

Land Use Regulations and Housing Development

Evidence from Tax Parcels and Zoning Bylaws in Massachusetts*

BRENDAN SHANKS[†]

February 10, 2021

JOB MARKET PAPER

[Click here for the most recent version.](#)

Abstract

Land use regulations come in a wide variety of forms and govern how development occurs. They restrict housing development resulting in housing supply being less responsive to demand shocks. Yet little is known on what facets of residential development are most impacted, hindered by lack of comprehensive data on land use regulation stringency. I address this shortcoming by compiling a novel measure of land use regulation based on applying natural language processing techniques to over 40,000 pages of zoning bylaw texts. Utilizing a spatial regression discontinuity design around municipal borders, I find that stringent land use regulations reduce housing supply primarily through increasing the land usage per house. Strongly regulated localities do not compensate by developing more land overall. These results highlight how regulations like minimum lot sizes and setback requirements pose barriers to housing development in high-growth regions.

JEL Classification: R14, R31, R52

Keywords: zoning, urban economics, housing supply, natural language processing, machine learning

*I am highly indebted to my advisors Joachim Winter, Stephan Heblich, and Davide Cantoni for their continued encouragement and guidance. I thank Vojtěch Bartoš, Nathaniel Baum-Snow, Lukas Buchheim, Mathias Bühler, Elisabeth Grewenig, Eleonora Guarnieri, Paul Hufe, Rainer Kotschy, Matthias Lang, Martina Magli, Lukas Mergele, Andreas Steinmayr, William Strange, Gregory Veramendi, and seminar participants in Munich and Toronto for helpful comments and suggestions. All remaining errors are my own.

[†]LMU Munich, Ludwigstr. 33, 80539 Munich. Email Address: brendan.shanks@econ.lmu.de

1 Introduction

Land use regulations (LUR) have been in place across US cities from the early 20th century, but did not take off until the 1970s (Gyourko et al., 2013). They dictate how land can be used, govern what can be built, where and how, and determine the role of local residents in the decision making process.¹ They are generally set at a local level, such as a township or municipality. These regulations are meant to address externalities from potential market failures: separating polluting sources from residential areas, reducing urban sprawl, coordinating development with transportation, amongst others. Recent work suggests, however, that the costs of restricting development outweigh the benefits (Turner et al., 2014; Hsieh and Moretti, 2019). These costs include reduced aggregate housing supply (Saiz, 2010; Hilber and Vermeulen, 2016) and housing markets that are less elastic to labour (Diamond, 2016) and immigration shocks (Saiz, 2007). However, there is a lack of evidence on how LUR impact housing development and therefore lessen supply. This is important for policy makers looking to understand the implications of LUR. Knowledge of these effects can also provide clues for *what* sorts of regulations are most consequential.

Moreover, it is infeasible to consider every potential regulation individually. First, the set of all possible regulations to choose from is large. Second, local stakeholders often have influence on the development process. Third, many regulations may seem different, but restrict development similarly. For example, restrictions on lot shape and lot width both serve to increase land usage per house. A common way to mitigate these issues is to create indices of LUR stringency meant to capture the overall regulatory burden. The most well-known of these indices is the Wharton Residential Land Use Regulation Index (WRLURI, Gyourko et al., 2008), derived from survey responses of 2649 US localities in 2005. Yet to answer questions on the impact of LUR that require measures of regulation stringency in out-of-sample jurisdictions, or in different years, there is no obvious solution.

This paper presents a novel measure of land use regulatory intensity derived from applying a machine learning method called the Latent Dirichlet Allocation model on over 40,000 pages of zoning bylaw texts from close to all municipalities in Massachusetts.² This paper focuses on the current regulatory environment in the state, but the technique can be applied in other jurisdictions on bylaws from different time periods. With this new index I then investigate how stringent LUR are manifested in housing development. First, I ask whether stringent LUR reduce the density of the housing stock at the municipal level. Next, I ask how restrictive regulations are reflected in the housing market, in housing attributes, and in land use. This allows me to provide evidence on

¹Some examples of LUR are minimum lot sizes, mandating a certain number of parking lots based on a building's size, establishing buffer zones around wetlands, and zoning (sorting usages across space).

²My measure has a coverage rate of 97.2% in Massachusetts compared to the 22.5% from the WRLURI.

what *sets* of rules are most likely to be the most restrictive. Finally, I investigate whether these restrictions are reflected in local house prices.

To answer these questions, I apply several natural language processing (NLP) techniques to municipal zoning bylaws in Massachusetts.³ To benchmark the resulting regulatory measures obtained from these text-based methods, I compare the NLP-derived results with the WRLURI, which is often used in the literature, as well as an index created from the Pioneer/Rappaport Housing Regulation Database (PRHRD). I find that an index created from a Latent Dirichlet Allocation (LDA) model (a latent finite-mixture machine learning model) best captures the variation in the land use regulatory environment. My LDA-based measure has a correlation coefficient with the other indices of around 0.6. This index is a natural complement to the existing measures: the current survey-based indices aid in interpreting the uncovered latent factors while the LDA method allows for expanding spatial and temporal coverage. Furthermore, it only requires the text from municipal zoning bylaws. I call this standardized measure the Natural Language Processing Zoning Stringency Index (NALPZ).

I then employ the NALPZ index in a spatial regression discontinuity design with municipal borders as cut-offs to evaluate the impact of stringent LUR on the pattern of housing development. This strategy exploits variation in the regulatory environment at 727 borders across 271 towns and controls for local housing demand, amenities, and tastes by comparing spatially close houses, only subjected to different LUR. Three main sources of data are utilized: i) housing characteristics from tax parcel data for all of Massachusetts, ii) Lidar data to calculate building heights,⁴ and iii) land use data.

As there are various facets of housing development that can be affected by LUR, I group my outcomes into three main categories to add structure. The first group of outcomes are related to the housing market: building age and the year a house was last sold. They capture the rate of new housing development and turnover in the housing market. The second group of outcomes reflect housing attributes: building (livable) area, lot size, and building height. This group addresses how the shape and size of the houses themselves are influenced by stringent LUR. The final group of outcomes concern land use: the rate of conversion of undeveloped land to residential usage as well as the share of residential land of all developed land. These outcomes speak to how LUR shape the spatial pattern of residential development.

Though I lack exogenous variation in *specific* LUR, by evaluating the impact of the *overall* restrictiveness of LUR on different features of housing development, I can speak to what types of regulations are most likely binding. For example, regulations that cap the overall rate of de-

³Massachusetts is an ideal location for this analysis as it has previous measures of LUR to aid in benchmarking and interpretation.

⁴Lidar (light detection and ranging) is a remote sensing method for measuring distances, often used to derive high-resolution maps.

velopment, such as growth controls, are more likely to impact the housing market rather than the attributes of the houses themselves. On the other hand, parcel-specific regulations, such as floor-area-ratios, are more likely to manifest themselves in the shape and appearance of houses.

This paper also speaks to the effect of LUR on local housing prices. Though LUR increase housing prices when considering average prices across municipalities in a regions (Hilber and Vermeulen, 2016), their effect at local levels depends on the degree of substitutability between nearby houses in different towns. From the tax parcel data, I observe the price the house was last sold for, as well as the most recent tax assessed value, broken down into building and land components. I use these outcomes in the spatial RDD to test how LUR impact house prices at the local level.

The results reveal several interesting findings. First, at the aggregate level, LUR strongly restrict development. The density of residential houses at town borders is lower in more restrictive municipalities. A standard deviation increase in NALPZ reduces housing density by 20% of the mean. Second, when considering the different groups of outcomes, I find that housing attributes react most strongly to restrictive regulations. For example, stringent LUR leads to significantly larger lot sizes (27% larger for each standard deviation increase in the regulatory index).⁵ House sizes and heights are essentially not impacted by these regulations. Third, though the housing market and land use respond to LUR, the effects are economically small compared with housing characteristics. Houses are slightly older and a marginally higher fraction of developed land is allocated to residential use in more regulated towns. The rate of land conversion (undeveloped to residential) is not influenced by LUR. Fourth, house prices, after controlling for building and lot sizes, are about 5–6% higher for every standard deviation increase in the index. However, this is mostly explained by school district quality. This provides evidence for nearby houses across municipal borders being substitutes for one another.

I show that controlling for school quality measures, such as per pupil spending and graduation rates, school district fixed effects, or municipal property tax rates does not change the interpretation of my results. Several tests confirm that the results are not being driven by unobserved amenities or pre-existing differences in demographic characteristics. To test for the role of unobserved amenities, I estimate local amenity values with a canonical urban general equilibrium model (Ahlfeldt et al., 2015) and examine whether these vary differentially across municipal borders. Highly regulated towns do have higher amenity levels on average, but this relationship disappears when comparing neighbouring census block groups across borders, as my identification strategy does. Further, the demographic composition of census blocks at town borders pre-widespread LUR does not vary with current regulatory stringency.⁶

⁵The average difference in the NALPZ across borders in Massachusetts is 0.74.

⁶A census block corresponds roughly to a city block. A census block group contains 31.6 blocks on average.

Taken together these results suggest that regulations that increase land usage *per* house are primarily responsible for constraining housing density and supply. For policy makers whose aims include increasing the availability of housing, regulations such as shape restrictions, setback requirements, and minimum lot sizes should be scrutinized.

This paper contributes to two strands of literature. The first strand of literature deals with the measurement of land use regulations.⁷ I contribute to this literature by using natural language processing techniques on zoning bylaws to measure regulatory stringency. The novel use of a machine learning algorithm to measure LUR builds on previous work. The two most relevant works for this paper are the Massachusetts Regulation Database compiled by the Pioneer Institute for Public Policy Research and Rappaport Institute for Greater Boston (PIRI, 2005) and Gyourko et al. (2008). LUR are high-dimensional, coming in various guises, making it difficult to summarize the regulatory environment with a comprehensive measure.⁸ This has made fully mapping the impact of more restrictive zoning on the pattern of housing development challenging. Remarkable attempts have been made to create measures of regulatory intensity. Researchers at the PIRI (2005) have compiled an extensive database of municipal land use regulations for towns around Boston (PRHRD). Gyourko et al. (2008) sent surveys to 6,896 jurisdictions across the US to establish the WRLURI. The thoroughness of these endeavours has resulted in a very detailed picture of their respective regulatory environments.

Though these have been necessary undertakings in helping our understanding of LUR, they are not without their disadvantages. First, the survey based methods suffer from nonresponse (Gyourko et al. (2008) had a 38% response rate). Even if the pattern of nonresponse is random, we lack measures of overall LUR stringency for many jurisdiction. Second, they are very resource intensive. Compiling the PRHRD required a research team consisting of a project manager, senior researcher, and twelve research assistants. This makes it difficult to administer in other localities in different time periods. Third, they are limited in the time dimension, usually depicting the regulatory environment at one point in time.⁹

I extend this previous work by creating an index of regulatory restrictiveness (NALPZ) that covers the vast majority of municipalities in Massachusetts (341 of 351, compared to 187 from the PRHRD and 79 covered by the WRLURI). Existing measures of LUR aid in the interpretation of the estimated latent categories, making my measure a complement to the previously established LUR measures.

⁷A broader overview of this literature is given in Section 3.

⁸An illustration of the numerous sorts of restrictions posed by LUR can be found in the PIRI (2005) database. They gather data on the regulatory environment around Boston through bylaws and surveys, and generate 119 variables meant to describe it. As many of these regulations affect development in similar ways, for example by reducing density, focusing on just a small subset of all regulations can lead to incorrect conclusions.

⁹It is possible to ask local officials about *when* regulations were implemented, as the PIRI (2005) do. But it is more challenging to gather the data repeatedly over several years.

As measures of land use restrictiveness are vital to many studies that require estimates of city-level housing supply elasticities, this technique provides a viable method to increase both the geographic coverage and time frequency of a LUR index. Researchers estimating Rosen-Roback style models (Diamond, 2016; Hsieh and Moretti, 2019) require measures of the responsiveness of local housing markets. Other papers use LUR measures to study their direct effects on welfare (Turner et al., 2014). A large body of research is interested in related questions that require measures of LUR or housing-supply elasticities derived from LUR measures (Dettling and Kearney, 2014; Hilber and Turner, 2014; Albouy, 2016; Aladangady, 2017; Stroebel and Vavra, 2019). Virtually all these papers use either the WRLURI as a measure of regulatory intensity, or housing supply elasticities from Saiz (2010), who in turn uses WRLURI to separate the contribution of geographic and regulatory restrictions on housing supply.

The second strand is concerned with the consequences of stringent land use or zoning regulations.¹⁰ This paper contributes to this literature in two dimensions. First, it identifies how restrictive LUR manifest themselves in housing development. Second, it speaks to the substitutability of nearby houses across municipal borders. Previous work has established a credible link between the implementation of stringent LUR and the reduction in aggregate housing supply and the increase in low-density developments (Mayer and Somerville, 2000; Saks, 2008; Glaeser and Ward, 2009; Turner et al., 2014; Diamond, 2016; Jackson, 2016; Hsieh and Moretti, 2019), and increasing house prices (Ihlanfeldt, 2007; Turner et al., 2014; Hilber and Vermeulen, 2016; Severen and Plantinga, 2018). Other research has found that LUR result in additional negative externalities: they exacerbate geographic sorting and inequality (Saks, 2008; Diamond, 2016; Ganong and Shoag, 2017; Hsieh and Moretti, 2019), increase price volatility (Glaeser et al., 2008; Jackson, 2018), as well as encourage land use conversion (Irwin and Bockstael, 2004; Sims and Schuetz, 2009; Shertzer et al., 2018).¹¹ This paper makes an important contribution by considering how stringent LUR are reflected in the pattern of housing development. The results on the substitutability of neighbouring houses subjugated to differing levels of land use restrictiveness are consistent with the model of Helsley and Strange (1995), where house price differences are insignificant between closely substitutable towns.

The remainder of the paper is structured as follows. Background on land use regulations in Massachusetts is given in Section 2. Section 3 outlines the construction of the regulation index. Section 4 describes the data, Section 5 the empirical strategy, and Section 6 the main results. Section 7 discusses balancing, specification, and robustness checks. The last section offers con-

¹⁰Though often used interchangeably, land use regulations subsume zoning regulations. LUR also include environmental regulations, for example.

¹¹In an address to the Urban Institute in 2015, Jason Furman, chair of then President Barack Obama's Council of Economic Advisers, claimed that "excessive or unnecessary land use or zoning regulations have consequences that go beyond the housing market to impede mobility and thus contribute to rising inequality and declining productivity growth." (Furman, 2015)

clusions. Specifics on the natural language processing techniques, model used to estimate local amenities, and additional details can be found in the Appendix.

2 Institutional Setting

Though municipalities in Massachusetts have utilized LUR since the 1930s, this was more the exception than the norm, as they needed to get approval from the state if they wanted to deviate from the state-level development guidelines. This changed with the introduction of the Massachusetts Zoning Act in 1975. The goal of this act was to “facilitate, encourage and foster the adoption and modernization of zoning ordinances and by-laws by municipal governments.” Effectively, it made it easier for municipalities to implement their own zoning bylaws without significant state interference. There is corroborating evidence that LUR were not restricting house construction nationwide until the 1970s (Gyourko et al., 2013).

After the Act was passed, municipalities began implementing their own LUR almost immediately and in quick succession. Survey data collected by the PIRI (2005), and plotted in Figure C.1, show the cumulative share of towns that have implemented a specific category of LUR by each year. Before 1975, only regulations regarding subdivisions (dividing a parcel of land into smaller parcels) were common. Afterwards, various types of LUR were introduced in different municipalities in quick succession.

Comparing zoning districts across municipalities is difficult, as there is no standardized classification of the district types. Furthermore, as municipalities set their own zoning regulations, land that is zoned for single-family residential in one town may be subjugated to a completely different regulatory environment than similarly zoned land in another town.

Figure C.1 also provides some examples of different categories of LUR.¹² For example, subdivision regulations are concerned with the division of a parcel of land into smaller units (eg a developer buying a parcel of land and building more than one house on the parcel with the intention of selling them as individual units) while wetland regulations deal with issues surrounding development around stagnate bodies of water (eg how close buildings can be to wetlands). Zoning bylaws generally cover some of the categories of LUR, but others (such as septic regulations) are often their own section in the municipal bylaw code.

3 Measuring Land Use Regulations

LUR are notoriously difficult to measure. This is primarily because they operate in a high-dimensional space; LUR are written in a variety of different ways, which are often not consistent across mu-

¹²There are various other categories that LUR can belong to, and Figure C.1 is in no way exclusive.

nicipalities, making direct comparison of bylaws across towns a challenging task.

Numerous strategies have been employed to measure the regulatory environment for land use or zoning. One strategy has been to focus on the implementation of or amendment to a single law (Zhou et al., 2008; Kahn et al., 2010; Severen and Plantinga, 2018). Other research has addressed this issue by calculating the share of LUR policies employed from a set of possible categories, and using this as a proxy for LUR intensity (Mayer and Somerville, 2000; Quigley and Raphael, 2005; Geshkov and DeSalvo, 2012).

Another approach is based on the assumption that the construction market is relatively competitive. These papers then use hedonic regressions or calculate price-to-cost ratios to infer the extent that LUR are impacting the housing market (Glaeser and Gyourko, 2003; Glaeser et al., 2005). Utilizing the same data source as I, other researchers gather information on LUR directly from zoning bylaws, normally by hand (Evenson and Wheaton, 2003; PIRI, 2005; Brooks and Lutz, 2019). Yet another more recent approach from Brueckner et al. (2017) combines data on prices for parcels of land in China with floor-area-ratio limits in an Alonso-Muth-Mills model to infer the stringency of regulation.

The most informative method—in terms of quantity and quality of information—has been to conduct surveys with local officials who are responsible for setting land use policy, often in combination with gathering data from primary sources (Levine, 1999; PIRI, 2005; Gyourko et al., 2008). These efforts to establish a unified database on LUR have been substantial undertakings; research teams had to either scour town websites and digitize the information or interview town officials in order to gather the necessary data to compile these resources, sometimes both.

Though these previous undertakings in measuring LUR provide an important and necessary starting point, the spatial regression discontinuity design described in Section 5 benefits from a measure of regulatory stringency that has close to universal coverage over a large region. I address this need by creating a regulatory index that is based on the text from zoning bylaw documents from nearly all the towns in Massachusetts.

Specifically, I compile a corpus of municipal bylaws by gathering the documents directly from the various municipalities’ websites. I then consider several different natural language processing (NLP) techniques and compare their results with the survey responses from the Pioneer/Rappaport Housing Regulation Database (PRHRD) and Gyourko et al. (2008).¹³ NLP are a natural fit in this particular setting, as they are primarily used on unstructured text data. Importantly, the set of NLP techniques considered can all be classified as “unsupervised” methods, as they use no information from the survey data used to assess fit. This is to avoid over-fitting any derived measure to the limited number of towns available for benchmarking.

¹³The measure of regulatory stringency (WRLURI) constructed by Gyourko et al. (2008) covers 79 towns in Massachusetts. Since the PIRI (2005) does not have a summary measure of the regulatory environment, I use their detailed data on the regulatory environment for 187 towns around Boston to create a measure of regulatory intensity.

As there is no *ex ante* best technique to measure the level of regulation, I consider several different NLP methods. The first method considered is the simplest: the number of meaningful¹⁴ words the zoning section of a bylaw contains. I refer to this as the “document length” measure. In addition to being the simplest, it also allows for testing the somewhat intuitive hypothesis: that the longer the zoning bylaw of a town is, the more stringent its LUR.

The second method employs a group of NLP techniques called either “dictionary methods” or “sentiment analysis”. These methods utilize pre-defined dictionaries of terms that belong to a specific category (eg the category “land” would contain words such as “valley”). The more words in a document that belong to a specific category, the more that category represents the document. The implicit hypothesis being tested here is that words from one (or more) of these dictionaries are found more often in more (or less) regulated towns’ bylaws.

The final method involves estimating a Latent Dirichlet Allocation (LDA) mixture-model, an unsupervised machine learning technique (Blei et al., 2003) that assumes that the distribution of observed words derives from latent, unobserved categories (normally called “topics”). Given a pre-defined number of latent categories, it estimates the probability a specific word was drawn from a specific category and assigns probabilities to each document over the distribution of the latent categories. Under the assumption that one (or more) of these latent categories determines regulatory stringency, I test whether the derived probabilities correlate with the survey based measures.

Of the NLP measures considered, I find the topic probabilities from the LDA model best capture the regulatory environment in Massachusetts, as judged by its high correlation to the two LUR existing indices. I call this the Natural Language Processing Zoning Stringency Index, or NALPZ.

This section is broken down as follows. First, I describe the data I need to employ the NLP techniques I use, as well as the data I later use to benchmark the measures I derive. Second, I characterize in greater detail the NLP methods I introduced above. Finally, I investigate how well the measures derived from NLP techniques are able to explain the variation in regulatory intensity found in the survey data.

3.1 Measurement Data

Zoning Bylaws

I compile a corpus of municipal bylaws by gathering zoning bylaw documents directly from the various municipalities’ websites. These documents are either in PDF or word format. Generally, the text is semi-machine readable, but in a few cases, the PDFs are scans of a paper version of the zoning bylaws. To extract the text from these documents I apply Optical Character Recognition

¹⁴ie not “filler” words such as conjunctions or pronouns.

software to the PDFs to extract the text. It is important to note that though the ordering of the text is not necessarily maintained, the methods I use do not utilize the ordering of the words, only their (relative) frequency. Of the 351 towns in Massachusetts, I am able to get bylaw documents for 341 of them, which are either up-to-date or only up to a couple of years old. The ten towns where zoning bylaw data is missing are smaller on average and have more rudimentary websites and digital services, and represent only 0.49% of the population of Massachusetts (2010 US Census).

Once the text is extracted, it needs to be preprocessed before it can be used. This involves, among other things, tokenizing the text (separating the text on whitespace and other characters into tokens), filtering (language-specific common words and stopwords are removed, such as “the”, “or”, “that”), and lemmatization (all words are reduced to their base form, “measurement” and “measured” become “measure”).¹⁵

Additionally, I filter out some custom, context-specific stopwords. This includes website specific stems from downloaded documents (e.g. “http”, “com”) as well as legal-text specific words (e.g. “doc”, “sec”), which do not provide much valuable information. Then, I filter out tokens that are either mentioned in almost every document (e.g. “month”, “equipment”, “waste”), or mentioned in just a couple (usually the names of towns). Finally, I weight the tokens using the term frequency-inverse document frequency method, which gives more weight to tokens that appear in fewer documents.¹⁶

Then I count the occurrences of each token (raw and weighted) in each document. This results in a document-term matrix, where each row is a document (here: town zoning bylaw) and each column a token. The cells refer to the number of times a word appears in a specific document (raw or weighted).

The top raw and weighted words are shown in Figure [A.1](#).

Dictionaries

With the processed documents, I apply various text analysis methods, investigating whether they map into the WRLURI or the PRHRD data. One common method is the “dictionary-based method” or “sentiment analysis”. Essentially, using predefined dictionaries consisting of words belonging to a specific topic, one counts the occurrences of the dictionary words appearing in each document (raw or weighted).

I use dictionaries from the Harvard IV-4 Categories to perform sentiment analysis on the zoning bylaw documents. I use nine dictionaries labelled “active”, “aquatic”, “building”, “land”,

¹⁵The individual words are referred to as “tokens” in keeping with the literature. This highlights the fact that they may have been altered (*ie* through lemmatization) and filtered, and thus do not correspond directly to the words in the original text.

¹⁶Details on the weighting scheme are described in Appendix [A.1](#)

“legal”, “nature”, “object”, “place”, and “region”. Examples of words belonging to each of these categories are given in Appendix [A.2](#).

Pioneer/Rappaport Housing Regulation Database

Though previous data on the regulatory environment for land use across Massachusetts, indeed across the US, has been scant, there have been a couple of notable undertakings. One such project was carried out through a joint project by the Pioneer Institute and the Rappaport Institute in Massachusetts.

They collected data on LUR for most of the towns in a 50 mile radius around Boston (187 of 351 towns in Massachusetts). To build their database, they gathered data in 2004 from municipal websites, through [Ordinance.com](#), and through phone calls if necessary. The data they gathered belonged to one of four categories: i) Zoning, ii) Subdivision, iii) Wetlands, or iv) Sewage Disposal. The end result is the Pioneer/Rappaport Housing Regulation Database (PRHRD)

Besides being a thorough description of the regulatory around Boston, it also allows me to test the external validity of the NLP-derived measures and to explore how the results I find later in the empirical section vary when using a different measure of regulatory stringency. As the PIRI (2005) do not create an index themselves, I use their data to construct a regulation index through principal component analysis.

Wharton Residential Land Use Regulation Index

Though the database compiled by the PIRI (2005) is very thorough for the area around Boston, it is not the most commonly used measure of land use regulatory restrictiveness, mainly due to its limited geographic coverage and the fact that it does not come with a uni-dimensional measure of LUR. Gyourko et al. (2008) address these two issues by focusing on breadth rather than depth.¹⁷

They sent surveys to 6,896 municipalities (with 2,649 responses) across the US to gather information on the state and local regulatory environment. Applying factor analysis to these responses, they create an index, the Wharton Residential Land Use Regulation Index (WRLURI), that is meant to capture the stringency of land use control.

