Why zoning is too restrictive*

Jack Favilukis†  Jaehee Song‡

May 21, 2024

Abstract

Zoning restrictions lowered aggregate growth by 36% (Hsieh and Moretti (2019)). If these restrictions are so costly, why do they exist? We propose a novel theory for why zoning is more stringent than the social optimum: the more zoning authorities a metro is fragmented into, the more restrictive zoning is in the metro. When zoning decisions are made locally, voters restrict development to avoid local congestion externalities, but fail to internalize worsened metro-wide housing affordability. Empirically, the HHI of local governments within a metro alone explains 12% of zoning variation across the U.S. Using lagged HHI in 1900 as an instrument, we show that fragmentation of zoning authorities increases zoning stringency and housing costs. Our results provide clear policy advice — zoning decisions should be made globally. Indeed, facing housing affordability crises, several states and nations have begun to adopt global zoning reforms.

Keywords: Determinants of zoning, Housing affordability

JEL classification: R21, R23, R52

---

*We thank seminar and conference participants at Baruch College, University of Alberta, University of Indiana conference on housing affordability, UC Berkeley housing politics and policy conference, and the Urban Economic Association.

†Sauder School of Business, University of British Columbia. Email: jack.favilukis@sauder.ubc.ca

‡Leeds School of Business, University of Colorado. Email: jaehee.song@colorado.edu
1 Introduction

Hsieh and Moretti (2019) estimate that restrictions on housing supply have lowered aggregate U.S. growth by 36% between 1964 and 2009. If zoning restrictions are so costly, why do they exist? We propose and empirically test a novel theory for why zoning restrictions exist, why they are more stringent than the social optimum, and how to relax them toward the social optimum. These questions are especially important at a time when many places around the world are struggling with housing affordability.

Our theory predicts that the more local entities making their own zoning decisions that a metro is fragmented into, the more restrictive zoning rules will be imposed across the entire metro. This is because when zoning decisions are made locally, voters choose more restrictive zoning than the social optimum to avoid local congestion externalities. However, they fail to internalize its negative effects on metro-level housing affordability due to limited housing supply from restrictive zoning.

We test our theoretical prediction – metros split into more zoning authorities have more density restrictions – in two ways. In our primary test, we find that in the cross-section, a metro-level Herfindahl–Hirschman Index (HHI) based on the number and size of zoning authorities that the metro is split into alone explains 12% of the variation in residential zoning, as measured by average minimum lot sizes (MLS), across Core-Based Statistical Areas (CBSA) in the U.S. We use HHI computed using municipal boundaries in 1900 as an instrument and find that lower HHI (more local decision-making) increases the stringency of zoning and housing prices.

We also conduct a case study of the amalgamation of Toronto in 1998, when six municipalities in central Toronto were dissolved and assembled into a single city of Toronto. We find that after 1998, construction permits grew significantly faster in these former municipalities than in neighboring municipalities that were not amalgamated. We show that this increase in housing supply is due to relaxed supply environments, consistent with our theory.

This theory provides clear policy advice — taking zoning decisions away from small localities and making the decisions more globally (e.g. state or metro level rather than municipality level) will lead to fewer zoning restrictions and bring zoning policy closer to the social optimum. Historically, in the U.S., local municipalities, such as cities, towns, and townships, have had the most authority in deciding zoning
regulations (Fischel, 2004). Facing housing affordability crises, several places have recently begun to take zoning authority away from small locales, consistent with our model. For example, California\(^1\) and Oregon,\(^2\) have adopted state-level relaxation of zoning, partly overriding existing local zoning regulations. Internationally, New Zealand passed a similar law in 2021.\(^3\) As such, our research is timely and relevant to the decisions being made by today’s policymakers.

We start by building a model of a metro with multiple municipalities populated by homeowners who vote for zoning restrictions.\(^4\) Because of a local congestion externality, the fewer people live within a particular municipality, the higher the utility of living in that municipality. Because of this, equilibrium house prices are higher in such a municipality. Therefore, all else equal, owners always prefer tighter zoning restrictions in their own municipalities for higher amenities and increased property values. However, tighter zoning across all municipalities raises metro-wide house prices, which discourages in-migration and reduces the metro’s population. This weakens agglomeration and leads to lower metro-wide wages, which is bad for local residents. When zoning decisions are made locally, local landowners do not internalize the effect of high house prices in their own municipalities on metro-wide wages, leading to zoning that is too restrictive.

We build a second model which motivates the same empirical prediction through an alternative channel. In this model, there is no migration into the metro, but local residents may need to move to another jurisdiction within the metro. Therefore, they are hurt by high metro-wide house prices, but they do not internalize this when making local zoning decisions.

We then empirically investigate the key implication of our model: when a metro

\(^1\)SB287, proposed in 2018, would have allowed five-story construction near all busy public transit stops across the state, though it failed to get out of committee. Subsequently, SB9, which took effect in 2022, mandated that local jurisdictions approve all projects that subdivide one lot into two and allow second dwelling units in a single-family residential zone across the state, as long as certain objective criteria were met. Additionally, several laws (SB897 in 2022; AB68, AB976, AB1033 in 2023) made it more difficult for jurisdictions to oppose construction of accessory dwelling units (ADUs), which are smaller units built in addition to the main unit on an existing lot.

\(^2\)HB 2001, which took effect in 2020, restricts local jurisdictions from preventing duplexes in all cities above 10,000 population; allows triplexes and fourplexes in all cities above 25,000 population.

\(^3\)The Resource Management (Enabling Housing Supply and Other Matters) Amendment Bill allows three homes of three stories on most parcels in New Zealand’s five largest cities.

\(^4\)Much of the rest of the literature relies on owner-renter conflicts to motivate zoning. While this channel is likely important, our mechanism works even without this conflict.
is fragmented into a larger number of smaller jurisdictions that each decide their own zoning regulation, then the metro will have more restrictive overall zoning, all else equal. We first construct a measure of how decentralized decision-making is within each CBSA in the U.S. To do so, we compute an HHI of the residential area of each zoning authority unit that makes up that CBSA — this is our independent variable of interest. Our dependent variable of interest is a CBSA-level measure of zoning stringency. We compute this by aggregating up the neighborhood-level zoning stringency measure from Song (2024) to the CBSA level. This zoning stringency measure estimates minimum lot sizes (MLS) at the neighborhood level and has uniquely broad geographic coverage, allowing us to analyze zoning stringency cross-sectionally across the U.S.

Zoning jurisdiction HHI may be endogenous and correlated with zoning stringency and housing costs. For example, places that wanted stricter zoning may have decided to incorporate into a separate municipality, resulting in reverse causality. To address such endogeneity, we adopt an instrumental variable (IV) approach while controlling for a rich set of location characteristics. Considering that comprehensive zoning was first adopted in 1916 in New York City and most actively adopted in the mid-20th century in the rest of the country, we construct the historical CBSA-level HHI using only municipalities that existed in 1900. We use this historical HHI as an IV. Additionally, we control for CBSA-level demographics in the 1900s, political lean, climate conditions, land use composition, land developability, and industrial compositions. The identification assumption is that the pre-1900 granularity of municipalities is as good as random conditional on observable location characteristics.

We find a strong positive relationship between the decentralization of jurisdictions making independent zoning decisions in a CBSA (low HHI) and the stringency of zoning in that CBSA (high MLS). The univariate OLS regression indicates that HHI alone explains 12% of the variation in zoning across CBSAs. The baseline IV regression implies that an increase in municipality concentration from the 25th to the 75th percentile would lead to zoning becoming 35% less stringent. This empirical finding is robust to using alternative zoning stringency measures such as the Density Restriction Index from the Wharton Land Use Survey (Gyourko et al., 2021) and using the number of municipalities instead of HHI as the independent variable. We also confirm other model predictions. Namely, we find that stringent zoning, driven by fragmented municipal institutions as of 1900, reduces population and increases
housing costs.

A case study of Toronto's amalgamation using a difference-in-difference design also supports our theory. The treatment group includes six municipalities amalgamated into the city of Toronto in 1998. Prior to amalgamation, each had its own government, and these governments were merged into a single one after. The control group includes five neighboring municipalities that experienced no administrative changes. We find that post-1998, housing starts in the treated group were 59% to 78% higher than they would have been had they followed the trend of the control group. The same analysis shows that there was little difference in prices or rents, making it unlikely that a demand shock was responsible for the increase in housing construction.

Primarily, our paper relates to the literature on the determinants of zoning. One strand of this literature argues that zoning is the result of a political conflict between pro-zoning homeowners, anti-zoning renters, anti-zoning developers or owners of vacant land, and anti-zoning business owners. Fischel (2004) writes that the original purpose of zoning was to protect homeowners in residential areas from devaluation by industrial and apartment use. Ellickson (1977) and Glaeser et al. (2005) argue that because owners are better organized, zoning tends to be too restrictive.\footnote{Brueckner (1995), Ortalo-Magne and Prat (2007), Hilber and Robert-Nicoud (2013), Parkhomenko (2018), and Parkhomenko (2021) make similar arguments. Ellickson (1977) postulates that it is easier for homeowners to become better organized in smaller municipalities, implying stricter zoning in smaller municipalities — similar to our paper. Owner-renter conflicts, while likely important, are orthogonal to our mechanism.}

Another strand argues that the primary purpose of zoning is to allow certain neighborhoods to stay homogenous or to exclude certain socioeconomic groups, often referred to as exclusionary zoning or fiscal zoning. This may be attributed to racial or social prejudice or to economic reasons such as minimizing the tax burden. For example, if lower-income people consume more public goods than the tax revenue that they pay, a neighborhood could keep them out by regulating large minimum lot sizes, which lower-income people cannot afford.\footnote{Along the same lines, Hamilton (1975) shows that in the presence of public goods, zoning is the mechanism through which the Tiebout (1956) hypothesis works — without zoning, anyone can move into a tiny house in a high public good neighborhood. Zoning allows citizens who prefer higher spending on public goods to sort themselves into the same neighborhood. Calabrese et al. (2007) numerically solve a more realistic version of Hamilton (1975) and show that zoning is likely to be strict, leads to aggregate welfare gains as in Tiebout (1956), but also to large welfare transfers, with poorer households suffering. Other papers in this strand of literature include Fischel (1978), Mills and Oates (1975), Erickson and Wollover (1987), and Wheaton (1993).}

---

5Brueckner (1995), Ortalo-Magne and Prat (2007), Hilber and Robert-Nicoud (2013), Parkhomenko (2018), and Parkhomenko (2021) make similar arguments. Ellickson (1977) postulates that it is easier for homeowners to become better organized in smaller municipalities, implying stricter zoning in smaller municipalities — similar to our paper. Owner-renter conflicts, while likely important, are orthogonal to our mechanism.

