The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement

Alexander W. Bartik, Arpit Gupta, and Daniel Milo*

March 20, 2025

Abstract

We present a novel method called "generative regulatory measurement" that uses Large Language Models (LLMs) to interpret administrative documents. We demonstrate its effectiveness in analyzing municipal zoning codes, achieving 96% accuracy in binary classification tasks and a 0.87 correlation for continuous questions. Applying this approach to a comprehensive sample of U.S. zoning regulations, we establish four facts about American zoning: (1) Housing regulations are multidimensional and can be clustered into two main principal components. (2) The first of which corresponds to *value capture*, indicating how municipalities extract economic benefits in areas of high housing demand. (3) The second principal component associates with *exclusionary zoning*, resulting in higher housing costs and socioeconomic exclusion. (4) Zoning follows a monocentric pattern with regional variations, with suburban regulations particularly strict in the Northeast. We develop a model of non-cooperative municipal government regulatory choice consistent with these facts.

JEL-Classification: R52, R58, K11, O38, R31, C81

Keywords: housing regulation, zoning codes, large language models, natural language processing, artificial intelligence, municipal ordinances, retrieval augmented generation

---

*Bartik: Department of Economics, University of Illinois at Urbana-Champaign, abartik@illinois.edu. Gupta: New York University, Stern School of Business, ag5808@stern.nyu.edu. Milo: New York University, Stern School of Business, dm4766@stern.nyu.edu This paper has benefited from conversations with David Albouy, Elliott Ash, Leah Brooks, Sophie Calder-Wang, Tom Cui, Nestor Davidson, Tatyana Deryugina, Jonathan Dingel, Giles Duranton, Ingrid Gould Ellen, Chris Elmendorf, Jason Faberman, Edward Glaeser, Joshua Gottlieb, Joseph Gyourko (discussant), John Horton, Greg Howard, Ben Hyman, Jeffrey Lin, Stijn Van Nieuwerburgh, Theresa Kuchler, Adam Osman, Franklin Qian, Julian Reif, Daniel Rock, Albert Saiz, Gregor Schubert, Allison Shertzer, Jaehee Song (discussant), Lena Song, Johannes Stroebel, Bryan Stuart, and comments from conference and seminar participants at NBER Summer Institute (Real Estate/Urban), NYU Stern Corporate Governance luncheon, NYU Law School, the Online Seminar in Economics and Data Science, Wharton real estate, UNC finance, AREUEA National Conference, the Dallas Fed, CREED Conference, UIUC economics, the Metric of Regulation, the New York Fed, University of Wisconsin Real Estate, the Chicago Fed, AirBnB, and the UEA. Preeyonuj Boruah, David Dai, Amber Gao, Grace Getman, Drew Harrington, Karin Hobelsberger, Nagharjun Mariappan, Scott Overbey, Alok Ranjan, Saul Richardson, Ritul Soni, Luofeng Zhou, and Kevin Zhou provided excellent research assistance. We gratefully acknowledge funding from Emergent Ventures and the NYU Stern Center for Global Economy and Business, and thank the Pioneer Institute for providing data. This work was supported in part through the NYU IT High Performance Computing resources, services, and staff expertise. Any errors or omissions are the responsibility of the authors. A replication package for the project can be found at: `https://github.com/dmilo75/ai-zoning`.

1 Introduction

Housing regulations govern the built environment of American cities, dictating the form and extent of residential development and housing affordability (Glaeser and Gyourko, 2018; Gyourko et al., 2008). The influence of zoning laws and land use policies extends far beyond the real estate market, with implications for social segregation, economic mobility, environmental sustainability, the growth of urban agglomerations, and construction sector productivity (Gyourko and Molloy, 2015; Kahn, 2000; Hsieh and Moretti, 2019; D'Amico et al., 2023). Despite their critical importance for such diverse outcomes, accurately measuring housing regulations remains a challenge due to the complexity and variation of municipal ordinances. These measurement gaps hinder our ability to understand the fundamental drivers of housing regulation and their broader impacts.

Our paper argues that advances in Large language Models (LLMs) enable scalable and accurate classification of regulatory documents, a task that we refer to as *generative regulatory measurement*. We obtain municipal codes for 63% of the population covered by local zoning ordinances, and develop an LLM-powered algorithm to estimate housing regulation on the full text of these documents. Our approach builds on the Retrieval Augmented Generation (RAG) architecture, adding prompt-engineering, prompt chaining, and detailed background information on zoning. We apply this methodology to a set of regulatory questions initially developed by the Pioneer Institute for Massachusetts (Glaeser and Ward, 2009), and benchmark our LLM-generated regulatory categorizations against human-coded measurements from this same study.

Our results indicate that LLMs have achieved near-human rates of precision in classifying regulation, with an accuracy rate of 96% for binary questions. LLMs also perform strongly on numerical questions with a correlation of 0.87 for continuous questions. We manually verify a subset of housing regulations in California to ensure that our results are not geographically biased. We also expand the original Pioneer classification by incorporating additional questions on housing process regulations, and verify high accuracy on this new set of measures. We use the resulting LLM-produced dataset on national housing regulations, along with other housing data, to establish four key facts about housing regulation and its impacts across the United States.

First, we find that housing regulation is multidimensional, and most variation cannot be captured by a single axis of housing stringency. This contrasts with prior analysis which has largely focused on a unidimensional distinction between regions with stricter or less strict land use regulation (Gyourko et al., 2021). We focus on the first two principal components of housing regulation, and find distinctive patterns of which municipalities adopt these regulations and the impacts they have on housing markets. The first principal component of our regulatory questions is associated with high prices and high construction, suggesting regulations that allow development but exact value in high-demand environments. In contrast, the second principal component associates with high prices but low construction, indicating regulations that restrict housing supply. Our ability to measure a more detailed, granular, and comprehensive set of regulations identifies distinct regulatory regimes that are associated with divergent housing market outcomes.

Second, we argue the first principal component can be interpreted as having a role of value capture in high-demand environments. This component loads heavily on regulations that allow local governments to extract and redistribute housing surplus, such as inclusionary zoning mandates that require developers to include affordable units. These regulations are typically found in densely populated, centrally located cities that tend to support Democratic political candidates. These regulations are higher in areas with higher amenities—as measured in terms of consumption (retail establishments), natural amenities (such as good weather or proximity to bodies of water), and productive spillovers (patents per capita and job density). The strong correlation with diverse amenity measures suggests these regulations emerge precisely in locations with the most valuable land and highest development potential. Rather than completely blocking all development, these regulations impose implicit taxes on new construction for the purposes of redistribution and funding local public goods.

Third, we connect the second principal component to exclusionary zoning practices aimed at limiting density and affordable housing options. This dimension relies heavily on bulk regulations, such as minimum lot size requirements, as well as procedural barriers to development. Among single-family zoned areas in municipalities across the country, 66% have town-wide minimum lot size requirements above 5,000 square feet, 17% of requirements are above 10,000 square feet, and 7% have requirements exceeding half an acre. These tools are particularly prevalent in affluent, predominantly white suburban areas that tend to lean Republican relative to other areas in their metropolitan area. We find a strong association between these regulatory measures and indicators of local school performance and social mobility, suggesting a role for educational sorting. Unlike value capture regulations that allow for development with conditions, these exclusionary practices directly left-truncate the housing distribution through regulations that effectively require households to consume a minimum quantity of housing. These regulations therefore intensify economic and racial segregation by effectively pricing out lower-income and minority households from neighborhoods with desirable public goods, particularly high-performing schools.

Fourth and finally, we find that housing regulation varies within metropolitan areas in ways that are broadly consistent with a monocentric city model (Alonso, 1964; Muth, 1971), while also highlighting significant deviations. As standard monocentric city models predict, denser building is generally allowed in city centers, with stricter bulk regulations and lower density requirements found in inner-ring suburbs. However, we also observe substantial sorting of high-income households into expensive suburbs with higher minimum size requirements. This is in contrast to the most basic models in which all households, regardless of income, are assumed to prefer central locations to minimize commuting costs. Our findings suggest a more complicated spatial pattern, with some peripheral suburban locations able to maintain high prices and sorting by affluent households in areas with stricter zoning. We find that this pattern is particularly pronounced in the Northeast, which has substantially more onerous bulk regulation requirements than other regions of the country. This pattern is largely driven by the fact that low-density areas around historic industrial centers in the Northeast (and to a lesser extent the Midwest) adopted disproportionately strict bulk regulations, generating greater persistence in urban form relative to the rest of the country. We suggest some possible drivers for this result, such as regulatory mitigations against pollutants and social exclusionary factors against workers in industrial hubs.