The WRLURI is also a measure of regulation intensity that is often used in the urban economics literature¹⁸ (See Saks, 2008; Glaeser and Kahn, 2010; Saiz, 2010; Turner et al., 2014; Diamond, 2016; Ganong and Shoag, 2017; Hsieh and Moretti, 2019, for example), making it a particularly useful

¹⁷These issues do *not* imply any major shortcomings of the work of the PIRI (2005) or Gyourko et al. (2008), but rather refer to a fundamental trade-off between scope and detail with limited resources, where the former focus on detail and the latter on scope. See Gyourko and Molloy (2015, pp. 1298) for a discussion on this point.

¹⁸It is also used in other fields, notably in labour and public economics, when modelling housing supply response is vital.

index to compare with. Gyourko et al. (2008) have provided their data along with this index for other researchers to use.

Though more commonly used than the PRHRD, only 79 municipalities in Massachusetts have a WRLURI value. Therefore, it is important to note that while a comparison with this index can be worthwhile, it is based on a (probably) non-random subsample of all towns in Massachusetts. Should the effect of more stringent LUR be heterogeneous, results using the WRLURI and any NPL-derived measure could differ without indicating a problem with the identification strategy.

A map of the data coverage of both the PRHRD and WRLURI coverage in Massachusetts can be seen in Figure 1, where the yellow (light) municipalities indicate data available from the PRHRD and purple (dark) WRLURI measures from Gyourko et al. (2008). Brown (grey) shaded towns are covered by both datasets, of which there are 47.

3.2 Natural Language Processing Methods

The following subsection will cover the various NLP techniques employed, and the measures derived from them that will be compared to the indices from the survey data mentioned in Section 3.1. In what follows, d will refer to a document (*ie* a town’s zoning bylaw text), v a unique token (processed and meaningful words), x_{vd} the count of tokens v in document d , and tf-idf_{vd} the term frequency-inverse document frequency weighted version.

Document Length

The first NLP technique consists of simply counting the number of tokens in each document. This is done for both the raw and tf-idf weighted tokens. It is conceivable that a longer bylaw document (*ie* more tokens) could be an indication of stricter LUR. Formally, these measures are calculated as follows:

$$\begin{aligned} L_d &= \sum_{v=1}^V x_{vd} \\ \tilde{L}_d &= \sum_{v=1}^V \text{tf-idf}_{vd} \end{aligned} \tag{1}$$

where V refers to the set of all unique tokens in the corpus.

As can be seen in the left panel of Figure A.2, there is a large variance in the length of these bylaw documents. Unsurprisingly, the largest municipality in the dataset, Boston, has the longest zoning bylaw. The distribution of the weighted counts also varies, though the normalization results in the distribution becoming closer to normal.

These two measures, raw token counts and tf-idf weighted token counts, are the first two NLP-derived candidate measures.

Dictionary Methods

The second technique involves tallying the number of tokens (raw or tf-idf weighted) that belong to a specific category (dictionary) for each document. For example, a dictionary labelled “Legal” contains tokens such as “advocate”, “prison”, and “ordinance”. This is similar to calculating document length, but only considers tokens in the respective dictionaries. Formally:

$$\begin{aligned} C_d^c &= \sum_{v \in \mathcal{D}_c} x_{vd} \\ \tilde{C}_d^c &= \sum_{v \in \mathcal{D}_c} \text{tf-idf}_{vd} \end{aligned} \tag{2}$$

where c indexes the chosen dictionary, and \mathcal{D}_c represents the tokens in the dictionary.

The distributions of the measures for the various dictionaries are shown in Figure A.3. In the left panel the raw counts are divided by the document length, so the values can be interpreted as the share of tokens in the document that belong to the respective category. The more tokens a document has in a particular category, the more that category represents the document. Though the scale changes, the results vary little between the raw and unweighted categories. Words from topics such as “object” or “land” appear little in the bylaws, whereas words belonging to “active” and “place” appear regularly. These topics with high occurrences are also the topics that display the most variation.

These category-specific token counts are the second group of NLP-derived measures I consider.

Latent Dirichlet Allocation

The next NLP method I consider is a multinomial mixed-membership model with latent topics, called the Latent Dirichlet Allocation (LDA) model, first described by Blei et al. (2003). This estimator assumes that each document is a mixture over K topics (document-topic distribution), and the topics in turn have a probability distribution over each word v (topic-word distribution). The dispersion of the probabilities from these two distributions are governed by priors. Concretely, the document-topic probabilities are assumed to be distributed as:

$$\theta_d \sim \text{Dirichlet}(\alpha) \tag{3}$$

and the topic-word probabilities as:

$$\beta_k \sim \text{Dirichlet}(\delta) \quad (4)$$

We observe the words in a document d with N_d words overall, resulting in $\mathbf{w}_d = (w_1, \dots, w_{N_d})$. Furthermore, we assume that each word was generated from one of the K latent topics. Conditional on the document-term and topic-word probabilities, we can now describe the generating process of each word (i) for every document (d):

$$z_{id} \sim \text{Multinomial}(\theta_d) \quad (5)$$

$$w_{id} \sim \text{Multinomial}(\beta_{z_{id}}) \quad (6)$$

In the simplest case of $K = 2$, a document d would belong to topics 1 and 2 with probabilities θ_{d1} and θ_{d2} , respectively. If one topic, k , is more indicative of more stringent LUR, then θ_{dk} would be a measure of this. A visual depiction of the assumed data generating process is shown in Figure A.4. It is important to stress that this is a unsupervised machine learning method, which means that there is no “outcome” variable. The LDA method finds patterns in the text in order to assign topic probabilities to each document.

Conditional on K , I derive a measure of regulation intensity, S_K , by choosing the vector of document-topic probabilities that correlates the strongest with the WRLURI. Specifically:

$$S_K = \arg \max_{\mathbf{x} \in \{\theta_1, \dots, \theta_K\}} \text{corr}(\tilde{\mathbf{x}}, \text{WRLURI}) \quad (7)$$

where \tilde{x} represents the normalization of the variable x to have the same moments as the standard normal distribution.¹⁹ I do this because the magnitude of the unstandardized measure is not very informative; rather, it is the relative rank of the S_K ’s that have meaning. Furthermore, though the ranking of the documents by topic probabilities is invariant to choices on the priors, the distance between these probabilities is not. Note that the vector θ is now index by k rather than d .

The LDA model has several parameters that must be set by the researcher. The parameters on the two prior (Dirichlet) distributions control the dispersion of the document-topic and topic-words probabilities. As I use the θ ’s to construct my regulation index, the prior on the document-topic distribution is more relevant (α). However, though the prior affects the dispersion of the topic probabilities, the relative topic probability rankings are stable. And since I standardize the measure in the end, the priors do not affect the creation of my index. Thus, for the parameter on the topic-word prior distribution I use the standard in the literature, $\delta = 0.1$

¹⁹This is done by ranking the vector \mathbf{x} and then using the quantile function of the standard normal to get a normalized score.

(Griffiths and Steyvers, 2004) and for the prior on the document-topic distribution I choose a prior that results in a wide coverage of probabilities.²⁰ I find that setting the Dirichlet parameter $\alpha = 1/K$ works well in practice.

The most important parameter to choose in my setup is the number of topics, K . Since the LDA is an unsupervised method, there is no outcome with which to measure goodness-of-fit. Thus I investigate how well the number of topics fits my data through 5-fold cross-validation. I first look at a measure called “perplexity”, which can be thought of as a likelihood of the estimated model *given* validation data not used in the estimation, where a lower number indicates better fit. Then I look at how well the document-topic probabilities correlate with the WRLURI; in other words, I calculate $\rho = \max_{\mathbf{x} \in \{\theta_1, \dots, \theta_K\}} \text{corr}(\tilde{\mathbf{x}}, \text{WRLURI})$ for each fold of each K considered. I plot the results from the cross-validation in Figure A.5. The first panel shows the perplexity measure, with each point indicating a fold and the solid line passing through the average of the five folds. In the second panel, I plot the ρ from above. Though the model is only estimated with the test data for each fold, I calculate the correlation for the entire sample.

Going only off of the perplexity measure, using a model with $K = 25$ would seem to be most appropriate. However, though introducing more topics may help with fit, the results become more difficult to interpret. Thus I focus on the results from the second panel to choose K . From it, an LDA model with three topics seems the most preferred. Not only is the average correlation with WRLURI the highest, but the variation among the different folds is lower.

Using the LDA model with three topics (i.e. $K = 3$), I obtain the distributions of the three topics (β_{1v} , β_{2v} , and β_{3v} , where β_{kv} is the probability that word v is drawn from topic k) over all tokens. The tokens with the highest probability per topic are shown in Figure A.6. It is important to note that though the model returns a distribution of words for three separate topics, the labelling and interpretation of the topics is left ambiguous. Though I refrain from attaching any labels to the topics, there are still evident patterns to the word groupings. For example, Topic 2 has words that deal with renewable energy sources (“photovoltaic” and “wind”) and deal with hedonic uses (“marijuana” and “adult”). Topic 1 concerns itself more with vocabulary describing cities (“urban” and “sidewalk”) as well as terms associated with development (“ratio” and “affordable”). Topic 3 is more mixed and is concerned with amenities (“entertainment”) and types of development (“mixed”, “cluster”). There is obviously lots of overlap, making it difficult *ex ante* to assign meaning to the latent topics.

Turning to the ability of the document-topic probabilities to capture variation in LUR, I find that Topic 2 is most correlated with a high level of restrictiveness (higher WRLURI) whereas Topics 1 and 3 are negatively correlated with regulatory stringency. This suggests using θ_{d2} as a

²⁰ *ie* not clustered around 0%, 100%, or 100/ K %. The last would imply that a document is nearly equally well described by any of the latent topics.

NLP-derived measure of LUR for comparison to both the other derived measures and the PRHRD regulatory index.

To get at the words that best differentiate between the most “regulated” topic (2) and the two less regulated “topics” 1 and 3, I plot the largest absolute logarithm differences between the probability that a word belongs to Topic 2 against Topic 1 or 3 in Figure A.7.²¹ Words at the top in blue are more likely to appear in Topic 2 (stronger regulations), whereas those at the bottom in red have a higher probability of being in Topic 1 or 3 (weaker regulations). For example, the words “city” and “mixeduse” are more descriptive of Topic 1 or 3, whereas “Annual Town Meeting” and “photovoltaic” are more indicative of Topic 2.

3.3 Natural Language Processing Zoning Stringency Index

With several NLP-derived measures in hand, I now turn to comparing them with the two regulatory indices described earlier. The results are shown in Table 2. It shows the correlation between the NLP-derived measures discussed in this section, namely the document length measures, the dictionary-based measures, as well as topic probabilities from the LDA model, and the regulatory indices WRLURI and the PCA index from the PRHRD data.

The top two rows correlate the length of the documents (based on the sum of raw or weighted tokens per document) with the regulatory measures. Somewhat surprisingly, the length of the bylaw documents does not seem to be strongly related to the regulation indices, either raw or unweighted. In fact, in the raw case, shorter zoning bylaw documents come from towns with more stringent LUR.

Turning to the middle section of rows, I then show the correlation of the tf-idf weighted counts of tokens belonging to one of the dictionaries with the regulatory measures.²² I find that none of the nine dictionaries used to score the documents results in a measure that is strongly correlated with the regulation indices. Towns that use words from the dictionaries “region”, “place”, and “building” appear to be somewhat *less* regulated, with correlation coefficients between -0.28 and -0.41 .

However, turning to the bottom row, it is apparent that the topic probabilities from the LDA model are strongly related to the two indices considered here. S_2 is strongly correlated with the aggregate LUR indices. I name this measure the Natural Language Processing Zoning Stringency Index (NALPZ). It has a correlation coefficient of 0.66 with the PCA index from the PRHRD data and 0.58 with the WRLURI. The mixture-model estimates latent topics that are strong predictors of LUR by finding patterns in the text contained in the bylaw documents.

²¹ $\log_2\{\beta_{2v}/(\beta_{1v} + \beta_{3v})\}$

²²The results using raw counts, unreported, are quite similar.

Given the ability of NALPZ to capture the regulatory environment seemingly well, I further examine the relationship between it and the other indices of regulation. These are shown in the scatter plots in Figure 2. Panel A and B display the relationship between NALPZ and the two survey-based indices. As can be seen, NALPZ does a good job of explaining these other two measures, especially given that it uses *none* of that information in its creation. Moreover, NALPZ is significantly easier to derive, requiring only the raw text given in the bylaws. The last plot, C, gives a sense of how the WRLURI measure varies with the PRHRD index.

It is also worth briefly discussing the disadvantages of using NALPZ as a measure of the regulatory environment. The primary concern is that it is a “black box”; what exactly results in a higher or lower index value is not entirely clear. A town may be regulated more stringently either through implementing more LUR or making the current LUR more intensive. However, I am able to investigate *which* words are more often found in documents with a higher regulation index value, given the estimated topic-specific word probabilities. Furthermore, the ease of implementation makes this measure a nice compliment to the survey-based measures to help expand geographic coverage and potentially add a time dimension to current regulation measures.

A map of the spatial distribution of NALPZ is shown in Figure 3. As has been found already by Gyourko et al. (2008), larger (in terms of land) municipalities with lower population densities tend to have a stricter regulatory environment than average, as can be seen by the cluster of dark blue towns in western Massachusetts. Larger metropolitan areas, such as those around Boston, Worcester, and Springfield, are less regulated in general.

This paper focuses on the *impact* of stringent LUR on housing development. However, Appendix B further explores the related question of the causes of stringent regulation. It discusses the spatial pattern of LUR in Massachusetts and what pre-existing town characteristics best predict a town’s regulation restrictiveness.

4 Data

The NALPZ from Section 3 is the main explanatory variable. It captures the relative stringency of a town’s LUR. To measure the impact of restrictive regulation, I compile data on current housing characteristics, land use over several decades in Massachusetts. Several additional data sources are included for the tests performed in Section 7. These include census data on demographics, bilateral travel times between areas, school district quality measures, and property tax rates.

Massachusetts Standardized Assessors’ Parcels

To effectively employ the spatial RDD design, the data I use need to have two vital pieces of information. First, they need to contain information on the characteristics of single-family houses

across Massachusetts, and second, the houses need to be precisely geocoded to compare units within close geographic proximity.

The Massachusetts Standardized Assessors' Parcels database fulfills these two criteria. It contains every parcel of taxable land in the state of Massachusetts, except for Boston (which maintains its own records and database). The data are collected from each municipalities' tax assessors, under guidelines set forth by the Massachusetts Bureau of Geographic Information (MassGIS), that are then compiled by MassGIS.

The database consists of several tables, two of which are used for this analysis. The first is the Assessor Data Extract that contains all the information that tax assessors gather on each taxable parcel in their respective municipality. Some data included in this table are the assessed tax value of the structures (buildings) and the land, along with some information about the characteristics of the structures (eg year built, number of rooms, lot size). It also has information on the size of the parcels of land and of the building itself. Importantly, the data also contain the "use code" for the parcel. This code classifies each property based on its primary use such as residential, commercial, agriculture, etc.

The second table is the Tax Parcel Attribute file. This file contains the polygon for each parcel of land in the state of Massachusetts, each with a unique location identifier, as well as a shapefile indicating its precise location within the state. The Assessor Data Extract can be matched to the Tax Parcel Attribute file through the location identifier. Though most of the matches between the two files are one-to-one, in some cases several assessment units are matched to one tax parcel. For example, units in a condominium may have different owners, and thus be taxed independently, but share the same geographic location.

For my analysis, I only consider single-family houses. This is primarily done to keep in line with previous literature, and due to the fact that they represent the vast majority of residential structures in Massachusetts (here: 73% of taxable residential parcels). I am able to identify these parcels by the use code. I also filter out parcels that are located in municipalities without a NALPZ value, as well as parcels that are bordering a town without a NALPZ value (as they would have no comparison units).

From the tax parcel data, I am also able to estimate whether it is owner occupied. The data contain the address of both the property itself, as well as that of the owner. From this, I compute the Levenshtein distance between them. Parcels where this value is less than eight I code as owner-occupied.²³

Of the 2,319,906 parcels overall, 951,971 remain after filtering. Summary statistics on the tax parcels are given in Table 3.²⁴ Last sale price is not reported for every tax parcel, thus I do not

²³Small differences arise between these two addresses even when they are identical. This is generally due to abbreviating words such as "street" (str.) or "avenue" (ave.) in one of the addresses but not the other.

²⁴The number of observations varies by attribute because not all municipalities gather information on every

exclude units from the baseline sample when this is the only detail missing. Furthermore, in analyses using last sale price, I exclude parcels where this is less than \$5,000 to include only arm's length transactions.

Building Heights

An important aspect of development is the height of buildings. This attribute is not contained in the tax assessment data, unfortunately. I overcome this lack by calculating building height from highly accurate lidar data in combination with a digital elevation model for Massachusetts and a shapefile containing all buildings in the state.

Lidar refers to light detection and ranging, a remote sensing method using reflected light to measure distances. The National Oceanic and Atmospheric Administration provides large quantities of lidar data covering all of Massachusetts. The main attributes of the data are its x, y, and z coordinates. This describes where the data point was measured in terms of longitude (x) and latitude (y), as well as its height above sea level (z). Together with the building shapefiles I am able to assign the highest point above sea level for each building. To derive elevation, I use the digital elevation model to get the elevation of the base of the houses. The height is then simply the difference between these two values.

I restrict the building height sample to buildings located on tax parcels that are coded as single-family residential. This allows the results for the tax parcel level and building level to be comparable.

Land Use

Data on land use is provided by MassGIS, which classifies land into 37 different categories based on aerial images. There is data available for the years 1971, 1985, and 1999. The change in the coverage of broad categories of land use from 1971 to 1999 can be seen in Figure C.2. It shows the change in land use within 1km of municipal borders. Motivated by the fact that a significant amount of forest and agriculture land has been converted to residential use from 1971 to 1999, I use the share of developed land as an outcome of interest. This enables me to speak to the impact of LUR on conversion of land use and the spatial class of restrictions. I also calculate the share of developed land used for residential purposes, to test whether LUR influence the spatial pattern of housing development.

US Census Data

Census Blocks: Demographic and housing attribute data is gathered from the US Census Bureau at the block level—the smallest level of aggregation available—for the years 1970, 1990, 2000, and

attribute.

2010. Data on demographics includes share of individuals under 18, over 64, non-white, married, female. Housing data consists of the average number of rooms, average rent, and average house price. This data is used to test whether the demographic composition of adjacent neighbourhoods in different towns are similar or not (pre-Massachusetts Zoning Act and widespread LUR). Thus, I only keep blocks that i) are in a town with a NALPZ measurement, ii) are neighbouring a town that also has a NALPZ value, and iii) are physically adjacent to a neighbouring town. This means that they share a border with a neighbouring town. I do this as my main empirical strategy uses houses in close to municipal borders. The assumptions regarding similar pre-treatment demographics thus must also hold at this level.

Census Block Groups: Additional census data is gathered at the block group level, which consists of a collection of blocks, which is used to create unobserved amenity values in Section 7.2. From the 2010 census I get the block group population.

American Community Survey: More block group level data in 2010 comes the ACS. This includes median household income and median housing prices.

Open Source Routing Machine/Open Street Maps

Bilateral travel times between all block group pairs is also required for the amenity estimations. These are calculated from Open Source Routing Machine (OSRM). Concretely, I generate the population-weighted centroids for each block group and calculate the bilateral commuting times between every pair using OSRM, which utilizes OpenStreetMap route networks.

School Districts

I estimate several specifications which control for one or more measures of school district quality. The first comes from [Niche.com](#). They are an organization that provides information and rankings on neighbourhoods, schools, universities, and workplaces across the US. Importantly for this paper, they rank each school district in Massachusetts based on a variety of criteria, such as grades, parent/student surveys, and facility quality.

Additional measures of school (district) quality come from [ClearGov.com](#). They aim to provide clarity to citizens on how tax revenues are spent within different communities. Notably, they also publish statistics at the school district level. From here I get measures of spending per pupil as well as graduations rates.

Since parents choosing a house based on school district quality will have access to the same, publicly available information, I am able to condition on the same information set as would be available these parents.

Finally, there are smaller towns which belong to a unified school district containing several

other towns. This allows me to compare towns *within* some school districts with different degrees of regulatory intensities, by including school district fixed effects.²⁵ This specification is quite demanding of the data as this reduces the sample size significantly.

Municipal Property Tax Rates

In another specification, I also control for municipal property taxes. Should housing characteristics change in response to higher or lower property taxes, which is in turn systematically related to LUR, controlling for the rate would rectify this.

This data is gathered from the Division of Local Services of the Massachusetts Department of Revenue. I use the tax rate from 2018. This is the same year most of the tax assessment data was gathered.

5 Empirical Strategy

My baseline empirical method is a spatial regression discontinuity design (RDD). I use this to estimate the effect of more stringent land use regulation on the development and characteristics of single-family houses.

To implement this method, I need two key pieces of information: i) the nearest neighbouring town for each house (and by extension the nearest border), and ii) the distance to that town. As the tax parcels are all precisely geotagged, I am able to calculate this information from the tax database when combined with the towns shapefile from MassGIS.

With this information I create a “segment” identifier. This will be used in the spatial RDD framework to ensure that I am comparing tax parcels with other parcels in the nearest neighbouring town. In other words, I am comparing parcels that share a municipal border and are arguably in the same neighbourhood (especially when considering smaller bandwidths).

The baseline specification for the spatial RDD is as follows:

$$y_{ism} = \beta \text{NALPZ}_m + f(\text{geography})_{ism} + \pi_s + u_{ism} \quad (8)$$

where y_{ism} is the outcome of interest for parcel i , bordering segment s , in municipality m . NALPZ_m is the standardized Natural Language Processing Zoning Stringency Index, π_s are segment fixed effects, and $f(\text{geography})_{ism}$ is a function of the geographic location, the running variable. As

²⁵It should be noted that because only smaller towns share school districts with neighbouring towns, the sample that allows for the identification of the effect of LUR in this specification is not representative of all towns in Massachusetts.

the treatment—the level of regulation—varies at the town level, I estimate cluster-robust standard errors at the town level.

The main coefficient of interest is β . This captures the effect of a one standard deviation increase in the NALPZ on the considered outcome (y_{ism}). The main outcomes at the tax parcel level are the year the house was build, year last sold, and the size of the building and of the lot; from the building data the outcome of interest is the height of the building.

Following Dell (2010) and Dell et al. (2018), I begin by modelling $f(\text{geography})_{ist}$ with the latitude and longitude of each parcel, additionally controlling for the distance to Boston and to the nearest coast. As high-order polynomials have been shown to be unstable in RDD settings (Gelman and Imbens, 2018), I restrict my sample to a particular bandwidth around each town border and run a kernel regression (with respect to the running variable) with triangular weights. This involves modelling the running variable flexibly, allowing the effect of the distance to the nearest border to vary differently to each side of a border segment. Putting this all together results in:

$$\begin{aligned} f(\text{geography})_{ism} = & \alpha_{sm} \text{distance to segment}_{ism} + \delta_1 \text{lat}_{ism} + \delta_2 \text{long}_{ism} \\ & + \delta_3 \text{distance to boston}_{ism} + \delta_4 \text{distance to coast}_{ism} \end{aligned} \quad (9)$$

where α_{sm} are the town-border segment specific coefficients on the running variable. In practice, I find that after conditioning on the segment fixed effects (π_s) and the town-border segment specific running variable, the additional geographic controls do not influence the results much. Thus, they are omitted from the reported results.

In the main results I show coefficient estimates with different bandwidths around town borders. This highlights the sensitivity of the estimates to including or excluding more observations further way from these borders. An example of this empirical strategy with a bandwidth of 500m is illustrated in Figure 4.

The empirical strategy here is similar to the one used by Turner et al. (2014) and Severen and Plantinga (2018), though I allow the distance gradient of the outcome variable to vary by town-border segment.

Aggregate Spatial Regressions

When investigating the impact of stringent LUR on the density of housing supply or on land use, where the unit of observation is a region or an aggregation of the tax parcel data, I estimate a simpler spatial regression. Here, the specification controls for segment fixed-effects to ensure comparisons are done between spatially near areas. Just as in the baseline specification, various

bandwidths are considered. This results in the following estimating equation:

$$y_{sm} = \beta \text{NALPZ}_m + \pi_s + u_{sm} \quad (10)$$

where the indices are the same as the main specification. y_{sm} is a summary measure for some outcome in segment s , municipality m . This is an aggregate version of the baseline empirical model. The unit of observation is now a town-segment area rather than tax parcels *within* these geographical areas.

In the case where the outcome is housing supply, y_{sm} corresponds to either i) the number of single-family homes within the considered bandwidth, or ii) the density of houses per square kilometer within the bandwidth. When the outcome is land use, y_{sm} corresponds to either i) the share of land developed for residential use from 1971–1999, or ii) the fractional of developed land that is residential.