6Along the same lines, Hamilton (1975) shows that in the presence of public goods, zoning is the mechanism through which the Tiebout (1956) hypothesis works — without zoning, anyone can move into a tiny house in a high public good neighborhood. Zoning allows citizens who prefer higher spending on public goods to sort themselves into the same neighborhood. Calabrese et al. (2007) numerically solve a more realistic version of Hamilton (1975) and show that zoning is likely to be strict, leads to aggregate welfare gains as in Tiebout (1956), but also to large welfare transfers, with poorer households suffering. Other papers in this strand of literature include Fischel (1978), Mills and Oates (1975), Erickson and Wollover (1987), and Wheaton (1993).
A third strand, including our paper, focuses on externalities as a rationale for zoning. The externality channel we study — local governments limit housing supply to avoid the local costs of extra housing (e.g., traffic, congestion, crowded schools, limited green space) but do not internalize the global benefits of higher density (e.g., agglomeration, lower inequality, lower climate emissions) — has been discussed qualitatively for many years, mostly by legal and political science scholars, but also by some economists. For example, Briffault (1990), Briffault (1996), Cashin (2000), and Fischel (2008) all consider downsides from local authority over zoning rules. Biber et al. (2022) provide an excellent recent summary of this literature. However, as far as we are aware, there were no formal models and few empirical tests of this hypothesis.

We are aware of only two papers, both theoretical, that consider the role of municipal boundaries for zoning. Hamilton (1978)’s model has the opposite prediction from ours — zoning should be least restrictive when there are many small jurisdictions within a larger metro. Helsley and Strange (1995) build a model in which decentralized decision-making can lead to inefficiently restrictive zoning due to congestion externalities. However, in their model, zoning is most restrictive when there are two competing jurisdictions, while zoning approaches an efficient level as the number of jurisdictions rises. This is the opposite of our model prediction and empirical findings. Our paper is also related to Albouy et al. (2019), who build a model in which metros are endogenously formed at locations that differ by their productivity, although they do not consider municipal boundaries within metros. Due to decreasing returns to scale, local governments would choose metro sizes that are sub-optimally small, leading to too many cities and lower aggregate welfare.

There is also related research studying housing production in different neighborhoods within a municipality instead of a metro fragmented into municipalities. Clingermayer (1994) finds that when elections within a municipality happen by ward

---

7 Examples include Cooley and LaCivita (1982), Brueckner (1990), Engle et al. (1992), Brueckner (1998), Rossi-Hansberg (2004), Allen et al. (2016), and Vermeulen (2016), though the externalities they consider are quite different from ours.

8 On page 427, he writes, “full internalization of all local actions and full participation for all those affected by local decisions would tend to require larger local units.”

9 He predicts that metros with more fragmented governments choose stricter zoning — exactly the same argument we make — however, he neither solves a full model nor provides empirical evidence.

10 In his model, more restrictive zoning raises house prices and reduces population, but also raises metro-wide wages due to decreasing returns to scale in labor. If the metro is made up of many small jurisdictions, owners do not internalize the effect of local zoning on metro-wide wages, giving them fewer incentives to restrict zoning.
as opposed to at large,\textsuperscript{11} then group homes are less likely to be built in such municipalities. Similarly, Mast (2022) finds that when a municipality switches from at large to ward elections, housing permits within the municipality fall. Khan (2022) conjectures that the costs of development are more local than the benefits, leading to externalities for any development near a border. He uses ward boundary changes in Chicago to measure this externality. Similar to our paper, he argues that making zoning more global should reduce this externality, but unlike our paper, he does not formally investigate a relationship between zoning stringency and the locality of the zoning decision process. Marantz and Lewis (2022) find that housing production per census tract is lower in less populated municipalities. However, they do not test whether this is because a large metro is fragmented into smaller municipalities – which would be consistent with our findings – or if it is due to smaller metros being less dense – which would be unrelated.

There is a large literature studying the determinants of zoning empirically, although this literature has not considered municipal boundaries. Empirically, stricter zoning has been found to be associated with higher income, productivity, fiscal health, education, amenities, share of whites, and liberal lean. Many of these empirical findings are consistent with the fiscal and exclusionary zoning theories. The relationship between homeownership and zoning appears somewhat ambiguous, inconsistent with the owner-renter conflict theories. The relationship between zoning and density also appears ambiguous.\textsuperscript{12} Most of these studies are correlational without addressing the

\textsuperscript{11}In an at large election, all voters in a municipality vote for all members of the city council. Alternately, a municipality may be split into wards, with voters in each ward sending their own representative to the city council. Note that usually, most decisions are still made at the municipal level, however in a ward system, politicians may be more responsive to the desires of their ward’s citizens.

\textsuperscript{12}Gyourko and Molloy (2015) provide a survey of the literature, and we highlight some of it here. Erickson and Wollover (1987) provide empirical support for the fiscal theory by showing that poorer communities in Philadelphia are more likely to zone for business for tax benefits despite the negative externality of business near residential. Lutz (2015) finds that New Hampshire communities in stronger fiscal health chose more restrictive zoning. Rolleston (1987) studies 185 New Jersey communities and also finds support for the fiscal theory, as well as for exclusionary zoning because zoning is stricter in communities with fewer minorities. Evidence for theories of political conflict is mixed — Khan (2022) shows that home-ownership is associated with stricter zoning, while Brueckner (1998) and Glaeser and Ward (2009) find no effect. Glaeser and Ward (2009) and Evenson et al. (2003) low past density is associated with low future density and stricter zoning. On the other hand, Hilber and Robert-Nicoud (2013) and Saiz (2010) find somewhat conflicting results. Evenson et al. (2003) also find that higher income towns allow less commercial development but do not have stricter overall zoning. On the other hand, Gyourko et al. (2008) show that higher income and education were associated with stricter zoning. Saiz (2010) also shows that fewer Christians and
potential endogeneity of zoning regulation. Also, we are not aware of any papers that have linked zoning restrictions to the way administrative boundaries are drawn within a metro, as we do.

The empirical part of our study relies on being able to measure zoning for different metros across the U.S. Quigley et al. (2008), Jackson (2018), Mawhorter and Reid (2018), Menendian et al. (2020), Bronin (2022), and Metropolitan Area Planning Council (2020) all use either surveys or manual collection of local zoning ordinances to measure zoning restrictions in individual cities or states, but not across the entire U.S. Gyourko et al. (2008) and Gyourko et al. (2021) use surveys to construct the commonly used Wharton Residential Land Use Regulatory Index (WRLURI) for 2450 jurisdictions across the U.S., Puentes et al. (2006) use similar methods to create an index for 1844 jurisdictions. Gyourko and Krimmel (2021) show that the difference between the average and marginal costs of land can proxy for zoning restrictions. We use an estimate of minimum lot sizes by neighborhood across the entire U.S. from Song (2024) as our measure of zoning restriction. Section 3 provides more detail on how this measure is constructed. One benefit of our measure is that it is available across all jurisdictions, whereas many important jurisdictions are missing from WRLURI. Additionally, our measure does not rely on subjective surveys. Instead, it quantitatively measures minimum lot sizes across different jurisdictions. Nevertheless, our results are robust when using the Density Restriction Index from WRLURI.

Our paper is also tangentially related to the literature on the effects of zoning on other quantities of interest. Quigley and Rosenthal (2005) and Gyourko and Molloy (2015) provide extensive surveys on this literature. More restrictive zoning has been shown to increase house prices, reduce development, increase segregation, gentrification, and inequality, and reduce welfare. For example, in the San Diego metro, the cities of La Mesa, National City, San Marcos, Vista, Chula Vista, El Cajon, and Poway are available. However, the city of San Diego, with 42% of the metro’s population, is missing.

Glaeser and Gyourko (2002), Green et al. (2005), Glaeser and Ward (2009), Huang and Tang (2012), Kok et al. (2014), and Landis and Reina (2021) study prices; Rosen and Katz (1981) and Brueckner (1990) provide several additional references. Mayer and Somerville (2000), Jack-
2 Model

We present two different models, each predicts that the more local entities deciding zoning restrictions a metro is fragmented into, the more restrictive zoning regulations are in that metro. In both models, more restrictive local zoning benefits local homeowners through lower congestion externalities and higher house prices. However, more restrictive metro-wide zoning raises metro-wide house prices, hurting residents metro-wide through two related but distinct channels.

In the first model, high metro-wide house prices hurt homeowners because high prices reduce migration from other metros, which reduces metro-wide wages by weakening the agglomeration externality. The lower wages make all metro residents worse off. When zoning is decided locally, voters do not internalize their effect on metro-wide wages and choose zoning that is too restrictive. We refer to this as the ‘Migration across metros’ channel.

In the second model we shut down migration from other metros and agglomeration. Instead, high metro-wide house prices hurt homeowners because they may need to move to another municipality within the metro.\(^\text{15}\) When zoning is decided locally, voters do not internalize their effect on metro-wide house prices and choose zoning that is too restrictive. We refer to this as the ‘Migration within metro’ channel.

Other externality channels would work in similar ways. For example, dense neighborhoods are more environmentally friendly, but small locales may not internalize the effect of their actions on climate change. Alternately, similar to the ‘Migration across metros’ channel, high house prices may discourage migration of low-skilled workers, increasing the price of many non-tradable services.\(^\text{16}\) Similarly, high house prices are especially costly to lower-income households, leading to higher inequality and societal unrest; again, small locales may not internalize this.