We develop a model of inter-municipal competition to interpret these facts. In the model, local governments strategically and non-cooperatively select between two sets of housing regulations: one associated with value capture and the other with exclusionary zoning through the imposition of minimum housing requirements. Our framework builds on a literature examining Tiebout sorting and local goods provision (Tiebout, 1956; Epple and Zelenitz, 1981), as well as traditional theories of zoning focused on housing sorting across public goods regimes (Fischel, 1987; Brueckner, 1995;

Hamilton, 1975), but generates distinct predictions about regulatory behavior that are consistent with our empirical results.

The model predicts that metropolitan cores with high amenities will implement value capture regulations. These regulations act as implicit taxes on housing development (similar to rent-seeking behavior in other municipal contexts as in Diamond (2017)), with the seized housing surplus redistributed to residents in the form of public goods provisioning. In contrast, the model predicts that suburban areas will adopt exclusionary regulations to generate more tax revenue from a concentrated pool of affluent households. Exclusionary zoning is less attractive to municipalities that are too large to be filled primarily with the most affluent households. This leads to a spatial equilibrium in which suburbs are home to higher-income residents, while poorer households concentrate in urban areas. The model thus generates clear spatial predictions about regulatory patterns: value capture in high-demand urban cores and exclusionary zoning in suburbs, matching the empirical patterns we document.

Our paper makes four primary contributions to the literature. First, our key methodological contribution lies in developing a general-purpose approach to measure the content of regulatory documents and quantify the accuracy and reliability of the resulting classifications. A long-standing literature has used text analysis to extract measures of sentiment from firm or policymaker communications (Romer and Romer, 2004; Tetlock, 2007; Hassan et al., 2019; Lopez-Lira and Tang, 2023). More recently, researchers have begun to use LLMs to collect more detailed data from literary, financial, regulatory, and legal documents (Dell, 2025; Giesecke, 2023; Lagakos et al., 2025; Jha et al., 2023; Yang, 2023; Bybee, 2023; Hansen and Kazinnik, 2023; Hoffman and Arbel, 2023). However, existing research on AI models emphasizes both their promise in analyzing textual data (Zhao et al., 2023), as well as challenges with undesirable AI features such as "hallucination" and manufactured model output (Azamfirei et al., 2023). We show that when coupled with careful text selection, prompting, and background information, that LLMs can be used to parse complicated regulatory documents with high fidelity.[1]

[1]Outside of our context of reading regulatory documents researchers have also highlighted the potential for using LLMs for a variety of purposes, including social science hypothesis generation (Horton, 2023). A growing literature also examines broader implications of Generative AI (Eisfeldt et al., 2023; Brynjolfsson et al., 2023), as well as the role of algorithms applied to real estate (Calder-Wang and Kim, 2023; Raymond, 2023).

In addition to high accuracy rates, our approach offers several other advantages for researchers. It provides unprecedented scalability at low cost: we successfully apply our regulatory classification measure across thousands of municipalities, a task that would be expensive and time-consuming for human analysts. This scalability opens up possibilities for comprehensive regulatory analysis across multiple domains. Our approach also ensures verifiability and auditability by prompting the LLM to provide specific supporting text from the regulatory documents, enabling independent verification of classifications. Finally, our approach is highly adaptable, allowing researchers to easily incorporate changes in regulatory interpretations or advancements in AI models, facilitating replication and refinement of measurements over time. The broad applicability of our approach extends to various domains where textual analysis is crucial, including building codes, tax regulations, legal cases, financial reports, newspapers, and other uses. This versatility is particularly valuable as the volume and complexity of regulations continue to increase (Singla, 2023).

Our second contribution is the production of a novel and comprehensive dataset on housing regulation across the United States. Our method improves on existing approaches in accuracy, granularity of regulatory measurement, and scope. Survey-based approaches to measuring zoning regulations nationally, such as the Wharton Regulatory Index (Gyourko et al., 2008, 2021; Huang and Tang, 2012), offer broad coverage on housing regulations, but are limited by low and potentially biased response rates, fixed questionnaires, and inaccuracy by respondents (Lewis and Marantz (2019)). Another approach imputes zoning regulations using either wedge-based approaches that measure the expected spatial macroeconomic distortions resulting from zoning (see Turner et al. (2014), Hsieh and Moretti (2019), Glaeser et al. (2005), Herkenhoff et al. (2018), Babalievsky et al. (2021), and Duranton and Puga (2019)) or use references to zoning regulations in court cases (Ganong and Shoag (2017)) or newspaper articles (Stacy et al. (2023)). These imputation approaches provide insights on the general importance of zoning regulations, but are more limited in estimating precisely which regulations matter or bind. Another branch of the literature instead conducts more detailed analysis of regulations at the local level, sacrificing generalizability for more granularity (see e.g. Quigley and Raphael (2005), Ihlanfeldt (2007), Glaeser and Ward (2009), Jackson (2016), Shanks (2021)).

Existing research therefore leaves significant gaps in our understanding of their measurement and impacts. National studies identify broad impacts of regulations on housing costs and construction but lack specificity on key drivers, while more detailed state-level analyses are geographically limited and may not be nationally representative. We show that challenges to the scaling more granular local approaches to the national level can be addressed through a combination of LLMs, RAG architecture, careful prompt-engineering and training, and background information. Our new approach provides both the comprehensiveness and granularity of the state-based approaches along with the scale of the national regulatory studies. Our approach provides detailed measures of zoning regulations for over 5,800 municipalities, more than twice as many municipalities as previous national studies such as the Wharton Regulatory Index, while providing as much granularity as local studies such as the Pioneer Institute study.[2]

Third, our key economic contribution is to exploit these rich data to provide a more nuanced portrait of zoning regulations than was previously possible, showing that zoning regulations are not well-summarized by one dimensional indices of stringency and that municipalities enact disparate regulatory regimes that vary dramatically across space depending on municipal circumstances. Previous research has often either focused on analysis of individual regulations (Gyourko and McCulloch, 2023; Cui, 2024) or has focused on uni-dimensional indices of regulatory stringency (Quigley and Raphael, 2005; Ihlanfeldt, 2007; Gyourko et al., 2008; Glaeser and Ward, 2009; Jackson, 2016).[3] Rather than relying on aggregate indices or indirect measures, we directly measure specific provisions, allowing us to distinguish between value capture tools in urban cores (like inclusionary zoning mandates) and exclusionary practices in suburbs (like minimum lot size requirements). The size and coverage of our data allows us to explore how regulations vary within municipality. Specifically, we show how these distinct regulatory approaches vary within metropolitan areas: high demand urban areas use value extraction regulations to extract some of the surplus generated by high amenities, while suburban regulations maintain exclusivity through density restrictions.

[2] Mleczko and Desmond (2023) use a non-LLM natural language processing (NLP) approach to measure a set of zoning regulations for most of the municipalities covered by the Wharton Regulatory Index (Gyourko et al., 2008), about 2600 in total, less than half the size of our sample. Their approach requires significantly more manual input than ours, reducing its scalability, and the accuracy level is unclear because they do not present results comparing their measures to a test dataset not used in training.

[3] One notable exception is Mayer and Somerville (2000) which classifies regulations based on whether they add costs, delays, or uncertainty development and then measure the extent to which these three indices of regulation have affected new construction using a panel dataset on construction for 44 metro areas using regulatory data from an early version of the Wharton Regulatory Index.

Fourth, our theory contribution is to develop a model of non-cooperative regulation choice by local governments who have access to multiple zoning instruments to explain the zoning patterns we observed in our data. This model shows that large, high amenity municipalities will be more likely to choose extractive zoning regulations, while small, outlying municipalities, with lower amenities will be more likely to choose exclusionary regulations. These predictions help unite disparate aspects of the theoretical literature on zoning, which has discussed how zoning regulations can be used to create efficient public goods provision and sorting (Tiebout, 1956; Fischel, 1987; Hamilton, 1975, 1976), lead to segregation (Rothstein, 2017; Cui, 2024), and function as distortionary taxes (Zodrow and Mieszkowski, 1986).