House Price Regressions

To test whether LUR affect house prices as well, I estimate a model similar to Equation 8, but I additionally control for lot sizes, and in some specifications for building sizes, for comparability. This results in the following specification:

$$y_{ism} = \beta \text{NALPZ}_m + \alpha_{sm} \text{distance to segment}_{ism} + \gamma_1 \text{building size}_{ism} + \gamma_2 \text{lot size}_{ism} + \pi_s + u_{ism} \quad (11)$$

where γ_1 and γ_2 are the price effects for building size and lot size respectively. Here, the outcomes of interest are the price the house was last sold for and the total tax assessed value of the property (building and land together). I also present results where the outcomes are the assessed value of the building per square meter of building and the assessed value of the land per square meter of land.²⁶

Identification Strategy

The empirical specifications above both exploit the same variation. Namely, the discrete change in the regulatory environment at the border between two municipalities. The primary assumption of this strategy is that no other observed or unobserved feature that varies at municipal borders affects housing development pattern *and* is systematically related to the stringency of LUR.

²⁶These two specifications do not add controls for the size of the building or lot as they are already normalized.

The main specification addresses several potential issues by only considering units that are geographically close. This controls for potential geographic confounders like local housing demand, access to well paying jobs, and preferences for certain regions.

In Section 7 I address other potential issues as follows. First, I test whether there existed differences in demographic characteristics in neighbouring blocks in different towns *before* the Massachusetts Zoning Act and widespread LUR. I supplement this by estimating amenity values by calibrating a standard urban spatial general equilibrium model in the spirit of Ahlfeldt et al. (2015), and testing whether these vary discretely at town borders systematically with the level of LUR. Second, I investigate whether my results are sensitive to alternative specifications of the running variable and weighting method. Third, I directly control for municipal characteristics, like school quality and property tax levels, that vary at town borders along with the regulatory environment.

Graphical Evidence

On account of there being multiple cut-offs (town border-segments) with a non-binary treatment variable (NALPZ index), it is not straightforward to graphically inspect the discontinuity. To nonetheless present suggestive evidence I plot the primary outcome variables residualized by border-segment fixed effects, binned into 100m intervals, by whether the housing unit belongs to the less or more regulated side of its matched border-segment.²⁷ The results are shown in Figure C.5.

Two things stand out. First, there is a significant discontinuity for some outcomes, such as for lot sizes and tax assessed land values. Second, the gradients for the outcome variables are not uniform. For example, there is a discontinuity for year built, but the distance to the border gradient appears quite similar. Looking at the house prices outcomes on the other hand, the gradients are quite different and in fact go in different directions. In less regulated towns the more expensive housing is located at the border region of a town whereas in more regulated towns the housing becomes more expensive as you approach the town centre. This fact highlights the importance of allowing the outcome gradients to vary by town border-segment.

6 Results

The results are broken down as follows. First, I present evidence for how the aggregate housing supply near the municipal borders is affected in towns with more restrictive development policies. This speaks to whether restrictive land use policies result in lower density housing overall.

²⁷ As towns mostly have multiple neighbours, it may both be considered “less regulated” and “more regulated” but each housing unit can only be in one category on account of it being matched exclusively to one border-segment.

Second, I highlight results from my baseline spatial RDD specification, looking at various housing development outcomes while considering different bandwidths around the borders. This consists of two primary groups of outcomes: i) housing market outcomes (the age of the structural as well as the year the house was last sold), and ii) housing characteristics (building size, building height, and lot size). Third, I test for the role of LUR in altering land use by showing results for the effect of stringent regulations on the conversion of undeveloped land to residential use as well as on the fraction of developed land used for residential purposes. Fourth, I investigate whether stringent LUR are capitalized into house prices at the local, neighbourhood level.

Overall Housing Supply

To get an overview of how stringent LUR affect the housing supply, I plot the results of estimating Equation 10, which captures the effect of these regulations on the number of single-family homes. I consider two outcome measures for overall housing supply: i) the density of houses per square kilometer, and ii) the raw count of houses within a specified distance to the border. The results of these regressions are shown in Figure 5. I estimate the model for bandwidths between 100m and 2km, at 100m intervals. The coefficients are interpreted as the respective change in the outcome variable for a standard deviation increase in the NALPZ.

As can be seen from the figure, the results are remarkably stable across the entire range of bandwidth options. They indicate (at a bandwidth of 100m) that the housing density is 14 houses per square kilometer lower and that the average number of houses within 1km of the border is 67 units less for every s.d. increase of regulatory stringency. These estimates correspond to an effect size of about 20% and 17% of the mean outcome level, respectively.

This provides evidence that these restrictive land use policies have led to a reduction in the housing supply, at least at the peripheries of these towns, by reducing the density of development.

Housing Market

I now turn to the results of the baseline spatial RDD. I have plotted the results for each outcome individually, considering various bandwidths for sample selection. The plots are given in Figure 6. I show the results for bandwidths varying from 100m to 2km. Each point represents the β estimate, along with its 95% confidence interval from clustering the standard errors at the town level, from Equation 8 for the tax parcel and building level regressions, and from Equation 10 for the land use regressions. NALPZ is standardized, so the coefficients correspond the change in the respective outcome variable to a standard deviation increase in regulatory stringency.

The results in Figure 6a speak to the effect of LUR on the housing market. The rate of development of new housing within a neighbourhood should be similar if the regulatory environment

is the same. The results for this outcome, shown in the left plot, indicate that houses in more regulated towns are older on average. At the smallest bandwidth, it suggests that a one standard deviation increase in NALPZ corresponds to homes being two years older on average. As the average house age in sample is 57 years, this represents 3.4% of the mean.

The right side of the panel shows results with the year the house was last sold as the outcome. This speaks to a related question of whether LUR reduce the efficiency of the housing development and the real estate market (Mayer and Somerville, 2000; Glaeser et al., 2008). For example, if LUR make the housing supply less elastic to changes in housing demand, the selling and buying of the current housing stock may take place less often as incumbent residents hold on to their homes. Overall, there is not a major difference in the last time a home was sold with respect to LUR. At the smallest bandwidth, the results suggest a one standard deviation increase in NALPZ corresponds to houses last being sold about 0.8 years further in the past on average. This is an effect size of roughly 4.3% of the mean.

Taken together the results provide some evidence for LUR resulting in a less responsive housing market, though the effects are economically small.

House Attributes

The results for the housing attribute outcomes are shown in Figure 6b. As seen in the upper-left plot, there is virtually no difference between the size of livable space in houses in more or less strongly regulated municipalities in the same neighbourhood. Regardless of the bandwidth considered, no coefficient is significantly different than zero, with the point estimate very close to zero, relatively precisely estimated. Turning to the upper-right plot, the lot sizes, on the hand, are strongly influenced by restrictive zoning policies. Considering the most narrow bandwidth, a standard deviation increase in land use restrictiveness results in lot sizes being roughly 27% larger.

This result provides evidence for a specific channel that leads to lower density development overall: stringent LUR increases the amount of land used per house, resulting in lower density and fewer houses. An important follow-up question, that will be addressed when looking at how land use has changed with respect to regulatory stringency, is whether municipalities compensate for the increase in land *per* lot by allocating more land overall to residential use.

The bottom two plots presents results for building height as the outcome. As there is no guidance from the literature on the functional relationship between land use regulation and building height, I present results with the height in level and logarithmic terms. However, regardless of the specification there is no strong relationship between the stringency of LUR and building height. Though most specifications are significantly different than zero, the results are precise enough to rule out a standard deviation increase in NALPZ increasing building height by 5% in the log-level specification. With an average house height of 11.5m, this is an economically insignificant result.

Considering all the results together, development in towns with stricter LUR tends to be on larger parcels of land, without correspondingly larger or taller buildings.

Land Use

Finally, the results testing for land use regulations that spatially restrict development are highlighted in Figure 6c. They plot the relationship between land use (development) patterns and NALPZ.

The left plot presents the effect of stringent LUR on the conversion of undeveloped land in 1971 to residential use in 1999. This tests whether restrictive development policies impact the rate of land being developed. Regardless of the bandwidth the estimated coefficients are essentially zero with the 95% confidence intervals also bound very close around null. This suggests that the rate of conversion of undeveloped land to residential use was not differential between municipalities with varying degrees of LUR intensity.

The results in the right plot tell a similar story. The estimated effect of restrictive LUR on the share of residential land of all developed land is marginally significant for all but the two smallest bandwidths, with effect sizes around 2.4% of the mean.

Two conclusions can be drawn from these results. First, stringent LUR do not appear to alter the amount of land allocated to residential purposes. Second, stringent municipalities do not develop more land for residential use to compensate for the lower housing density due to larger lot sizes.

Spatial RDD by Density Group

Motivated by the the large effects of LUR on parcel lot sizes, I further investigate how the distribution of lot sizes changes with respect to the different quartiles of the NALPZ index, shown in Figure C.3. As is evident, there is a significant degree of heterogeneity in the distribution of lot sizes over the NALPZ quartiles. To aid in interpreting the distribution of lot sizes, the cut-offs between high, mid, and low density lot sizes are shown with the dashed vertical lines, as defined by MassGIS.

To better understand how LUR impact the lot sizes *within* the density groupings, I re-estimate the baseline spatial RDD empirical model for each density subsample with lot size as the outcome. The results are shown in Figure C.4.

A word of caution interpreting these results. Most likely, LUR result in more development happening in one density category rather than another (eg more regulated municipalities may encourage the building of more low-density homes). Therefore, these should be considered descriptive rather than causal. What is immediately clear, however, is that conditional on density

grouping, the effect of more stringent LUR is strongest amongst low-density housing. This suggests that the results are being driven by larger lots becoming even larger.

House Prices

Figure 7 shows the results of the house price regressions. These results speak to the question of whether more stringent regulation is reflected in the house prices, which depends on the degree of substitutability between neighbouring houses.

The top panel (7a) displays the results when the outcome is the total tax assessed value of the property or the last sale price of the house. The estimated effect of LUR on house prices is quite similar regardless of the metric used to measure house prices: a standard deviation increase in the NALPZ increases house prices roughly 5–6%.

However, these regressions do not identify the impact of LUR separately for land and building prices. Thus, I leverage the fact that the tax assessment values are broken down into building and land components. I then use these outcomes, normalized by the size of the building and lot respectively, in place of the total house prices. The results for these regressions are given in Figure 7b. Unlike with the total house prices, here the evidence for stringent LUR being capitalized into house values is weaker. The results for both the building and lot values are not significantly different from zero across all the bandwidths.

Together, these results provide evidence for houses in neighbouring towns being substitutes for one another. Thus restrictive building policies in one town do not necessarily lead to higher prices overall if there is sufficient housing in neighbouring towns. In other words, population mobility may arbitrage away price differentials *across* municipalities within a region, but increase the price level of the region as a whole.

7 Balancing, Specification, and Robustness Checks

Having established a strong relationship between land use regulations and lower density development, primarily via larger lot sizes, I now turn to investigating whether these results could be driven by other factors. First, I present a pair of balancing checks, where I test whether the Natural Language Processing Zoning Stringency Index can predict neighbourhood demographic characteristics in 1970, before LUR were commonplace, and whether the NALPZ is related to unobserved neighbourhood amenities. I recover these unobserved neighbourhood amenities by calibrating an urban spatial general equilibrium model. Second, I conduct a series of specification and robustness checks to test whether different sources of variation may partially explain the results.

7.1 Neighbourhood Demographic Characteristics Pre-Massachusetts Zoning Act

One is typically concerned that the forcing variable in an RDD design is manipulated by the units of observation. As I am considering tax parcels of land, and the forcing variable is the distance to the town border, this is not problematic in my setting. However, there is still the concern that towns which implemented stronger/weaker regulations varied significantly, even within a neighbourhood that belongs to two or more towns.

Since the tax assessors database is constantly updated, I am unable to obtain data before and after the introduction of municipal-level LUR. Therefore, to investigate whether there were any pre-existing differences before the Massachusetts Zoning Act, I compare the demographic composition of US Census Blocks directly to either side of the borders (*ie* blocks touching the border). I test whether *future* regulatory intensity predicts these demographic outcomes in the past. I use data from the 1970, 1990, 2000 and 2010 censuses and run a modified version of Equation 8:

$$y_{bsmt} = \beta_1 \text{NALPZ}_m + \beta_2 \text{NALPZ}_m \times \{year_t > 1970\} + \pi_{st} + \alpha_s \text{wdist}_{bsm} + u_{bsmt} \quad (12)$$

where the outcomes are now indexed by time, π_{st} are year-specific segment fixed effects, and I no longer need to control for geographic distance (as the distance to the border is zero by construction). However, I do control for how far the average cell of a block group is away from the comparison border. This is the wdist_{bsm} term. It refers to the block grid-cell average distance to the border. I allow its effect to vary at the border-segment level (α_s). This controls for segment-specific gradients in demographics characteristics. For example, larger block tend to cover space further away from a municipal border (thus having higher wdist_{bsm} values), making them less representative of the area immediately surrounding municipal borders. The unit of observation, b , is now the census block. The coefficient of interest is β_1 : if there are no pre-existing differences in demographics and housing, it should be zero.

β_2 may be non-zero if the implementation of LUR changed the demographic composition through modified residential development and residential sorting. This is an interesting outcome in and of itself. If, for example, existing residents lobbied for more restrictive land use policies to stem the flow of in-migration to the municipality, barring individuals from certain demographics disproportionately from moving to the town, that would be captured by β_2 .

The coefficients β_1 and β_2 are plotted in Figure 8. The purple circles represent β_1 and the green triangles β_2 . For comparability across the various outcomes, I report standardized coefficients. As can be seen in the figure, the estimated β_1 coefficients are not statistically different

from zero. None of the estimated effects have standardized coefficient estimates over 0.05 in absolute terms. In addition the 95% confidence intervals reject any absolute effect size above 0.15 standard deviations.

Looking at the same demographic composition after the introduction of the Massachusetts Zoning Act, I find that the share of non-white residents is significantly lower in highly regulated towns. Taken literally, it implies that a one standard deviation increase in *NALPZ* corresponds to roughly a 0.24 standard deviation decrease in the share of the population that is non-white. This provides evidence for the theory that restrictive zoning policies have resulted in residential sorting, either deliberately or as a side effect.

7.2 Amenities

Given the observable data from the census pre-Massachusetts Zoning Act, I am able to test for demographic differences across town borders that may confound my estimates. However, there may still be unobservable differences in local amenities that may induce sorting by preferences for housing with specific characteristics, or for low-density neighbourhoods. Furthermore, as I am looking at small geographic regions, simply controlling for *some* observable amenities (eg parks, waterfront locations, access to playgrounds, etc.) does not vary the empirical estimates much.

To address this concern, I estimate local amenities as implied by the canonical quantitative spatial equilibrium model of a city (Ahlfeldt et al., 2015) at the census block group level (a small collection of city blocks). To calibrate the model, I use the parameter values from several recent papers that have structurally estimated the model: Ahlfeldt et al. (2015), Tsivanidis (2018), and Heblich et al. (2020). I then test whether these amenity values vary in adjacent census block groups to either side of municipal borders.

Model Intuition

For the purposes of estimating implied amenities, it is sufficient to discuss and parameterize the worker demand system. Details on the model are given in Appendix Section E. Intuitively, workers trade-off wages, commuting costs, housing costs, and local amenities when choosing residence (*i*) and workplace (*j*) locations.²⁸ Workers are heterogeneous with regards to residence-workplace location pairs.

²⁸Here I use the term “block” to refer to a location, as it corresponds nicely with the us Census geographic terminology as well as my empirical geographic unit.

Solving the model results in the following expression for residential amenities:

$$\frac{B_i^*}{\widetilde{B_i^*}} = \left(\frac{H_i}{\widetilde{H_i}} \right)^{1/\epsilon} \left(\frac{q_i}{\widetilde{q_i}} \right)^{1-\beta} \left(\frac{\text{CMA}_i}{\widetilde{\text{CMA}_i}} \right)^{-1/\epsilon} \quad (13)$$

where B_i^* are residential block-specific amenities, H_i refers to block resident population, q_i is the price of housing, and $\text{CMA}_i = \sum_{j=1}^S (w_j / e^{\kappa \tau_{ij}})^\epsilon$ is a measure Commuting Market Access, which captures how closely in terms of commuting time (τ_{ij}) a residential block is located to well paying jobs (w_j).²⁹ $\widetilde{X} = \left(\prod_i X_i \right)^{1/N}$ denotes the geographic mean of the respective variables, which is included to remove terms that are invariant across residential locations. I use $B_i^* / \widetilde{B_i^*}$ as my measure of local amenities. As this term has no natural scale, I standardize it to ease interpretation of the results.

Intuitively, if a residential block has a high population in spite of high housing costs and poor access to well-paying employment opportunities, there must be high local amenities to compensate the residents.

In addition to the observable data $\{H_i, q_i, \tau_{ij}, w_j\}$, I require estimates of the models parameters $\{\epsilon, \beta, \kappa\}$ to back out implied amenities. I consider various parameter combinations from recent contributions to the literature to obtain several different measures of amenities. The exact parameter values are given in the Appendix, Table E.1.

Estimating Amenities

To estimate the amenity values I use data from the US Census Bureau at the block group level and the Open Source Routing Machine as described in Section 4.

I restrict my sample to census block groups that are adjacent to *one* municipal border (*ie* I drop block groups that border several towns or are in the interior). This is to ensure that the census block groups I am comparing across borders are as similar to one another as possible.

The estimation equation is similar to the one used to test for demographic balance in census blocks in 1970 (Eq. 12), but without the time dimension (I am only considering data from 2010):

$$y_{bsm} = \beta_1 \text{NALPZ}_m + \pi_s + \alpha_s \text{wdist}_{bsm} + u_{bsm} \quad (14)$$

where NALPZ_m and π_s are defined the same as previously. y_{bsm} refers to the standardized measure of residential amenities ($B_i^* / \widetilde{B_i^*}$) as implied by the model under four different parameter combinations (see Table E.1). The wdist_{bsm} term is the same as used in Equation 12, but is calculated at the block group level. This controls for segment-specific gradients in amenities.

²⁹More details in Section E

Amenities and Land Use Regulation

I first show results only including $NALPZ_m$, to get an idea of the raw relationship between residential amenities and land use regulation. Then, I subsequently include segment fixed effects, and finally the segment-specific grid-cell average distances.

These are shown in Figure 9. The raw coefficients (green circles) show that there is a very strong relationship between land use regulations and implied amenities on average. This suggests that households are willing to pay higher housing costs and live further away from well-paying employment opportunities to live in communities that are more stringently regulated.

However, this is comparing block groups in Massachusetts that are located far from one another. As highlighted in Figure 3, a large cluster of highly-regulated towns is located in the North-West section of Massachusetts, isolated from the major employment centres of Boston, Springfield, and Worcester, whereas the towns closer to these centres are more relaxed on average. There are significant differences between the households, and therefore the sort of housing demanded, that choose to live in these different regions.

Once I include segment fixed effects, and more so by additionally including segment-specific mean grid cell distance linear gradients, the estimated relationship between LUR and implied amenities drops significantly. In two of the four sets of parameter combinations (values for both high- and low-skilled workers from Tsivanidis (2018)), I cannot reject the null hypothesis that there is no relationship between LUR and amenities. The standardized effect sizes for the other two parameter combinations are never larger than 0.15. This provides evidence that the main empirical strategy allows me to compare units (houses, parcels, buildings) that are in similar neighbourhoods and have similar levels of local amenities.

7.3 Specification and Robustness Checks

To validate my main findings, I also consider alternative specifications and controls to explore the robustness of my results. The first change made is to rerun the baseline regression without the triangular weights (*ie* the weight given to units closer to and farther from municipal borders is the same). Next, I model the function of geography as a quadratic in each term (but still allowing the effect of the running variable to differ by border-segment, also as a quadratic).

Next, I control for the presence of school districts. As school district borders generally align with municipal borders, it is not directly possible to control for the districts for every observation (parcel). To overcome this, I control for measures of school quality. These school quality measures are described in Section 4. I also estimate a model controlling for school district fixed effects, which identifies the effect of LUR in smaller municipalities that share a school district. This specification is quite demanding as many municipalities have their own school district, so

the effective sample size reduces significantly.

To the extent that property taxes systematically relate to LUR as well as impact housing development, I control for the residential property tax rate in a further robustness specification. I then include all of school district quality and rank measures together with the property tax rate control.

There is also a concern that new development is mostly infill development: *ie* new houses are built on the sites of demolished previous houses and that this is more important in determining lot sizes than LUR.³⁰ Thought of another way, there may be path dependence when redeveloping land for new residential use. To address this issue, I re-estimate my primary specification on the subsample of houses that were built on previously undeveloped land, where a developer would only be restricted by LUR and geography. Specifically, I look at development taking place *after* 1971 on land that was previously vegetation (*eg* forest, bush) or agricultural.

The results of these robustness checks for the housing market, attributes, and price outcomes are shown in Figures 10a, 10b, and 11, respectively. The baseline specification, along with all the robustness checks, are with a bandwidth of 100m.

Looking at the housing market outcomes, the estimates for the age of the house are quite stable across the various specifications. The results change the most when conditioning on only new development post-1971. The small, but statistically significant, results confirm previous evidence that LUR have a modest but persistent effect on the rate of new development. The results for year last sold also do not vary much. The baseline estimates were already small, and become insignificant especially with the school district controls.

Turning to the housing attributes outcomes, lot and building size, again there is little change in the estimated coefficient on NALPZ. The effect of stringent LUR on lot sizes remains large and statistically significant across the estimated equations, except when including school district fixed effects, which results in imprecise estimates. Nonetheless, most of the 95% confidence interval lies above zero. The effects on building size remain close to zero. Overall, these results confirm the role LUR play in increasing the amount of land used per house.

Finally, the robustness results for the housing price outcomes are shown. Here, the estimated relationship between housing prices and LUR shrinks when school quality is controlled for. This suggests that the price differentials across municipalities with different regulatory environments is due to access to better schools, rather than LUR. This all supports the view that houses that are nearby but in different towns (with similar quality schools) are close substitutes.

³⁰Of course, this process would have to vary with the overall restrictiveness of LUR as well to be an issue.

8 Conclusion

Land use regulations are ubiquitous across the US, but their causes and impacts are not fully understood. I create a new regulation index, called the Natural Language Processing Zoning Stringency Index (NALPZ), by applying a machine learning algorithm, a Latent Dirichlet Allocation model, to over 40,000 pages of zoning bylaw documents. This method builds off previous work to greatly increase spatial coverage. This lets me address the question of *how* stringent land use regulations are manifested in housing development.

By exploiting the variation in the regulatory environment at municipal borders in Massachusetts, I confirm previous studies showing that stringent land use regulations reduce housing density. Moreover, I extend these findings by showing that lot sizes are most responsive to LUR, being considerably larger in more regulated jurisdictions. This provides evidence for parcel-specific land use regulations—regulations that encourage higher land usage per house—being most responsible for restricting housing supply. These include regulations such as minimum lot sizes, setback requirements, strict floor-area-ratios.

Furthermore, the results suggest that spatially close houses in differing towns are highly substitutable. This provides evidence for restrictions in one locality leading to price increases in neighbouring towns as well, if overall housing supply is not responsive enough, as previous work has shown.

My findings suggest that land use regulations encourage less dense development on larger parcels of land, without the compensation of allocating more land to residential use overall, effectively limiting the supply of housing available. For policy makers looking to increase available housing in high-growth regions, scrutinizing these local constraints to development is a promising avenue.

9 References

- Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). “The economics of density: Evidence from the Berlin Wall”. *Econometrica*, 83 (6), pp. 2127–2189.
- Aladangady, A. (2017). “Housing Wealth and Consumption: Evidence from Geographically-Linked Micro-data”. *American Economic Review*, 107 (11), pp. 3415–46.
- Albouy, D. (2016). “What are Cities Worth? Land Rents, Local Productivity, and the Total Value of Amenities”. *Review of Economics and Statistics*, 98 (3), pp. 477–487.
- Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). “CEO Behavior and Firm Performance”. *Journal of Political Economy*, 128 (4), pp. 1325–1369.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). “Latent Dirichlet Allocation”. *Journal of Machine Learning Research*, 3 (Jan), pp. 993–1022.
- Brooks, L. and Lutz, B. (2019). “Vestiges of Transit: Urban Persistence at a Microscale”. *Review of Economics and Statistics*, 101 (3), pp. 385–399.
- Brueckner, J. K., Fu, S., Gu, Y., and Zhang, J. (2017). “Measuring the Stringency of Land Use Regulation: The Case of China’s Building Height Limits”. *Review of Economics and Statistics*, 99 (4), pp. 663–677.
- Dell, M. (2010). “The Persistent Effects of Peru’s Mining Mita”. *Econometrica*, 78 (6), pp. 1863–1903.
- Dell, M., Lane, N., and Querubin, P. (2018). “The Historical State, Local Collective Action, And Economic Development In Vietnam”. *Econometrica*, 86 (6), pp. 2083–2121.
- Dettling, L. J. and Kearney, M. S. (2014). “House Prices and Birth Rates: The Impact of the Real Estate Market on the Decision To Have a Baby”. *Journal of Public Economics*, 110, pp. 82–100.
- Diamond, R. (2016). “The determinants and welfare implications of US workers’ diverging location choices by skill: 1980-2000”. *American Economic Review*, 106 (3), pp. 479–524.
- Evenson, B. and Wheaton, W. (2003). “Local Variation in Land Use Regulations”. *Brookings-Wharton Papers on Urban Affairs*, 4, pp. 221–260.
- Furman, J. (2015). “Barriers to Shared Growth: The Case of Land Use Regulation and Economic Rents”. Remarks. Urban Institute.
- Ganong, P. and Shoag, D. (2017). “Why has regional income convergence in the US declined?” *Journal of Urban Economics*, 102, pp. 76–90.
- Gelman, A. and Imbens, G. (2018). “Why High-order Polynomials Should Not Be Used In Regression Discontinuity Designs”. *Journal of Business & Economic Statistics*, pp. 1–10.
- Geshkov, M. V. and DeSalvo, J. S. (2012). “The Effect of Land-Use Controls on the Spatial Size of US Urbanized Areas”. *Journal of Regional Science*, 52 (4), pp. 648–675.