\(^\text{15}\)Even if they do not need to move to another municipality, they may dislike high metro-level house prices because they want their children to be able to afford a home.

\(^\text{16}\)Eeckhout et al. (2014) show that complementarity of high and low skilled workers is important.
2.1 Migration across metros channel

There are two periods, \( t = 0 \) and \( t = 1 \). There are \( M \) landowners living in \( m \) identical municipalities in the metro at \( t = 0 \). There are \( N_i \) immigrants in municipality \( i \), for a total of \( N = \sum N_i \). The total number of people in municipality \( i \) is \( \frac{M}{m} + N_i \).

Municipalities matter only in that they determine their own local zoning rules. There is a fixed amount of land evenly distributed and normalized to one unit per landowner implying that density in municipality \( i \) is \( \frac{M/m + N_i}{M/m} = 1 + \frac{m N_i}{M} \). Let \( h_0 \) represent the amount of housing (e.g., housing units or floor space) to be built per unit of land. The inverse of \( h_0 \) will also represent how restrictive zoning is since less housing per unit of land represents lower housing density. The cost of constructing \( h_0 \) housing units, expressed in utility units, is \( \lambda h_0 \).

There is a congestion externality \( \phi(1 + \frac{m N_i}{M}) \), which is local and depends on the density in the municipality. In equilibrium, all municipalities are identical and density \( 1 + \frac{N}{M} \) is identical everywhere. The cost \( \phi \) may represent higher traffic due to density, lack of green space, pollution, or blocked view lines. In equilibrium, a lower \( h_0 \) leads to a smaller population and a lower congestion cost. All construction happens at \( t = 0 \), and no new housing is built at \( t = 1 \).

At \( t = 0 \), the landowners decide the zoning rules in their own municipality, solving the following problem:

\[
 u_0 = \max_{h_0} u_1(h_0) - \lambda h_0
\]

where \( u_1(h_0) \) is the \( t = 1 \) utility of an owner with \( h_0 \) units of housing and \( \lambda h_0 \) is the cost of building the housing. A low \( h_0 \) is interpreted as large minimum lot sizes, prohibition to subdivide lots such as single-family-only zoning, low maximum floor-to-area ratio, low maximum height restriction, or any other density restrictions.\(^{17}\)

At \( t = 1 \), we assume that landowners who own in municipality \( i \) stay in the same municipality. Each landowner in municipality \( i \) is endowed with housing \( h_{0,i} \) and inelastically supplies one unit of labor minus the congestion cost \( \phi_i \) at endogenous wage \( w \). The owner chooses non-housing consumption \( c \) and housing consumption \( h \) to maximize utility, taking wage \( w \), house price \( p_i \), and congestion \( \phi_i \) as given. The

\(^{17}\) An alternative is to interpret \( h_0 \) as land allowed for development. Households jointly own all land, which is plentiful. At \( t = 0 \), they decide how much of the total land to parcel out for private ownership — this is \( h_0 \) per household, with the remainder being public land for parks, roads, etc. In this case, a low \( h_0 \) represents a low ratio of developable land to total land.
landowner’s problem is:

\[ u_1(h_0) = \max_{c,h} c^{\alpha_c} h^{\alpha_h} \]

s.t. \( p_i h + c = p_i h_{0,i} + w(1 - \phi_i) \) \hspace{2cm} (2)

This problem can be solved analytically:

\[ c = \alpha_c (h_{0,i} p_i + (1 - \phi_i) w) \]
\[ h = \frac{\alpha_h (h_{0,i} p_i + (1 - \phi_i) w)}{p_i} \]
\[ u_1(h_{0,i}) = \alpha p_i^{1 - \alpha_h} (h_{0,i} p_i + (1 - \phi_i) w) \]

where \( \alpha = \alpha_c \alpha_h \). Plugging this into the owner’s problem at \( t = 0 \), we can rewrite it as:

\[ u_0 = \max_{h_0} \alpha p_i^{1 - \alpha_h} (h_0 p_i + (1 - \phi_i) w) - \lambda h_0 \]

where \( w, p_i, \) and \( \phi_i \) are beliefs about metro-wide wages, local house prices, and local congestion. These beliefs depend both on the zoning choices made in the owner’s municipality, as well as on the zoning choices made in other municipalities and the choices made by immigrants.

### 2.1.1 Immigrants

There are \( N \) additional immigrants that move to the metro at \( t = 1 \); \( N \) is endogenous and determined by indifference between living in the metro and receiving reservation utility \( u_m \). Their utility function is identical to the owners, but they have no housing endowment. Similar to equation 3, immigrants’ optimal choices and utility from living in municipality \( i \) of the metro are:

\[ c = \alpha_c (1 - \phi_i) w \]
\[ h = \frac{\alpha_h (1 - \phi_i) w}{p_i} \]
\[ u_{m,i} = \alpha p_i^{1 - \alpha_h} (1 - \phi_i) w \]

### 2.1.2 Equilibrium

We search for a symmetric Nash equilibrium to solve this problem. Owners in municipality \( i \) take the zoning choice \( \hat{h}_0 \) of all other municipalities as given and choose \( h_{0,i} \) to maximize their utility in equation 4. As an intermediate step to solve this problem,
we must also solve for the housing price in one’s own municipality \( p_i \), housing price in the other municipalities \( \hat{p} \), congestion in one’s own municipality \( \phi_i \), congestion in the other municipalities \( \hat{\phi} \), immigrants in one’s own municipality \( N_i \), immigrants in each of the other municipalities \( \hat{N} \), and the metro-wide wage \( w \), all as functions of \( \hat{h}_0 \) and \( h_{0,i} \).

The equilibrium is characterized by nine equations and nine unknowns. The unknowns are \( h_{0,i} \), \( \hat{h}_0 \), \( p_i \), \( \hat{p} \), \( \phi_i \), \( \hat{\phi} \), \( N_i \), \( \hat{N} \), and \( w \). The nine equations are:

\[
\begin{align*}
\underline{u}_m &= \alpha p_i^{-\alpha_p} (1 - \phi_i) w \\
\bar{u}_m &= \alpha \hat{p}^{-\alpha_p} (1 - \hat{\phi}) w \\
\frac{M}{m} h_{0,i} &= \frac{\alpha h_i}{p_i} \left( \frac{M}{m} h_{0,i} p_i + \left( \frac{M}{m} + N_i \right)(1 - \phi_i) w \right) \\
\frac{M}{m} \hat{h}_0 &= \frac{\alpha \hat{h}_0}{\hat{p}} \left( \frac{M}{m} \hat{h}_0 \hat{p} + \left( \frac{M}{m} + \hat{N} \right)(1 - \hat{\phi}) w \right) \\
\phi_i &= \phi_0 \left( 1 + \frac{mN_i}{M} \right)^{\phi_1} \\
\hat{\phi} &= \phi_0 \left( 1 + \frac{m\hat{N}}{M} \right)^{\hat{\phi}_1} \\
w &= \omega_0 \left( M + N_i + (m - 1)\hat{N} \right)^{\omega_1} \\
h_{0,i} &= \arg \max_{h_{0,i}} \frac{\alpha h_{0,i} p_i}{h_{0,i}} + (1 - \phi_i) w - \lambda h_{0,i} \\
\hat{h}_0 &= \hat{h}_0
\end{align*}
\]

The first and second equations above are the immigrants’ indifference conditions between living in municipality \( i \), living in any other municipality, or living outside of the metro; both equations use the immigrants’ utility derived in equation 5. The third equation equates the supply of housing in municipality \( i \) on the left-hand side with demand on the right-hand side. To obtain the aggregate demand, we sum the demand of owners in equation 3 and the demand of immigrants in equation 5, noting that the numbers of owners and immigrants in municipality \( i \) are \( \frac{M}{m} \) and \( N_i \) respectively. Similarly, the fourth equation equates the supply of housing with the demand for housing in any other municipality. The fifth and sixth equations define the congestion externality in municipality \( i \), and in any other municipality, the externality depends on the number of immigrants in the municipality. The seventh equation defines the metro-wide wage; it is increasing in total population where total immigrants are \( N = N_i + (m - 1)\hat{N} \). The parameter \( \omega_1 \) determines the strength of the agglomeration externality and is crucial for our main result.\(^{18}\) The eighth

\(^{18}\)For an equilibrium to exist, the agglomeration parameter must be sufficiently weak. In this case, as the number of immigrants rises, wages rise due to the agglomeration externality, but prices rise
equation is the Nash equilibrium condition implying that conditional on their beliefs about all other endogenous variables, the owners in municipality $i$ choose the housing supply (equivalently, the zoning restrictions) to maximize their own utility, derived in equation 4. Finally, the ninth equation specifies that the equilibrium is symmetric.

### 2.2 Migration within metro channel

This model is very similar to the model in section 2.1. To avoid repetition, rather than presenting the full model, we summarize the key differences between the two models here.