2 Construction of National Housing Regulatory Database

In this section, we describe the institutional background and data sources on municipal zoning codes, describe our approach to generative regulatory measurement in classifying these codes at scale, provide evidence on the accuracy of our approach, and discuss the scalability and replicability of our method.

2.1 Municipal Codes and Zoning: Institutional Background and Data

In the United States, local governments are "creatures of the state" subordinate to state control. Municipal corporations, which include cities, towns, villages, and other local government units, are authorized by state law to organize and function as local governing bodies. This concept largely overlaps with the Census definition of "incorporated place", which we use to organize our analysis.[4] In most states, one of the powers granted to municipalities by the state government is control over local zoning decisions; indeed, the desire to control local zoning is a primary motive for incorporation.

Zoning, broadly, consists of two key sets of regulations: land use regulations, which partition local land into distinct use classes, and bulk regulations, which regulate the physical dimensions and

[4] In several states the "township" form of government also has jurisdiction in zoning which aligns with the Census County Subdivision definition.

density of buildings in different land use classes. Examples of bulk regulations include: minimum lot sizes (specifying the smallest allowed area for a buildable lot), lot coverage requirements, front and side setbacks (mandating minimum distances buildings must be from property lines), height restrictions, and floor area ratio caps (regulating total floor area relative to lot size). Zoning codes also specify the process through which development is approved, such as whether development can be done by right and which governing bodies must approve developments. These regulations combine to regulate how densely areas can be developed, the size and spacing of buildings, the overall built form of neighborhoods, and the process through which construction is approved. Other mandates and requirements, such as parking minimums, further constrain both commercial and residential development in different areas.[5]

Municipalities enforce laws by issuing municipal codes which outline local regulation in different domains. Some regulations apply broadly to all land within a jurisdiction; other regulations (such as minimum lot sizes) typically vary depending on the specific use class and district (i.e., single-family zoning, commonly referred to as R-1, or commercial or industrial). These ordinances are typically updated over time to reflect changes in local regulations and often aggregated by different companies online.

We source the text of municipal codes from a variety of municipal aggregators, including American Legal Publishing, Municode, and `Ordinance.com`. These municipal aggregators post municipal codes and other information on local governments, often at the behest of the local government to fulfill public posting requirements. Table 1 illustrates the breadth of our sample coverage. In total, we cover 25% of all municipalities in the U.S. and 6% of all townships. This coverage is skewed to larger cities, and so of the 76% of the population in the U.S. that live in either a municipality or a township, we have relevant municipal documents for 63% of the population. Panel B shows our underlying sources for the ordinances in our sample. American Legal Publishing provides significant numbers of records in the Northeast and Midwest, Municode provides especially good coverage in the South and the Midwest, and `Ordinance.com` provides substantial coverage of the West and Northeast.[6]

[5] States and municipalities also enact building codes, which govern the building and safety standards that new construction needs to adhere to.

[6] When a municipality hosts its ordinance on multiple aggregators, we prioritize Ordinance.com, and then Muni-

We combine data on these municipal ordinances with a variety of municipal level demographic, housing, and economic information such as building permits data from the Census Building Permits Survey, rent and home price data from the American Community Survey (ACS), and other sources. We describe the data sources we use in greater detail in Table A1.

2.2 Large Language Models

Large Language Models (LLMs) are a form of artificial intelligence that primarily handle sequential data such as sequences of words in textual data. LLMs are based on the deep learning transformer architecture as introduced in Vaswani et al. (2017). The key innovation is the attention mechanism, enabling the model to focus on multiple words of the input text at once. This helps the model understand words in context, such as sentences or paragraphs. Transformers also represent a significant advancement in terms of both accuracy and runtime over previous models like Recurrent Neural Networks, which processed sequences linearly. LLMs are trained with semi-supervised learning, first pre-training the model on a large corpus of text and subsequently fine-tuning the model with human feedback. After training, LLMs can generate human-like text, answer questions, summarize text, and generalize from their training to perform tasks they were never explicitly trained for, a concept known as zero-shot learning. This means that the model does not require explicit examples of additional training to perform well in an out-of-sample exercise, a key advantage we use in our analysis.

LLMs have several advantages and disadvantages in our setting. The central advantage is scalability at low cost: we are able to load large quantities of municipal code data for classification and analysis, which exceeds the capacity of a typical human team to analyze at reasonable cost. Other advantages include the prospect for additional training, allowing for increased accuracy over time as LLMs improve in quality and additional training data is incorporated into the analysis. Another key benefit is creating a comprehensive and nationwide dataset resulting from the application of a uniform and standard set of criteria for analysis, rather than relying on a group of human analysts who may employ idiosyncratic legal interpretation.

Drawbacks in using LLMs for this purpose include potentially inaccurate measurement and the

code over American Legal Publishing.

need for manual sourcing of relevant documents. Inaccurate measurement stems from a failure to locate or interpret the relevant sections of legal code. LLMs can only process a limited amount of text at once, and so they require a reliable process to locate the most relevant parts of the ordinance for a given question. Additionally, legal interpretation requires many assumptions and nuances, and even though LLMs are likely exposed to legal interpretation in their training, they may need to be reprompted on them to ensure greater focus for the questions at hand. Even current state-of-the-art LLMs may inadvertently produce incorrect information, generate information with an incorrect degree of certitude, and potentially manufacture data output ("hallucination"). Possible biases in the responses are linked to the quality of training data, prompting, and multi-step processing steps, and so measurement error may or may not be classical depending on the explanatory variable of interest. Finally, relevant information to answer zoning regulation questions may be found in other legal documents. We provide the LLM with the municipal zoning ordinance for each municipality but not other potentially useful documents (i.e., state or county laws). We address these drawbacks by focusing on questions that can be answered with only a municipal zoning ordinance, developing a process to isolate relevant text along with background information about interpretation, and measuring performance against human-defined categorizations of regulation to quantify sources of errors.

Despite these drawbacks, several applications of LLMs for legal analysis suggest considerable promise. Private sector firms specializing in legal-focused LLM applications have achieved high valuations[7] and several local governments have found success using LLMs to aid residents in understanding laws and procedures.[8] Moreover, randomized controlled trials have shown that LLMs causally increase lawyer productivity (Schwarcz et al., 2025).

2.3 Processing Municipal Codes Using LLMs

In this subsection, we outline the general process we use to generate our housing regulatory dataset, a process we refer to as generative regulatory measurement. Figure 1 also illustrates the overall approach.

[7] Harvey AI, which uses LLMs to assist lawyers, reached a $3 billion valuation after a Series D funding round.

[8] For example, Williamsburg Virginia found that AI powered chatbots answered 79% of user queries without the need for human assistance.

The first step of our process is to download the sources of municipal codes listed in Table 1, which provides us with a large corpus of zoning documents relevant for our analysis. We collected these ordinances throughout Fall 2023, providing us a snapshot of zoning ordinances at that time. These municipal codes contain detailed housing and zoning regulations relevant for our study.[9] Any images of tables are transcribed using Amazon Textract.

The lengths of many zoning documents exceed the context windows usable by current LLMs[10] (see Appendix Figure A1 for a histogram of token length across our entire sample compared with the maximum token length for various LLMs). As a result, it is either impossible or cost-prohibitive to simply upload the entirety of municipal code documents into standard LLM services and ask our questions directly. Moreover, LLM performance significantly declines with context length, rendering the effective context length of LLMs well below the length of municipal codes (Modarressi et al., 2025). This is due to limitations in the attention mechanism, which struggles to retrieve relevant information as context length increases.

To address this challenge, our second step is to use a standard framework in computer science known as "retrieval-augmented generation" (RAG) (Lewis et al., 2020). The basic objective of this approach is to combine a large pre-trained language model with external information retrieval, in order to give the LLM the ability to "look up" information from a vast corpus of text during the generation process. We describe in subsection 2.6 the accuracy and cost implications of this choice.

