes from July 1999 January 2009. Specifically, for each change, this file includes the old parcel number(s), the new parcel number(s), and the effective date of the change. The County has electronic records for parcel changes starting in July 1999 and continuing to the end of our data.

This change database allows us to isolate land assembly and disassembly. The California Assessor's Handbook mentions only one reason for a parcel number to change: if the physical boundaries of a parcel are modified (California State Board of Equalization, 1997, page 26).

We purchased this change database again in July 2009 and 2010 (covering all changes in the past two years) to allow us to link all later parcels with previous parcels.

## 1.4 Digital Parcel Maps

For each year since 2006, we have an electronic map of all parcels that exist in that year. These maps have a boundary (a polygon) for each individual parcel. For each parcel, we use ArcGIS to calculate the x- and y-coordinates (latitude and longitude) of the polygon's geographic center (centroid).

## 1.5 Census Tract and Block Group Indentification

The Census provides census tract and block group boundaries in shapefile format online.[1] We use ArcGIS to intersect the 2000 census boundaries and the 2006 parcel boundaries to assign each parcel to a census block group.

The majority – 96% – of ever-existing parcels have block group identifiers.

## 1.6 Block Group Data

We use block group level data from the 2000 Decennial Census (ICPSR file 13346, summary level 150), and from the 1990 Decennial Census (ICPSR 9782, summary level 150, but California file is damaged so we used a similar file downloaded from UCLA ATS).

We use ArcGIS and the Census's electronic maps to make a linkage between 1990 and 2000-based block groups, where relationships are based on land area overlap.

## 1.7 Assorted Non-Parcel Digital Maps

- Parks: Information from 2008 ESRI files of local and national parks for California
  - parks displayed on maps are only those more than 1.25 square miles
- Freeways: Data from State of California Cal-Atlas Geospatial Clearinghouse
  - website is http://www.atlas.ca.gov/download.html
  - transportation → Census 2000 → state_highways.* and us_highways.*
- Freeway Entrances and Exits
  - Tele-Atlas US Data, contains federal interstate highway entrances and exits
- Coastline
  - layer of points every 1000 feet along LA County coastline
  - created by taking Census 2000 county map and deleting non-coastline portions
  - used X-tools feature to points to convert coastline line to points

[1] Files are at http://www.census.gov/geo/www/cob/bdy_files.html.

- Metrolink Stations
  - Commuter rail stations
  - File received from Javier Minjares, Southern California Association of Governments, 2010
- Metro Rail Stations
  - Intra-urban rail stations
  - File received from Javier Minjares, Southern California Association of Governments, 2010
- Major Roads
  - Tele-Atlas US Data, version 9.3
  - Major roads only

# 2 Initial Data Linking

Each parcel is identified by a 10-digit number: MMMM-PPP-XXX. The first four digits are the "map book" number – literally the number of the "book" in which the parcel appears. The second three digits are the map book page. Each map book page contains a set of geographically contigious parcels. The last three digits identify individual parcels on the map book page.

We began the data assembly by attempting to link all the annual cross-sections described above (1999-2009, without the missing 2003) using the parcel change database. Panel A of Table 1 presents the results of this original linkage. Slightly fewer than 2 million parcels never change their number throughout the sample (column 2). Column 3 reports the number of ever-changing parcels; this number varies by year as a changer could be 3 parcels in 1999, but 12 (or 1) parcels in 2009. The number of changes in any given year varies from 28,996 to 58,557. Column 4 reports the number of parcels that exist in this year of the sample, but not in all years of the sample. Column 4 shows a striking number of parcels that exist

after 2004, but not before. These missing, or "phantom," observations are due to DataQuick editing. We discuss how we deal with these discrepancies in the next section.

We also analyze the total amount of land area, measured in square feet, in each category as a check on our data assembly process. We report annual totals in Panel B of Table 1. Columns 2 through 4 report the total amount of parcel square footage in each of the parcel assembly categories by year. Sadly, for 2004 and 2005 we do not observe parcel size in the data. It is very clear that across all categories, but particularly in the "phantom" category, that the total land area is drastically smaller in the earlier data source (DataQuick). The California Department of Water Resources measures the total land area of Los Angeles County as 132,487,077,888 square feet, which is clearly much closer to the later data source (143 billion square feet) than the earlier one (22 billion square feet).[2]

Table 2 shows the total number of parcels (Panel A) and total land area (Panel B) by four categories of use types and by year. The Assessor assigns each property a "use code" that describes how the property is currently used (as distinct from how it is zoned). We use this code to make four major categories of use: single-family, multi-family, vacant, and other. The number of vacant parcels by year (column 4) shows a large break concomitant with the break in data sources, and similar to the break for the "phantom" parcels of the previous table.[3] The bottom panel of the table shows that this discrepancy appears across types.