In this model, there are no migrants from other metros, and as a result, the wage is constant and normalized to one. The congestion externality function $\phi(h_0) = \phi_0 h_0^{\phi_1}$ is still local but now depends explicitly on the zoning rule $h_0$, with lower congestion when $h_0$ is low and zoning is stricter. We set the construction cost $\lambda = 0$ — the results are similar with positive costs but the model works even with a zero cost.\(^{19}\)

The key difference between the models is that between $t = 0$ and $t = 1$, with probability $q$ homeowners stay in their own municipality, and with probability $1 - q$ they receive a random moving shock and must move to one of the other $m-1$ municipalities with probability $q_{ij} = \frac{1-q}{m-1}$ when $i \neq j$. Conditional on moving from municipality $i$ to municipality $j$ at $t+1$, the household solves:

$$u_{1,j}(p_i h_{0,i}) = \max_{c_{ij}, h_{ij}} c_{ij}^{\alpha_c} h_{ij}^{\alpha_h}$$

$$\text{s.t. } p_j h_{ij} + c_{ij} = p_i h_{0,i} + w(1 - \phi(h_{0,j}))$$

This equation is identical to equation 2 but makes clear that the household’s net worth depends on the value of housing in $i$, but its spending on the value of housing

\(^{19}\)Because of the congestion externality, there is an interior solution for $h_0$ even when $\lambda = 0$. The migration across metros model in section 2.1 has a different specification for the congestion externality. As a result, it does require a construction cost to get an interior solution for $h$ if utility is Cobb-Douglas because, without construction costs, optimal structures $h$ are infinite. However, construction costs are not important for the mechanism. For example, if the utility function is CES with stronger complementarity than Cobb-Douglas, then the model has an interior solution for $h$ even with zero construction costs. In this case, our key results on the relationship between the number of voting municipalities $m$ and the restrictiveness of zoning $h$ are the same as in the model presented in Section 2.1.
in $j$. This problem can be solved analytically:

$$
c_{ij} = \alpha_c(x_i + (1 - \phi)w) $$
$$
h_{ij} = \frac{\alpha_h(x_i + (1 - \phi)w)}{p_j} $$
$$
u_j(x_i) = \bar{\alpha}p_j^{-\alpha_h}(x_i + (1 - \phi_j)w) $$

(8)

The ex-ante expected utility of a household who owns in neighborhood $i$ is:

$$
u_i = \sum q_{ij}u_j(x_i) = \bar{\alpha}(qp_i^{\alpha_h}(x_i + (1 - \phi(h_i))w) + (1 - q)\hat{\alpha}^{\phi_h}(x_i + (1 - \hat{\phi})w)) $$

(9)

As in section 2.1, we search for a symmetric Nash equilibrium to solve this problem, which leads to six unknowns ($h_{0,i}, \hat{h}_{0,i}, \hat{p}, \phi_i, \hat{\phi}$) and six equations:

$$
h_{0,i} = \frac{\alpha_h}{p_i} \left( q(h_{0,i}p_i + (1 - \phi_i)w) + (1 - q)(\hat{h}_{0,i}\hat{p} + (1 - \phi_i)w) \right) $$
$$
\hat{h}_0 = \frac{\alpha_h}{\hat{p}} \left( \hat{h}_{0,i}\hat{p} + (1 - \hat{\phi})w \right) $$
$$
\phi_i = \phi_i h_{0,i} $$
$$
\hat{\phi} = \hat{\phi}_0 \hat{h}_{0,i} $$
$$
h_{0,i} = \arg \max_{h_{0,i}} \bar{\alpha}p_i^{-\alpha_h}(q(h_{0,i}p_i + (1 - \phi_i)w) + (1 - q)(\hat{h}_{0,i}\hat{p} + (1 - \phi_i)w)) $$
$$
h_{0,i} = \hat{h}_0 $$

(10)

The first equation equates the aggregate supply of housing in municipality $i$ with the aggregate demand, which includes demand from current residents as well as residents moving from other neighborhoods. Similarly, the second equation equates the supply and demand of housing in any other municipality. The third and fourth equations define the congestion externality in municipality $i$, and in any other municipality, the externality depends on the strictness of zoning in the municipality. The fifth equation is the Nash equilibrium condition, implying that conditional on their beliefs about all other endogenous variables, the owners in municipality $i$ choose housing supply to maximize their expected utility from equation 9. Finally, the sixth equation specifies that the equilibrium is symmetric.

The six equations above can be reduced to two equations and two unknowns. It is also easy to check that when there is a single municipality — that is $q = 1$ and all moves are to places within the same jurisdiction — then the solution is equivalent to one chosen by a social planner.
2.3 Results

We use a numerical example to illustrate the model’s results. The model’s qualitative implications are robust to all parameter combinations we have tried.\textsuperscript{20} We are interested in how the model’s outcomes change as we vary the number of municipalities \( m \) that the metro is split into.\textsuperscript{21}

The panels on the left of Figure 1 present results from the migration within metro model as we vary \( m \), the number of decision making municipalities that the metro is fragmented into. A metro where all zoning decisions are made at the metro level corresponds to \( m = 1.0 \) and \( q = 1.0 \). For higher \( m \), the probability \( q \) of staying in the same municipality is lower. An increase in \( m \) implies that zoning decisions shift from global to local.

As zoning decisions become more local, zoning for the entire metro becomes more restrictive — \( h \) falls, which corresponds to lower density, such as larger lot sizes or less flexibility to subdivide. This leads to higher house prices \( p \) and lower congestion \( \phi \) for the entire metro. However, utility is maximized when \( m = 1 \); when \( m > 1 \), there is too little congestion and not enough housing. One can show that the utility maximizing planner’s solution — regardless of the number of jurisdictions \( m \) — is identical to the decentralized solution when \( m = 1 \). This is because when decisions are made globally, households fully internalize the trade-off between too little congestion and not enough housing. When \( m > 1 \) and decisions are made locally, voters choose too much zoning and too little housing because they benefit from low local congestion and high local house prices, but they do not internalize their effect on high metro-wide house prices. High metro-wide house prices are bad for households because they may need to move to another municipality and will be forced to pay a high price.

The panels on the right of Figure 1 present results from the migration across metro model as we vary \( m \). Similar to the previous model, as \( m \) rises, zoning becomes more

\textsuperscript{20}For the within metro channel, we set \( \alpha_h = 0.25, \alpha_c = 1 - \alpha_h, \phi_0 = 0.1, \phi_1 = 2.0, \) and \( w = 1 \). For the across metro channel, we choose the same parameters. Additionally, \( w \) is now endogenous and we set \( M = 1, \gamma_m = 0.65, \lambda = 0.1, \omega_0 = 1, \) and \( \omega_1 = 0.13 \). Quantitatively, the agglomeration parameter \( \omega_1 \) is crucial to strengthen the channel. We choose it based on estimates from Glaeser and Gottlieb (2009).

\textsuperscript{21}For the within metro channel, we assume that the probability of staying is inversely related to the number of municipalities \( q = 1/m \). This is sensible because the smaller each municipality, the higher the likelihood that any move will bring you to another municipality. When \( m = 1 \), any move within the metro leaves a mover in the same municipality.
restrictive, and $h$ falls. This leads to higher house prices $p$ and lower congestion $\phi$. But it also leads to too few migrants $N$. Because of the agglomeration externality ($\omega_1 > 0$), this results in lower metro-wide wages $w$. This increase in house prices and reduction in housing, congestion, population, and wages is suboptimal — as $m$ rises, utility falls. The intuition is very similar to the previous case. When decisions are made globally, households fully internalize the trade-off between quantity of housing, congestion, and wages. When $m > 1$ and decisions are made locally, voters choose too much zoning and too little housing because they benefit from low local congestion and high local house prices, but they do not internalize their effect on high metro-wide house prices, low metro-wide population, and low metro-wide wages. The difference between the two models is that in the migration within metro model, high metro-wide prices were bad because there was a possibility of moving and having to pay those prices. In the migration across metros model, high metro-wide prices were bad because they discouraged migrants and lowered wages.

The main message of our models is that if high metro-wide house prices are bad for residents, and if zoning decisions are made by fragmented jurisdictions within the metro then metro-wide zoning will be more restrictive than the social optimum. We focused on two reasons for why metro-wide house prices are bad: possibly needing to move to another part of the metro — this required $q < 1$, or discouraging migration into the metro and reducing metro-wide productivity — this required $\omega_1 > 0$.

2.4 Heterogeneity

In our model, all households are identical. As emphasized by Tiebout (1956), real-world households exhibit significant heterogeneity, which may be why some places have stricter zoning and less housing supply than others. For example, some may be willing to pay more to live in a town with large lots and green space, while others are not bothered by density. We choose to ignore heterogeneity in our model for two reasons: first, focusing on homogeneous households allows us to highlight the effects of jurisdictional fragmentation; second, there are many different dimensions of heterogeneity and many potential ways for governments to respond — it is unclear which are the most relevant to include in the model. Nevertheless, it is useful to discuss how including heterogeneity may affect our results.

We conjecture that even with heterogeneity, a metro with fragmented jurisdictions
Figure 1: Model results

On the left, figures present results from the ‘Migration within metros’ model in section 2.2 as we vary the probability of staying in one’s own neighborhood \((q)\). This is equivalent to increasing the number of equal-sized jurisdictions \((m = 1/q)\) that make independent zoning decisions; we plot \(m\) on the x-axis. On the right, figures present results from the ‘Migration across metros’ model in section 2.1 as we vary the number of neighborhoods \((m)\) that make independent zoning decisions. For each model, we show results for metro-wide housing quantity \(h\) (equivalently, the inverse of zoning restrictiveness), house prices \(p\), congestion \(\phi\), number of immigrants \(N\) (perfectly correlated with population \(M + N\)), and utility \(u\) in vertical panels.
would choose stricter zoning and less housing supply than socially optimal. The intuition is the same as in our decentralized model with local voting — even when local governments consider heterogeneity, they would still not internalize the effect of high house prices outside of their jurisdiction.

We also conjecture that if a global government was able to take heterogeneity into account, it could achieve a global optimal. It would allow more overall supply than fragmented governments, however, it would also account for heterogeneity. For example, it could allow some neighborhoods to have lower supply and less density for households who prefer this.

Of course, it may be too optimistic to believe that a global government would be capable of doing this. Perhaps a more realistic benchmark is a global government whose toolbox includes a one-size-fits-all solution of the same zoning everywhere. In this case, neither local nor global control would achieve the first best. Whether local or global control leads to higher welfare would be determined by a trade-off between having aggregate housing supply that is too low, versus aggregate housing supply that is just right, but some fraction of households having too little or too much density relative to their tastes.

3 Empirical Analysis: the U.S. Cross-Section of Metros

In this section, we empirically examine the relationship between the concentration of zoning authorities and zoning restrictiveness. To do so, we compile a map of local municipalities to compute the Herfindahl-Hirschman index (HHI) of governments in each metro, our independent variable of interest. We also construct a CBSA-level zoning stringency measure, our dependent variable of interest, using geographically detailed minimum lot size (MLS) data from Song (2024).