In order to implement RAG, we first partition each ordinance into small chunks of text taking into account the hierarchical section structure of the ordinance.[11] Next, we map each chunk of text into a vector representation called an embedding. Embeddings are vector representations of text trained to minimize the distance between semantically similar content, allowing efficient comparison of text meaning (Reimers and Gurevych, 2019). We vectorize each subsection of the ordinance document using the OpenAI 'text-embedding-3-large" algorithm.

[9] Ordinances sourced from Municode and American Legal Publishing are general municipal ordinances covering topics ranging from permits to purchase a pet to local election processes. Some of these general municipal ordinances do not include a zoning section, instead referencing separate zoning ordinance. We filter out such ordinances which do not contain zoning information by searching for key phrases in zoning documents, like common table headers (i.e. "Table of Uses") or zoning district names (i.e., R-1 for the first residential zoning district)

[10] Several large cities exceed the maximum commercial model limit of 2 million tokens: for example, New York City (17.1 million), Detroit (4.3 million), and Atlanta (3.6 million).

[11] We download each section within an ordinance separately, and then further split or combine sibling sections so that their length ranges between 50 and 1,000 of text. One token is roughly 3/4 of a word.

The third step in our process is to prepare a set of questions that we want to ask of our sample. We begin with the question base already used by the Pioneer Institute (i.e., "Is multifamily zoning allowed in this area as-of-right?"). We add to these questions four additional questions about the process determining construction permitting and approvals. In the initial step of our process, we simply use the text of the questions verbatim as first posed by the Pioneer Institute (or the first draft of the additional process question). The text of the questions is also run through an embedding process to generate its own vector representation.

The fourth step of our process is to identify the most likely relevant information from the ordinance to show the LLM. We use cosine similarity, a standard measure of vector distance, to rank each text chunk by proximity to the question.

The fifth step is to refine the initial ranking of the most relevant text produced by cosine similarity and double-check for accuracy. We do so by using a cross-encoder reranking model on the top 50 chunks of text, which processes the question and section text pairs simultaneously to determine the most semantically similar sections.[12] Reranking has been shown to increase retrieval performance (Anthropic, 2024). We then select text to show the LLM in order of highest relevance until a threshold. We choose 4,000 tokens (~3,000 words) as the threshold, since LLM performance significantly degrades beyond this length (Modarressi et al., 2025). The final output of this step is the set of 4,000 tokens representing the text in the initial document that is relevant to answer each question.

The sixth step in the process is the LLM query itself. We provide two key pieces of information to the LLM through an API call. First, we include 4,000 tokens of relevant text. Second, we provide the zoning question. Both pieces of information are provided in a single call to the LLM, in order to produce model output which is our answer. Each answer consists of an open-ended argument followed by a parsable answer (i.e. "Yes" or "No"). The open-ended answer allows for humans to audit the reasoning path of the LLM and has been shown to increase performance by providing space for the LLM to think out loud (Zhang et al., 2022).

These six initial steps are sufficient to produce an answer to each regulatory question sourced

[12] We use the Cohere reranking model for this step. For some questions, when double-checking the answer we instead use keyword inclusion to rerank. See Appendix Section C for more details on which questions we do this for and which keywords we check for.

from municipal documents. However, they are not necessarily very accurate. Subsection 2.6 describes in more detail the accuracy for this specification and each iterative improvement. As a result, we followed a data-driven process to iterate and improve the accuracy of our approach. We used three distinct strategies which we describe in more detail below: prompt chaining, prompt engineering, and providing detailed question background information. To avoid overfitting, we conducted all such training on a distinct subsample of the Pioneer data, and then performed our final validity checks only once on a leave-out sample of the Pioneer data.

The first strategy is prompt chaining, which we used to produce additional information necessary to help answer each question. Prompt chaining breaks down LLM queries into multiple steps (prompts) where the output from one prompt is used as input in the next.[13] For instance, when asking about the largest frontage requirement for all single-family residential districts, we first ask the LLM to name all districts which allow single-family housing. We do this as a separate step because the relevant text defining allowable uses in a district and the text defining frontage requirements for districts are typically in different sections of the ordinance with distant embedding vectors. Additionally, LLM performance is enhanced when it is only required to answer a direct single step question in each call (Khot et al., 2023).

Another use of prompt chaining is through the post-processing of certain questions, which functions to double-check answers. For instance, the answer "Yes" to a question about whether townhouses/attached housing is allowed typically means the LLM has found affirmative evidence that such housing typologies are allowed, while an answer of "No" signifies either a lack of approval, or a lack of sufficient context for the LLM to answer the question. In such cases where an answer could indicate lack of information, we reprompt the LLM and directly use keywords like "townhouse" or "attached" to refine and rerank our search (instead of the reranking algorithm).

A second strategy of model improvement is to generate additional background information to provide to the LLM. The background information and model assumptions were initialized based on the Pioneer study (their "Issue Overview" and "Research Coding" sections for each question) when possible and were LLM generated otherwise. We manually refined this background information to address areas of misinterpretation by investigating cases in which regulations were misclassified.

[13] See Anthropic Prompt Chaining Guide for further details on prompt chaining, as well as Wu et al. (2022).

Appendix C contains full information on the original Pioneer questions, our rephrased questions, as well as the additional background information and assumptions provided.

The third dimension of improvements comes from prompt engineering. We include a "system prompt" which tells the LLM that it is a municipal zoning expert, details what the structure of the prompts for particular questions will be, and instructs the LLM to think "step by step" to induce chain of thought reasoning (Zhang et al., 2022) [14]. Additionally, we rephrase the questions from the original wording provided by the Pioneer Institute in order to produce a more simplified version which is easier for the LLM to parse. This primarily consists of breaking down compound questions. We did so through a data-driven approach in which we categorized model errors through more simple questions, and iterated on simplifications of questions to produce more accurate results. As mentioned above, our final validation estimates are free of any overfitting bias because they were estimated on a different set of municipalities than the ones which we used for training in this purpose.

2.4 Simple Example to Illustrate Approach

To help illustrate our process more concretely, we go through our general procedure for one question in our sample in Arlington, Massachusetts on the presence of inclusionary zoning mandates or incentives. In this case, the question used by the Pioneer Institute was sufficiently concise and clear that we did not need to modify it. This question reads:

> **Question:** Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

To answer this question in this municipality, we first partition the ordinance into chunks. We follow the hierarchical structure of the ordinance document when possible, which means that chunks are typically discrete sub-sections of the ordinance as seen in Appendix Figure A2 for Arlington. Second, we embed each chunk as well as the question. Panel A of Figure 2 visualizes a two-

[14] See Appendix Section C for exact system prompts for each type of question.

dimensional approximation of this embedding space. Each dot represents one chunk of text and is colored by its broader article in the ordinance. Reassuringly, we find that chunks within the same article cluster together, which suggests that embeddings are picking up on the underlying meaning of the text. The black star in the upper right quadrant represents the embedding for the question itself. RAG selects the chunks of text closest to the question in embedding space for the LLM.

Panel B of Figure 2 zooms in on the embedding space around the question and shades each dot by its cosine similarity with the question vector. Our approach first isolates the nearest chunks of text, and then reranks these results. The closest subsections to the inclusionary zoning question are indeed the relevant sections of the ordinance on affordable housing requirements:

> "Section 11.08 - Affordable Housing Requirements Part 1: ...1. Fifteen percent (15%) of the Residential Units in new Projects shall be Affordable Units...
>
> Section 11.08 - Affordable Housing Requirements Part 2: where more than 50% of the units are affordable to households ... the number of parking spaces may be reduced to 50% of the requirements."

Third, we provide these chunks of text, our system prompt described above, along with the question shown above, and additional pieces of background information and assumptions to whichever LLM model that we're using. In this case, the background information and assumptions given to the model are as follows:

> **Background Information:** Inclusionary zoning requires or encourages developers to include affordable dwelling units within new developments of market rate homes. Some municipalities call it "incentive zoning" ... (omitted for brevity)
>
> **Assumptions:** Do not include provisions for entirely affordable, subsidized housing development by public or non-profit corporations. Also do not include provisions under

"rate of development" headings that exempt affordable units from project phasing and growth caps.