## 3 Cleaning and Consistency Changes

This section details the work we did to make the property data consistent across data sources and time. The main issues in cleaning the data were how to give a measure of land area to

[2]Parcels cover all land and water area of the county with the exception of public roadways. We discuss later how the latter datasets "overcount" the county's land area.

[3]Vacant parcels may be coded vacant residential, vacant commercial, or simply vacant. "Vacant" appears to refer to habitation or use, and not to the presence of structures.

parcels in 2004 and 2005, how to count the land area of parcels that were vertically stacked, how to account for parcels present in later years and missing in the earlier years, how to account for lack of consistency between the change database and the cross-sectional data, how to deal with reported changes in lot size for parcels that do not change, how to deal with consistently reporting land area for parcels that make complicated changes, and how to impute geographic information for parcels pre-dating our electronic maps.

## 3.1 Defining the Parcel Group

For ease of analysis, we define a "parcel group" to include all parcels linked by a change. For example, consider a change where parcels A and B combine to C, parcel C splits to D and E, and E splits to F and G. All parcels A through G would have one parcel group. We create a unique identifer for each group.

## 3.2 No Observation of Land Area in 2004 and 2005

The data source we use for 2004 and 2005 does not contain information on parcel land area. We assume that all 2004 parcels existing in 2002 have the same land area that they did in 2002, and all 2005 parcels existing in 2006 have the same land area they will have in 2006. These are very reasonable assumptions, as lot size changes only when parcel number does. As we describe in greater detail below, we use a variety of methods to evaluate and clean lot size across time.

If a 2004 or 2005 parcel is also missing lot size information in 2002 or 2006, we take lot size from the closest year for which it is available.

### 3.3 Vertically Stacked Parcels

The assessor draws the boundary for a piece of land with multiple owners, such as a condominium, so that each owner's parcel has the land area of the entire lot. This type of separate ownership of the same piece of land is distinct from joint ownership of a single property. Condominium-type ownership parcels are "stacked" vertically on the Assessor's map, in the same way you would stack checkers. In this case, think of each checker as the condominium of an individual owner. A 100-unit condominium will consist of 100 parcels, one for each legally distinct unit of ownership. Stacked parcels are usually, but not always, condominiums.

In the DataQuick data, the land area of each condominium unit is the land area of the lot divided by the number of parcels covering that lot. In the remaining datasets, when units and parcels are co-terminous, each unit is assigned the land area of the entire lot. This leads to much of the discrepancy in land area between the 1999 to 2002 and 2003 to 2010 periods in Tables 1 and 2. We correct for this problem by modifying all parcels to have a land area consistent with the method in the DataQuick data (land area of a parcel = total land area of parcel / number of parcels occupying that lot). We prefer this method of measuring land area because the net land area remains unchanged even if one parcel becomes three condos. The method used in the later datasets suggests that the total land area of the county increases when condominiums are created.

We identify vertically stacked parcels by calculating the centroid of every parcel on each map we have (annual maps, 2006-2010). In any given year, all parcels that share the same x- and y-coordinates are "stacked," and their land area requires correction. On average, a map-year has 250,562 unique parcels that are stacked, 14,400 stacks, and a mean of 17 parcels per stack. Unfortunately, we do not have maps for years before 2006, so this correction is incomplete for 2004 and 2005. Specifically, we can correct all stacked parcels that exist in 2004 and 2005 and which continue to exist in 2006. We cannot correct stacked parcels that

exist in 2004 or 2005 and not in 2006.

### 3.4 "Phantom" Parcels

The third data challenge we faced was the presence of parcels that did not appear in all years, and the disappearance of which could not be explained by any observation in the parcel change database. We call these appearing and disappearing parcels "phantoms."

Table 1 suggests that the DataQuick data do not include all parcels that exist in the county. We strongly suspect that, in an effort to satisfy customers primarily interested in residential (or at least tradeable property) property, DataQuick deletes parcels that will not be transacted.