One major empirical challenge is the endogeneity of zoning jurisdiction boundaries. On the one hand, more exclusive communities may choose to incorporate in order to create their own local municipalities and set stringent zoning. This may be the case as local zoning was most actively adopted in the mid- to late-20th century, coinciding with postwar suburbanization. On the other hand, a metro with lax zoning may see its population grow faster, leading to the creation of new municipalities. As such, the
current HHI of municipalities may be endogenous and related to zoning stringency. Therefore, we instrument HHI with lagged HHI as of 1900, which is constructed in the same way as the current HHI but only using municipalities that existed in 1990. The identification relies on the assumption that the granularity of local governments as of 1900 is as good as random and not correlated with the stringency of zoning, which was set many years later, controlling for the observable location characteristics. In particular, we control for the political lean, climate conditions, the land area by land uses (commercial, industrial, and agricultural), land developability, and the proportion of residential development built after 1940 and 1970, when the zoning laws were most actively adopted. We also control for demographic, housing, and industrial characteristics in the mid–to late–1900s and the early 2000s. Online Appendix A.1 provides more detail on the data sources of the control variables.

3.1 Measuring zoning stringency

Our dependent variable of interest is the restrictiveness of local zoning in each metro. To measure zoning stringency, we use neighborhood-level minimum lot size estimates (MLS) from Song (2024). MLS is the most common type of density restriction in residential zoning across the U.S., and it requires the lot size to be no smaller than the MLS. A larger MLS means more restrictive zoning as they limit the number of housing units that can be constructed on a unit of land. Song (2024) looks for a structural break in the distribution of lot sizes for newly-built single-family houses using CoreLogic property tax data at the zoning district level, when available, or at the Census Block Group level to estimate neighborhood-level MLS across the U.S. This measure of local zoning stringency has a unique advantage in its comprehensive nationwide coverage and is suitable for our cross-sectional analysis of zoning.

We first take the median of these neighborhood-level MLS estimates in each municipality, weighted by the number of parcels in each neighborhood, and aggregate the municipality median MLS estimates at the CBSA level by taking the average. For robustness, we aggregate these neighborhood-level MLS estimates differently, for example, taking the 25th or 75th percentiles in each municipality (See Section 3.7).

\[\text{In Online Appendix ??, we investigate the number and length of river streams in each metro as instruments as they are commonly used in the literature. However, we find these instruments alone are too weak in the first stage and do not predict HHI. Therefore, nothing is significant in the second stage.}\]
Furthermore, we consider the Density Restriction Index in 2018 Wharton Land Use Survey as an alternative measure of zoning stringency and discuss the consistent result in Section 3.7.

### 3.2 Herfindahl-Hirschman index of zoning jurisdictions

We construct the zoning jurisdiction Herfindahl-Hirschman index (HHI) to characterize how decentralized zoning decisions are in each metro, our independent variable of interest. The zoning jurisdiction HHI is defined as:

$$HHI_i = (s_1 * 100)^2 + (s_2 * 100)^2 + \ldots (s_{n_i} * 100)^2$$  \hspace{1cm} (11)

where $i$ is the metro, $j$ is the zoning jurisdiction in metro $i$, $n_i$ is the number of zoning jurisdictions in metro $i$, and $s_j$ is the share of residential area of jurisdiction $j$ within the metro. Metros are defined to be CBSA. This index may range from 0 (infinitely many small jurisdictions) to 10,000 (a single large jurisdiction); if all jurisdictions were equal-sized, then the index would be equal to $HHI_i = 10,000/n_i$. Our theoretical model implies that there is a negative relationship between zoning jurisdiction HHI and zoning restrictiveness: when zoning decisions are more decentralized (low HHI), zoning regulations are more strict (high MLS).

To identify local governmental units that can set zoning laws, we follow Song (2024) to compile the 2019 Census County Subdivision and Census Place maps and refer to the 2010 Census Guide to State and Local Census Geography. Local zoning authority depends on the state institution and may be an incorporated place (often a city or town), a minor civil division (often a town or township), or a county. For example, in Connecticut, incorporated places are dependent of any minor civil divisions and have no zoning authority. Hence, all minor civil divisions are zoning jurisdictions in Connecticut. For another example, in California, only incorporated places have zoning authorities, and counties set zoning for unincorporated areas. In this case, all incorporated places and counties with unincorporated areas are zoning jurisdictions. As such, in each state, a combination of incorporated places, minor civil divisions, and counties are functioning local governments and are assumed to be zoning authorities in our empirical analysis.

We identify 21,067 municipalities (1,155 counties, often unincorporated areas gov-
erned by counties, 9,583 incorporated places, and 10,329 minor civil divisions) with functioning local governments in 909 CBSAs of the contiguous United States and construct a map of them in a shapefile format. We then merge the map of municipalities with CoreLogic Tax Assessor data. CoreLogic Tax Assessor data includes detailed parcel-level information on land uses and building characteristics for the near-universe of residential and non-residential properties. We use the data to obtain residential areas under each municipality and compute the zoning jurisdiction HHI in each metro.

3.3 Instrumental variable: historical HHI

We construct the HHI of local governments as of 1900 in each CBSA as an instrumental variable. It is analogous to the HHI described in Section 3.2 but uses only jurisdictions that existed in 1900. To identify municipalities that existed in 1900, we collect the year of municipality establishment — year of incorporation for incorporated places and year of creation for minor civil divisions and counties — from the following sources.

First, we use the Municipal Incorporation Data compiled by Goodman (Goodman, 2023). This data has nearly complete coverage of incorporation years of incorporated places in most states except for NE, OK, SD, and UT. To complement this, we use IPUMS Census sample data from 1850, 1860, 1870, 1880, and 1900 and determine whether incorporated places appear in each sample year. These 1% and 5% sample Censuses include the INCORP variable, describing which incorporated place the household lives in. We match the state and name of incorporated places with the IPUMS dictionary for the INCORP variable to decide whether incorporated places appeared in the Census data before 1900.\textsuperscript{23}

Similarly, we use the MCD variable in IPUMS Census data, which describes which minor civil division the household lives in, to determine the existence of minor civil divisions prior to 1900. Since the IPUMS data is not a full-count sample, we further digitize 242 historical maps that were drawn before 1900 and detect names of minor civil divisions on the maps. We assume the minor civil division was created before

\textsuperscript{23}The INCORP data field does not exist in the full-count Census. Hence, only using the IPUMS approach to determine the incorporation of municipalities prior to 1900 may result in false negatives. Similarly, the MCD data field below does not exist in the full-count Census and suffers from false negatives.
1900 if and only if a minor civil division appears in the IPUMS Census data or on the historical maps.

### 3.4 Descriptive statistics

We construct a sample of 19,153 municipalities in 834 Core-Based Statistical Areas (CBSA) for analysis with existing MLS estimates. Note that 75 CBSAs are excluded from the sample because of missing MLS estimates. The municipalities in the sample cover 84.6 million single-family residential tax parcels and 11.8 million multi-family residential tax parcels in CoreLogic, the near universe of residential properties in all CBSAs.

Table 1 presents the summary statistics of the analysis sample. The sample shows a large variation in its zoning stringency, measured by median minimum lot sizes. The municipality-level median of minimum lot size has a mean of 46,703 square feet (1.07 acres) and a standard deviation of 72,149 square feet (1.66 acres). The 10th percentile municipality median is 7,498 square feet (0.17 acre), the 50th is 18,034 square feet (0.41 acre), and the 90th is 114,824 square feet (2.64 acres). Across CBSA, the median minimum lot size is 27,332 square feet (0.63 acres), and the standard deviation is 27,715 square feet (0.64 acres).

Figure 2 illustrates the univariate relationship between zoning jurisdiction HHI and minimum lot size at the CBSA level. Zoning jurisdiction HHI and minimum lot size show a strong negative correlation, with the coefficient estimates of $-0.222$ from the log-log regressions. The $R^2$ implies that approximately 12% of the cross-sectional variation in minimum lot sizes across CBSAs is explained by zoning jurisdiction HHI alone.

---

24 The missing MLS data is due to incompleteness of underlying CoreLogic data. In estimating MLS, Song (2024) restricts to single-family construction built after 1940, which leads to missing MLS estimates in some counties where the construction year variable is sparsely filled in the CoreLogic data. In Section 3.7, we present the results using another version of MLS estimates from Song (2024), which use all single-family home construction to infer MLS and thus have full coverage. The empirical results are consistent.

25 According to 2020 American Community Survey, there are 86 million single-family housing units and 42 million multi-family housing units in CBSAs in the contiguous United States.

26 The municipality-level median is weighted by the number of parcels in each zoning district.

27 In each CBSA, we again take the median the municipality-level minimum lot size to compute its zoning stringency.
Table 1: Summary statistics

This table reports the mean, standard deviation, 10th percentile, 25th percentile, 50th percentile, 75th percentile, and 90th percentile of the minimum lot sizes, the zoning jurisdiction HHI, the number of municipalities, and the area of residential land at the municipality level (panel A) and at the CBSA level (panel B). In panel A, we compute the median of minimum lot size at each municipality weighted by the number of parcels in each zoning district and present the summary statistics. In panel B, we compute the mean of municipality-level minimum lot size in each CBSA and present the summary statistics. We also compute the zoning jurisdiction HHI, the number of municipalities, and the residential area in each CBSA to present the summary statistics.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. municipality-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median MLS (in sqft)</td>
<td>46,703</td>
<td>72,149</td>
<td>7,498</td>
<td>10,359</td>
<td>18,034</td>
<td>43,560</td>
<td>114,824</td>
</tr>
<tr>
<td>B. CBSA-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>median MLS (in sqft)</td>
<td>27,332</td>
<td>27,715</td>
<td>8,800</td>
<td>12,071</td>
<td>18,729</td>
<td>37,462</td>
<td>43,560</td>
</tr>
<tr>
<td>zoning jurisdiction HHI</td>
<td>4,485</td>
<td>3,285</td>
<td>527</td>
<td>1,199</td>
<td>4,084</td>
<td>7,738</td>
<td>9,178</td>
</tr>
<tr>
<td>HHI as of 1900</td>
<td>4,886</td>
<td>3,520</td>
<td>543</td>
<td>1,247</td>
<td>4,662</td>
<td>8,574</td>
<td>9,770</td>
</tr>
<tr>
<td># municipality</td>
<td>23.0</td>
<td>40.0</td>
<td>4</td>
<td>6</td>
<td>12</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td># municipality as of 1900</td>
<td>16.5</td>
<td>31.3</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td>residential area (in acre)</td>
<td>245,350</td>
<td>570,453</td>
<td>18,677</td>
<td>39,253</td>
<td>100,754</td>
<td>215,994</td>
<td>511,744</td>
</tr>
</tbody>
</table>
**Figure 2:** Correlation between zoning jurisdiction HHI and minimum lot size

The figure depicts the scatter plots (black dots) and their fitted lines (blue solid lines) where the x-axis is zoning jurisdiction HHI, and the y-axis is minimum lot size, aggregated at the CBSA level by taking the mean of median municipality-level MLS. Both axes are on a logarithmic scale. It also depicts binned data points with 20 bins (red crosses) and shows the coefficient estimate and $R^2$ from the univariate regression of log minimum lot size on log HHI at the CBSA level ($N = 834$).