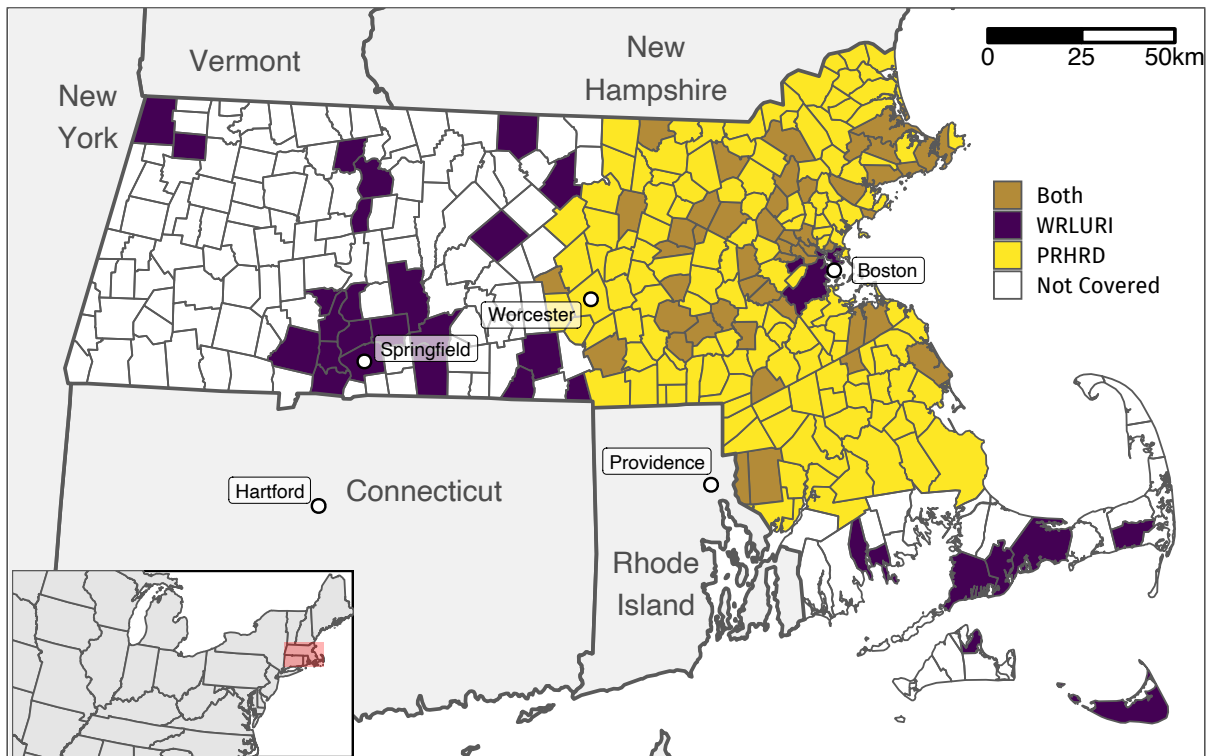
- Glaeser, E. L. and Gyourko, J. (2003). "The Impact of Building Restrictions on Housing Affordability". *Economic Policy Review*, 9 (2).
- Glaeser, E. L., Gyourko, J., and Saks, R. (2005). "Why is Manhattan so Expensive? Regulation and the Rise in Housing Prices". *The Journal of Law and Economics*, 48 (2), pp. 331–369.
- Glaeser, E. L. and Kahn, M. E. (2010). "The greenness of cities: carbon dioxide emissions and urban development". *Journal of Urban Economics*, 67 (3), pp. 404–418.
- Glaeser, E. L., Gyourko, J., and Saiz, A. (2008). "Housing supply and housing bubbles". *Journal of Urban Economics*, 64 (2), pp. 198–217.
- Glaeser, E. L. and Ward, B. A. (2009). "The Causes and Consequences of Land Use Regulation: Evidence from Greater Boston". *Journal of Urban Economics*, 65 (3), pp. 265–278.
- Griffiths, T. L. and Steyvers, M. (2004). "Finding Scientific Topics". *Proceedings of the National Academy of Sciences*, 101 (suppl 1), pp. 5228–5235.
- Gyourko, J., Mayer, C., and Sinai, T. (2013). "Superstar Cities". *American Economic Journal: Economic Policy*, 5 (4), pp. 167–99.
- Gyourko, J. and Molloy, R. (2015). "Regulation and Housing Supply". *Handbook of regional and urban economics*. Vol. 5. Elsevier, pp. 1289–1337.
- Gyourko, J., Saiz, A., and Summers, A. (2008). "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index". *Urban Studies*, 45 (3), pp. 693–729.
- Heblich, S., Redding, S. J., and Sturm, D. M. (2020). "The making of the modern metropolis: evidence from London". *The Quarterly Journal of Economics*, 135 (4), pp. 2059–2133.
- Helsley, R. W. and Strange, W. C. (1995). "Strategic Growth Controls". *Regional Science and Urban Economics*, 25 (4), pp. 435–460.
- Hilber, C. A. and Turner, T. M. (2014). "The Mortgage Interest Deduction and its Impact on Homeownership Decisions". *Review of Economics and Statistics*, 96 (4), pp. 618–637.
- Hilber, C. A. and Vermeulen, W. (2016). "The impact of supply constraints on house prices in England". *The Economic Journal*, 126 (591), pp. 358–405.
- Hsieh, C.-T. and Moretti, E. (2019). "Housing constraints and spatial misallocation". *American Economic Journal: Macroeconomics*, 11 (2), pp. 1–39.
- Ihlanfeldt, K. R. (2007). "The effect of land use regulation on housing and land prices". *Journal of Urban Economics*, 61 (3), pp. 420–435.
- Irwin, E. G. and Bockstael, N. E. (2004). "Land Use Externalities, Open Space Preservation, and Urban Sprawl". *Regional Science and Urban Economics*, 34 (6), pp. 705–725.

- Jackson, K. (2016). “Do land use regulations stifle residential development? Evidence from California cities”. *Journal of Urban Economics*, 91, pp. 45–56.
- Jackson, K. (2018). “Regulation, land constraints, and California’s boom and bust”. *Regional Science and Urban Economics*, 68, pp. 130–147.
- Kahn, M. E., Vaughn, R., and Zasloff, J. (2010). “The Housing Market Effects of Discrete Land Use Regulations: Evidence From the California Coastal Boundary Zone”. *Journal of Housing Economics*, 19 (4), pp. 269–279.
- Levine, N. (1999). “The Effects of Local Growth Controls on Regional Housing Production and Population Redistribution in California”. *Urban Studies*, 36 (12), pp. 2047–2068.
- MassGIS (Bureau of Geographic Information), Commonwealth of Massachusetts EOTSS (2020). URL: <https://docs.digital.mass.gov/dataset/massgis-data-layers>.
- Mayer, C. J. and Somerville, C. T. (2000). “Land Use Regulation and New Construction”. *Regional Science and Urban Economics*, 30 (6), pp. 639–662.
- OCM Partners (2020). *Lidar Data*. NOAA National Centers for Environmental Information. URL: <https://coast.noaa.gov/dataviewer/#/> (visited on 02/16/2020).
- Open Source Routing Machine (2020). OpenStreetMap. URL: <http://project-osrm.org> (visited on 09/18/2020).
- Pioneer Institute for Public Policy Research and Rappaport Institute for Greater Boston (2005). *Massachusetts Housing Regulation Database*. Prepared by Amy Dain and Jenny Schuetz. URL: <http://www.masshousingregulations.com>.
- Quigley, J. M. and Raphael, S. (2005). “Regulation and the High Cost of Housing in California”. *American Economic Review*, 95 (2), pp. 323–328.
- Saiz, A. (2007). “Immigration and Housing Rents in American Cities”. *Journal of Urban Economics*, 61 (2), pp. 345–371.
- Saiz, A. (2010). “The Geographic Determinants of Housing Supply”. *The Quarterly Journal of Economics*, 125 (3), pp. 1253–1296.
- Saks, R. E. (2008). “Job creation and housing construction: Constraints on metropolitan area employment growth”. *Journal of Urban Economics*, 64 (1), pp. 178–195.
- Severen, C. and Plantinga, A. J. (2018). “Land-use Regulations, Property Values, and Rents: Decomposing the Effects of the California Coastal Act”. *Journal of Urban Economics*, 107, pp. 65–78.
- Shertzer, A., Twinam, T., and Walsh, R. P. (2018). “Zoning and the economic geography of cities”. *Journal of Urban Economics*, 105, pp. 20–39.
- Sims, K. R. and Schuetz, J. (2009). “Local Regulation and Land-use Change: The Effects of Wetlands Bylaws in Massachusetts”. *Regional Science and Urban Economics*, 39 (4), pp. 409–421.

- Stroebel, J. and Vavra, J. (2019). “House Prices, Local Demand, And Retail Prices”. *Journal of Political Economy*, 127 (3), pp. 1391–1436.
- Tax Rates by Class* (2020). Division of Local Services, MA Department of Revenue. URL: https://dlsgateway.dor.state.ma.us/reports/rdPage.aspx?rdReport=PropertyTaxInformation.taxratesbyclass.taxratesbyclass_main (visited on 07/20/2020).
- Tsivanidis, N. (2018). “The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá’s transmilenio”. Working Paper.
- Turner, M. A., Haughwout, A., and Van Der Klaauw, W. (2014). “Land use regulation and welfare”. *Econometrica*, 82 (4), pp. 1341–1403.
- Zhou, J., McMillen, D. P., and McDonald, J. F. (2008). “Land values and the 1957 comprehensive amendment to the Chicago zoning ordinance”. *Urban Studies*, 45 (8), pp. 1647–1661.

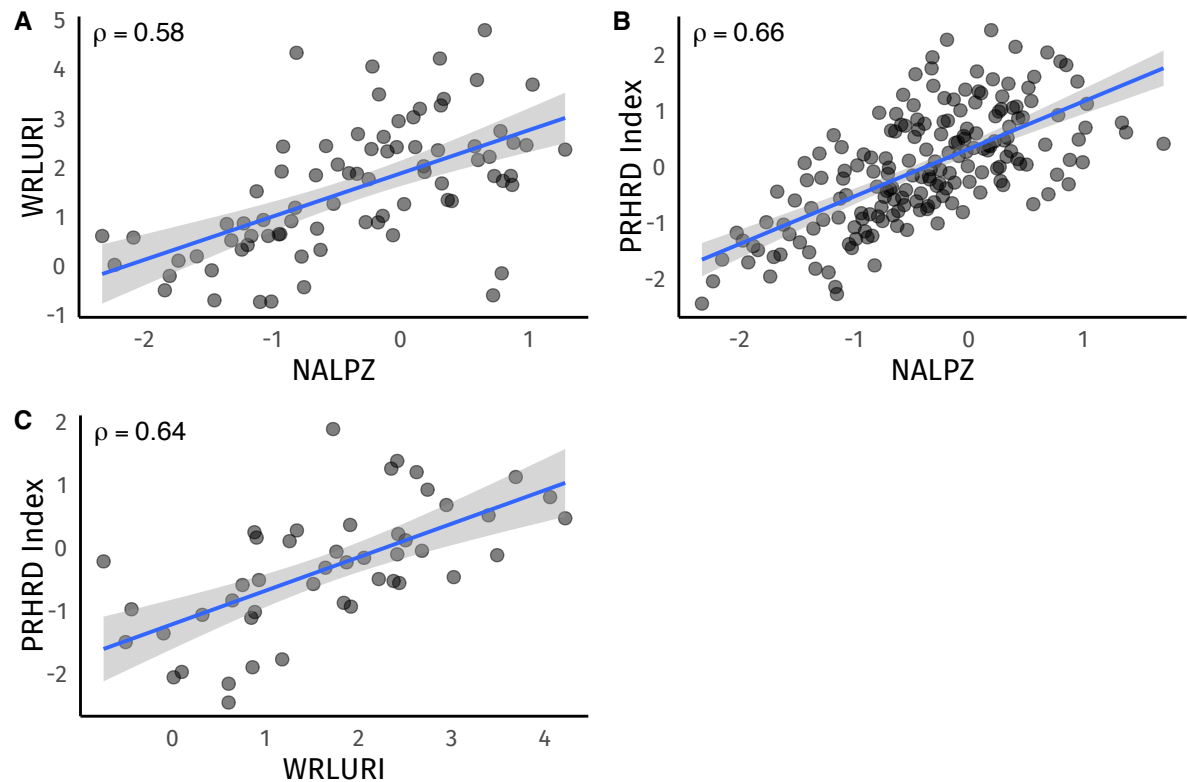
10 Figures

Figure 1: Current Data on Land Use Regulations in Massachusetts



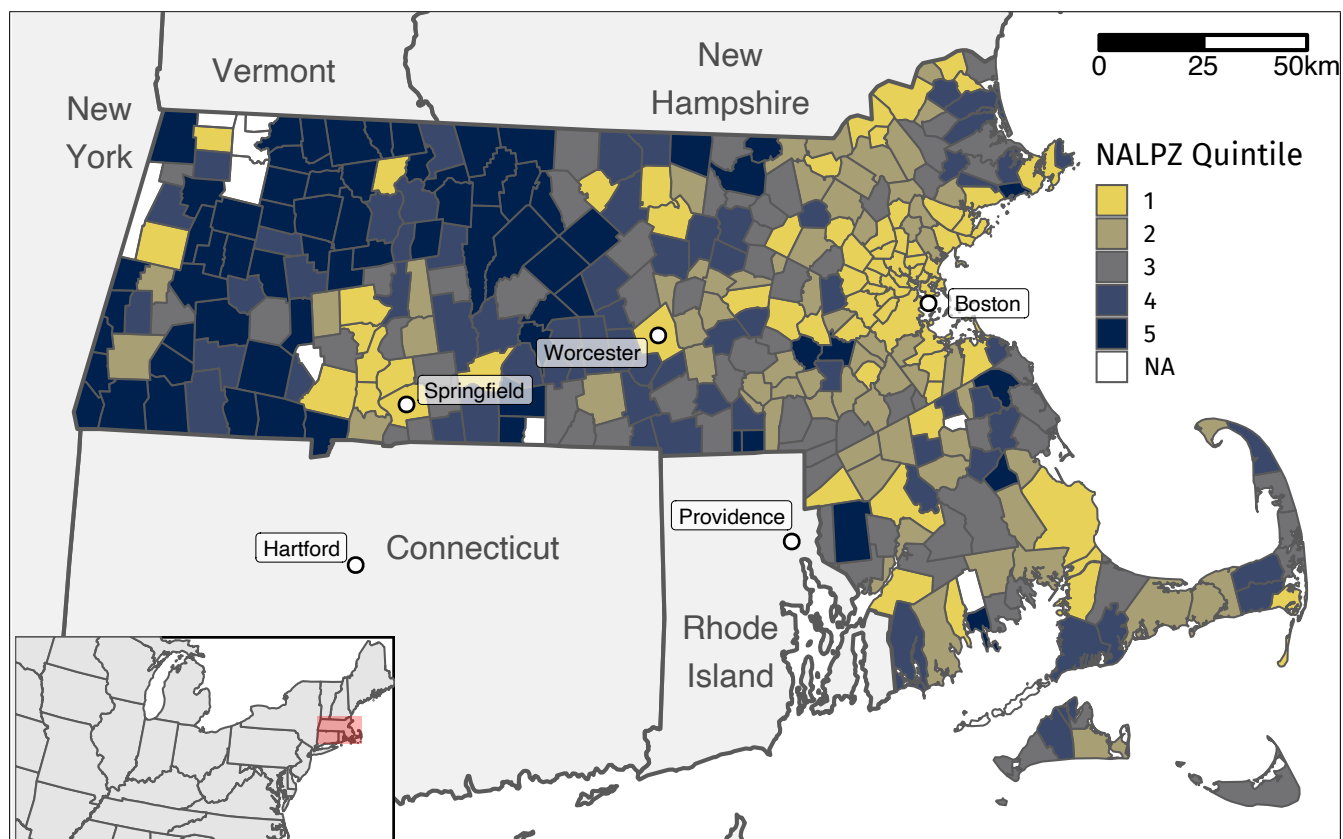
Notes: Data come from Gyourko et al. (2008) and the PIRI (2005). WRLURI is the Wharton Residential Land Use Regulation Index from 2005. PRHRD is the Housing Regulation Database of Massachusetts compiled by the Pioneer Institute and the Rappaport Institute in 2004. I create an index from their data using Principal Component Analysis.

Figure 2: Comparison of Natural Language Processing Zoning Stringency Index (NALPZ) with Existing Measures of Land Use Restrictiveness



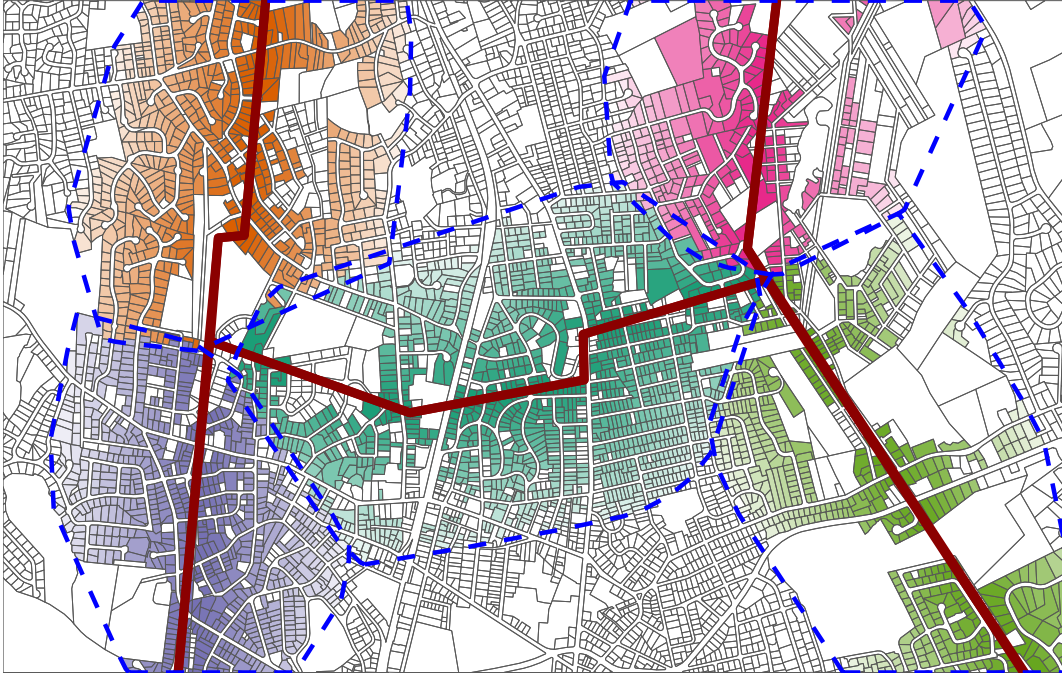
Notes: WRLURI (Gyourko et al., 2008) is the Wharton Residential Land Use Regulation Index from 2005. PRHRD (PIRI, 2005) is the Housing Regulation Database of Massachusetts compiled by the Pioneer Institute and the Rappaport Institute in 2004. I create an index from their data using Principal Component Analysis.

Figure 3: Natural Language Processing Zoning Stringency Index (NALPZ) Across Massachusetts



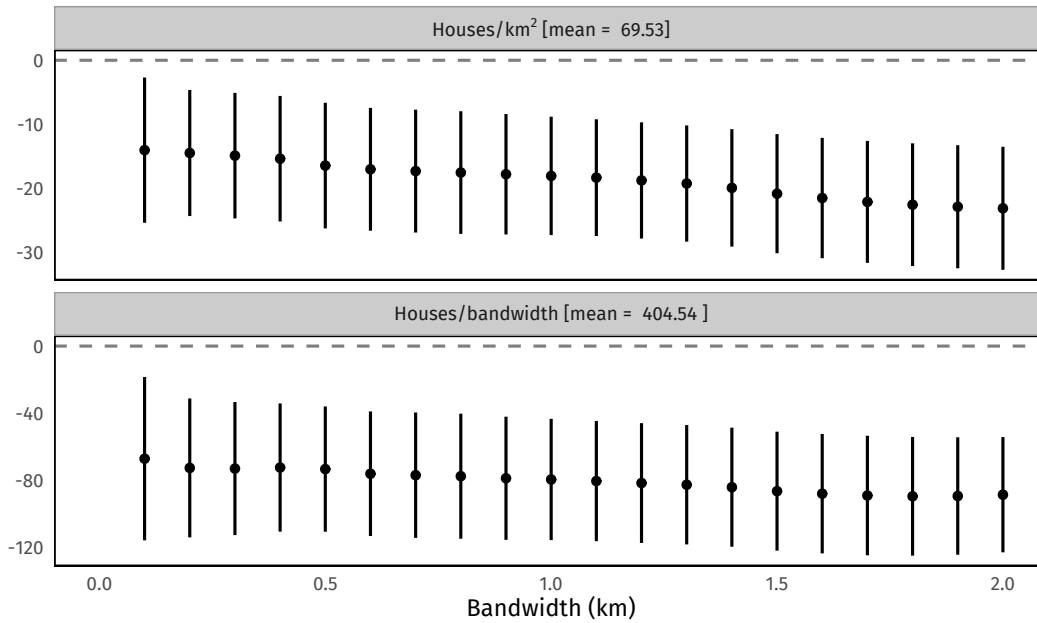
Notes: NALPZ is an index of land use regulation stringency derived from a Latent Dirichlet Allocation model applied to the zoning bylaw text of towns across Massachusetts.

Figure 4: Parcel Spatial RDD Example



Notes: Example of the spatial RDD strategy with a bandwidth of 500m. Colours correspond to the matched town border. Red lines indicate town borders. Colour intensity indicates weight value. Uncoloured parcels are either not in the sample (eg not residential parcels) or outside of the bandwidth.

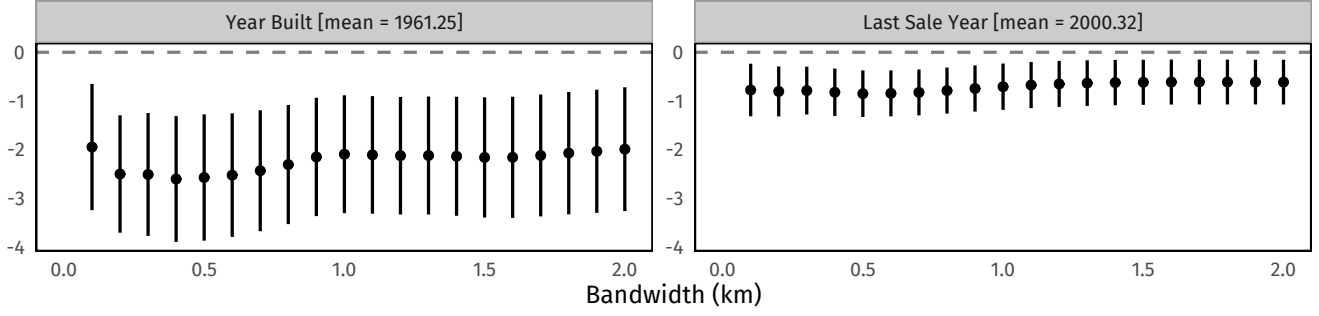
Figure 5: Spatial Regression of Housing Supply and Density on NALPZ



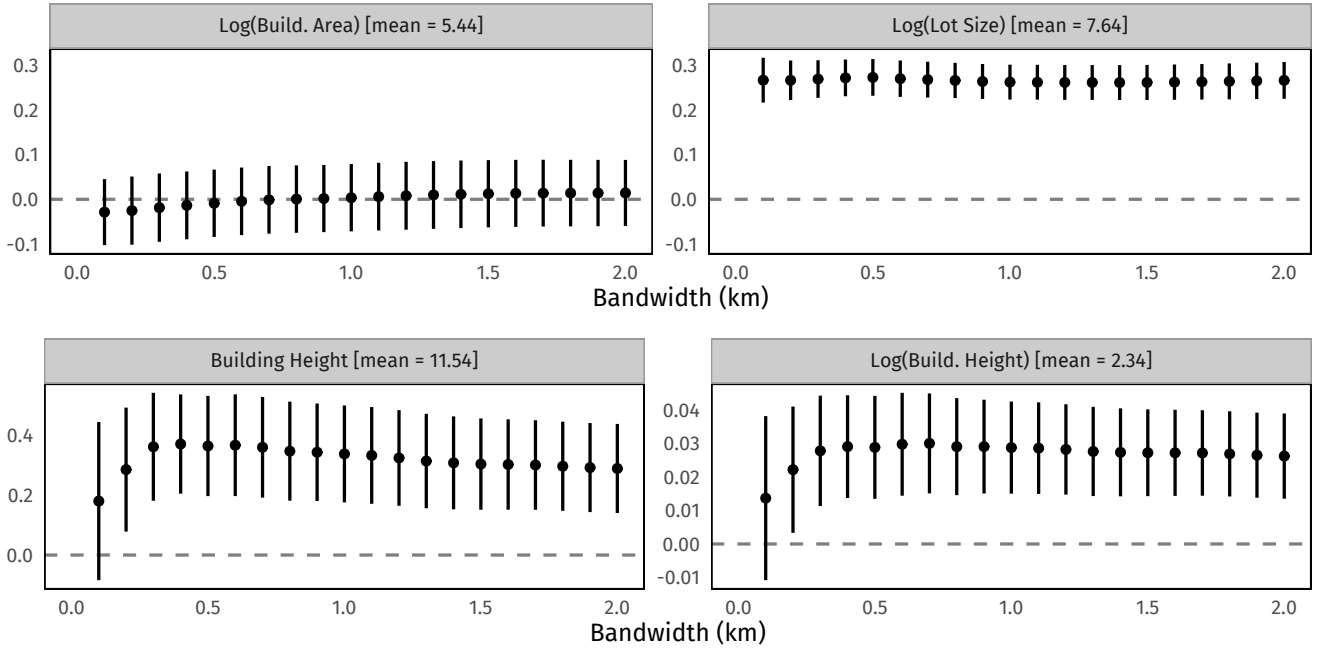
Notes: β coefficient from Equation 10 for various bandwidths plotted. Respective outcome is regressed on Natural Language Processing Zoning Stringency Index NALPZ and border segment fixed effects. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level. Mean outcome refers to the average outcome value with a bandwidth of 100m.