Table 3 describes the appearance of phantoms across the years of the dataset. The horizontal axis has one column for each year of the dataset. The vertical axis has one row for each year of phantom non-appearance. The (1999,2000) cell says that 516 parcels that exist in the 1999 dataset do not exist in the 2000 dataset and are not accounted for by any parcel changes. The diagonal axis is by definition zero, since a parcel cannot be missing in the year in which it exists. The clear and striking pattern in this table is that the vast majority of phantom parcels are missing in in the years 1999 to 2002, and come from datasets in years 2004 to 2010.

To resolve this problem, we say that any parcel that ever exists in the dataset is a "true" parcel. Because these parcels are not in the change database, we assume that their shape is constant over the period of interest and that they exist in all years. Roughly 200,000 of these parcels truly exist from 2004 to 2010 and do not appear from 1999 to 2002. Only about 20,000 parcels appear from 1999 to 2002 and not from 2004 to 2010. Were we not to make this adjustment, we would have an severely unbalanced panel, and the early years of the panel would not account for a substantial amount of the land area of the County.

## 3.5 Relating Parcel Change to Parcel Cross-sectional Databases

We use the assessor's parcel change database to link parcels across years. The assessor reports on changes of four types: one to many, one to one, many to one, zero to one and one to zero. These last two changes (zero to one and one to zero) are the least frequent of all change types. They are used to move land in and out of the county. Land in use by roads does not have parcels, so the addition of a road from existing land could cause a zero to one change. A one to zero change could occur when a parcel becomes part of a road.

Table 4 reports on the quality of the match between the datasets. Across all years, we are able to match 89 percent of parcels in the change database with parcels in the cross-sectional databases (column 9). The table reports statistics for all changing parcels, and by change type. Columns 2 and 3 show that more parcels enter the county than leave over the period, according to the change database. Columns 4 and 5 report the number of parcels from the change database that find a match in the cross-sectional database, and shows whether these parcels are entering or leaving the county. Columns 7 through 10 report the share of parcels from the change database that find a match in the cross sectional databases.

Panel B reports these same statistics for a later year in the sample. In general, match rates between the parcel match dataset and the cross-sectional datasets improve over time, but fall off in the final year. In the final year, all new parcels may not yet have appeared in the cross-sectional dataset. This seems reasonable, as column 10 tells us that it takes an average of a year from the date of appearance in the change database for the new parcel to appear in the cross-sectional data.

To give a specific date to a parcel change, we must make a choice. We decided that in cases where the parcel change database and the cross-sectional databases did not agree, we would defer to the change database. The change database records an administrative, or policy change, while the cross-sectional data more likely reflects what is actually physically changing. Since we are most interested in policy choices regarding land use, the date of the

decision to change a parcel seemed more relevant than the physical change itself.

### 3.6 Reported Changes in Land Area for Non-Changing Parcels

We check the internal consistency of the data by examining whether there are changes in land area for parcels that do not change identification number across the years of the sample. By construction, such parcels should have the same land area for all years of the sample. For each parcel, we calculate the percentage difference between the largest land area and the smallest land area reported (relative to the mean lot size of the entire period). The median change is 0.7 percent, but changes at the ninetieth percentile and above can be quite large.

Given this, we make two of adjustments to clean the land area variable. First, if a parcel's land area changes by less than 20 percent from 1999 to 2010, we give the parcel the 2010 land area value for each year. We rely on data from the later years of the sample because by all measures we describe here it is more reliable than the DataQuick data. This change modifies the values for virtually all of the parcels that remain in the dataset. Of the modified parcels, slightly more than 20 percent of them have very minor differences of less than 0.1 percent, which seem to be due to rounding differences between the different datasets. Another 61 percent have differences between 0.1 and 5 percent of land area over time, and the remaining 18 percent have differences of between 5 and 20 percent.

The second adjustment to clean the land area variable uses the same rule to clean parcels that are vertically stacked (if the percentage change in land area for the same parcel is less than 20 percent, replace the area for all years with the 2009 value).

Finally, we delete 109,923 unique parcels, between 2 and 4 percent of the observations in each cross-section, that have land area changes of more than 20 percent, or which have always-missing values for land area. These parcels account for roughly three percent of each cross-section

## 3.7 Land Area Adjustments for Parcel Changes

When parcels change, the total lot initial lot size of changing parcels should equal the final total lot size of the changing parcels. Unfortunately, this is not always the case.