By providing relevant text and guidance to the LLM, we receive a final classification along with references for the source of this information:

GPT-4 Turbo Output: The zoning bylaw/ordinance provided does indeed include mandates and incentives for the development of affordable units. Here's a breakdown of the evidence supporting this conclusion:

1. **Mandates for Affordable Housing Development**: The ordinance mandates that 15% of the residential units in new projects be affordable units (Section 11.08, Part_1, d. REQUIREMENTS, 1.). This requirement applies to all new residential projects, including phased or segmented developments, with six or more units subject to Environmental Design Review (Section 11.08, Part_1, b. APPLICATION). The affordable units must conform to specific affordability criteria, ensuring they are accessible to households earning no more than 60% of the median income for rental units and no more than 70% of the median income for homeownership units (Section 11.08, Part_1, c. DEFINITIONS)...*(rest of answer omitted for brevity)*

In this case, the model's output matches the Pioneer Institute classification of inclusionary zoning in Arlington, MA.

2.5 Model Validation with Pioneer Data

A critical step in assessing the performance of LLM-based approaches lies in comparing model-generated classifications against a ground truth benchmark. To do so requires a high-quality annotated reference dataset. The Pioneer dataset serves as an excellent starting point for our purposes, as previously mentioned, due to the expert classification of a large number of municipalities. The main drawback in using this dataset is the staleness of responses—with responses categorized as of

2004. Some regulations have changed in the intervening twenty years, and we have access only to the most recent zoning ordinances, not the ones that prevailed in that time period. Additionally, the Pioneer Institute relied on some outside information (i.e., directly contacting local regulatory bodies) in addition to municipal ordinance text.

To address these issues, we construct a testing dataset based on 30 randomly chosen municipalities from the Pioneer Institute dataset. Importantly, we made sure to leave out these municipalities from any prior training exercises to avoid overfitting. We also recode these municipalities to 1) exclude question responses which relied on outside context, and 2) hand-correct inaccuracies in the original classification.[15]

Table 2 shows the performance results of our baseline GPT-4 Turbo model against the testing sample in Massachusetts. Among continuous questions (Panel A), our generated answers have an average correlation of 0.87 with the ground truth of expert classifications, after winsorization of our results at the 1% level and corrections of errors in the Pioneer sample. This represents a quite high benchmark and also incorporates substantial heterogeneity. When asking about the number of zoning districts in the municipality, we obtain a correlation of 0.98. When asking about the lowest of residential min lot sizes (i.e., the lot size requirement for R-1 zoned single family homes, an important zoning question determining allowable density), we find a high 0.92 correlation. These results suggest we are able to reach high model performance when matching against continuous numerical outcomes.

We also find high model accuracy when measuring binary questions (i.e., those with a yes or no answer like whether "multi-family housing is allowed" which we measure perfectly across all municipalities). As shown in Panel B of Table 2, we observe a model accuracy of 96% across all binary questions. Because the raw accuracy measure may be biased depending on the base rate of answers, we also provide a Relative Squared Error (RSE) that compares each model result to a naive model which guesses the sample mode. We observe quite small RSEs as well.

[15]Due to the time-intensive nature of the expert correction step, we only check responses in which our LLM approach disagrees with the Pioneer Institute classification. This means that we potentially overstate model accuracy in cases in which the LLM agrees with the Pioneer Institute original classification; but that original classification was wrong.

2.6 Sources of Model Improvements

In this section, we provide a detailed decomposition of where the accuracy gains come from when using our approach, both overall and question-by-question. We do so to provide additional intuition for which aspects of our procedure are most important in improving final accuracy. Additionally, we also demonstrate robustness regarding important features such as the choice of LLM model. We run each model five times to calculate the mean and standard deviation of accuracy for each specification. The results of this investigation are shown in Table 3. The first two columns report the mean and standard deviation of our accuracy measures, while the third and fourth columns report the mean and standard deviation of the "I don't know" rate. Panel A reports results for continuous questions where accuracy is measured using the correlation with the Pioneer answers while Panel B reports results for binary questions where accuracy is measured using the share of classifications that match Pioneer. The different sub-panels report results exploring different aspects of our procedure, with the first sub-panel exploring the role of different RAG strategies, the second sub-panel exploring the roles of different prompting strategies, and the final sub-panel exploring the effects of different choices of LLMs.

The first basic choice we face is whether to use RAG models at all or not. Over our entire national sample, many municipal codes are simply too long for even the LLM models with the largest context windows. For these municipalities, RAG models are necessary. Because the municipalities that make up our core validation sample in Massachusets are shorter on average, we are able to compare the accuracy of specifications with and without RAG in Table 3. Here, we use Gemini Flash 1.5, a long context window model.[16] We test three specifications: Full RAG, Basic RAG, and No-RAG. Full RAG follows our main specification, while basic RAG uses a smaller, less powerful, embedding model[17] and does not use reranking. No-RAG provides the entire ordinance to the LLM as context. For each specification we only provide the question and ordinance text to the LLM, omitting any of our refinements like prompt chaining or background information.

Ex-ante it is unclear whether using RAG will improve performance. Because RAG only provides

[16] Gemini Flash 1.5 has a context window of one million tokens, or roughly 750k words. The primary model we use in our analysis, GPT-4 Turbo only has a context window of 128k tokens or 96k words, which is too short even for many of the relatively shorter Massachusetts codes we use for validation.

[17] Specifically we use `multi-qa-mpnet-base-dot-v1`.

a subset of the overall ordinance to the LLM, it may omit some critical information. On the other hand, without RAG the LLM may struggle to interpret the critical sections of text correctly as performance degrades with input size (Modarressi et al., 2025). In our environment, we find that RAG approaches demonstrate improved overall performance over No-RAG models for both continuous questions (Table 3 Panel A) and binary questions (Panel B). For continuous questions, we observe a dramatic reduction in "I don't know" responses when implementing RAG (dropping from 38.7% with No-RAG to 12.9% with full RAG, with particularly strong improvements from basic to full RAG). Continuous questions typically involve information that is localized to specific sections of the ordinance, such as in tables. In our setting, therefore, we conclude that refinements to RAG, such as re-ranking and the size of the embedding model, meaningfully improve retrieval (Anthropic, 2024); and that the benefits from smaller context windows outweigh the costs of possible information omission. These findings, combined with RAG's feasibility benefits and cost advantages, justify our choice to use RAG for the remainder of the paper.

Next, we examine the impact of iteratively layering the prompting strategies discussed in subsection 2.3. For this analysis, we switch to our primary model, GPT-4 Turbo, and use full RAG throughout. We start with a "no prompting" specification that simply feeds the questions as originally phrased by the Pioneer Institute along with the context, achieving a correlation of 0.72 for continuous questions and an accuracy rate of 83.5% for binary questions. We then layer in additional prompting strategies: first adding prompt engineering, then incorporating background information, and finally implementing prompt chaining. For binary questions, incorporating background information yields the largest improvement (from 86.0% to 93.9%), while for continuous questions, prompt chaining provides the biggest boost (increasing correlation from 0.73 to 0.87). Our main specification, which uses prompt chaining, reaches 0.87 correlation for continuous questions and 96.7% accuracy for binary questions.

In Table A2, we disaggregate these results to examine how different prompting strategies affect each question. The impact varies substantially across questions. Some questions, like those about flexible zoning, achieve high accuracy even without prompting. Others show marked improvements from specific strategies—wetland restrictions in lot size calculations sees substantial gains from prompt engineering alone, while questions about affordable housing incentives benefit from both rephrasing and background information. For questions like longest frontage requirements, where relevant information often appears in semantically distant parts of the ordinance, prompt chaining yields the largest improvements (correlation increasing from 0.46 to 0.70). We also contrast the baseline approach with estimates drawn from other models in the model selection panel of Table 3. GPT-4 Turbo, using RAG and our full set of prompting, has the highest correlation for continuous questions (0.87) as well as the highest accuracy for binary questions (96.7%). GPT 3.5 Turbo is similar in accuracy for continuous questions, though considerably worse for binary questions, while Gemini Flash 1.5 is worse for both binary questions and continuous ones.