Figure 6: Main Results

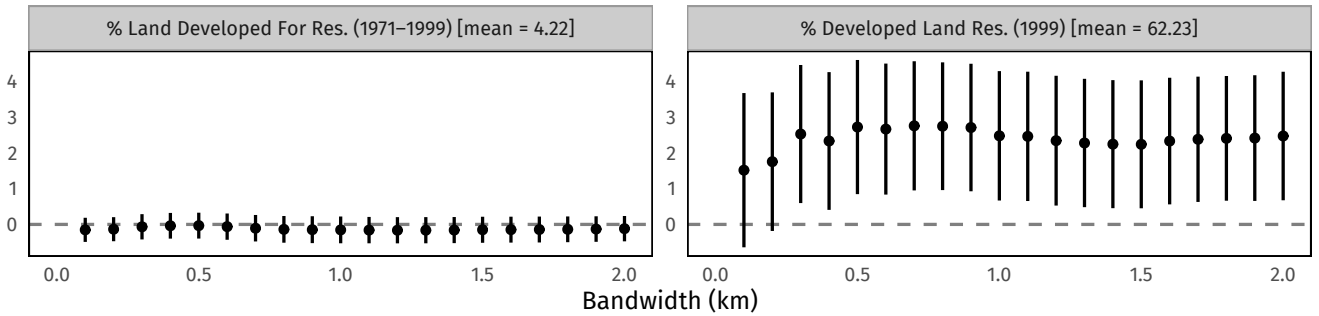
(a) Housing Market Outcomes: Spatial RDD of Respective Outcome on NALPZ



(b) Housing Attributes Outcomes: Spatial RDD of Respective Outcome on NALPZ



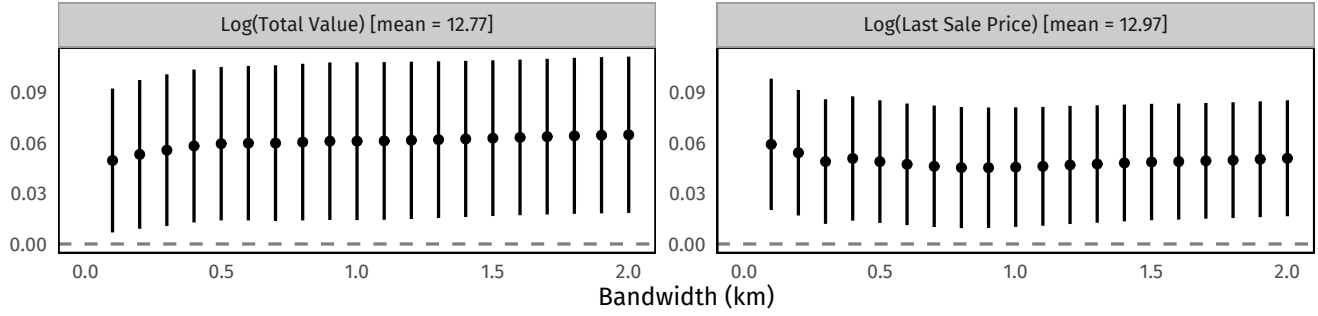
(c) Land Use Outcomes: Spatial Regression of Respective Outcome on NALPZ



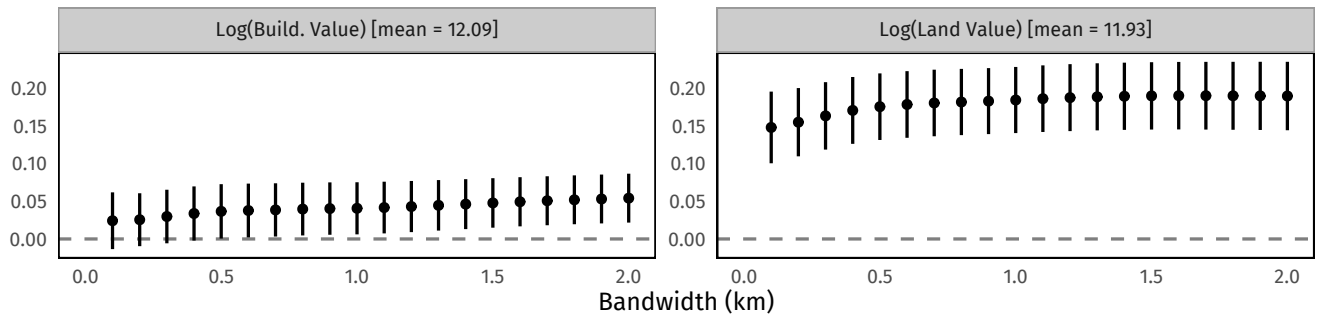
Notes: Panel (a) and (b): plots the estimated β coefficient from Equation 8. Respective outcome is regressed on the Natural Language Processing Zoning Stringency Index (NALPZ), border segment fixed effects, and border segment-specific distance controls. Panel (c): plots the estimated β coefficient from Equation 10. Respective outcome is regressed on the NALPZ and border segment fixed effects. Both: Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level. Mean outcome refers to the average outcome value with a bandwidth of 100m.

Figure 7: House Price Outcomes: Spatial RDD of Respective Outcome on NALPZ

(a) Logarithm of Total Assessed Value and of Last Sale Price (2018 prices)

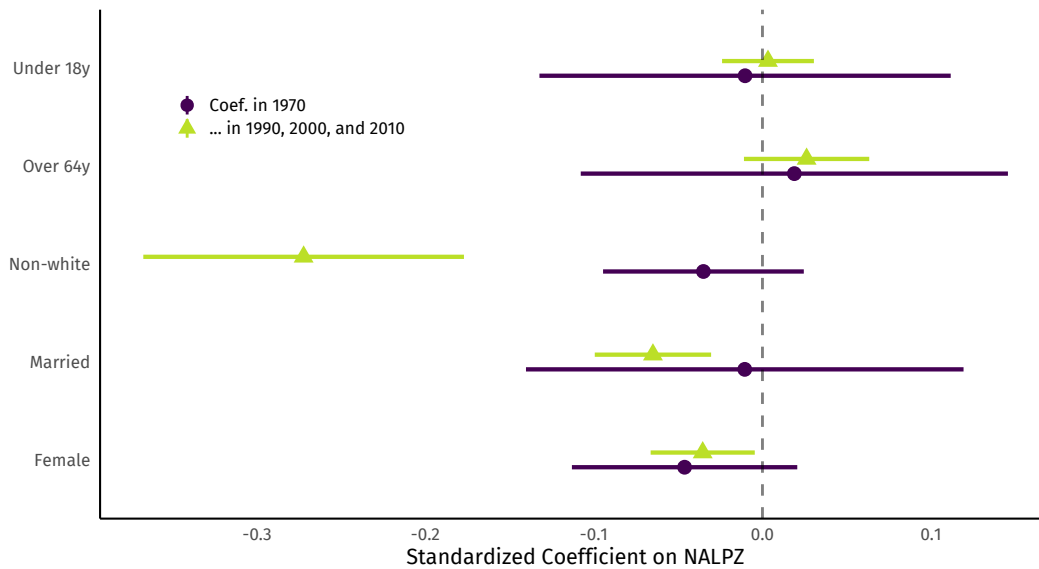


(b) Assessed Value of Building per m² and of Land per m²



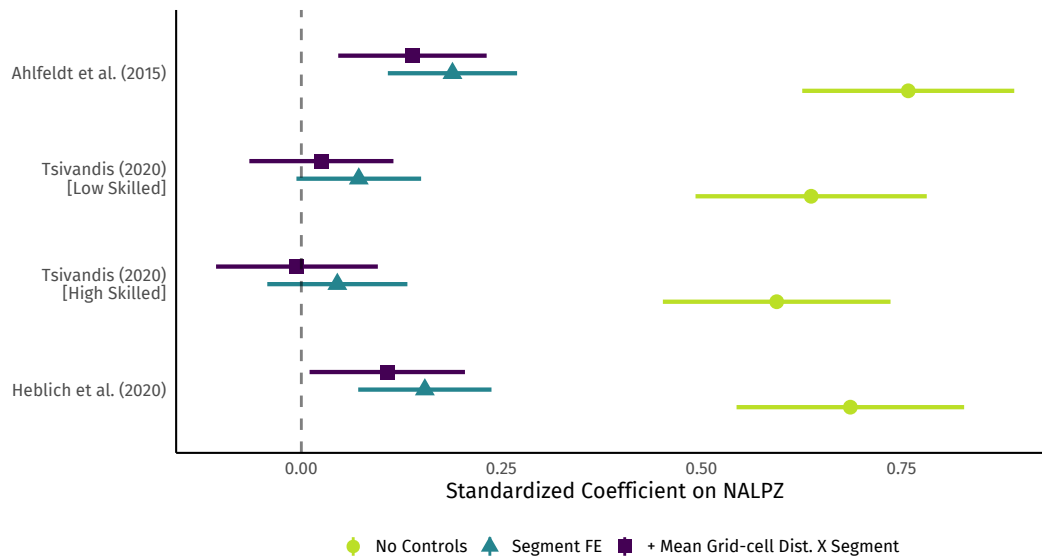
Notes: Plots the estimated β coefficient from Equation 11. Respective outcome is regressed on the Natural Language Processing Zoning Stringency Index (NALPZ), border segment fixed effects, and border segment-specific distance controls. Panel (a) additionally controls for building size and lot size. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level. Mean outcome refers to the average outcome value with a bandwidth of 100m.

Figure 8: Demographic, Housing Characteristics and Land Use Regulation



Notes: Plots the estimated β_1 and β_2 values from estimating Equation 12. Respective demographic or housing characteristic is regressed on the NALPZ interacted with 1970 and post-1970 dummies, and segment by year fixed effects. β_1 captures the ability of the NALPZ to predict census block demographic characteristics in 1970. β_2 captures the ability of the NALPZ to predict census block demographic characteristics after 1970 (1990, 2000, 2010). Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level.

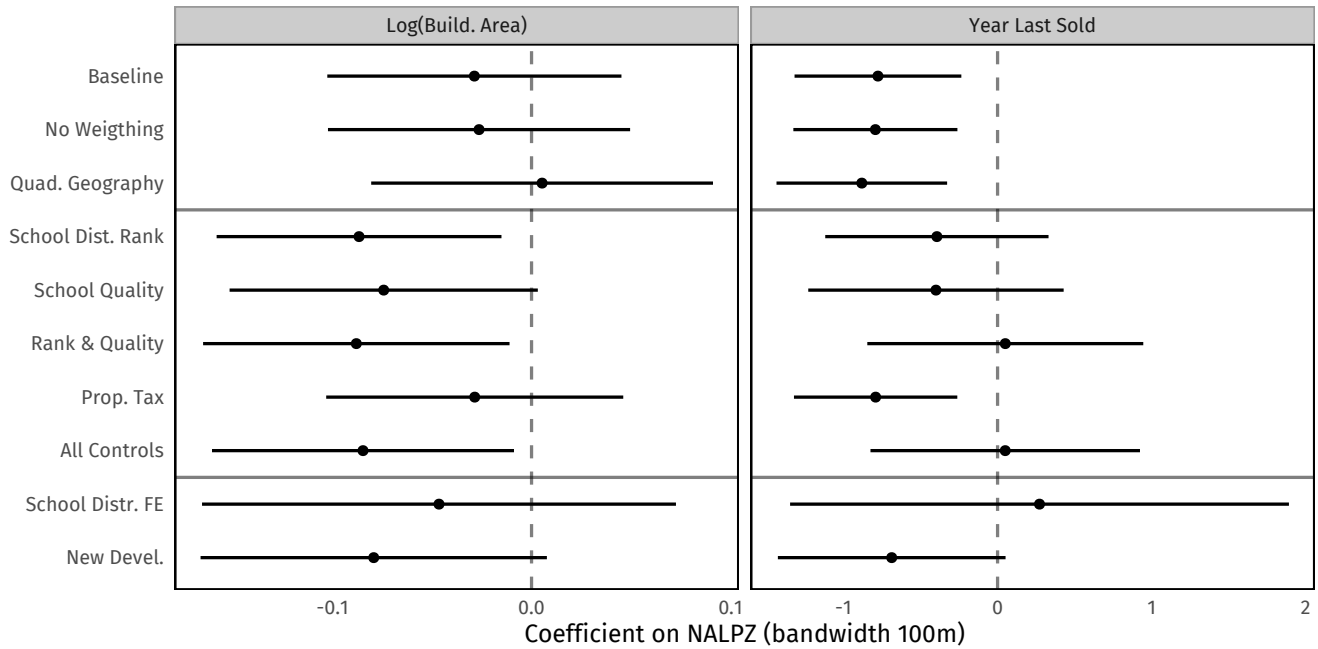
Figure 9: Local Amenities and Land Use Regulation



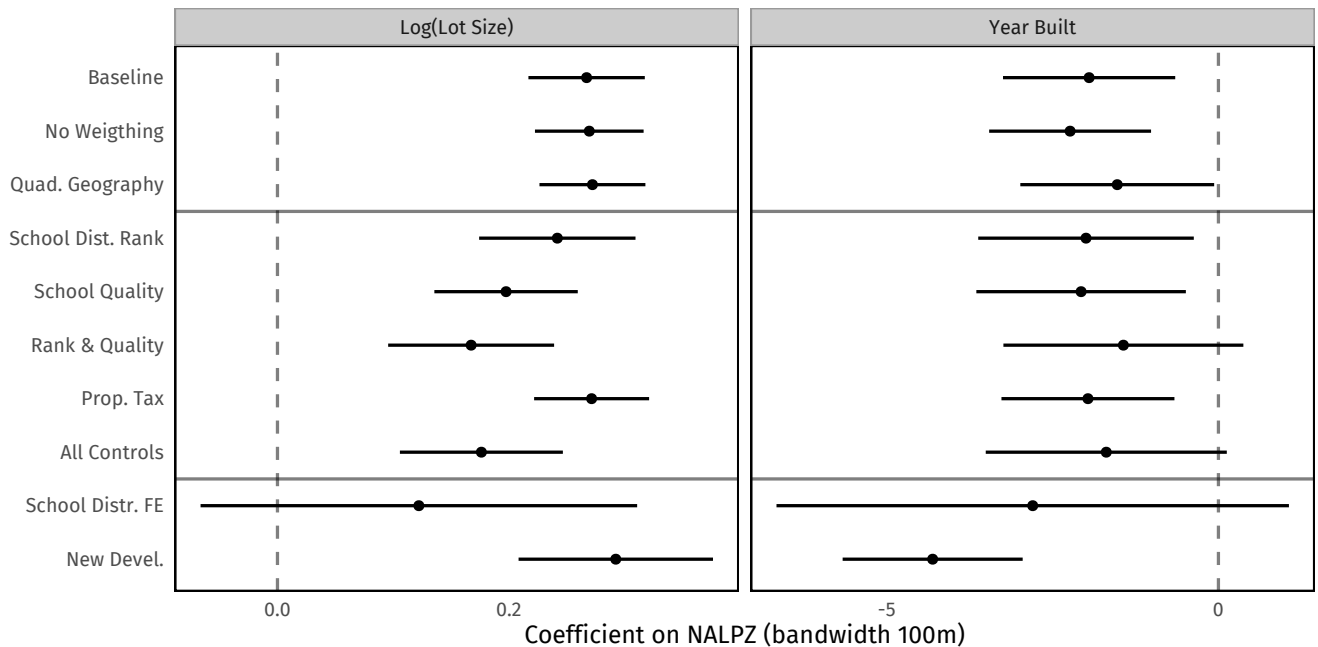
Notes: Plots the estimated β value from estimating Equation 14 under different sets of controls. Unobserved amenity values, calculated from an Ahlfeldt et al. (2015) spatial model under different parameter combinations, are regressed on the NALPZ and the noted controls. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level.

Figure 10: Robustness Checks Baseline

(a) Housing Market Outcomes

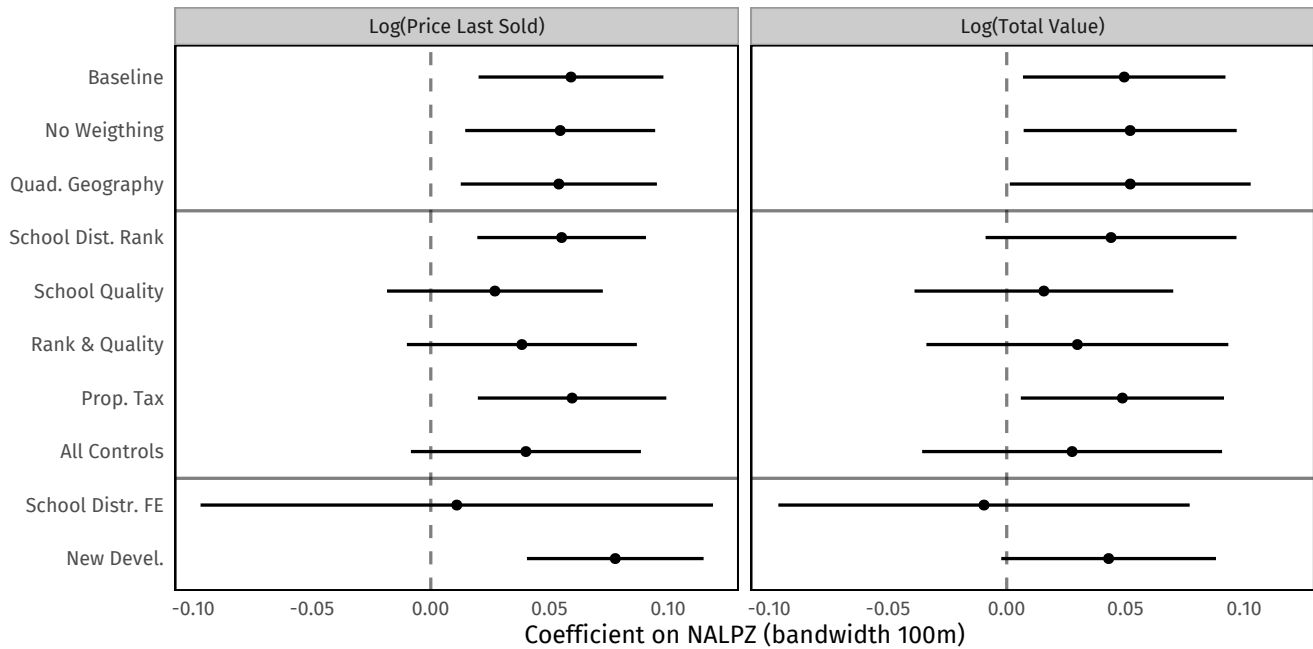


(b) Housing Attributes Outcomes



Notes: Plots the estimated β value from estimating different specifications of Equation 8. “No Weighting” removes triangular weights and assigns uniform weights. “Quad. Geography” adds a square term of the town border segment-specific distance controls, as well as quadratic controls for longitude, latitude, distance to coast, and distance to Boston. “School Dist. Rank” adds a quadratic for school district rank taken from [Niche.com](#). “School Quality” adds per pupil expenditure and graduation rate controls from [ClearGov.com](#). “Prop. Tax” controls for municipal property tax rates. “All Controls” includes all school district controls and the residential property tax rate. “School Distr. FE” controls for school district fixed effects. “New Devel.” estimates the model on the subset of houses that were developed after 1971. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level.

Figure 11: Robustness Checks: House Prices



Notes: Plots the estimated β value from estimating different specifications of Equation 11. “No Weighting” removes triangular weights and assigns uniform weights. “Quad. Geography” adds a square term of the town border segment-specific distance controls, as well as quadratic controls for longitude, latitude, distance to coast, and distance to Boston. “School Dist. Rank” adds a quadratic for school district rank taken from [Niche.com](#). “School Quality” adds per pupil expenditure and graduation rate controls from [ClearGov.com](#). “Prop. Tax” controls for municipal property tax rates. “All Controls” includes all school district controls and the residential property tax rate. “School Distr. FE” controls for school district fixed effects. “New Devel.” estimates the model on the subset of houses that were developed after 1971. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level.

11 Tables

Table 1: Data Sources

Source	Data
Pioneer Institute and Rappaport Institute	Land-Use Regulation Data
Wharton Survey	Wharton Residential Land Use Regulatory Index
Massachusetts Towns' Websites	Zoning/Protective Bylaw Documents
US Census (via IPUMS:NHGIS)	Census Data (Block Level)
Harvard IV-4 Dictionary Categories	Dictionaries
Mass GIS	Tax Assessment Database (Tax Parcels)
	Geographic Data

Table 2: Correlations of Natural Language Processing Regulation Indices Candidates and Existing Indices

Document Length	-0.14	-0.41
Document Length (tf-idf)	0.05	-0.00
Active Dict. (tf-idf)	0.06	0.02
Aquatic Dict. (tf-idf)	-0.08	-0.14
Building Dict. (tf-idf)	-0.34	-0.32
Land Dict. (tf-idf)	0.04	0.07
Legal Dict. (tf-idf)	-0.20	-0.36
Nature Dict. (tf-idf)	0.04	0.21
Object Dict. (tf-idf)	-0.01	0.02
Place Dict. (tf-idf)	-0.31	-0.28
Region Dict. (tf-idf)	-0.41	-0.34
LDA 3 Topics	0.58	0.66
	WRLURI	PRHRD Index

Notes: Data come from Gyourko et al. (2008) and the PIRI (2005). WRLURI is the Wharton Residential Land Use Regulation Index from 2005. PRHRD is the Housing Regulation Database of Massachusetts compiled by the Pioneer Institute and the Rappaport Institute in 2004. I create an index from their data using Principal Component Analysis.

Table 3: Summary Statistics

	Mean	SD	Min	Max	N
<i>Panel A: Tax Parcels</i>					
Tax Assessment Data					
Building Value (\$'000)	202.26	187.16	0.10	15,931.00	1,192,943
Land Value (\$'000)	186.56	208.71	0.10	23,337.90	1,192,943
Other Value (\$'000)	5.74	18.20	0.00	1,965.80	1,192,943
Total Value (\$'000)	394.57	349.69	4.50	27,573.40	1,192,943
Other Parcel Characteristics					
Lot Size (m^2)	3,737.40	13,894.39	37.23	4,318,363.84	1,192,943
Year Built	1,955.96	38.31	1,800.00	2,018.00	1,192,943
Build Area (m^2)	254.64	148.15	50.07	6,611.17	1,192,943
No. of Rooms	6.78	7.99	1.00	8,020.00	1,099,449
Owner Occupied	0.84	0.36	0.00	1.00	1,192,827
Last Sale Year	2,000.22	15.05	1,900.00	2,019.00	1,192,943
Last Sale Price (\$'000, 2018 prices)	482.63	416.02	50.01	18,508.42	713,602
Geographic Characteristics					
Dist. to Boston (km)	47.56	38.01	0.00	190.45	1,192,943
Dist. to Coast (km)	26.94	37.46	0.00	180.50	1,192,943
Dist. to Nearest Border (km)	1.45	1.31	0.00	13.52	1,192,943
<i>Panel B: Buildings</i>					
Building Height (m)	11.53	5.17	3.00	25.00	1,302,984
Dist. to Nearest Border (km)	1.54	1.37	0.00	13.56	1,302,984
<i>Panel C: Towns</i>					
Regulation Index					
NALPZ ¹	-0.03	0.91	-2.33	2.33	303
2010 Census Characteristics					
Population ('000)	17.88	22.79	0.17	181.47	303
Housing Units ('000)	7.61	9.57	0.11	74.64	303
Housing Density (units/ km^2)	217.40	392.59	2.48	3,147.92	303
Rural (%)	32.43	37.61	0.00	100.00	303
Female (%)	51.70	2.07	39.81	60.17	303
Under 18 Years (%)	22.13	4.04	6.83	31.98	303
Over 64 Years (%)	15.00	4.44	7.89	39.80	303
Non-White (%)	7.43	9.06	0.45	56.87	303
Married (%)	68.59	7.12	26.16	83.27	303
School Quality Characteristics					
Expenditure Per Pupil (\$'000)	15.90	4.85	9.52	59.81	296
Graduation Rate (%)	92.79	5.99	65.50	100.00	291
Municipal Taxes					
Residential Property Tax Rate (%)	15.24	3.85	2.75	24.34	303

Notes: Last sale price data drops observations for which the price is under \$50,000, as it is most likely reflects the transfer of property rather than the true market price. Number of tax parcels and buildings differs as some properties have multiple buildings and some buildings are missing height data.