For example, if one parcel changes number and becomes only one new parcel, the parcel's geographic area does not change by construction. However, due to dataset breaks and other inconsistencies, differences in lot size sometimes appear. These create inconsistencies in geographical areas over time that do not reflect actual changes to parcels.

We resolve this problem by inflating or deflating each parcel's lot size so that the total lot size for parcels that change is the same for each year of data. Let $g$ be the parcel group, which is all parcels ever associated with a given change. The adjustment is

$$\text{lot size}_{i,g,t} = \text{reported lot size}_{i,t} * \frac{\text{average total lot size}_g}{\text{total lot size}_{g,t}}.$$

The final value for parcel $i$ in year $t$ (lot $\text{size}_{i,t}$) is equal to parcel $i$'s reported square footage in year $t$ (reported lot $\text{size}_{i,t}$) multiplied by the ratio of the average of the parcel group's total square feet (average total lot $\text{size}_g$) divided by the current year total square feet (total lot $\text{size}_{g,t}$). By adjusting at the parcel group level, this method allows us to account for both simple and very complex changes.

The result of this adjustment is that each parcel group has a constant lot square footage. This better reflects the reality of a parcel change. Square footage at the beginning should equal square footage at the end, regardless of how the land is parceled.

## 3.8 Missing Values for Parcel Attributes

In addition to the work with lot square feet that we do above, we fill in missing values for parcel attributes to a limited extent. For non-geographic information, we edit parcels that do not change number, and for which a later observation of the parcel exists. In these cases,

we attribute the later parcel characteristics to the earlier parcel if the structure on the parcel was built before the year of the data. For example, imagine parcel A exists from 1999 to 2009, but we observe attributes starting only in 2005. If the structure on parcel A was built in 1995, we fill in the missing values for parcel A from 1999 through 2004 with the 2005 values. If the structure on parcel A was built built in 2005, we do not fill in missing values (save lot square footage as described above).

Geographic identifying information – tract and block group information, and latitude and longitude information – come from merging the panel data with information generated from electronic maps. We have these electronic maps starting only in 2006. For parcels that do not change, the physical border has not changed and therefore the 2006 geographic information is still valid.

For for parcels that existed before 2006 and not afterward we impute geographic information. This is relatively straightforward. For parcels that change, we assign old parcels to the same block group or tract as the later-existing parcels (no parcel change groups cross block groups). For the latitude and longitude, we give the old parcels the average of the latitude and longitude of the later-existing parcels. This should be a very close approximation to the actual location.

For parcels that do not have a block group through these methods (changing or not), we interpolate to find the block group. Specifically, if a parcel is missing a block group, we assign the parcel to have the block group of all other parcels on that map book page. A map book page is identifiable by the first seven digits of the parcel number. Map book pages have very few parcels and very rarely cross block group boundaries.

## 3.9 Preparing Sales Data

To prepare the sales data, we restrict our analysis to arms-length transactions. We drop sales with dates not on any calendar (e.g., day 45; this is a problem only for the 2006

dataset), sales with values of 10 or less (which the assessor uses as a special code for adjoining pools or structures), and sale amounts equal to 999,999,999. Because data appear to be collected electronically only from 1964 onward, we drop sales before 1964. We keep only one observation if the data report two sales on the same day of the same amount. A small share of observations have two sales on the same day of differing amounts. In this case, we drop both observations.

# 4 Cleaning and Consistency Results

Tables 5 and 6 present the results of the data cleaning. In the first of these tables, Panel A shows that the overall pattern of the number of parcels over time no longer exhibits any breaks when we change data sources. The number of phantom parcels and never-changing parcels is now consistent across all years. The number of changing parcels increases over time: there are roughly 30,000 of these parcels in 1999 and 55,000 of them in 2010. This means that, on net, more parcels have split than merged. Panel B shows that this increase is not driven by substantial changes in land area. This table no longer shows discontinuous patterns across dataset breaks.

Panel A of Table 6 shows that, from 1999 to 2009, the number of single- and multi-family parcels increased, while the number of vacant and "other" parcels decreased. This pattern holds for land area as well, as shown in Panel B, though the percentage changes are not as large. As in the previous table, these results no longer show sharp discontinuities related to dataset breaks. Total land area increases over time, since individually-owned properties in condominiums are separately parceled, and each parcel contains the lot size of the entire lot.