The key takeaway from our approach towards generative regulatory parsing is that, at least with models available at the time of writing, model accuracy improves substantially above simple "zero shot learning" examples given additional human input. We provide substantial human input in the areas of prompt engineering and providing background information, which helps to direct the LLM on the relevant focus of the text. Additionally, we design a multi-step reasoning chain for each question to simplify the tasks required by the LLM in each sub-step. Such additional human processing is likely necessary in other contexts as well, at least until further advances in LLMs are made.

2.7 Understanding Model Errors

To better diagnose reasons for model error in our baseline approach, in Figure 3 we provide a complete decomposition of all of the reasons for disagreement between GPT-4 Turbo and the original Pioneer Study on binary questions in our testing sample. We categorize disagreements into whether the Pioneer study was itself outdated or inaccurate, the LLM was incorrect, or the answer is ambiguous. While ideally municipal regulations would identify a clear and unambiguous answer, we observe differences even among legal experts hired for the task of hand-classifying regulations. In principle, the ambiguous or unclear aspects of regulations can also be systematically classified through LLM-based approaches. Finally, we further divide cases the LLM was incorrect into those where it missed the relevant text chunks and those where it misinterpreted the context.

Largely, answers from the Pioneer Institute that our model did not match were due to changes in the underlying ordinance since the Pioneer Institute study roughly 20 years ago. LLMs missed the relevant text chunks in two cases, while in four cases the answer itself was ambiguous. The most important category for our purposes is cases in which the LLM misinterpreted the context—this happens in nine cases, most often with respect to whether townhouses are allowed and with permit caps or phasing. Six questions do not exhibit this type of error at all. When considered over a large sample, these results appear promising in suggesting that errors are typically quite rare.

Importantly, the errors also appear balanced across false positives and false negatives. Appendix Table A3 provides a confusion matrix comparing our baseline GPT-4 Turbo model against the Pioneer classifications, separating true positives, false positives, true negatives, and false negatives. Our errors are equally represented among false positives as well as false negatives (six each), suggesting no obvious bias in our classification.

2.8 Additional Validation Checks Beyond the Pioneer Sample

Additional Hand Validation: To assess accuracy outside the Pioneer study Massachusetts sample, we conducted additional manual validation. First, we reviewed four process questions on a nationwide random sample of 30 municipalities (Panel A of Table A4). This component also enables us to augment the initial Pioneer set of questions to incorporate additional regulatory questions related to housing process. After removing ambiguous cases, the model achieved accuracy rates ranging from 89% to 100% across these questions.

Second, we assessed accuracy on bulk zoning regulations by randomly selecting 30 municipalities in California (Panel B of Table A4). After dropping ambiguous cases—such as zoning rules that varied between interior and corner lots—the model achieved 89% accuracy for both measures. These results suggest that the model generalizes well beyond the Massachusetts-based training data, performing reliably across diverse regulatory contexts.

Comparison Against WRLURI: We next compare our analysis to another commonly used dataset of national housing regulation: the WRLURI from Gyourko et al. (2021); we describe this contrast in more detail in Appendix D. The key takeaways from this investigation are that our approach improves on WRLURI in both coverage and accuracy. Additionally, our indices consist of only hard regulations, while WRLURI loads heavily on housing market outcomes.

A positive, though not perfect, correspondence between generated data and surveys is consistent with prior literature (Lewis and Marantz, 2019) investigating the reliability of survey-based responses on land use regulation. This paper finds planners have an incomplete understanding of their own municipality's regulation, and survey responses across years are inconsistent. Additionally, our data rely on zoning codes from 2023 whereas (Gyourko et al., 2021) uses survey questions from 2018 and our wording on overlapping questions differs in some cases. Consequently, some of the divergence likely also reflects changes in the underlying zoning codes over time or differences in survey wording.

2.9 Scalability and Replicability of LLM Regulatory Analysis

While LLMs hold promise for regulatory analysis based on accuracy, their key advantage lies in scalability and cost-effectiveness, which enable substantially larger systematic analysis of unstructured textual datasets.

Appendix Figure A3 illustrates the cost comparison between human analysis and two LLM-based approaches: one using RAG and another without RAG. Our analysis assumes several key parameters: lawyers require an average of five minutes per question-municipality pair at a rate of $50 per hour. For the RAG approach, each LLM API call involves 4,000 tokens of text at $0.03 per query using GPT-4 Turbo pricing. For the No-RAG approach, each API call requires processing 330,000 tokens, reflecting the average length of ordinances in our sample. We analyze costs across 20 questions per municipality. Initial setup costs include $390 per question for model training and $2.80 per municipality for ordinance retrieval and cleaning.

The RAG-based approach becomes substantially more cost-effective than the human-based analysis at larger scales, while the No-RAG approach exhibits a cost structure more similar to human analysis. Human-based analysis shows a linear cost increase as the number of municipalities grows. The RAG approach has a higher setup cost, but demonstrates significantly better scalability. The cost curves for RAG and human analysis intersect at approximately 300 municipalities, beyond which the RAG method becomes increasingly more cost-effective. While both LLM-based approaches require similar setup costs, the No-RAG method's heavy token usage results in steeper cost scaling, making it less advantageous for large-scale analysis compared to the RAG approach.

Moreover, the LLM approach offers additional benefits not captured in pure cost comparisons. These include faster processing times (the entire dataset for this study can be generated in under three days), consistent application of criteria across all documents (while human research assistants might vary in their interpretation of law), and the ability to easily update analyses as regulations change.

Another important consideration in adopting LLM-based approaches for regulatory analysis is the replicability of results. Here we summarize our key findings on LLM replicability, with complete details provided in Appendix Section B. The key issue is that frontier LLMs are based on a Mixture-of-Experts architecture that does not provide fully deterministic responses. Using a sample of 30 municipalities and two questions (one binary question about permit caps and one continuous question about the number of districts), we experiment with various approaches to improve replicability, including ensemble methods (requesting multiple responses and aggregating by majority rule) and varying LLM parameters designed to reduce random variation.

We produce multiple iterations of each model specification to assess its variance across runs. Specifically, we use the pairwise matching rate across runs as our preferred measure of replicability. There are two ways to request multiple iterations of output from LLMs: requesting multiple parallel responses per LLM generation, or requesting the same LLM generation many times. This leads to two consistency metrics: internal consistency (agreement within a single LLM generation between its parallel responses) and external consistency (agreement between majority rule answers across separate LLM generations). External consistency is our primary measure of interest as it captures replicability across independent runs of the model.

A key choice under the control of the researcher is the temperature, which is a hyperparameter controlling the randomness of the model's output. A lower temperature (closer to 0) makes the model's responses more deterministic by picking the most probable next token, while a higher temperature (closer to 1) increases randomness. Not surprisingly, we find that lower temperature settings generally lead to more consistent responses. With an ensemble size of 10, at temperature 0, we observe an internal consistency score of 90% for the continuous question about number of districts, compared to an internal consistency of 60% at temperature 1, with similar patterns for external consistency. LLMs also allow users to set random seeds for replicability, though we find these do not significantly affect consistency scores.

External consistency grows with ensemble size, especially for the continuous question, though at least 5% of pairwise comparisons do not match even with temperature 0 and ensemble size of 10. Interestingly, internal consistency scores are more predictive of external consistency for higher temperature models, suggesting that within generation variation at low temperatures may not reflect the same sources of randomness as across-generation runs.

To maximize replicability, we find a trade-off between a low-temperature single-shot approach with its high external consistency and cost-effectiveness, and a high-temperature ensemble approach with at least five iterations for improved reliability. These strategies balance the trade-offs between consistency, accuracy, and computational cost. Our main specification implements a low-temperature single-shot approach, prioritizing external consistency and cost-effectiveness. This approach also reflects our interest in creating a single national dataset, for which consistent interpretation of zoning regulations across municipalities is important. However, for other applications requiring more creative or diverse output, the high-temperature ensemble approach may be more appropriate.

3 Characterizing Housing Regulations

In this section, we summarize our comprehensive housing regulations dataset and examine regulatory patterns within metropolitan areas and across municipalities. This analysis reveals insights into both the drivers and impacts of different zoning approaches.