¹ Natural Language Processing Zoning Stringency Index

Appendix A Natural Language Processing Details

A.1 Term Frequency-Inverse Document Frequency Weighting

tf-idf weighting assigns more weight to tokens that appear more often in fewer documents, under the presumption that these tokens are better able to discriminate between different documents. Conversely, tokens that appear seldomly in almost all documents do not tell us much. Formally, it is the product between a term frequency (tf) part, and an inverse document frequency (idf) part. There are different ways to measure both of them, and those used in this paper are described below.

The formula for the term frequency part is given as:

$$\text{tf}_{vd} = x_{vd} / \sum_{v \in \mathcal{V}} x_{vd} \quad (\text{A.1})$$

where the term frequency for token v in document d depends on the count of that token in the document (x_{vd}) divided by the total length of the document for all tokens in \mathcal{V} .

The expression for the inverse document frequency is as follows:

$$\text{idf}_v = \log(D/df_v) \quad (\text{A.2})$$

where D refers to the number of documents in the corpus. This term is term specific, but is the same for every document.

The product of these two terms is the tf-idf weighted token count:

$$\text{tf-idf}_{vd} = \text{tf}_{vd} \times \text{idf}_v \quad (\text{A.3})$$

A.2 Dictionary Methods

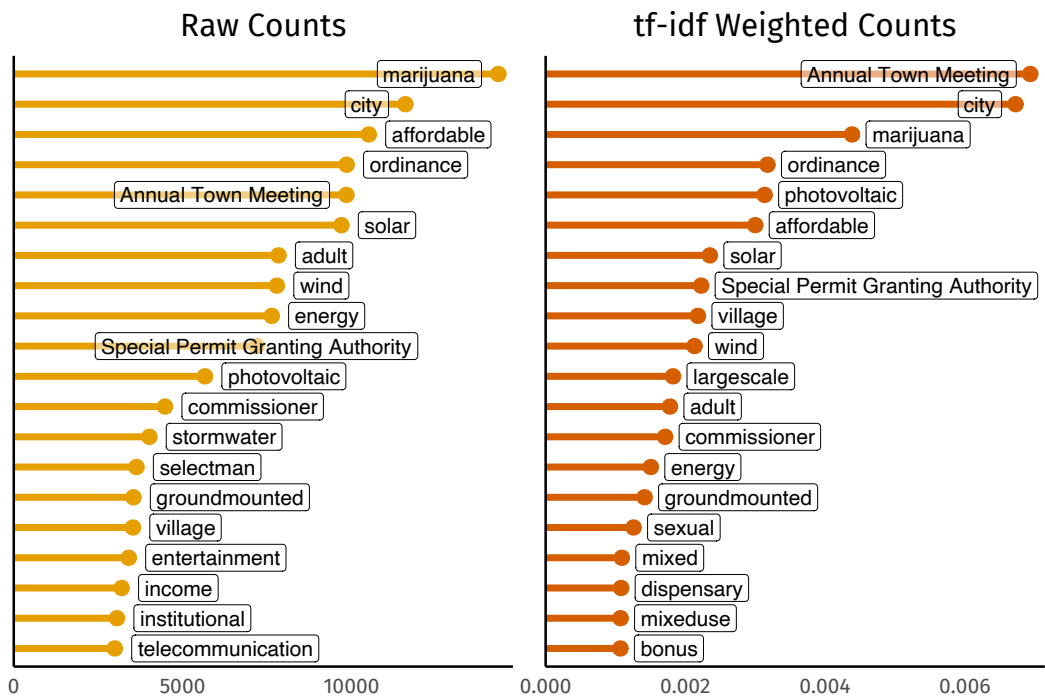
Table A.1: Harvard IV-4 Category Dictionary Examples

Active (2045)	Legal (192)	Place (318)	Region (61)
damage	truant	range	municipality
investor	counsel	province	rural
aid	law	ridge	slum
argue	disputable	skyline	spot
protect	eye	arena	city
Aquatic (20)	Land (63)	Building (46)	Object (104)
wave	mainland	construction	constitution
breaker	cave	shutter	board
creek	scene	bedroom	paint
channel	border	bathroom	stamp
water	island	chamber	notice
Nature (61)			
maple	grass	manure	
calcium	flower		

Notes: Each sub-table is a dictionary category used when investigating dictionary NLP methods. The number in parentheses indicates the total number of words belonging to that topic. Five words, chosen at random, are listed under the heading.

A.3 Figures

Figure A.1: Top Word Counts from Municipal Bylaws



Notes: TF-IDF weights are calculated corpus-wide rather than per document.

Figure A.2: Histogram of Raw and tf-idf Weighted Token Counts

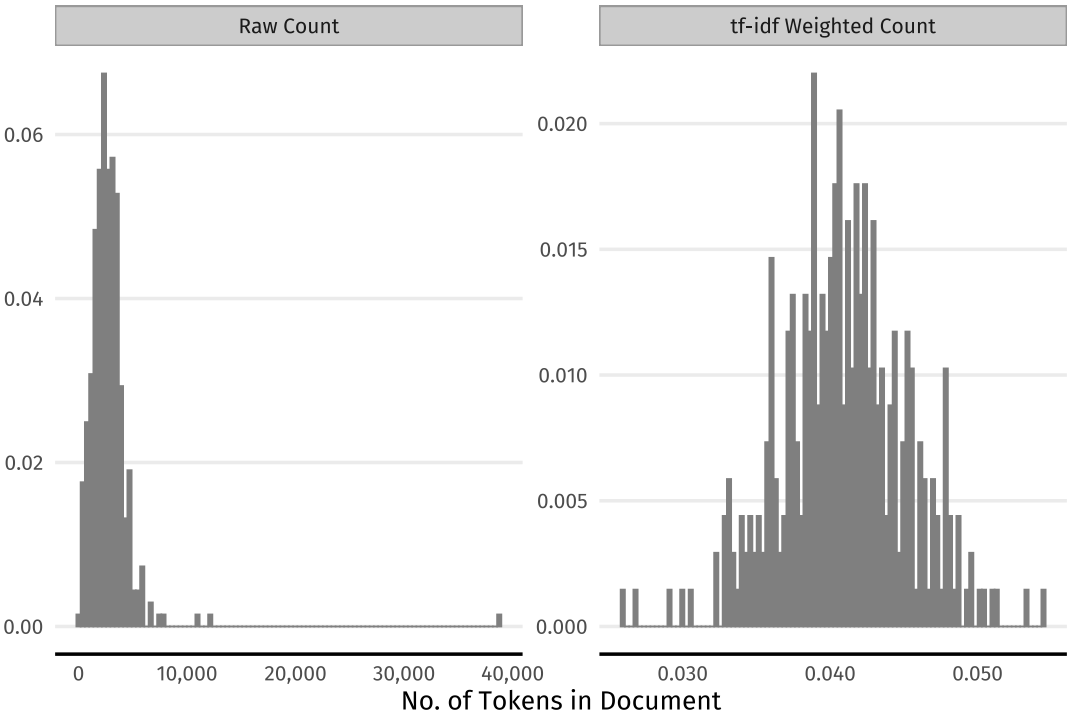
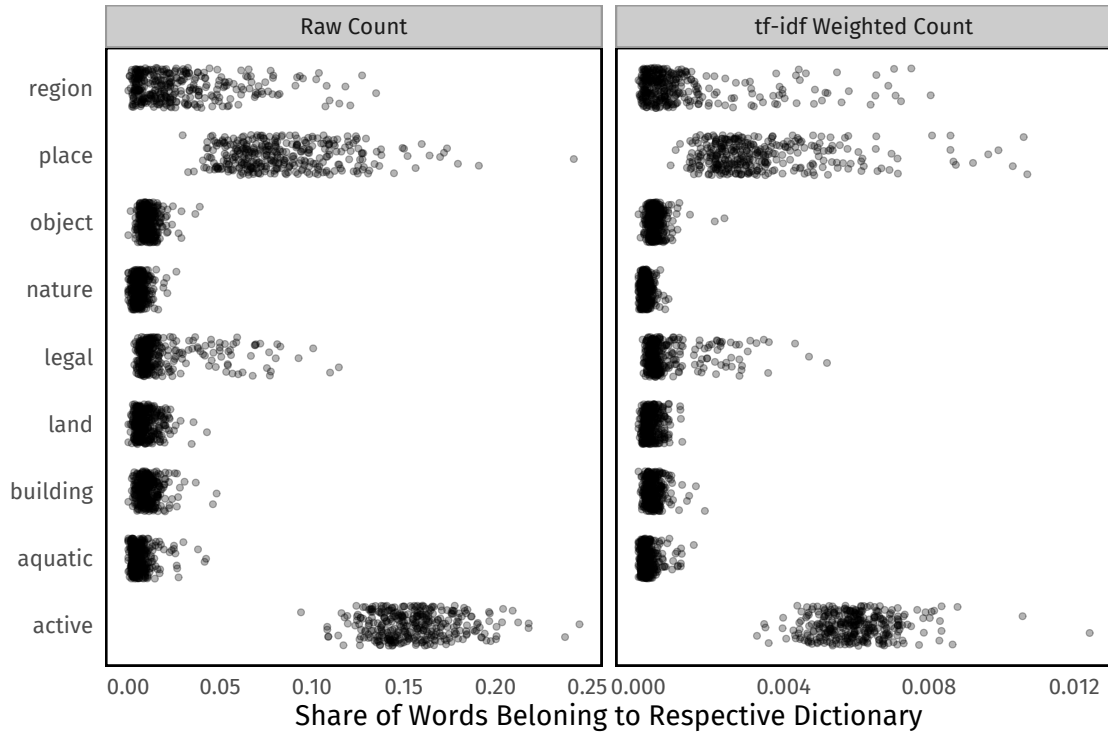
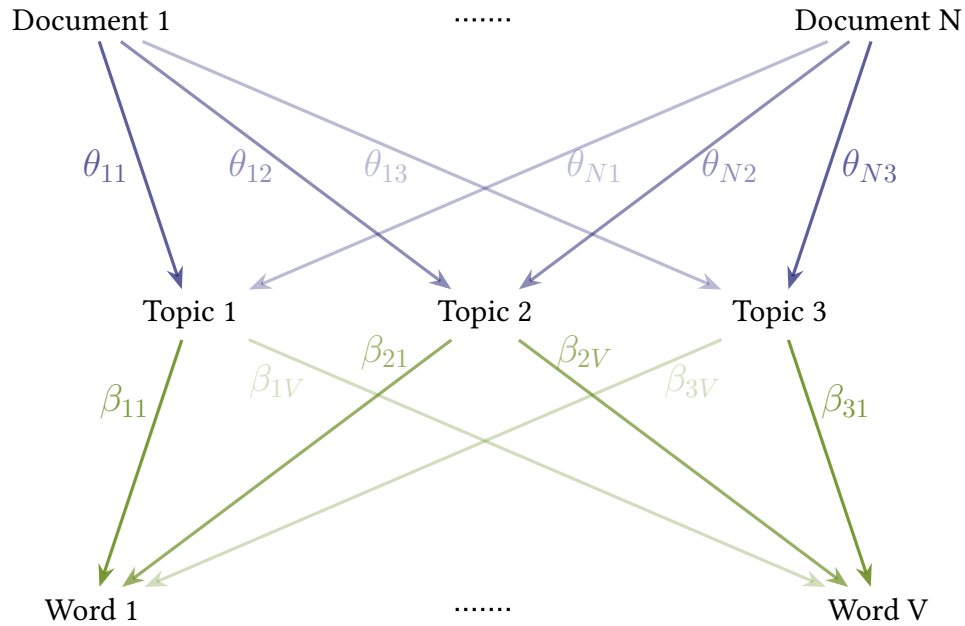


Figure A.3: Distributions of Raw and tf-idf Weighted Dictionary Scores



Notes: Raw counts are divided by their document length for comparability. The measures can be interpreted as the share of document tokens belonging to the respective category.

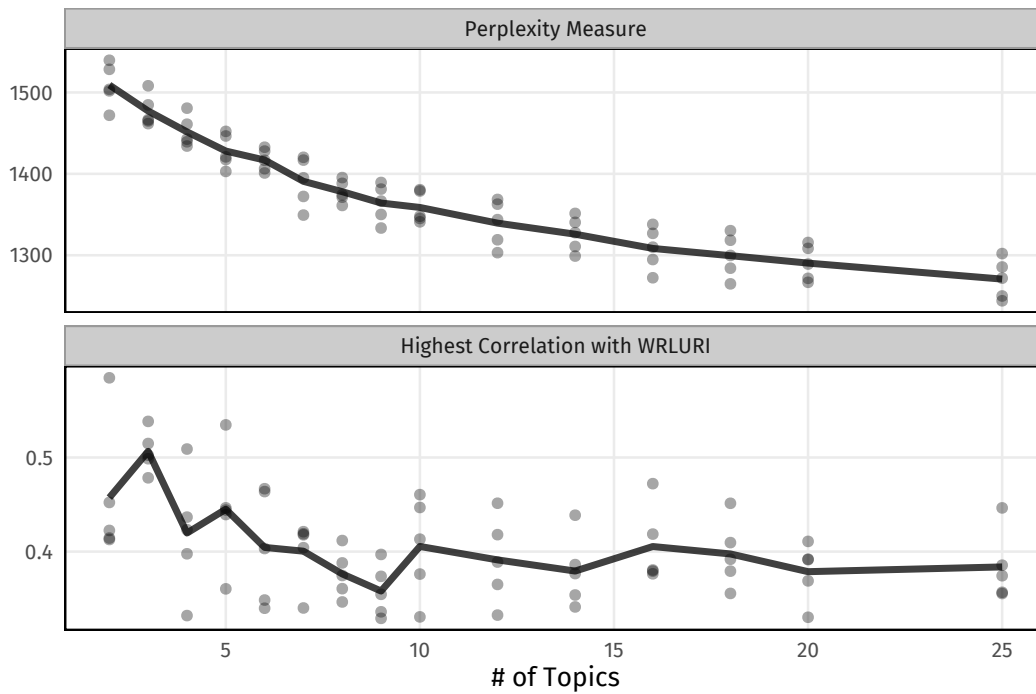
Figure A.4: LDA Data Generating Process



Inspired By: Bandiera et al. (2020).

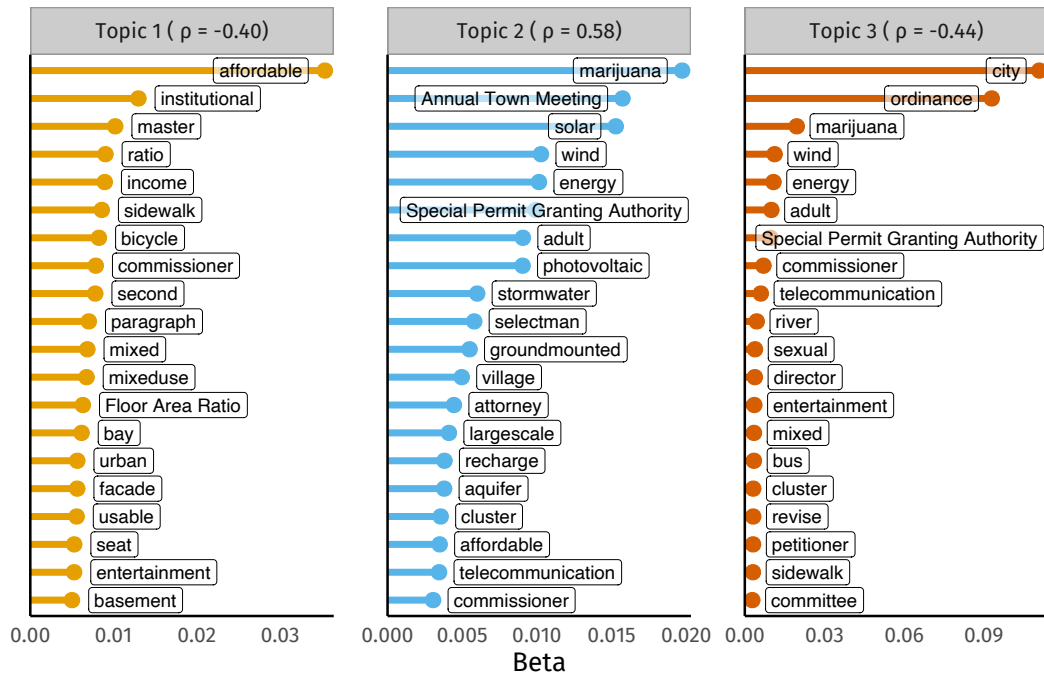
Notes: Example assumes three latent topics. θ are the parameters of the true posterior distribution, while γ , as used in the text, are the parameters from the approximating distribution.

Figure A.5: LDA: Cross-Validation for Number of Latent Topics



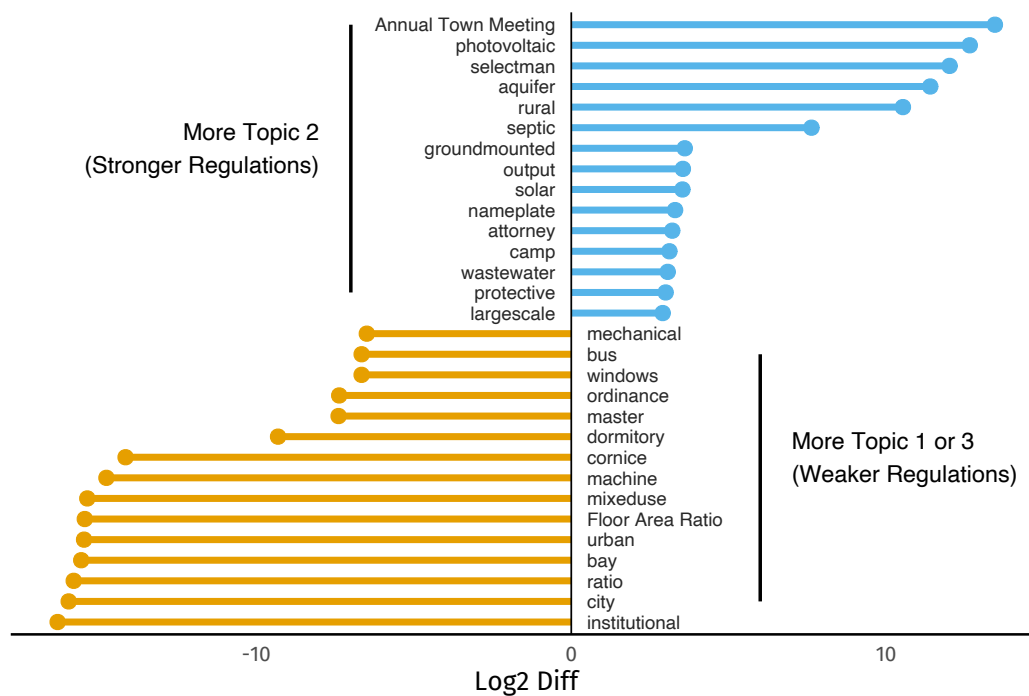
Notes: Lower is better in the first panel and higher is better in the second. Each point is a fold.

Figure A.6: LDA: Top Words per Latent Topic



Notes: Each value corresponds to the β value from the estimated LDA model (ie the estimated probability of a word being drawn from that specific topic).

Figure A.7: LDA: Words that Most Discriminate Between Topic 2 and Topics 1 or 3



Notes: Tokens with a β value less than 0.03 are excluded. Each point represents the base two logarithm of the ratio of β_{2v} and $\beta_{1v} + \beta_{3v}$.

Appendix B Characteristics of Highly Regulated Towns

With a measure of land use regulatory intensity in hand, that covers almost the entirety of one state, I now turn to discussing the geographic patterns of the Natural Language Processing Zoning Stringency Index, as well as town-level predictors of stringent LUR.

The near-universal coverage, at small geographic units, of the regulation measure enables me to present, to the best of my knowledge, new stylized facts about the geographic distribution of LUR. The first is that LUR are highly correlated over space. This has already been noted for larger geographic regions when looking at the Wharton Residential Land Use Regulation Index (eg cities in California and New England are highly regulated while those in the sunbelt are not). But even *within* these regions, there is a high degree of geographic clustering.

Next, I show that though more regulated municipalities allocate less overall land to development, of the land developed is given to residential purposes and less towards commercial or industrial uses.

B.1 Geographic Clustering of Land Use Regulation

To get an idea of the geographic clustering of land use regulatory intensity, I plot the relationship between a town's own NALPZ and the average of their direct neighbours. This is shown in Figure B.1. The blue line indicates the linear fit between the two, while the dashed grey line indicates 45°. It shows a clear pattern between the regulatory environment of neighbouring towns. On average, towns with more strict zoning regulations have neighbours with similarly strict land use policies. This aligns well with the geographic distribution of NALPZ shown previously in Figure 3.

This fact, to the best of my knowledge, has not been shown at this geographic detail. This has already been noted at the Metropolitan Statistical Area level using the WRLURI measure, but not *within* these units. Figure B.1 uses NALPZ for the entire sample in Massachusetts, but the relationship remains when considering the subsample of towns that surround Boston (specified as those that are part of the Pioneer/Rappaport Housing Regulation Database) and those further from the state's economy centre.

Though there is a high degree of geographic clustering of land use stringency, there also exists a great deal of heterogeneity with respect to the differential zoning policies between neighbouring towns. Figure B.2 highlights this distribution. The average differential of NALPZ between any two neighbouring town pairs is 0.74 of a standard deviation (median is 0.58 of a s.d.). This is the key variation that my empirical strategy exploits to estimate the impact of more restrictive zoning.

B.2 Predictors of Restrictive Land Use Regulation

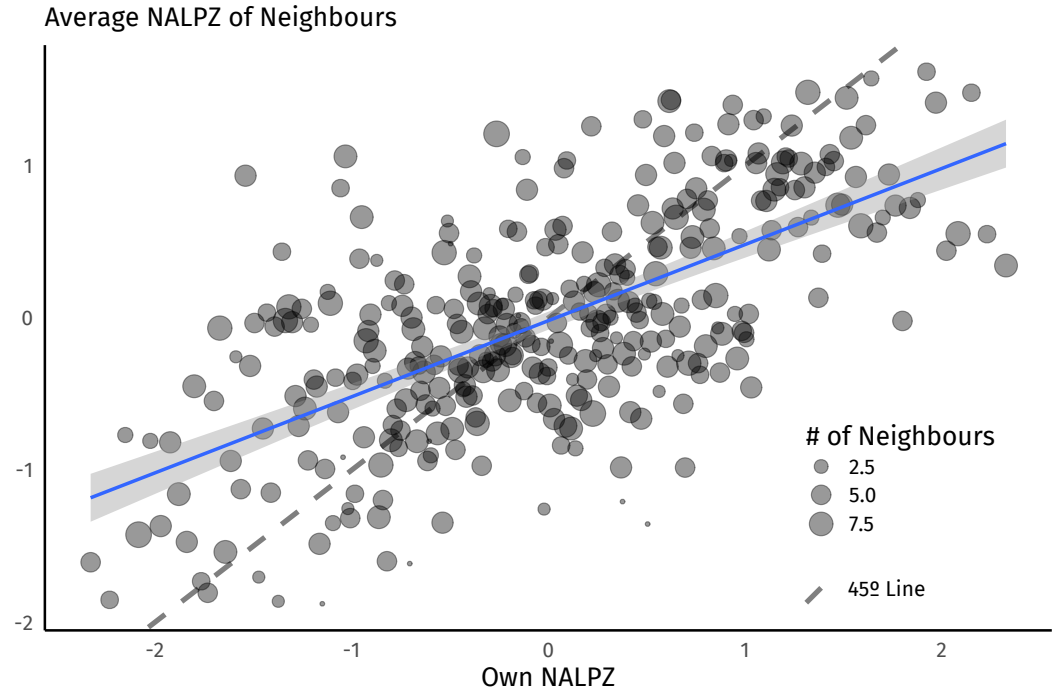
I now explore what town-level characteristics in 1970, pre Massachusetts Zoning Act, best predict future NALPZ levels. Specifically, I look at 1970 town-level census data and 1971 land use data from MassGIS to investigate the sorts of towns that implement stringent LUR.

The two strongest predictors of NALPZ are shown in Figure B.3: the percentage of land covered in forests (in the first panel) and the natural logarithm of housing units per square kilometer (in the second). Both of these variables correlate very strongly with NALPZ, with coefficients of correlation of -0.81 and 0.74 respectively. These results align with those found by Glaeser and Ward (2009), who consider minimum lot sizes (one aspect of zoning regulations). Remarkably, these patterns persist well into the future; corresponding relationships for 2010 are shown in the appendix, Figure B.4. This fact may point to restrictive land use policies being used to preserve the contemporaneous city shape, makeup, and characteristics, once individual towns were given the legislative ability to implement their own zoning policies.