# References

California State Board of Equalization, 1997. "Assessor's Handbook 215: Assessment Map Standards." Tech. rep., California State Board of Equalization.

**Table 1: Original Dataset: By Type of Appearance in Dataset**

| | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| | | **Panel A: Parcels** | | | |
| Year | Source | Total | Never-Changers | Changers | Phantoms |
| 1999 | DQ | 2,054,630 | 1,974,322 | 28,996 | 6,118 |
| 2000 | DQ | 2,062,666 | 1,974,322 | 30,067 | 13,115 |
| 2001 | DQ | 2,087,854 | 1,974,322 | 32,516 | 33,727 |
| 2002 | DQ | 2,087,416 | 1,974,322 | 36,928 | 29,120 |
| 2004 | AG | 2,311,387 | 1,974,322 | 37,083 | 247,115 |
| 2005 | AG | 2,319,526 | 1,974,322 | 40,027 | 250,095 |
| 2006 | AG | 2,337,233 | 1,974,322 | 43,856 | 254,809 |
| 2007 | LR | 2,354,414 | 1,974,322 | 51,827 | 254,104 |
| 2008 | LR | 2,366,832 | 1,974,322 | 56,164 | 251,744 |
| 2009 | LR | 2,376,360 | 1,974,322 | 57,947 | 251,048 |
| 2010 | SB | 2,380,386 | 1,974,322 | 58,757 | 249,355 |
| | | **Panel B: Land Area, Square Feet** | | | |
| Year | Source | Total | Never-Changers | Changers | Phantoms |
| 1999 | DQ | 27,764,230,819 | 19,645,546,954 | 977,750,276 | 448,460,263 |
| 2000 | DQ | 22,403,685,543 | 19,645,922,629 | 846,283,090 | 601,903,930 |
| 2001 | DQ | 23,929,788,784 | 19,645,923,464 | 854,822,086 | 1,931,672,544 |
| 2002 | DQ | 22,411,729,309 | 19,645,957,949 | 869,685,087 | 540,404,227 |
| 2004 | AG | . | . | . | . |
| 2005 | AG | . | . | . | . |
| 2006 | AG | 138,025,141,417 | 49,445,296,763 | 3,040,095,062 | 77,568,380,188 |
| 2007 | LR | 141,249,361,785 | 50,825,703,096 | 3,478,861,722 | 77,157,371,619 |
| 2008 | LR | 142,126,764,819 | 50,824,750,529 | 4,026,210,983 | 76,883,380,642 |
| 2009 | LR | 142,791,412,096 | 50,824,661,151 | 4,284,048,431 | 76,667,836,617 |
| 2010 | SB | 143,083,855,131 | 50,800,757,455 | 4,624,977,833 | 76,579,664,196 |

Note: Columns 2, 3 and 4 do not add up to the total in column 1, as we drop some parcels as described in this appendix. Because these parcels are dropped, they are not classified under this scheme.