![Figure 2](image-url)

\[ b = -0.222^{***} \ (R^2 = 0.122) \]

3.5 More fragmented metros have stricter zoning

Our baseline IV regression takes the form

\[
\log MLS_i = \beta_0 + \beta_{HHI} \cdot \log \widehat{HHI}_i + X_i \beta_X + \epsilon_i
\]

(12)

where $i$ is a Core-Based Statistical Area (CBSA), $MLS$ is the CBSA-level average of municipality-level median minimum lot sizes, $\widehat{HHI}$ is the predicted CBSA-level zoning jurisdiction HHI (instrumented by historical HHI as of 1900), and $X$ is a vector of location characteristics. We examine the effects of log HHI on the stringency of zoning, measured by log MLS. Section 3.7 discusses robust results under alternative functional form assumptions.

Columns (1) and (2) in Table 2 report the results from the CBSA-level OLS regressions using current $HHI$ without instruments. Column (2) includes the rich set of location controls, including 1940 and 1969 demographics, land use compositions, industrial compositions, weather, and political lean. As predicted by our model, HHI is strongly negatively related to zoning stringency measured by minimum lot sizes;
This table presents the results of the regression in equation 12. The dependent variable is a measure of zoning stringency, defined as the log of min lot size aggregated at the CBSA level. Column (1) is a univariate regression, and Column (2) includes the full set of control variables. Columns (3)-(4) report the coefficients of CBSA-level IV regressions. All columns include the full set of controls, which are log residential area; developable land in 2001; log population and fraction white in 1969; log population, fraction white, fraction black, fraction hispanic, average rent, average home value, average home ownership rate, average income, and average non-wage income in 1940; percent republican vote in 2000; average temperature, average temperature in January, and average precipitation; and a flag if there is no municipality in the MSA in 1900. Due to space limitations, we omit the coefficients for the controls.

<table>
<thead>
<tr>
<th></th>
<th>Outcome variable: log minimum lot size</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>IV: HHI as of 1900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log HHI</td>
<td>-0.2222***</td>
<td>-0.2011***</td>
<td>-0.2576***</td>
<td>-0.2189***</td>
</tr>
<tr>
<td></td>
<td>(0.0205)</td>
<td>(0.0309)</td>
<td>(0.0213)</td>
<td>(0.0328)</td>
</tr>
<tr>
<td>Observations</td>
<td>834</td>
<td>834</td>
<td>834</td>
<td>834</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.1223</td>
<td>0.3577</td>
<td>0.1182</td>
<td>0.3574</td>
</tr>
</tbody>
</table>

Significance: ***: 0.01, **: 0.05, *: 0.1

metros where zoning decisions are more dispersed (low HHI) have more strict zoning (high MLS). HHI’s t-statistic is $-3.79$, which is higher than any of the t-statistics among our large set of controls. HHI’s univariate $R^2$ is 0.122, more than a third of the $R^2$ in the full multivariate regression.

To address the endogeneity of HHI, our baseline regression adopts HHI as of 1900 as an instrument. In 28 CBSAs, there were no municipalities before 1900. Hence, we include the indicator variable of whether CBSAs had established municipalities before 1900 as a control variable while setting their HHI as of 1900 to 10,000. Columns (3) and (4) in Table 2 report the results from the CBSA-level IV regressions. First, we find a strong first-stage result: HHI as of 1900 is positively related to current HHI, controlling for observables, with a coefficient of 0.963 and a t-statistic of 79.62 (see Online Appendix Table A.2 for the full regression table of the first stage). Second, the main IV regression result shows strong support for our theory. The slope coefficient of $-0.219$ from the full regression in Column (4) indicates that if a CBSA with the median level of HHI had a centralized zoning authority, zoning would have been
18% less stringent. Similarly, an increase in concentration from the 25th to the 75th percentile of HHI would lead to a reduction in minimum lot sizes of 35%.

### 3.6 More fragmented metros have lower population, higher cost

We next investigate whether more stringent zoning driven by more local zoning authorities leads to higher housing costs or lower density, as predicted by our model. In particular, we run the following CBSA-level IV regression

\[
y_i = \beta_0 + \beta_{MLS} \cdot \log MLS_i + X_i \beta_X + \epsilon_i
\]

where \( y_i \) is a CBSA-level outcome variable that is either population or housing cost. We instrument \( \log MLS \) with \( \log HHI \) as of 1900 to construct the predicted value, \( \log MLS \), in the first stage. We include the full set of location controls \( X \). Table 3 reports the regression results. Columns (1) and (2) report the IV regression results with the CBSA population as the outcome variables in different functional forms. Although the coefficient of \( \log MLS \) loses its statistical significance in one of the specifications, we find a consistent negative relationship between \( \log MLS \) and population. Columns (3) and (4) report the IV regression results with the CBSA housing cost as the outcome variable in different functional forms. We find a positive and significant relationship between \( \log MLS \) and housing costs. The results imply that more stringent zoning due to the granular structures of municipalities as of 1900 decreases population while increasing housing costs, as predicted by our theory.

### 3.7 Robustness checks

#### Alternative minimum lot size estimates

Panels A and B in Table 4 present the robustness checks using alternative CBSA-level minimum lot size measures. Panel A uses alternative MLS estimates from Song (2024), where neighborhood-level minimum lot size estimates are estimated from

\[
28^{\text{Going from the median HHI to one entity implies HHI changing from 4,084 to 10,000 and MLS changing by } 1 - e^{-0.219 \log(10000)/e^{-0.219 \log(4084)}} = 0.178; \text{going from the 25th to the 75th percentile implies HHI changing from 1,199 to 7,738 and MLS changing by } 1 - e^{-0.219 \log(7738)/e^{-0.219 \log(1100)}} = 0.348.}
Table 3: Estimated effect of zoning stringency on other outcomes

This table presents the results of the regression in equation 13. The dependent variable in Columns (1) and (2) is the CBSA-level total population, and the dependent variable in Columns (3) and (4) is the CBSA-level median monthly housing. Both are from the 2020 ACS 5-year estimates. Columns (1) and (3) are level regressions, and Columns (2) and (4) are log regressions. All columns report the coefficients of CBSA-level IV regressions and include the full set of controls, which are log residential area; developable land in 2001; log population and fraction white in 1969; log population, fraction white, fraction black, fraction hispanic, average rent, average home value, average home ownership rate, average income, and average non-wage income in 1940; percent republican vote in 2000; average temperature, average temperature in January, and average precipitation; and a flag if there is no municipality in the MSA in 1900. Due to space limitations, we omit the coefficients for the controls.

<table>
<thead>
<tr>
<th>log MLS</th>
<th>(1) pop in mn.</th>
<th>(2) log pop</th>
<th>(3) $ monthly cost</th>
<th>(4) log cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.5809**</td>
<td>-0.0625</td>
<td>103.1**</td>
<td>0.0989**</td>
</tr>
<tr>
<td></td>
<td>(0.2441)</td>
<td>(0.0647)</td>
<td>(45.09)</td>
<td>(0.0438)</td>
</tr>
</tbody>
</table>

| Observations | 815 | 815 | 815 | 815 |
| Controls     | Yes | Yes | Yes | Yes |
| Adjusted R²  | 0.4249 | 0.9666 | 0.6095 | 0.6376 |

Significance: ***, 0.01, **, 0.05, *, 0.1
all single-family construction instead of restricting them to construction after 1940. These alternative MLS estimates cover all 909 CBSAs, while the precision may be lower than the baseline estimates due to including older single-family construction. Panel B uses the baseline MLS estimates from Song (2024), thus covering 834 CBSAs, but adopts a different aggregation scheme. We repeat the analysis using the median of municipality-level median MLS in each CBSA instead of taking the average. Both alternative minimum lot size measures do not affect the results, with small coefficient changes in the full specifications (8).

Using Wharton index as zoning stringency measure

Panel C in Table 4 presents the robustness checks with the more commonly used Wharton indices as the outcome variable (Gyourko et al., 2021). The Density Restriction Index (DRI) in the Wharton Residential Land Use Survey characterizes the stringency of density restrictions and reflects the largest minimum lot size required in each municipality.\(^{29}\) Although the survey does not have as many municipalities as our baseline minimum lot size estimates, the survey reports DRI in 2,434 municipalities of 560 CBSAs in the contiguous United States.\(^{30}\) We take the average of the DRI at each CBSA and repeat the CBSA-level OLS and IV regressions. The results are consistent with our prediction, with a coefficient of \(-0.1791\) and a t-statistic of \(-2.68\) in the full IV regression.

Number municipalities as a measure of the granularity of zoning authority

One alternative measure of the granularity of the zoning authority institution is simply counting the number of municipalities. Panel D in Table 4 presents the robustness checks using the number of municipalities instead of HHI as the independent variable. In the IV regressions in Columns (7) and (8), we instrument the number of municipalities with the historical number of municipalities as of 1900. Although HHI may better characterize the concentration of zoning authorities in each metro, the number of municipalities is robust to expansions and contractions of municipality areas.

\(^{29}\)DRI reports whether no MLS is imposed in the jurisdiction (DRI=0) and, if so, whether the largest MLS is no larger than 0.5 acres (DRI=1), between 0.5 and 1 acres (DRI=2), between 1 and 2 acres (DRI=3), or larger than 2 acres (DRI=4).