3.1 Principal Component Analysis of Housing Regulatory Dataset

Our nationwide dataset consists of 20 questions and covers nearly 6,000 local governments. To provide context for our analysis of regulatory dimensions, Table 4 presents the key descriptive statistics for our nationwide housing regulation dataset. The continuous measures in Panel A reveal substantial variation in regulatory scope and complexity across municipalities. The average municipality has 14 zoning districts, showing considerable regulatory sophistication. Bulk regulations also demonstrate significant stringency, with lowest minimum lot sizes averaging 10 thousand square feet and longest frontage requirements averaging 92 feet—well above typical urban lot dimensions. Process regulations typically involve extended review periods, with maximum potential waiting times averaging around 7 months. We also show variation in housing process regulations in Appendix Figure A4, finding that the American West and California in particular appear to have particularly onerous requirements in terms of waiting times, public hearings, and mandatory approval steps.

Panel B of Table 4 highlights the prevalence of binary regulatory controls. A notable feature of housing regulations across municipalities is restrictions on housing density: multifamily housing is prohibited in 5% of jurisdictions (rising to 10% in high-income areas), and 86% of municipalities restrict conversions to multifamily units. About 37% ban mixed-use developments combining residential and commercial uses. These restrictive measures generally intensify in higher-income and more rural areas. The overall pattern suggests significant constraints on housing density and development flexibility throughout the U.S., with noticeable variation across socioeconomic and geographic dimensions.

To summarize our nationwide measure of housing regulations, we perform a PCA analysis. This technique reduces the dimensionality of our dataset by identifying key components that capture a large fraction of variation across our regulatory questions. Table 5 provides the loadings of each question on the first five principal components.

The first key takeaway from our analysis is that housing regulations are not well-summarized by a simple unidimensional level of stringency. The first principal component explains just 13% of overall variation, while the second principal component explains 11% of the variation. The fraction of variation explained goes down substantially after this point, justifying our focus on the first two main principal components. This suggests a more nuanced covariance structure of housing regulations. In Figure 4 we plot the pairwise correlations between all zoning questions as a heatmap. Many questions are positively correlated, especially similar types of regulation (i.e., comparing different bulk regulations). However, the maximum correlation between any pair of regulations is below 0.5 and many other correlations are low or even zero; for example, minimum lot size requirements show near zero correlation with flexible zoning policies. Moreover, some of the correlations have surprising signs. For example, allowing accessory dwelling units (i.e., a less restrictive regulatory environment) is actually positively correlated with the lowest minimum lot size and permit and development caps (which both correspond to a more restrictive regulatory environment). These complicated covariance patterns reflect the diversity of zoning codes created by local governments, and the wide variety of objectives these governments pursue.

To better understand the economic interpretations behind the two main principal components, and to disentangle the relative roles of demand and supply in housing production, we show in Figure 5 the associations between housing construction (building permits), housing cost (median house prices), and our two key principal components. We interpret the association of these two principal components with housing construction and cost in light of a simple framework of supply and demand for housing. In places with rising demand for housing and inelastic supply, home prices will be high and construction low (the lower-right quadrants of Panels A and B of Figure 5. In areas with rising demand for housing and elastic supply, there will be high house prices and high construction (i.e., the upper right quadrant). In places with falling housing demand, there will be both low construction and prices. In places with elastic housing supplies and constant or moderately rising demand, there will be low home prices and high construction.

Panel A of Figure 5 highlights that areas with a high value for the first principal component generally have high house prices as well as construction, while areas low in this dimension typically have both low prices and construction. This association suggests that the first principal component generally coincides with high housing demand environments. This interpretation is supported by the main regulatory loadings on this principal component (Table 5), which loads heavily on measures that are typically associated with more developed, high-demand housing markets. For instance, affordable housing and age restricted provisions are policy tools that are more likely to be implemented in areas with significant housing pressure and the administrative capacity to manage complex policies. Other associates of the first principal component relate to additional layers of local government: maximum review wait times and public hearing requirements. However, areas heavy in this component are much less likely to have bans or limits on multifamily housing, townhouses, accessory apartments (ADUs), or flexible zoning.

By contrast, the second principal component (Panel B, Table 5) associates highly with areas that have high house prices but low construction, while being negatively associated with areas that have low prices and high construction. This suggests that while the second principal component is also associated with higher costs, these higher costs appear to be more related to housing supply shifts rather than higher housing demand (as appears to be the case with the first principal component).

Figure A5 provides a visual scatterplot representation of the relationship between these two principal components across U.S. municipalities. The plot also highlights the heterogeneity in zoning practices across municipalities. Suburban areas around high-demand metropolitan areas (such as Darien, CT or Billerica, MA) rank highly on both PCs. High-demand urban areas (like Santa Ana, CA) score low in PC2, but are relatively high in PC1. Low-demand urban areas (like Cleveland, OH) rank low on both PCs. Finally, exclusionary suburbs of low-demand areas (like Novi, MI) rank high in PC2 but low in PC1. Figure 6 maps the first two principal components across the nation, while Appendix Figure A6 shows variation across counties.

We summarize the multidimensional nature of zoning codes as our first fact:

Fact 1. *Zoning codes are multidimensional and can be clustered into two main principal components.*

3.2 Value Capture and Housing Markets

We next focus on the economic interpretation of these two principal components. We argue the first principal component can be interpreted as value capture. Value capture in the housing context refers to mechanisms by which local governments extract a portion of the increased land value or housing surplus that arises from high market demand. This typically occurs through regulatory tools that allow municipalities to convert some of the economic gains from private development into public benefits or revenue. As discussed in the previous section, factor loadings reveal that this component associates most strongly with regulatory tools like inclusionary zoning mandates (where developers must include affordable units), affordable housing incentives, and age-related provisions, which directly reflect tools that exact concessions from developers. Housing process regulations, such as public hearing requirements and waiting times, also reflect additional wedges or distortions which are only possible in higher demand environments.

We further support this interpretation by linking this regulatory dimension to its economic associates in Table 6, focusing first on the relationships with county-level amenities. These measures include: an index of retail establishments per capita across industries (constructed by taking the first PC of industry-level establishment counts), a natural amenity index (capturing features like temperature and sunlight hours), patents per capita (measuring productivity), and local employment density (job access benefits). We find that the first PC has a large and statistically significant positive relationships with all four measures, while the second PC has either small or statistically insignificant relationships with each. The relationship becomes even stronger for the first PC when we construct a combined amenity index by taking the first principal component across these four measures. These results support a housing demand interpretation of this regulatory regime: areas with higher amenities feature higher housing demand, which generates a surplus that can be potentially captured through regulation.

We explore a more granular comparison across different types of retail establishments in Appendix Figure A7. Counties scoring high on the first PC have higher concentrations of consumer retail outlets (like apparel stores and restaurants) and professional services (including educational institutions, healthcare facilities, and cleaning services) per capita, while having fewer establishments typically associated with negative externalities, such as gas stations, utility services, and truck transportation businesses per capita. This pattern suggests these areas are high-demand locations where businesses can command premium rents (also see Couture et al. (2024)). Further reinforcing the interpretation of the first PC as reflecting demand-driven amenities, Appendix Figure A8 shows that the first PC exhibits a high degree of spatial autocorrelation, as captured by Moran's I. In contrast, the second PC has very low spatial autocorrelation, suggesting that it is less influenced by spatially clustered amenities and more likely driven by other economic forces. This supports the notion that the first PC captures factors that vary smoothly across space, such as quality-of-life amenities, which naturally exhibit spatial dependence.

We then examine associations of this first PC on other variables, controlling in column 2 for metro fixed effects. Again consistent with a demand interpretation of this principal component: we find areas that are high on this dimension have a high college share, more young households, have lower poverty rates, and have substantially higher shares of Democrats. However, they have higher density on some measures (especially multifamily structures; i.e., those with two or more units), and are also generally larger areas in land units, with more municipalities as neighbors, and closer to the city center. These associations together suggest that areas high in PC1 are high-demand areas which may be prone to extraction of value by local governments (Diamond, 2017).