Two other interesting features of Massachusetts towns in 1970 that relate to regulatory intensity are shown in Figures B.5 and B.6. In the first panel of Figure B.5, the fraction of land being used for residential purposes is plotted against NALPZ, while in the second, the fraction of *developed* land being used for residential purposes is plotted instead. This highlights that, on average, towns that are more strictly regulated have less land overall for residential use, but of the land they already developed, slightly more is residential. This reversal pattern when considering all land compared with only developed land is not apparent when considering the share of land allocated to commercial or industrial uses as shown in Figure B.6. Unlike with residential land coverage, the implication does not vary if one considers all land, or only already-developed land. Towns with stricter zoning regulations allocate less absolute and relative land to industrial purposes.

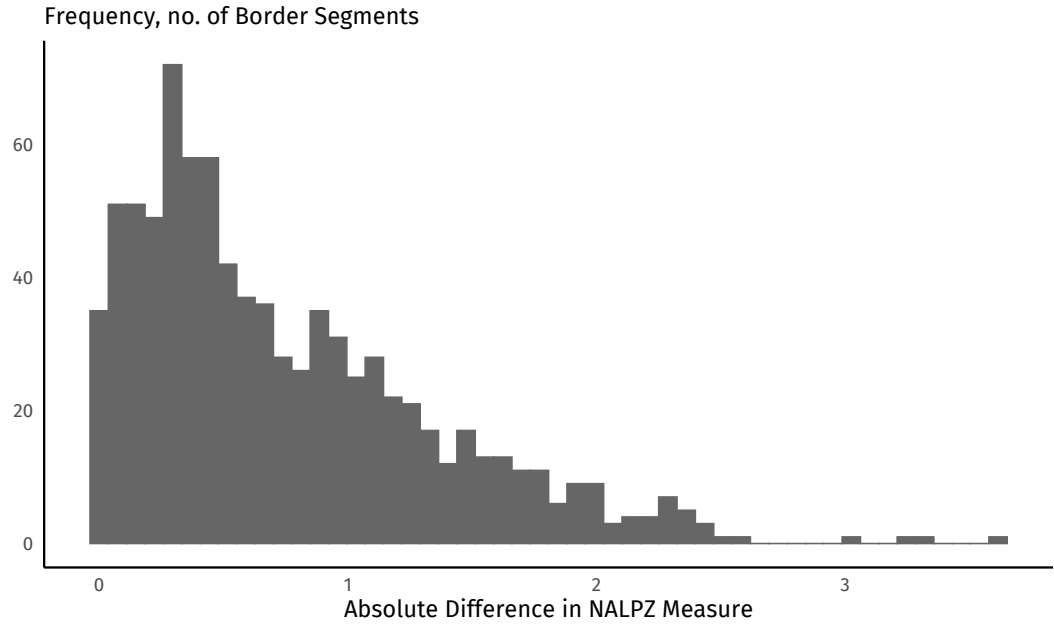
The trends shown in the last two figures are also virtually unchanged when one looks at the data in 2010, as shown in the appendix Figures B.7 and B.8.

Figure B.1: Relationship Between Own NALPZ and Average of Neighbouring Towns



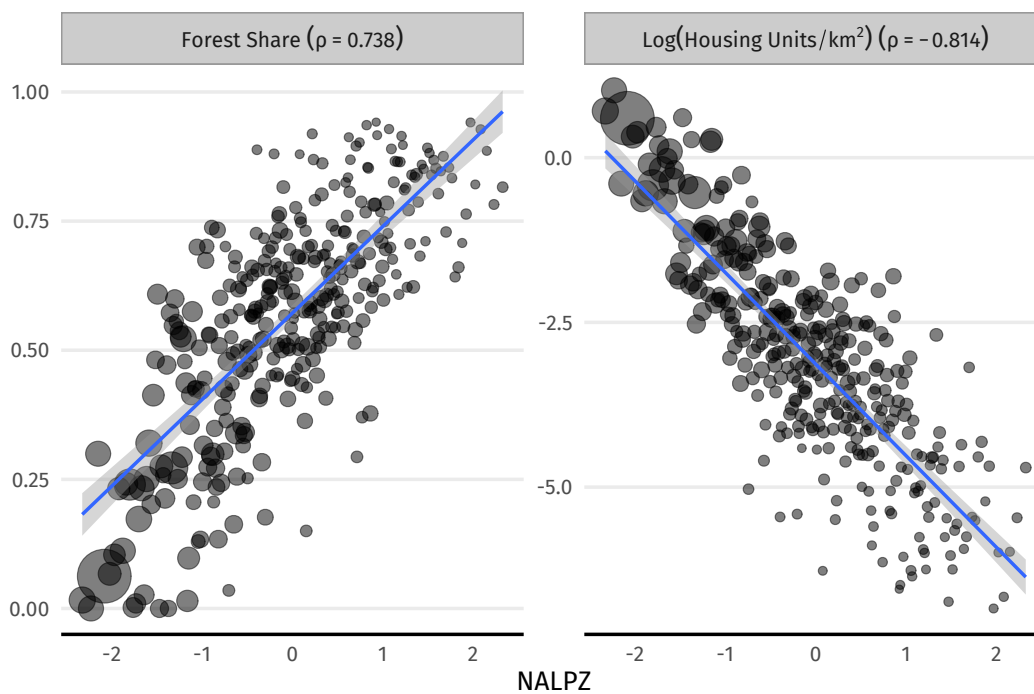
Notes: Each observation is a Massachusetts town. Size of dot corresponds to number of neighbours.

Figure B.2: Distribution of Absolute Differential of NALPZ at Borders



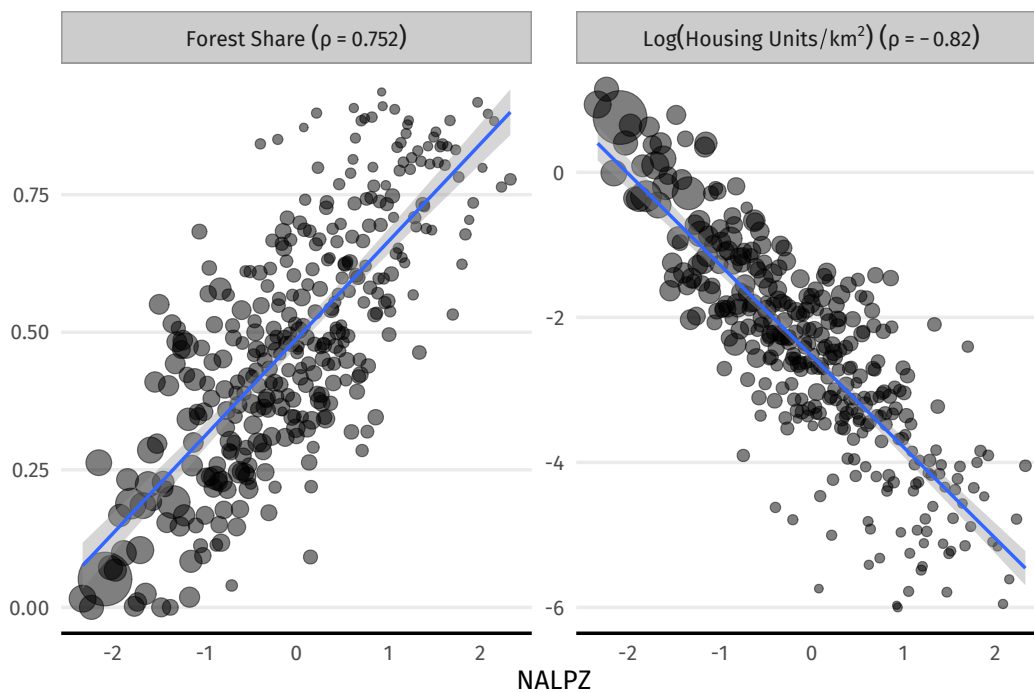
Notes: Each observation is a shared border between two Massachusetts Towns.

Figure B.3: 1970 Town Predictors of NALPZ



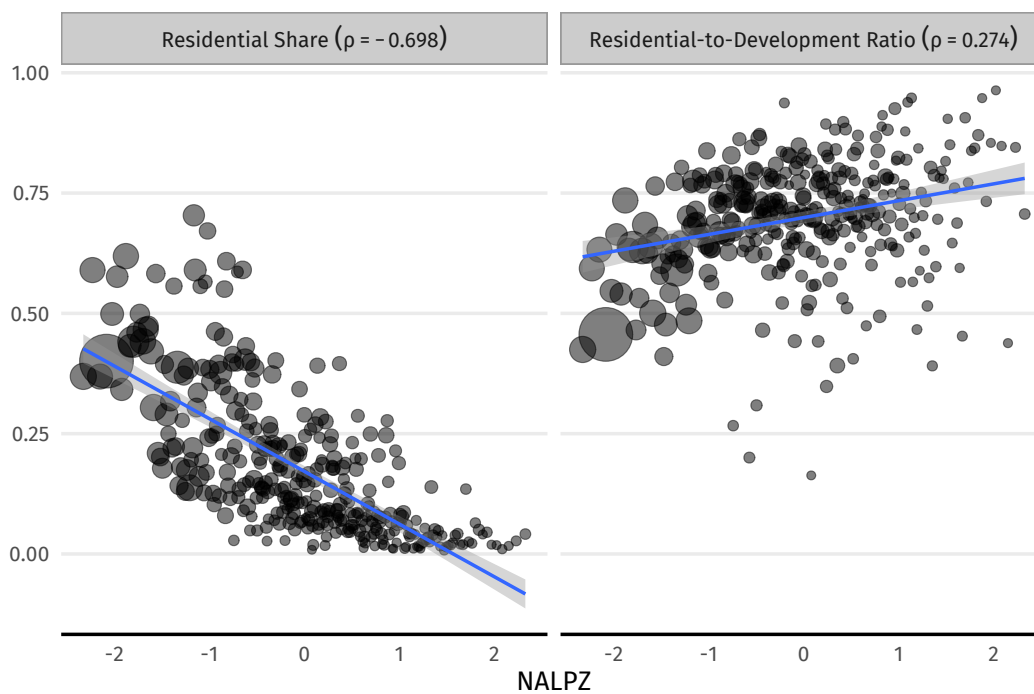
Notes: Each observation is a Massachusetts town in 1970. Size of dot corresponds to population.

Figure B.4: 2010 Town Predictors of NALPZ



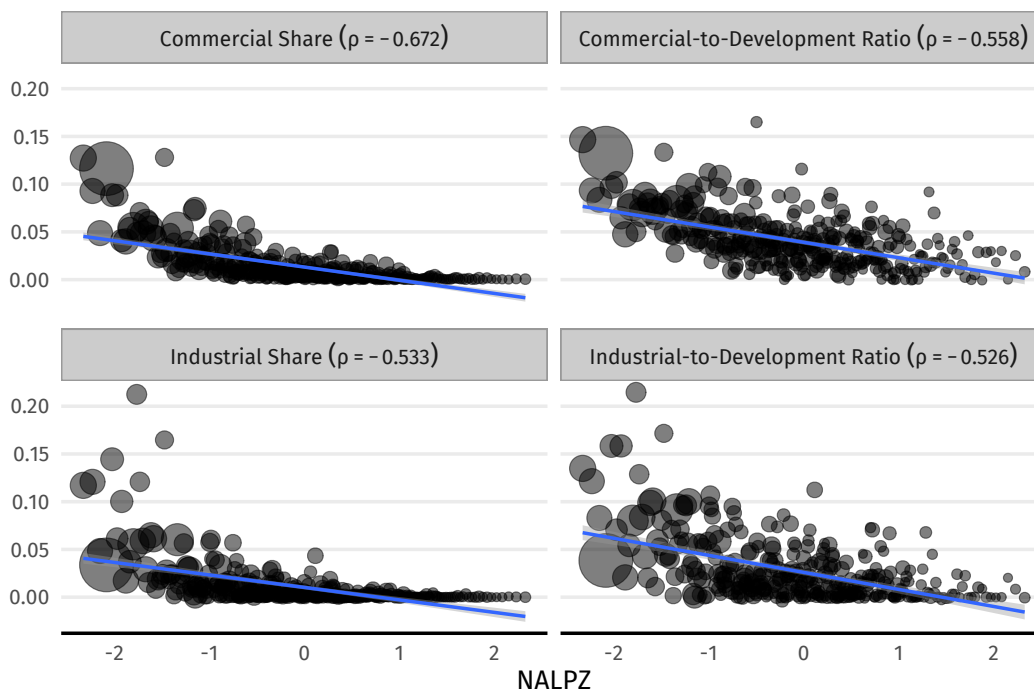
Notes: Each observation is a Massachusetts town in 2010. Size of dot corresponds to population.

Figure B.5: 1970 Town Residential Development and NALPZ



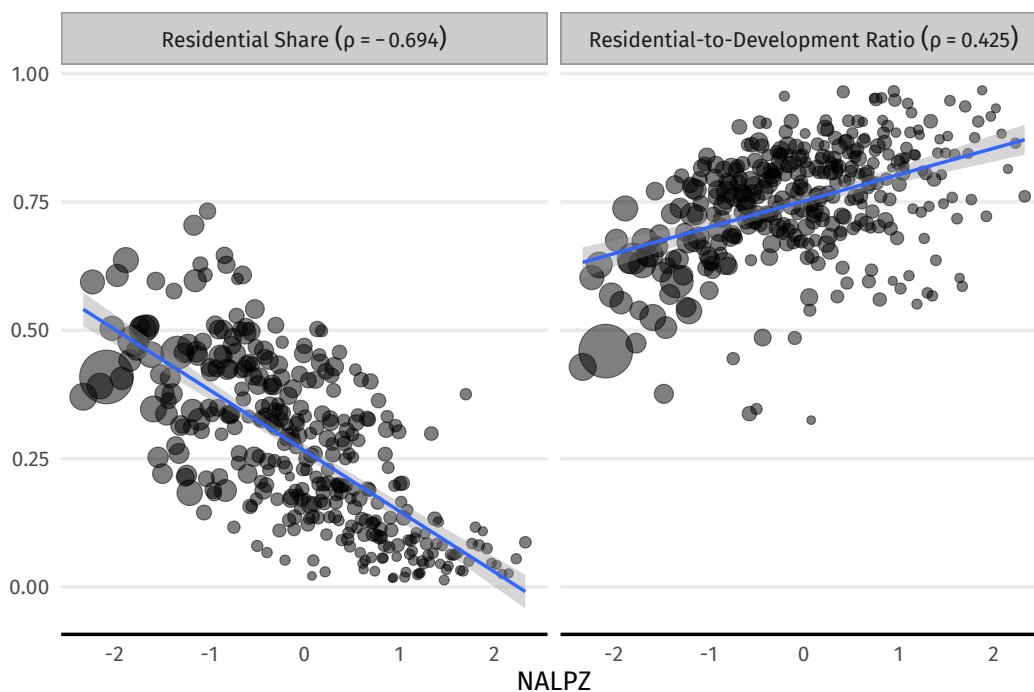
Notes: Each observation is a Massachusetts town in 1970. Size of dot corresponds to population.

Figure B.6: 1970 Town Industry Development and NALPZ



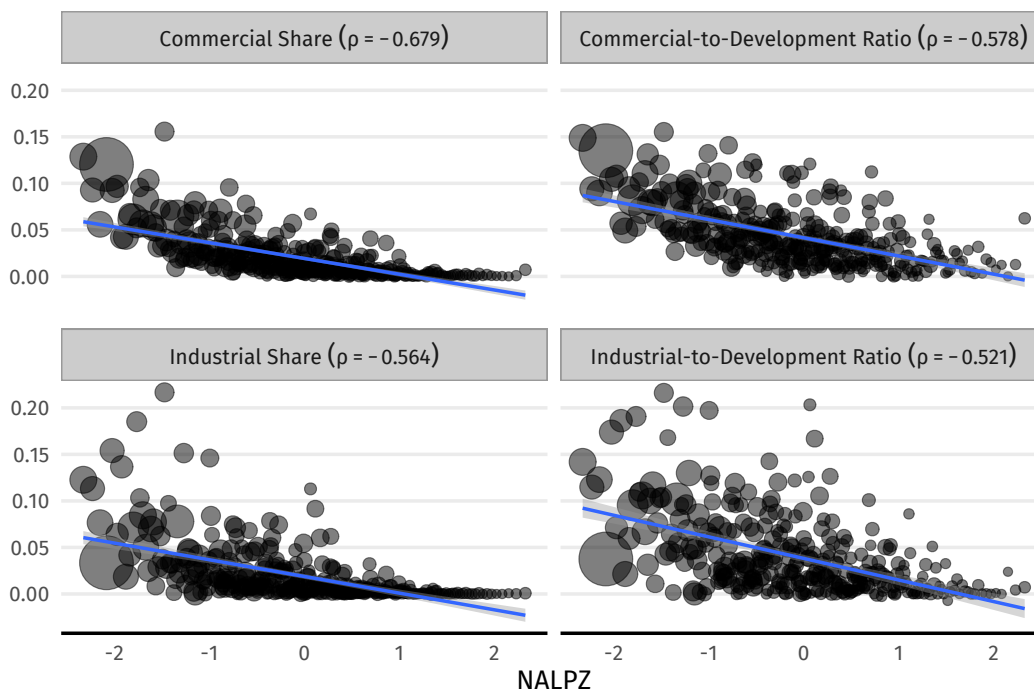
Notes: Each observation is a Massachusetts town in 1970. Size of dot corresponds to population.

Figure B.7: 2010 Town Residential Development and NALPZ



Notes: Each observation is a Massachusetts town in 2010. Size of dot corresponds to population.

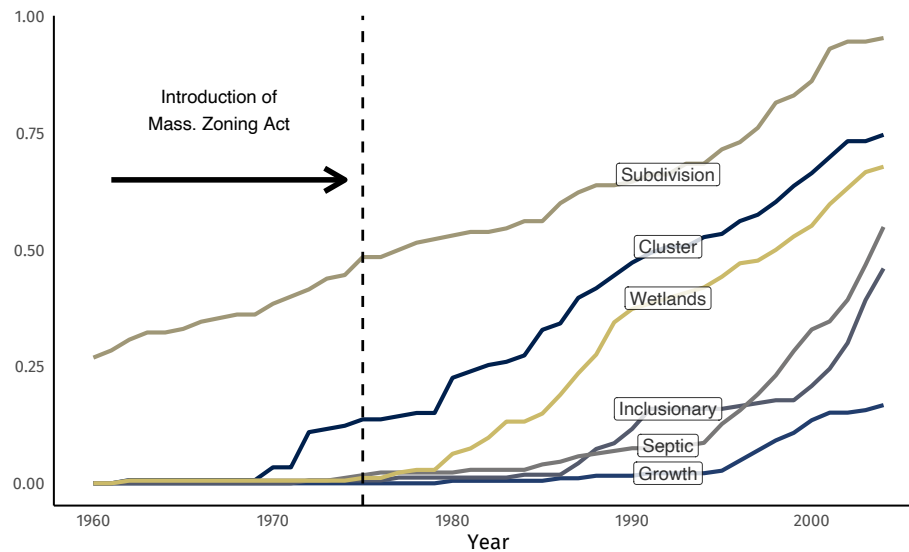
Figure B.8: 2010 Town Industry Development and NALPZ



Notes: Each observation is a Massachusetts town in 2010. Size of dot corresponds to population.

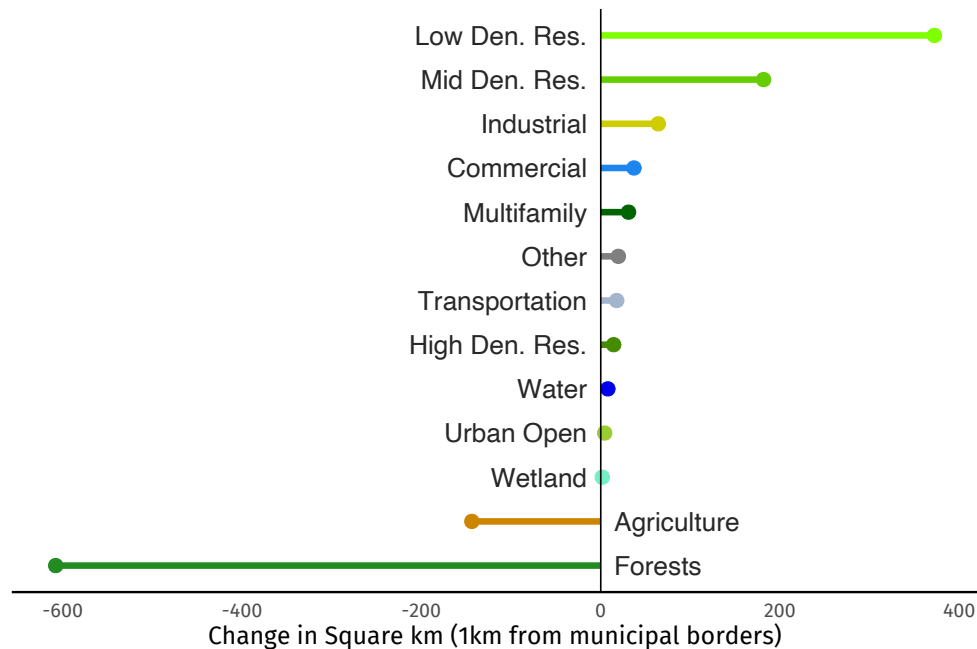
Appendix C Additional Figures

Figure C.1: Share of Towns Adopting Land Use Regulations by Year, Cumulative



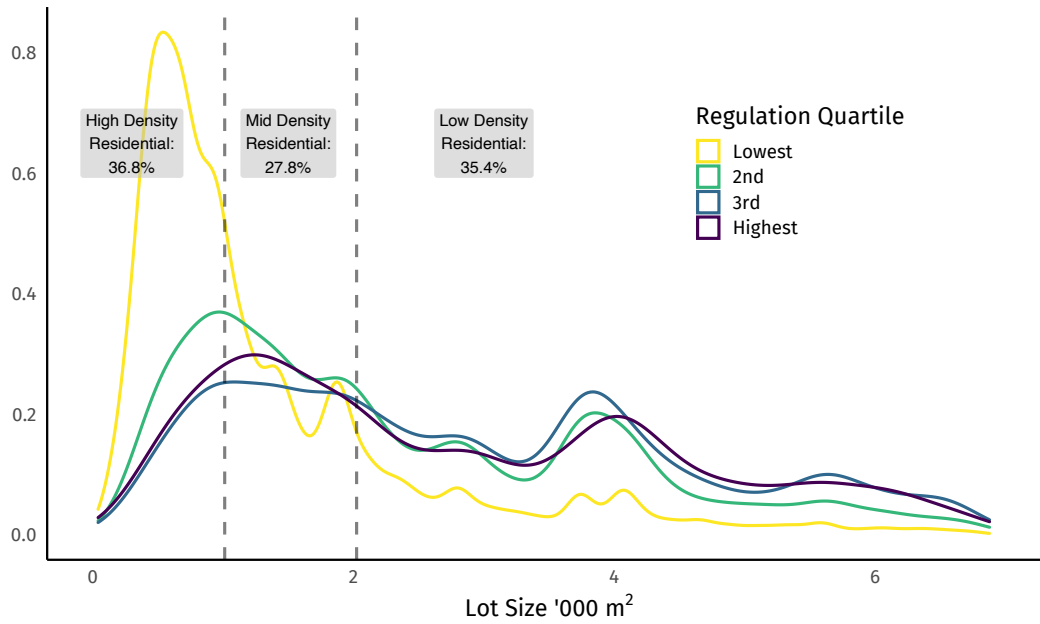
Notes: Some LUR enable development and others inhibit it. Data comes from the Pioneer Institute/Rappaport Institute (PIRI, 2005) Housing Regulation Database for Massachusetts.

Figure C.2: Land Use Conversion from 1971 to 1999



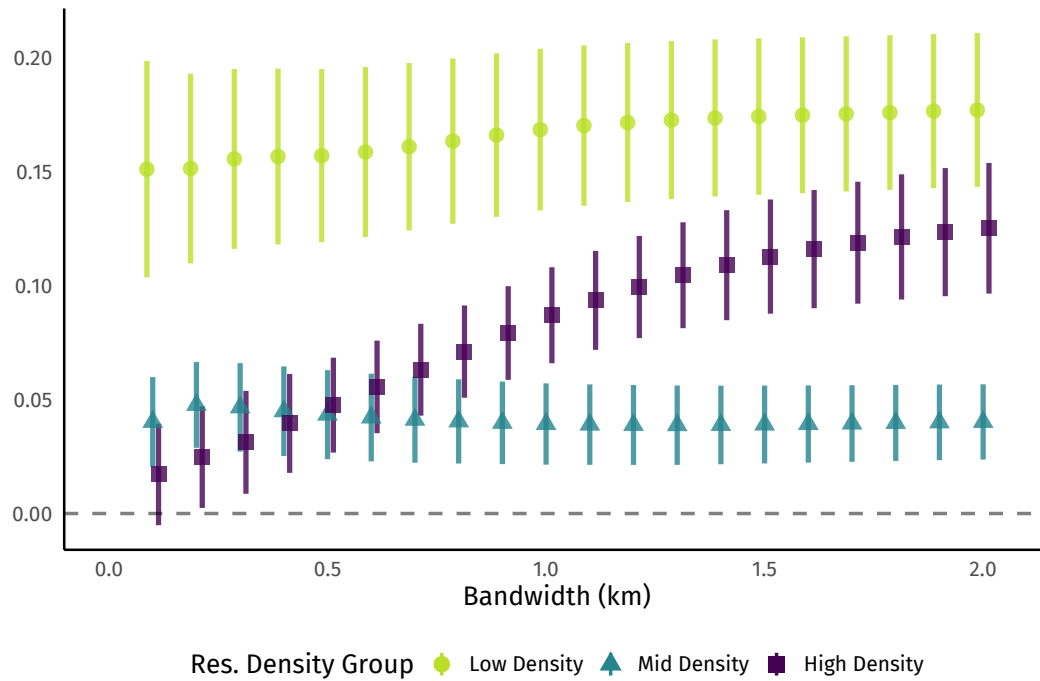
Notes: Each bar indicates the change in area in km² for the respective land use category from 1971 to 1999. Data from MassGIS

Figure C.3: Distribution of Lot Sizes by Regulation Quartile



Notes: Each line represents a density polygon plotted for each quartile of the NALPZ. The dashed lines separate the density classes according to the land use categories from MassGIS.