## Table 2: Original Dataset: Use Type By Year

| | | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|---|
| | | **Panel A: Parcels** | | | | |
| Year | Source | Total | Single-Family | Multi-Family | Vacant | Other |
| 1999 | DQ | 2,054,630 | 1,659,917 | 243,386 | 135,153 | 16,069 |
| 2000 | DQ | 2,062,666 | 1,666,775 | 243,564 | 136,015 | 16,205 |
| 2001 | DQ | 2,087,854 | 1,675,331 | 243,991 | 139,845 | 28,354 |
| 2002 | DQ | 2,087,416 | 1,686,514 | 244,490 | 137,462 | 18,589 |
| 2004 | AG | 2,311,387 | 1,705,031 | 245,497 | 226,280 | 131,992 |
| 2005 | AG | 2,319,526 | 1,713,035 | 245,518 | 226,943 | 132,034 |
| 2006 | AG | 2,337,233 | 1,731,829 | 245,862 | 227,397 | 129,918 |
| 2007 | LR | 2,354,414 | 1,749,070 | 246,454 | 227,303 | 129,600 |
| 2008 | LR | 2,366,832 | 1,761,984 | 247,136 | 228,564 | 127,129 |
| 2009 | LR | 2,376,360 | 1,772,071 | 247,755 | 228,860 | 125,511 |
| 2010 | SB | 2,380,386 | 1,777,436 | 248,097 | 228,791 | 123,751 |
| | | **Panel B: Land Area, Square Feet** | | | | |
| Year | Source | Total | Single-Family | Multi-Family | Vacant | Other |
| 1999 | DQ | 27,764,230,819 | 16,456,132,388 | 3,164,288,059 | 7,314,122,805 | 809,916,239 |
| 2000 | DQ | 22,403,685,543 | 14,688,828,290 | 2,647,621,527 | 4,610,112,501 | 450,069,675 |
| 2001 | DQ | 23,929,788,784 | 14,754,694,550 | 2,662,660,428 | 5,242,808,095 | 1,258,558,401 |
| 2002 | DQ | 22,411,729,309 | 14,806,880,026 | 2,677,849,069 | 4,481,695,438 | 437,417,990 |
| 2004 | AG | | | | | |
| 2005 | AG | | | | | |
| 2006 | AG | 138,025,141,417 | 49,477,666,919 | 3,358,135,088 | 46,790,566,956 | 13,670,675,727 |
| 2007 | LR | 141,249,361,785 | 51,697,344,419 | 3,373,264,623 | 47,139,052,348 | 13,895,134,582 |
| 2008 | LR | 142,126,764,819 | 52,429,447,269 | 3,404,275,216 | 47,380,322,248 | 13,770,460,036 |
| 2009 | LR | 142,791,412,096 | 53,111,200,572 | 3,419,191,304 | 47,470,576,152 | 13,625,541,866 |
| 2010 | SB | 143,083,855,131 | 53,320,804,684 | 3,421,321,560 | 47,600,112,590 | 13,653,776,642 |

Note: Columns 2 through 5 do not add up to the total in column 1, as a very few parcels do not have an intelligible use code.

## Table 3: Phantom Parcels Across Years

| | | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Year of Data | | | | | | | | | | |
| | | 1999 | 2000 | 2001 | 2002 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| Parcel takes a value of 1 in this year if the parcel exists in the original data | 1999 | 0 | 7,497 | 25,417 | 20,629 | 223,634 | 222,409 | 219,972 | 217,992 | 215,349 | 213,974 | 212,214 |
| | 2000 | 516 | 0 | 20,870 | 17,839 | 222,629 | 221,331 | 218,896 | 216,920 | 214,248 | 212,873 | 211,116 |
| | 2001 | 866 | 390 | 0 | 11,743 | 208,998 | 207,866 | 205,786 | 204,002 | 201,481 | 200,200 | 198,503 |
| | 2002 | 5,221 | 4,913 | 16,068 | 0 | 220,238 | 218,917 | 216,606 | 214,648 | 212,033 | 210,670 | 208,903 |
| | 2004 | 1,752 | 1,319 | 1,157 | 1,112 | 0 | 5,376 | 6,981 | 7,428 | 7,662 | 7,901 | 7,936 |
| | 2005 | 1,758 | 1,329 | 1,176 | 1,134 | 102 | 0 | 8,916 | 9,333 | 9,557 | 9,821 | 9,833 |
| | 2006 | 1,761 | 1,333 | 1,183 | 1,146 | 248 | 164 | 0 | 2,405 | 2,789 | 3,108 | 3,200 |
| | 2007 | 1,766 | 1,340 | 1,196 | 1,164 | 361 | 301 | 309 | 0 | 1,018 | 1,424 | 1,589 |
| | 2008 | 1,766 | 1,341 | 1,197 | 1,166 | 481 | 435 | 475 | 228 | 0 | 1,159 | 1,382 |
| | 2009 | 1,766 | 1,341 | 1,198 | 1,167 | 517 | 473 | 529 | 310 | 149 | 0 | 576 |
| | 2010 | 1,766 | 1,341 | 1,199 | 1,170 | 583 | 544 | 614 | 427 | 299 | 180 | 0 |

Notes: This table reports a count of only phantom parcels. Each cell reports the number of parcels in the year of the horizontal axis (year of data) that are "phantoms" that we have added in other years. For example, cell (20004,1999) has a value of 230,407, which means that 230,407 parcels that at some point started (or stopped appearing) without a matching change in the change database appear in the original data from year 2004. By construction, the diagonal is zero – a parcel that we add to the database because it does not appear (a phantom) cannot exist in the original data in that year.