\(^{30}\)DRI is 0 for 160 of the 2,434 municipalities, 1 for 946 municipalities, 2 for 379 municipalities, and 4 for the rest 627 municipalities.
Table 4: Robustness checks

This table presents robustness checks to alternative zoning stringency measures and granularity measures in equation 12. Panel A uses alternative minimum lot size estimates from Song (2024) using all single-family construction instead of post-1940 with complete coverage. Panel B aggregates municipality-level median minimum lot sizes at the CBSA level by taking the median instead of the mean. Panel C uses the Density Restriction Index (DRI) in 2018 Wharton Land Use Survey. Here, we construct the outcome variable by taking the unweighted average of municipality-level DRI at each CBSA. Panel D uses the same outcome variable as the baseline regression but uses # of municipalities instead of HHI, instrumented by # of municipalities as of 1900, to define the independent variable. The coefficients of control variables are omitted due to space limitations.

<table>
<thead>
<tr>
<th>A. Outcome: MLS inferred from all construction</th>
<th>B. Outcome: Median of median</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) OLS IV</td>
<td>(5) OLS IV</td>
</tr>
<tr>
<td>log HHI</td>
<td>log HHI</td>
</tr>
<tr>
<td>-0.2718*** -0.2525*** -0.3025*** -0.2713***</td>
<td>-0.1031*** -0.2115*** -0.1230*** -0.2317***</td>
</tr>
<tr>
<td>(0.0193) (0.0287) (0.0201) (0.0305)</td>
<td>(0.0189) (0.0276) (0.0190) (0.0293)</td>
</tr>
<tr>
<td>Obs. 909 909 909 909</td>
<td>Obs. 834 834 834 834</td>
</tr>
<tr>
<td>Controls No Yes No Yes</td>
<td>Controls No Yes No Yes</td>
</tr>
<tr>
<td>Adj. $R^2$ 0.1791 0.3987 0.1760 0.398</td>
<td>Adj. $R^2$ 0.033 0.337 0.036 0.3362</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Outcome: DRI in Wharton Survey</th>
<th>D. Treatment: # of municipalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) OLS IV</td>
<td>(5) OLS IV</td>
</tr>
<tr>
<td>log HHI</td>
<td>log HHI</td>
</tr>
<tr>
<td>-0.1591*** -0.1925*** -0.1485*** -0.1791***</td>
<td>0.00042*** 0.0010 0.0038*** 0.0017***</td>
</tr>
<tr>
<td>(0.0376) (0.0624) (0.0393) (0.0667)</td>
<td>(0.0006) (0.0007) (0.0006) (0.0008)</td>
</tr>
<tr>
<td>Obs. 560 560 560 560</td>
<td>Obs. 834 834 834 834</td>
</tr>
<tr>
<td>Controls No Yes No Yes</td>
<td>Controls No Yes No Yes</td>
</tr>
<tr>
<td>Adj. $R^2$ 0.0294 0.0882 0.0284 0.0881</td>
<td>Adj. $R^2$ 0.0172 0.3169 0.0112 0.3160</td>
</tr>
</tbody>
</table>

Significance: ***: 0.01, **: 0.05, *: 0.1

due to boundary changes, such as annexation. The results are still consistent; the positive slope estimates imply that the more municipalities individually make zoning decisions, the more stringent minimum lot size regulations are set across the CBSA.

4 Case study: 1998 Toronto Municipality Amalgamation

In this section, we present a case study of municipality amalgamation in Toronto, Canada. In Canada, provinces have the power to create and dissolve municipalities. In January 1998, the previously incorporated municipalities of York, North York, East York, Scarborough, Etobicoke, and old Toronto were dissolved by the Government of Ontario. The six dissolved municipalities were replaced by the new municipality of Toronto, referred to as the Amalgamation of Toronto. No changes were made to
the other municipalities in the Toronto area. We exploit this municipality boundary change to provide more empirical evidence for our theory.

In Garrett (1999), Toronto’s chief administrative officer states that the primary goal of the amalgamation was a reduction of administrative and operating costs, as well as increased efficiency. Projections included a 10% reduction of the workforce and annual savings of $150 million. The move was unpopular — in 1997, a large majority of participating voters rejected amalgamation in a non-binding referendum. Of relevance to housing and development, post-amalgamation, a new Department of Urban Planning and Development Services was created. The city also introduced a new Strategic Plan with a vision for the city’s future.

Our model predicts that when several fragmented municipalities are amalgamated into a single larger one, the new city government will more optimally trade off the costs and benefits of density, leading to an easing of the regulatory burden for development. While we cannot measure regulatory burdens directly, we find that consistent with our model, housing starts grew faster in Toronto than in neighboring municipalities without substantial price responses, indicating supply regulations were relaxed after the amalgamation.

We apply a standard difference-in-difference (DiD) analysis with the six former municipalities making up the city of Toronto (treatment group) and the five municipalities that share a border with the city of Toronto but were not amalgamated (control group), including Mississauga, Brampton, Vaughan, Markham, and Pickering pre- and post-1998. Individual data for the six amalgamated municipalities is unavailable post-1998, therefore we always aggregate them into a single treated region; we use the term amalgamated Toronto to refer to the entire treated region, both before and after 1998. At the time of amalgamation, there were approximately 2.4 million people living in amalgamated Toronto, compared to 1.3 million in the five neighboring municipalities. Online Appendix figure A.1 presents a map of the region.

We begin by plotting the raw data. Data on housing starts, rents, and prices of new construction is from the Canada Housing and Mortgage Corporation (CHMC). Data on income, population, and density is from Statistics Canada. All data is publicly available.
begin to grow much faster. By 2010, starts are more than twice the level they would have been if Toronto stayed on the same trend as its neighbors.

We next test the statistical significance of this difference. Our baseline regression takes the form

$$\log Y_{i,t} = \alpha_t + Treated_i + \beta Treated_i \times Post_t + \gamma X_{i,t} + \epsilon_{i,t}$$ (14)

In this regression, \( t \) represents the year, and \( i \) represents the municipality. Our variable of interest \( Y_{i,t} \) is either housing starts, average rent, or average price, scaled by its 1997 level. The independent variables include time fixed effects \( \alpha_t \), a treatment group dummy \( Treated_i \), which equals one for amalgamated Toronto, a post dummy \( Post_t \), which equals one for time periods after amalgamation (\( t \geq 1998 \)). \( X_{i,t} \) represents income controls (each municipality’s median wage for each lagged year) and density controls (computed in 1989 for each municipality). Our coefficient of interest, \( \beta \), measures the impact of amalgamation on the variable of interest.

In panel A of Table 5, we present the results of the DiD analysis by comparing amalgamated Toronto to the five neighboring municipalities, thus including one treated unit and five control units. In panel B, we aggregate the five control municipalities into a single control as an alternative specification, thus including one treated unit and one combined control unit. We do this because the analysis in panel A implicitly treats all five control municipalities equally, but they differ in size, and it is informative to compare amalgamated Toronto to the aggregate control. The results are mostly consistent across the specifications.

The first two columns of Table 5 shows that the number of Toronto’s housing permits was 43%-78% higher than it would have been had Toronto stayed on the same trend as its neighbors in the 22-year period after amalgamation, consistent with our model and earlier empirical results.\(^{32}\) These estimates are statistically significant, with t-statistics between 2.8 and 3.5.

There are some limitations to this empirical approach. Although both treated and control areas contain urban and suburban neighborhoods, amalgamated Toronto tends to be more urban, more dense, and lower earning, with smaller households and

\(^{32}\) These estimates are for housing starts. They are not quite comparable to the estimates in Table 2, which imply that going from six to one jurisdictions causes MLS to change by

$$e^{-0.215 \log(10000/6)} / e^{-0.215 \log(10000)} - 1 = 46.7\%.$$
The figure depicts the 1990-2020 time series of total housing starts, average rent, and average price for newly built properties. The solid line represents the current Toronto municipality; prior to 1998 the same geographical area was covered by the municipalities of Etobicoke, North York, York, East York, old Toronto, and Scarborough. The dashed line represents the aggregate of the cities of Mississauga, Brampton, Vaughan, Markham, and Pickering, which each share a contiguous boundary with the City of Toronto. When aggregating, average rents and prices were weighted by each city’s number of rental and newly built properties, respectively. The data comes from the Canada Mortgage and Housing Corporation (CMHC).
Table 5: Estimated effect of Toronto’s amalgamation in the housing market

This table presents the results of the regression in equation 14. The dependent variable in Columns (1) and (2) is the number of housing starts; in Columns (3) and (4), it is the rent; and in Columns (5) and (6), it is the price of newly built units, all in logs. The data comes from Canada Mortgage and Housing Corporation (CMHC), with the sample period from 1990 to 2020. In addition to the fixed effects and treated dummy, which are controlled in all columns, Columns (2), (4), and (6) control for median income in each municipality in each year and for density in each municipality as of 1989. The controls come from Statistics Canada, are available in 5-year periods, and are interpolated for years in between.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log # starts</td>
<td>log # starts</td>
<td>log rent</td>
<td>log rent</td>
<td>log price</td>
<td>log price</td>
<td>log price</td>
</tr>
<tr>
<td>Panel A: Treated versus individual untreated municipalities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated x Post</td>
<td>0.7453***</td>
<td>0.7846***</td>
<td>0.0474</td>
<td>0.0214</td>
<td>0.4260</td>
<td>0.3771</td>
</tr>
<tr>
<td>(0.2566)</td>
<td>(0.2572)</td>
<td>(0.0297)</td>
<td>(0.0218)</td>
<td>(0.3013)</td>
<td>(0.3010)</td>
<td></td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Income, Density controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Geographies (N)</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Years (T)</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Observations (N x T)</td>
<td>186</td>
<td>155</td>
<td>186</td>
<td>186</td>
<td>155</td>
<td>186</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.3213</td>
<td>0.3252</td>
<td>0.9168</td>
<td>0.9558</td>
<td>0.0677</td>
<td>0.0796</td>
</tr>
<tr>
<td>Panel B: Treated versus single merged untreated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treated x Post</td>
<td>0.5989***</td>
<td>0.4250***</td>
<td>0.0693***</td>
<td>0.0643***</td>
<td>0.1074</td>
<td>0.1147</td>
</tr>
<tr>
<td>(0.1748)</td>
<td>(0.1133)</td>
<td>(0.0080)</td>
<td>(0.0073)</td>
<td>(0.0722)</td>
<td>(0.0752)</td>
<td></td>
</tr>
<tr>
<td>Year FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Income, Density controls</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Geographies (N)</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Years (T)</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Observations (N x T)</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.7223</td>
<td>0.8894</td>
<td>0.9962</td>
<td>0.9970</td>
<td>0.9230</td>
<td>0.9207</td>
</tr>
</tbody>
</table>

Significance: ***: 0.01, **: 0.05, *: 0.1

fewer children. Therefore, there may be confounding factors. For example, if a demand shock made urban living more attractive, or a productivity shock made urban areas higher earning, and if this shock happened around 1998, then housing development in amalgamated Toronto would have likely increased for reasons unrelated to supply relaxation due to Amalgamation of Toronto. The literature typically deals with this by a discontinuity analysis in a relatively small area along the border of treated and control regions, where housing and household characteristics are similar.33

Unfortunately, our data is not sufficiently granular to adopt a boundary discontinuity design. Instead, we analyze the response of rents and prices to the amalgamation.