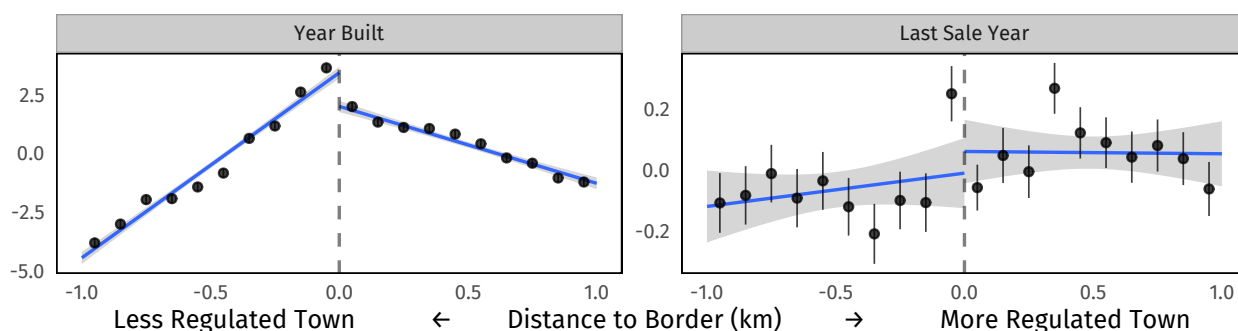
Figure C.4: Spatial RDD by Residential Density Grouping: Logarithm of Lot Size on NALPZ



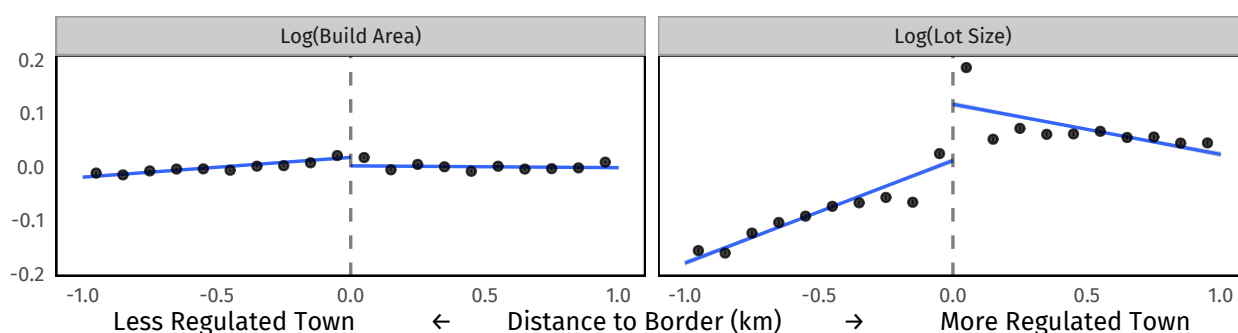
Notes: Plots the estimated β value from estimating Equation 8 separately for each group of parcels according to their density grouping. Respective outcome is regressed on the Natural Language Processing Zoning Stringency Index (NALPZ), border segment fixed effects, and border segment-specific distance controls. Density groupings are defined by MassGIS. Point estimates and 95% confidence intervals are given. Standard errors clustered at the town level.

Figure C.5: Residualized Outcomes by Less/More Regulated Town for Every Town Border

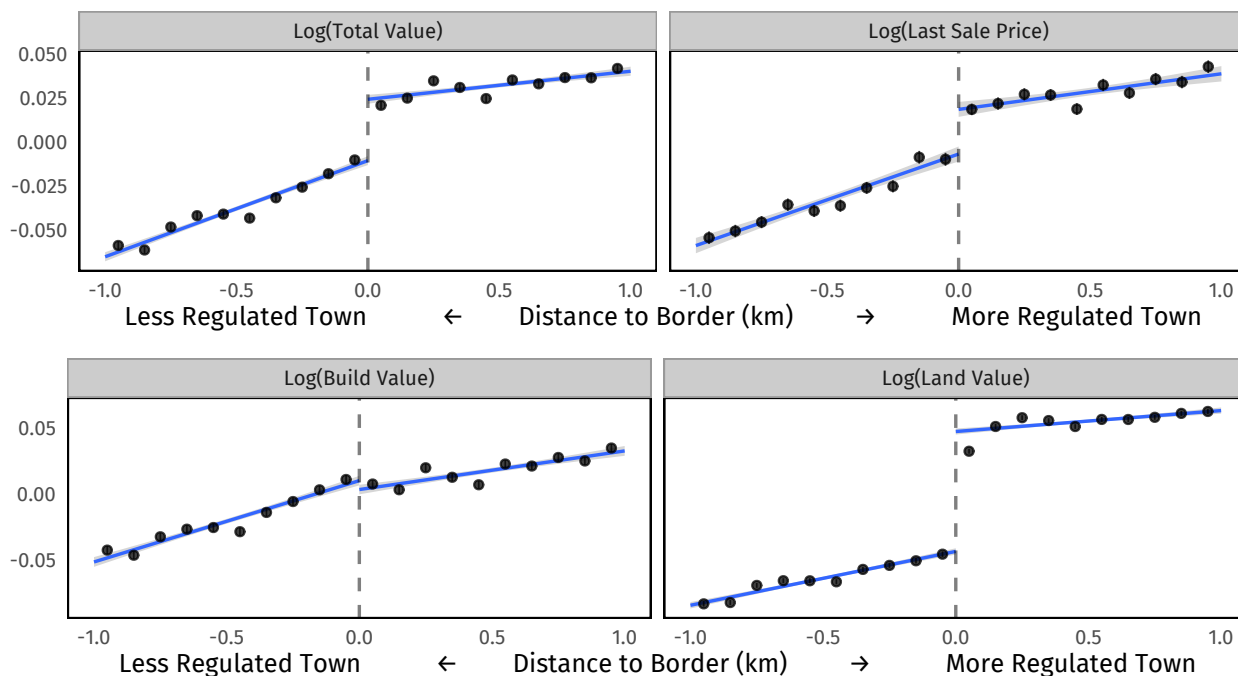
(a) Housing Market Outcomes



(b) Housing Attributes Outcomes



(c) House Price Outcomes



Notes: Each panel plots the respective outcome residualized by border-segment fixed effects. Residuals are binned by every 100m. Mean and the respective standard error plotted for each bin. Towns may have housing units in both the “less regulated” and “more regulated” categories, but as each unit is matched to a unique border an observation can only belong to one category.

Appendix D Spatial RDD Regression Tables

Table D.1: Spatial RDD: Main Results

	Log(Lot Size)	Year Built	Log(Building Size)	Year Last Sold	Build. Height	Log(Build. Height)
	(1)	(2)	(3)	(4)	(5)	(6)
0.1km	0.267***	−1.950***	−0.029	−0.776***	0.179	0.014
	[0.217,0.317]	[-3.248,-0.651]	[-0.103,0.045]	[-1.317,-0.235]	[-0.084,0.442]	[-0.011,0.038]
	56,196/552/278	56,196/552/278	56,196/552/278	56,196/552/278	42,428/408/226	42,428/408/226
0.2km	0.267***	−2.504***	−0.025	−0.806***	0.284***	0.022**
	[0.223,0.311]	[-3.711,-1.297]	[-0.102,0.051]	[-1.320,-0.292]	[0.078,0.490]	[0.003,0.041]
	105,726/632/290	105,726/632/290	105,726/632/290	105,726/632/290	94,422/537/250	94,422/537/250
0.3km	0.270***	−2.514***	−0.019	−0.788***	0.360***	0.028***
	[0.227,0.312]	[-3.777,-1.251]	[-0.095,0.058]	[-1.280,-0.296]	[0.181,0.539]	[0.011,0.044]
	158,384/676/293	158,384/676/293	158,384/676/293	158,384/676/293	149,965/598/259	149,965/598/259
0.4km	0.272***	−2.607***	−0.014	−0.823***	0.369***	0.029***
	[0.231,0.313]	[-3.900,-1.314]	[-0.090,0.062]	[-1.310,-0.335]	[0.204,0.534]	[0.014,0.044]
	212,819/708/295	212,819/708/295	212,819/708/295	212,819/708/295	208,260/629/264	208,260/629/264
0.5km	0.273***	−2.576***	−0.009	−0.851***	0.363***	0.029***
	[0.232,0.314]	[-3.875,-1.277]	[-0.084,0.067]	[-1.330,-0.373]	[0.196,0.529]	[0.013,0.044]
	266,765/722/297	266,765/722/297	266,765/722/297	266,765/722/297	266,388/647/268	266,388/647/268
0.6km	0.270***	−2.528***	−0.005	−0.845***	0.365***	0.030***
	[0.230,0.311]	[-3.799,-1.257]	[-0.080,0.071]	[-1.317,-0.373]	[0.196,0.534]	[0.014,0.045]
	319,649/737/297	319,649/737/297	319,649/737/297	319,649/737/297	323,052/661/269	323,052/661/269
0.7km	0.268***	−2.437***	−0.001	−0.826***	0.358***	0.030***
	[0.228,0.308]	[-3.681,-1.192]	[-0.077,0.075]	[-1.298,-0.354]	[0.191,0.526]	[0.015,0.045]
	372,480/751/298	372,480/751/298	372,480/751/298	372,480/751/298	379,757/680/270	379,757/680/270
0.8km	0.266***	−2.310***	0.000	−0.788***	0.345***	0.029***
	[0.227,0.306]	[-3.535,-1.085]	[-0.075,0.076]	[-1.263,-0.314]	[0.181,0.510]	[0.015,0.044]
	424,921/759/299	424,921/759/299	424,921/759/299	424,921/759/299	436,222/692/270	436,222/692/270
0.9km	0.264***	−2.153***	0.002	−0.745***	0.342***	0.029***
	[0.225,0.303]	[-3.370,-0.936]	[-0.074,0.077]	[-1.221,-0.270]	[0.180,0.504]	[0.015,0.043]
	476,181/767/299	476,181/767/299	476,181/767/299	476,181/767/299	491,889/701/270	491,889/701/270
1km	0.263***	−2.097***	0.004	−0.708***	0.336***	0.029***
	[0.224,0.302]	[-3.309,-0.885]	[-0.072,0.079]	[-1.184,-0.232]	[0.175,0.497]	[0.015,0.042]
	525,329/772/299	525,329/772/299	525,329/772/299	525,329/772/299	545,086/706/270	545,086/706/270

continues

continued

	Log(Lot Size)	Year Built	Log(Building Size)	Year Last Sold	Build. Height	Log(Build. Height)
	(1)	(2)	(3)	(4)	(5)	(6)
1.1km	0.262*** [0.223,0.302] 572,867/778/299	-2.111*** [-3.320,-0.902] 572,867/778/299	0.006 [-0.070,0.082] 572,867/778/299	-0.674*** [-1.149,-0.200] 572,867/778/299	0.331*** [0.170,0.492] 597,265/714/270	0.029*** [0.015,0.042] 597,265/714/270
1.2km	0.262*** [0.223,0.301] 618,422/783/300	-2.127*** [-3.336,-0.918] 618,422/783/300	0.008 [-0.068,0.084] 618,422/783/300	-0.651*** [-1.124,-0.179] 618,422/783/300	0.323*** [0.164,0.482] 647,031/719/270	0.028*** [0.015,0.042] 647,031/719/270
1.3km	0.262*** [0.223,0.301] 661,056/786/300	-2.124*** [-3.337,-0.910] 661,056/786/300	0.010 [-0.066,0.086] 661,056/786/300	-0.635*** [-1.104,-0.165] 661,056/786/300	0.313*** [0.156,0.470] 693,683/723/271	0.028*** [0.014,0.041] 693,683/723/271
1.4km	0.262*** [0.223,0.301] 702,007/789/301	-2.139*** [-3.364,-0.915] 702,007/789/301	0.011 [-0.065,0.087] 702,007/789/301	-0.623*** [-1.090,-0.157] 702,007/789/301	0.307*** [0.152,0.461] 738,117/724/272	0.027*** [0.014,0.040] 738,117/724/272
1.5km	0.262*** [0.223,0.301] 740,910/793/301	-2.163*** [-3.399,-0.927] 740,910/793/301	0.012 [-0.063,0.088] 740,910/793/301	-0.617*** [-1.082,-0.153] 740,910/793/301	0.302*** [0.151,0.454] 781,053/726/272	0.027*** [0.014,0.040] 781,053/726/272
1.6km	0.263*** [0.223,0.302] 777,388/795/301	-2.161*** [-3.408,-0.914] 777,388/795/301	0.013 [-0.062,0.089] 777,388/795/301	-0.613*** [-1.076,-0.150] 777,388/795/301	0.301*** [0.151,0.451] 821,162/728/272	0.027*** [0.014,0.040] 821,162/728/272
1.7km	0.263*** [0.224,0.303] 811,506/799/301	-2.123*** [-3.376,-0.869] 811,506/799/301	0.014 [-0.061,0.089] 811,506/799/301	-0.611*** [-1.073,-0.149] 811,506/799/301	0.299*** [0.150,0.448] 858,107/734/272	0.027*** [0.014,0.040] 858,107/734/272
1.8km	0.264*** [0.224,0.305] 843,251/800/301	-2.075*** [-3.334,-0.816] 843,251/800/301	0.014 [-0.061,0.089] 843,251/800/301	-0.612*** [-1.073,-0.152] 843,251/800/301	0.295*** [0.147,0.443] 892,419/736/272	0.027*** [0.014,0.040] 892,419/736/272
1.9km	0.265*** [0.225,0.306] 872,953/801/301	-2.037*** [-3.303,-0.770] 872,953/801/301	0.014 [-0.060,0.088] 872,953/801/301	-0.613*** [-1.073,-0.153] 872,953/801/301	0.291*** [0.143,0.439] 924,831/738/272	0.026*** [0.014,0.039] 924,831/738/272
2km	0.267*** [0.225,0.308] 900,484/801/301	-1.993*** [-3.267,-0.720] 900,484/801/301	0.014 [-0.060,0.088] 900,484/801/301	-0.614*** [-1.074,-0.154] 900,484/801/301	0.288*** [0.140,0.436] 955,040/738/272	0.026*** [0.014,0.039] 955,040/738/272

Notes: Each cell is a separate regression of the outcome variable (column name) on the NALPZ variable. The respective bandwidth is indicated in each row. 95% confidence intervals indicated in brackets. Last row in each cell indicates the sample size (number of single-family tax parcels), the number of town borders (segments), and number of towns included in the regression, respectively.

Table D.2: Spatial RDD: Robustness and Specification Checks

	Log(Total Value)	Log(Lot Size)	Year Built	Log(Building Size)	Year Last Sold	Log(Last Sale Price)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	0.049** [0.007,0.092] 56,196/552/278	0.267*** [0.217,0.317] 56,196/552/278	-1.950*** [-3.248,-0.651] 56,196/552/278	-0.029 [-0.103,0.045] 56,196/552/278	-0.776*** [-1.317,-0.235] 56,196/552/278	0.059*** [0.020,0.098] 32,857/476/252
No Weigthing	0.052** [0.007,0.097] 56,196/552/278	0.269*** [0.222,0.316] 56,196/552/278	-2.236*** [-3.457,-1.016] 56,196/552/278	-0.026 [-0.102,0.050] 56,196/552/278	-0.793*** [-1.325,-0.261] 56,196/552/278	0.055*** [0.015,0.095] 32,857/476/252
Quad. Geography	0.052** [0.001,0.103] 56,196/552/278	0.272*** [0.226,0.318] 56,196/552/278	-1.526** [-2.987,-0.064] 56,196/552/278	0.005 [-0.081,0.091] 56,196/552/278	-0.881*** [-1.435,-0.327] 56,196/552/278	0.054** [0.013,0.095] 32,857/476/252
School Dist. Rank	0.044 [-0.009,0.097] 47,395/479/214	0.242*** [0.174,0.309] 47,395/479/214	-1.996** [-3.621,-0.371] 47,395/479/214	-0.087** [-0.158,-0.015] 47,395/479/214	-0.394 [-1.118,0.331] 47,395/479/214	0.055*** [0.020,0.091] 28,099/419/204
School Quality	0.016 [-0.039,0.070] 55,239/544/269	0.197*** [0.136,0.259] 55,239/544/269	-2.070** [-3.650,-0.491] 55,239/544/269	-0.074* [-0.152,0.003] 55,239/544/269	-0.400 [-1.228,0.429] 55,239/544/269	0.027 [-0.018,0.072] 32,300/473/247
Rank & Quality	0.030 [-0.034,0.093] 46,692/477/211	0.167*** [0.096,0.239] 46,692/477/211	-1.433 [-3.242,0.376] 46,692/477/211	-0.088** [-0.165,-0.011] 46,692/477/211	0.050 [-0.845,0.945] 46,692/477/211	0.038 [-0.010,0.087] 27,640/418/201
Prop. Tax	0.049** [0.006,0.092] 56,196/552/278	0.271*** [0.222,0.321] 56,196/552/278	-1.967*** [-3.271,-0.663] 56,196/552/278	-0.029 [-0.103,0.046] 56,196/552/278	-0.791*** [-1.321,-0.261] 56,196/552/278	0.060*** [0.020,0.099] 32,857/476/252
All Controls	0.028 [-0.036,0.091] 46,692/477/211	0.176*** [0.106,0.247] 46,692/477/211	-1.691* [-3.508,0.126] 46,692/477/211	-0.085** [-0.161,-0.009] 46,692/477/211	0.050 [-0.825,0.924] 46,692/477/211	0.040 [-0.008,0.089] 27,640/418/201
School Distr. FE	-0.010 [-0.096,0.077] 56,196/552/278	0.122 [-0.066,0.311] 56,196/552/278	-2.800 [-6.663,1.063] 56,196/552/278	-0.047 [-0.166,0.073] 56,196/552/278	0.272 [-1.346,1.891] 56,196/552/278	0.011 [-0.097,0.119] 32,857/476/252

continues

continued

	<u>Log(Total Value)</u>	<u>Log(Lot Size)</u>	<u>Year Built</u>	<u>Log(Building Size)</u>	<u>Year Last Sold</u>	<u>Log>Last Sale Price)</u>
	(1)	(2)	(3)	(4)	(5)	(6)
New Devel.	0.043*	0.292***	−4.309***	−0.079*	−0.687*	0.078***
	[−0.002,0.088]	[0.208,0.376]	[−5.666,−2.951]	[−0.166,0.008]	[−1.426,0.052]	[0.041,0.115]
	21,093/440/247	21,093/440/247	21,093/440/247	21,093/440/247	21,093/440/247	12,264/329/208

Notes: Each cell is a separate regression of the outcome variable (column name) on the *NALPZ* variable (except for the last two rows, where the outcome is regressed on the variable in the first column). The type of robustness check is named in the respective row of the first column. 95% confidence intervals indicated in brackets. Last row in each cell indicates the sample size (number of single-family tax parcels), the number of town borders (segments), and number of towns included in the regression, respectively.

Appendix E Spatial General Equilibrium Model and Amenities

Model Setup

A city³¹ embedded in a larger economy is assumed to comprise of a set of discrete locations ($\mathbb{S} = \{1, \dots, S\}$), which differ in terms of housing supply, local amenities, and access to workplaces. A Worker (o) chooses a residence-workplace pair (i, j) that maximizes their utility. They derive utility from consuming a freely-traded numeraire good (c_o), housing (h_i), residential amenities (B_i), and dis-utility from commuting ($e^{\kappa\tau_{ij}}$) that depends on travel time (τ_{ij}). Workers are heterogeneous with respect to resident-workplace pairs, which is captured by z_{ijo} . Utility takes the Cobb-Douglas form:

$$U_{ijo} = \frac{B_i z_{ijo}}{e^{\kappa\tau_{ij}}} \left(\frac{c_o}{\beta} \right)^\beta \left(\frac{h_i}{1 - \beta} \right)^{1-\beta} \quad (\text{E.1})$$

where β governs the share of income on the numeraire good. As is standard, the idiosyncratic shocks, z_{ijo} , are assumed to be Fréchet distributed:

$$F(z_{ijo}) = e^{-T_i z_{ijo}^{-\epsilon}} \quad (\text{E.2})$$

where $T_i > 0$ determines the average utility derived from living in block i , and $\epsilon > 1$ governs the dispersion of the shock. Maximizing utility given workplace wage (w_j), housing costs (q_i), results in the following form for indirect utility:

$$V_{ijo} = \frac{z_{ijo} B_i w_j q_i^{1-\beta}}{e^{\kappa\tau_{ij}}} \quad (\text{E.3})$$

The properties of the Fréchet distribution imply that the probability that a worker lives in block i is given as:

$$\pi_i = \frac{\sum_{j=1}^S T_i (e^{\kappa\tau_{ij}} q_i^{1-\beta})^{-\epsilon} (B_i w_j)^\epsilon}{\underbrace{\sum_{r=1}^S \sum_{s=1}^S T_r (e^{\kappa\tau_{rs}} q_r^{1-\beta})^{-\epsilon} (B_r w_s)^\epsilon}_{\Phi}} = \frac{T_i q_i^{(1-\beta)-\epsilon} B_i^\epsilon \text{CMA}_i^\epsilon}{\Phi} \quad (\text{E.4})$$

³¹Here I consider the “city” to be Massachusetts. As there is only one Combined Statistical Area as defined by the US Census Bureau, this assumption is not unwarranted.

where

$$\text{CMA}_i = \sum_{j=1}^S (w_j / e^{\kappa \tau_{ij}})^\epsilon \quad (\text{E.5})$$

denotes the Commuting Market Access of block i . Intuitively, this term is higher when a block is located close to well paying jobs.

The population is assumed to have full mobility, implying that residents will move until expected utility is equalized across residence-workplace pairs, as well as to the reservation level of utility in the larger economy (\bar{U}):

$$\bar{U} = \mathbb{E}[U] = \underbrace{\Gamma\left(\frac{\epsilon-1}{\epsilon}\right)}_{\gamma} \underbrace{\left[\sum_{r=1}^S \sum_{s=1}^S T_r (e^{\kappa \tau_{rs}} q_r^{1-\beta})^{-\epsilon} (B_r w_s)^\epsilon \right]}_{\Phi}^{1/\epsilon} \quad (\text{E.6})$$

where $\gamma = \Gamma(\frac{\epsilon-1}{\epsilon})$ is the Gamma function. Given the residential choice probabilities (Eq. E.4) and population mobility (Eq. E.6) we arrive at the following relationship:

$$\frac{B_i T_i^{1/\epsilon}}{\bar{U}/\gamma} = \left(\frac{H_i}{H}\right)^{1/\epsilon} \frac{q_i^{1-\beta}}{\text{CMA}_i} \quad (\text{E.7})$$

where H_i is the population of block i and H is the population of the city. We can remove the block-invariant components by dividing the equation by its geometric mean. This makes it possible to write block-level amenities in terms of observable characteristics and model parameters:

$$\frac{\widetilde{B_i^*}}{\widetilde{B_i^*}} = \left(\frac{\widetilde{H_i}}{\widetilde{H_i}}\right)^{1/\epsilon} \left(\frac{\widetilde{q_i}}{\widetilde{q_i}}\right)^{1-\beta} \left(\frac{\widetilde{\text{CMA}_i}}{\widetilde{\text{CMA}_i}}\right)^{-1/\epsilon} \quad (\text{E.8})$$

where $B_i^* = B_i T_i^{1/\epsilon}$ is the composite residential amenity term and $\tilde{X} = \left(\prod_i^N X_i\right)^{1/N}$ denotes the geometric mean of the respective variable over all residential locations. I take parameter values of ϵ , β , and κ from recent literature (described more in the main text). Then I calculate the normalized, composite residential amenity term. I use this in my regressions to test differences in amenity values across municipal borders.

Parameter Values

The parameter values used to estimate the implied amenity values are given in the following table:

Table E.1: Model Calibration: Parameters

Source	ϵ	$\nu = \epsilon\kappa$	$1 - \beta$	κ
Ahlfeldt et al. (2015)	6.6190	0.0951	0.25	0.01537
Tsivanidis (2018) [Low-skilled]	2.840	0.0336	0.24	0.012
Tsivanidis (2018) [High-skilled]	2.054	0.0242	0.24	0.012
Heblich et al. (2020)	5.25	0.05203	0.25	0.0099