## Table 4: Cleaned Dataset: By Type of Appearance in Dataset

| | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| | | **Panel A: Parcels** | | | |
| Year | Source | Total | Never-Changers | Changers | Phantoms |
| 1999 | DQ | 2,238,337 | 1,974,322 | 28,994 | 235,021 |
| 2000 | DQ | 2,245,080 | 1,974,322 | 29,276 | 241,482 |
| 2001 | DQ | 2,250,420 | 1,974,322 | 28,590 | 247,508 |
| 2002 | DQ | 2,257,824 | 1,974,322 | 29,944 | 253,558 |
| 2004 | AG | 2,266,664 | 1,974,322 | 35,541 | 256,801 |
| 2005 | AG | 2,272,612 | 1,974,322 | 37,851 | 260,439 |
| 2006 | AG | 2,275,776 | 1,974,322 | 42,504 | 258,950 |
| 2007 | LR | 2,281,750 | 1,974,322 | 50,702 | 256,726 |
| 2008 | LR | 2,283,392 | 1,974,322 | 54,832 | 254,238 |
| 2009 | LR | 2,284,011 | 1,974,322 | 56,554 | 253,135 |
| 2010 | SB | 2,281,702 | 1,974,322 | 56,778 | 250,602 |
| | | **Panel B: Land Area, Square Feet** | | | |
| Year | Source | Total | Never-Changers | Changers | Phantoms |
| 1999 | DQ | 101,141,388,368 | 20,104,375,427 | 1,521,589,117 | 79,515,423,823 |
| 2000 | DQ | 101,139,463,687 | 20,104,375,427 | 1,409,035,164 | 79,626,053,096 |
| 2001 | DQ | 101,123,335,380 | 20,104,375,427 | 1,574,991,288 | 79,443,968,664 |
| 2002 | DQ | 101,087,643,094 | 20,104,375,427 | 1,678,513,494 | 79,304,754,172 |
| 2004 | AG | 101,106,061,329 | 20,104,375,427 | 2,012,482,403 | 78,989,203,498 |
| 2005 | AG | 101,094,721,056 | 20,104,375,427 | 2,122,474,399 | 78,867,871,230 |
| 2006 | AG | 101,092,256,000 | 20,104,375,427 | 2,888,466,390 | 78,099,414,183 |
| 2007 | LR | 101,091,124,053 | 20,104,375,427 | 3,341,035,624 | 77,645,713,001 |
| 2008 | LR | 101,090,900,770 | 20,104,375,427 | 3,846,494,494 | 77,140,030,848 |
| 2009 | LR | 101,092,046,083 | 20,104,375,427 | 4,087,266,349 | 76,900,404,307 |
| 2010 | SB | 101,091,736,639 | 20,104,375,427 | 4,413,714,704 | 76,573,646,507 |