---

33For example, see Diamond and McQuade (2019).
to rule out many potential confounders. If a demand shock around 1998 was responsible for the increase in housing starts in amalgamated Toronto relative to neighboring municipalities, then it is likely that rents and prices would have risen at the same time. Reasonable estimates of supply elasticities imply that the rise in prices should be similar in magnitude to the rise in housing supply. Furthermore, the rise in rents and prices would likely precede the rise in housing starts because market prices can react more quickly to news than real estate investment decisions.

The bottom two panels in Figure 3 show no obvious signs of price growth and a quantitatively small increase in rent growth in the treated region compared to control regions. This is especially true in the period preceding the increase in housing starts—the small increase in rent growth that did happen in the treated region occurred after 2015. For comparison, by 2015 housing starts in treated were nearly triple that of the control trend.

Columns (3) and (4) show the effect of the amalgamation on rents. They are statistically insignificant in panel A but significant in panel B where much of the time series variation in rent growth across municipalities in the control group is smoothed out by aggregation. Despite some estimates being statistically significant, the point estimates in both panels are an order of magnitude smaller than for housing starts, which suggests that a supply rather than a demand shock is driving the large response in housing starts. Columns (5) and (6) show the effect on prices of newly built units. None of the estimates are statistically significant. The point estimates in panel A are larger, though still statistically insignificant, but this is driven by smaller municipalities. The point estimates for the aggregated control in panel B are also close to zero. Again, this suggests that it is unlikely that the rise in housing starts in amalgamated Toronto after 1998 was driven by demand shocks.

Saiz (2010) estimates the housing supply elasticity to be 1.75 for the average U.S. metro, and around 0.8 for larger metros like Toronto. This implies that a 1% rise in prices or rents (capitalized into prices) should be associated with a 0.8% increase in supply.

If construction costs were low and delays short, then a demand shock would result in a rise in construction but little effect on prices or rents as new supply would offset new demand. However, in the real world, construction delays are significant. Delays, even with low costs, lead to a temporary rise in prices in response to demand shocks. Furthermore, real-world construction costs exhibit increasing return to scale. For example, once all buildable land has been used, it becomes increasingly expensive to flatten hills or dry swamps. Similarly, building tall skyscrapers is more expensive per square foot than single-family homes. Such costs lead to permanent increases in prices in response to demand shocks. On the other hand, an increase in housing starts but little change in prices or rents is consistent with a case where individuals are relatively indifferent between living in treated
5 Conclusion

We identify a novel reason why zoning is too strict: if zoning rules are determined locally, decision-makers do not internalize the effect of these rules on global house prices. We build two models illustrating that when larger metros are fragmented into smaller decision-making municipalities, then local zoning is set too strictly. The fragmented municipalities minimize local congestion externalities by choosing strict zoning, but this hurts residents because it leads to too little aggregate housing supply. We find strong empirical support for this mechanism in the cross-section of metros — the more subdivided a metro is into local governments, the more stringent its zoning rules are. We also find that when several municipalities in the Toronto metro merged, housing starts in those municipalities rose significantly faster than in neighboring municipalities. Our research suggests clear policy advice — policymakers worried about housing affordability should redirect zoning decisions from municipalities to state or national levels. Since 2019, several places have done exactly this.

References


Albouy, David, and Gabriel Ehrlich, 2018, Housing productivity and the social cost of land-use restrictions, Journal of Urban Economics 107, 101–120.


Goodman, Christopher B., 2023, Municipal Incorporation Data, 1789-2020.


Hamilton, Bruce W., 1975, Zoning and property taxation in a system of local governments, *Urban Studies* 12, 205–211.


Landis, John, and Vincent J. Reina, 2021, Do restrictive land use regulations make housing more expensive everywhere?, *Economic Development Quarterly* 35, 305–324.

Lens, Michael C., and Paavo Monkkonen, 2026, Do strict land use regulations make metropolitan areas more segregated by income?, *Journal of the American Planning Association* 82, 6–21.


Mawhorter, Sarah, and Carolina Reid, 2018, Terner california residential land use survey, University of California, Berkeley.


Metropolitan Area Planning Council, 2020, Zoning atlas.


Parkhomenko, Andrii, 2018, The rise of housing supply regulation in the u.s.: Local causes and aggregate implications, Working paper.


## Appendix

### A Supplementary Figures and Tables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land area by uses (total residential land area, % residential, % agricultural, % commercial, % industrial, % multifamily among residential)</td>
<td>CoreLogic property tax records in 2018</td>
</tr>
<tr>
<td>% residential properties built after 1940 and 1970</td>
<td>CoreLogic property tax records in 2018</td>
</tr>
<tr>
<td>1969 demographics (total population, % white)</td>
<td>Survey of Epidemiology and End Results</td>
</tr>
<tr>
<td>1940 housing and financial characteristics (total population, % white, % Black, % Hispanic, % homeownership, avg home value, avg rent, avg income, % with nonwage income)</td>
<td>IPUMS complete-count Census</td>
</tr>
<tr>
<td>Presidential election returns (% rep., % dem. from 2000)</td>
<td>MIT Election Data + Science Lab</td>
</tr>
<tr>
<td>Weather condition (avg temperature year-around and in Jan, avg precipitation)</td>
<td>PRISM Weather data</td>
</tr>
<tr>
<td>Land developability index in 2021</td>
<td>National Land Cover Database</td>
</tr>
<tr>
<td>1990 industry shares (by # of establishments)</td>
<td>County Business Patterns</td>
</tr>
</tbody>
</table>
Table A.2: First Stage Regressions

This table presents the coefficients of the 1st stage IV regression in equation 12. The dependent variable is a current HHI, and the instrument is HHI as of 1900. Column (1) is a univariate regression, and Column (2) includes the full set of control variables, which are log residential area; developable land in 2001; log population and fraction white in 1969; log population, fraction white, fraction black, fraction hispanic, average rent, average home value, average home ownership rate, average income, and average non-wage income in 1940; percent republican vote in 2000; average temperature, average temperature in January, and average precipitation.

<table>
<thead>
<tr>
<th>Outcome variable: log HHI</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log HHI as of 1900</td>
<td>0.9712***</td>
<td>0.9629***</td>
</tr>
<tr>
<td></td>
<td>(0.0078)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>1(No municipality before 1900)</td>
<td>-0.3041***</td>
<td>-0.2132***</td>
</tr>
<tr>
<td></td>
<td>(0.0493)</td>
<td>(0.0476)</td>
</tr>
<tr>
<td>log Area of residential land</td>
<td>0.0236*</td>
<td>(0.0133)</td>
</tr>
<tr>
<td>2001 land developability</td>
<td>0.0003</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>log 1969 population</td>
<td>-0.1950***</td>
<td>(0.0307)</td>
</tr>
<tr>
<td>1969 % white</td>
<td>-10.90***</td>
<td>(3.439)</td>
</tr>
<tr>
<td>log 1940 population</td>
<td>0.0924***</td>
<td>(0.0294)</td>
</tr>
<tr>
<td>1940 % white</td>
<td>-0.0004</td>
<td>(0.0027)</td>
</tr>
<tr>
<td>1940 % Black</td>
<td>-0.0009</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>1940 % Hispanic</td>
<td>0.0006</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>log 1940 avg. rent</td>
<td>0.0122</td>
<td>(0.0140)</td>
</tr>
<tr>
<td>log 1940 avg. home value</td>
<td>0.0427</td>
<td>(0.0367)</td>
</tr>
<tr>
<td>1940 ownership rate</td>
<td>-0.0035***</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>log 1940 avg. income</td>
<td>0.0374</td>
<td>(0.0346)</td>
</tr>
<tr>
<td>1940 % with nonwage income</td>
<td>0.0019</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>2000 % republican votes</td>
<td>0.0004</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>avg. temperature</td>
<td>-0.0033</td>
<td>(0.0088)</td>
</tr>
<tr>
<td>avg. temperature in Jan</td>
<td>-0.0010</td>
<td>(0.0066)</td>
</tr>
<tr>
<td>avg. precipitation</td>
<td>0.0383***</td>
<td>(0.0115)</td>
</tr>
</tbody>
</table>

Observations: 834 834
Adjusted R²: 0.94979 0.96233

Significance: ***: 0.01, **: 0.05, *: 0.1
Figure A.1: Toronto map of treated and control areas

The figure depicts a map of the city of Toronto, as well as several neighboring municipalities, all in the Toronto metropolitan area. The municipal borders are drawn by the thicker lines. Mississauga, Brampton, Vaughan, Markham, and Pickering are in the control group, while Toronto is in the treated group. Within Toronto, the former municipal borders (dissolved in 1998) of Etobicoke, North York, York, East York, Scarborough, and old Toronto are outlined by the thinner lines.