## Table 5: Cleaned Dataset: Use Type By Year

| | | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|---|
| | | | **Panel A: Parcels** | | | |
| Year | Source | Total | Single-Family | Multi-Family | Vacant | Other |
| 1999 | DQ | 2,238,337 | 1,418,826 | 463,900 | 176,344 | 179,180 |
| 2000 | DQ | 2,245,080 | 1,422,936 | 466,436 | 177,197 | 178,392 |
| 2001 | DQ | 2,250,420 | 1,426,788 | 467,734 | 179,080 | 176,712 |
| 2002 | DQ | 2,257,824 | 1,432,413 | 469,829 | 178,679 | 176,788 |
| 2004 | AG | 2,266,664 | 1,439,064 | 473,384 | 188,326 | 165,885 |
| 2005 | AG | 2,272,612 | 1,443,011 | 475,036 | 188,302 | 166,255 |
| 2006 | AG | 2,275,776 | 1,449,644 | 475,649 | 184,932 | 165,548 |
| 2007 | LR | 2,281,750 | 1,455,515 | 477,137 | 183,788 | 165,307 |
| 2008 | LR | 2,283,392 | 1,457,763 | 478,612 | 181,672 | 165,345 |
| 2009 | LR | 2,284,011 | 1,459,218 | 479,000 | 180,202 | 165,591 |
| 2010 | SB | 2,281,702 | 1,459,232 | 478,676 | 178,400 | 165,394 |
| | | | **Panel B: Land Area, Square Feet** | | | |
| Year | Source | Total | Single-Family | Multi-Family | Vacant | Other |
| 1999 | DQ | 101,141,388,368 | 15,220,405,675 | 3,329,124,339 | 42,481,518,355 | 40,051,723,892 |
| 2000 | DQ | 101,139,463,687 | 15,738,700,778 | 3,357,013,177 | 42,288,988,008 | 39,701,786,975 |
| 2001 | DQ | 101,123,335,380 | 16,850,694,605 | 3,440,125,295 | 41,110,698,633 | 39,668,229,961 |
| 2002 | DQ | 101,087,643,094 | 16,561,419,196 | 3,739,242,848 | 41,034,330,369 | 39,706,498,431 |
| 2004 | AG | 101,106,061,329 | 16,802,756,545 | 3,767,608,880 | 41,408,767,780 | 39,125,157,636 |
| 2005 | AG | 101,094,721,056 | 16,870,449,564 | 3,758,195,623 | 41,314,648,352 | 39,148,160,367 |
| 2006 | AG | 101,092,256,000 | 16,864,368,220 | 3,811,098,267 | 41,264,820,826 | 39,148,716,613 |
| 2007 | LR | 101,091,124,053 | 16,975,553,582 | 3,922,783,968 | 41,010,639,560 | 39,178,894,869 |
| 2008 | LR | 101,090,900,770 | 17,046,810,448 | 3,933,788,985 | 40,895,764,867 | 39,214,536,470 |
| 2009 | LR | 101,092,046,083 | 17,072,064,613 | 3,961,136,451 | 40,779,823,552 | 39,279,021,466 |
| 2010 | SB | 101,091,736,639 | 17,082,857,623 | 3,972,777,866 | 40,760,459,053 | 39,275,642,097 |

Note: Columns 2 through 5 do not add up to the total in column 1, as a very few parcels do not have an intelligible use code.

## Table 6: Parcels in Change Database by Type of Appearance

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|
| | Parcels in Change Database | | Parcels in the Cross-Section with match in parcel change database | | Share of all parcels in Change Database | | | | |
| Change Type | Entering County | Leaving County | Entering County | Leaving County | Match in Cross-sectional database | Temporary parcels, not matched, but subsequent parcel in change sequence is | Successfully Matched Parcels | No match in Cross-sectional database | Matches: Average years for change to appear in cross-section |
| **Panel A: Full Sample** | | | | | | | | | |
| All types | 184,860 | 87,339 | 149,800 | 51,268 | 0.74 | 0.15 | 0.89 | 0.11 | 1.00 |
| Zero to One | 2,205 | 1 | 637 | 0 | 0.29 | 0.60 | 0.89 | 0.11 | 1.51 |
| One to Zero | 1 | 860 | 0 | 425 | 0.49 | 0.00 | 0.49 | 0.51 | |
| One to One | 27,919 | 27,919 | 19,747 | 19,340 | 0.70 | 0.15 | 0.85 | 0.15 | 1.13 |
| One to Many | 139,814 | 16,230 | 119,523 | 8,026 | 0.82 | 0.10 | 0.92 | 0.08 | 0.96 |
| Many to One | 14,921 | 42,329 | 9,893 | 23,477 | 0.58 | 0.27 | 0.86 | 0.14 | 1.09 |
| **Panel B: Restricted sample: Cross-sectional years 2007-2008, same headings** | | | | | | | | | |
| All types | 35,262 | 14,129 | 31,514 | 10,953 | 0.86 | 0.10 | 0.96 | 0.04 | 0.94 |
| Zero to One | 265 | 1 | 86 | 0 | 0.32 | 0.62 | 0.94 | 0.06 | 1.00 |
| One to Zero | 1 | 52 | 0 | 44 | 0.83 | 0.00 | 0.83 | 0.17 | |
| One to One | 4,116 | 4,116 | 3,409 | 3,668 | 0.86 | 0.07 | 0.93 | 0.07 | 0.85 |
| One to Many | 28,548 | 2,811 | 26,185 | 1,845 | 0.89 | 0.07 | 0.96 | 0.04 | 0.95 |
| Many to One | 2,332 | 7,149 | 1,834 | 5,396 | 0.76 | 0.20 | 0.96 | 0.04 | 0.93 |