To be sure, such regulations may also affect housing supply. We explore this variable in more detail in columns 1–2 of Appendix Table A5, adding additional topographical and land availability controls such as the fraction of land developed in 2001, the squared fraction of land developed in 2001, and the fraction of land with a flat topography. The first principal component is associated with lower housing elasticities, though the significance becomes marginal once metro fixed effects are included. This suggests that, despite the raw association of this PC with higher construction, it represents a bundle of higher taxation which weakly associates with lower housing supply elasticity. We summarize these patterns as our second fact:

Fact 2. *The first principal component corresponds to value capture in high-demand locations. This component loads heavily on regulations that allow municipalities to extract and redistribute housing value, such as inclusionary zoning mandates that require developers to include affordable units. These regulations enable local governments to convert market-rate housing demand into public benefits, and are most prevalent in densely populated cities with high consumption and production amenities.*

3.3 The Role of Exclusionary Zoning

Exclusionary zoning refers to land use regulations that limit housing density and types, often with the effect of excluding lower-income residents from certain areas. The role of exclusionary zoning in shaping socioeconomic patterns has been a subject of significant research and debate in urban economics and policy circles, but accurately defining these practices has been a challenge for the literature. Our analysis provides new insights into this phenomenon, leveraging our comprehensive dataset to examine the prevalence, distribution, and impacts across the United States of exclusionary zoning practices. Regulations commonly linked to exclusionary zoning practices include measures such as large minimum lot sizes, restrictions on multi-family housing, and other bulk regulations that increase the cost of housing in a given area by effectively mandating a minimum amount of housing or land consumption.

The second principal component of housing regulations identified in this study correlates strongly with intuitive measures of exclusionary zoning, particularly minimum lot sizes and other bulk regulations that limit density, the loadings for which we show in Table 5. In fact, the highest loading for this principal component is the smallest residential minimum lot size, a commonly used proxy for exclusionary practices in general. This regulation sets a lower bound on how small a piece of land can be for a single housing unit, impacting the potential density of an area. Larger minimum lot sizes result in fewer, more spread-out homes, while smaller minimums allow for denser development. They have been frequently estimated in prior research through bunching methods (Cui, 2024; Song, 2021) as important drivers of housing regulations; the contribution of our approach is to measure these regulations directly from municipal documents, rather than indirectly through their effects on observed housing development. Figure 7 demonstrates the high frequency of these minimum lot size regulations specifically; two-thirds of local governments prohibit residential lots smaller than 5,000 square feet across the entire town (Panel A), and nearly half enforce minimum lot sizes of at least half an acre in some part of their jurisdiction (Panel B). This mandated minimum amount of housing consumption should effectively truncate the left tail of the housing value distribution by preventing the construction of houses on small lots of land or apartments.

We then validate our economic interpretation of the second PC by highlighting its effects on housing affordability. We show municipalities high in the second principal component have substantially fewer housing units affordable to the state median income household (Panel A of Figure 8). We focus on three key mechanisms driving this relationship. First, exclusionary zoning shifts the entire housing price distribution rightward, with significantly higher median home values in high second PC municipalities compared to low second PC areas (Panel B). Second, exclusionary zoning shifts mass away from the left tail and towards the median of the housing price distribution; i.e., a left-truncation of the house price distribution corresponding to missing units made unviable through the presence of minimum housing requirements. In Panel C, we compare the distribution of home values across granular home value bins for low and high second PC areas. High second PC areas have relatively more housing units at or right below the median home price and less well below the median, suggesting a leaner left tail. We further show in Appendix Figure A9 that this left truncation is more pronounced in higher median home values areas, and that areas higher in the second PC have smaller shares of low-income and younger households. The third dimension by which exclusionary zoning practices limit affordability is on the rental margin. Panel D reveals that high second PC areas limit the degree of rental housing with a strong correlation of 0.40 between the second PC and the share of owner-occupied housing. For rental housing, the second PC is more strongly associated with overall rents than with rent conditional on housing characteristics (i.e., number of bedrooms, as shown in Appendix Figure A10), suggesting that it further raises rental prices by shifting the composition of available units. The resulting lack of affordable housing options results in a strong negative correlation between the second PC and the share of low-income families (-0.26) and young households under 35 (-0.39).

We further validate that density restrictions impact observed density in Appendix Table A6. We examine associations between our measured regulations and housing values, building permits, rents, density, and affordability of local housing units using three approaches: bivariate regressions, LASSO, and XGBoost. Density restrictions (multifamily bans, minimum lot sizes, frontage requirements) consistently associate with lower observed density. In the XGBoost analysis, frontage requirements score highest (100), followed by highest (86) and lowest (84) residential lot size requirements. These variables also strongly associate with higher house prices and rents, and reduced affordable housing availability. This rules out a potentially competing hypothesis that lower construction in these areas is associated with lower demand. When predicting median home values, multifamily housing prohibition emerges as the most important predictor (100), followed by affordable housing mandates (75) and townhouse prohibition (56).

Appendix Table A7 confirms the predictive power of including housing regulations on other outcomes, such as home value, building permits, rents, density, and affordable housing shares. Including regulatory variables improves model fit, even with other controls present. XGBoost consistently outperforms Lasso and OLS, particularly with all variables included, achieving the lowest RMSE (0.66) for median home value.

We also explore the resulting distributional consequences of these quantity-based housing restrictions. Areas higher in the second PC are associated with demographic patterns that suggest exclusionary effects. In Table 6 we find that PC2 corresponds to a higher proportion of white, medium-to-high income, and college educated residents. This suggests racial and ethnic segregation effects of these policies. This is consistent with prior analyses of zoning as a tool for maintaining racial homogeneity in the absence of explicit racial covenants (Rothwell, 2011; Cui, 2024). Areas higher in the second PC also have a relatively higher Republican share of the population, relative to their metro.

Areas characterized by more exclusionary zoning practices also show higher average math test scores. This pattern suggests that exclusionary zoning effectively creates enclaves of educational privilege, where resources and positive peer effects are concentrated. Consistent with a public goods motive, we also observe higher local revenue per student. These results are consistent with classic theories of fiscal zoning which emphasize the role of zoning and property taxes in conjunction with support of local public goods (Fischel, 1987; Hamilton, 1975, 1976).These areas also tend to have lower property tax rates, suggesting that by limiting the share of cheaper housing, they can maintain a lower tax burden while still sustaining high-quality and well funded public education (Table A8). Areas higher in the second PC also show higher causal measures of economic opportunity (drawn from Chetty et al. (2014)). The presence of such public goods may help to explain the sorting of higher-income residents in these areas. However, they also point to the socioeconomic costs of excluding lower-income residents from access to areas with such public goods. We summarize these associations as our third fact:

Fact 3. *The second principal component captures exclusionary zoning practices aimed at restricting housing density and affordability through bulk regulations. This dimension loads heavily on minimum lot size requirements and procedural barriers to development, which are concentrated in affluent suburban areas. These regulations associate with a truncation of the left tail of house prices, lower construction, and strong sorting on education and income.*

3.4 Monocentric City Model and Zoning Gradients

We next interpret municipal regulations in the context of the monocentric city model (Alonso, 1964; Mills, 1967; Muth, 1971). In these models, there is a central location in each city where production is concentrated and rents decay as one moves away from this productive center, with the rate of decay governed by transportation costs. These dynamics may also affect the benefits and costs of zoning regulations at different distances from the city center. Housing regulations, in turn, may then affect the rent gradient as one moves away from the city center.

We show various regulatory variables along the dimension of distance to city center in Figure 9. Affordable housing mandates are decreasing in distance from the center of the city, illustrating that these regulations are most commonly found at the centers of cities.

Minimum lot size requirements show a different pattern, and vary markedly across regions. While minimum lot size requirements are higher in suburban areas on average, this relationship is particularly driven by the Northeast and Midwest regions. To further illustrate these patterns at the metropolitan level, Figures 10 and 11 show maps of minimum lot sizes and affordable housing incentives, respectively, for jurisdictions within the metropolitan areas surrounding four select cities in the U.S., Atlanta, Chicago, Philadelphia, and San Francisco.

These graphs document substantial variation in both minimum lot sizes and affordable housing mandates and incentives within metropolitan areas across municipalities, with the central city and inner suburbs having lower minimum lot sizes and higher rates of affordable housing policies than in jurisdictions farther from the central city. We observe strongly increasing minimum lot sizes away from the city center towards suburban areas in Chicago and Philadelphia, but this pattern is less pronounced in other regions of the country. This figure also illustrates a key advantage of our approach: the ability to construct measures of zoning ordinances at the level of the municipality across a wide variety of municipalities and regions in the United States.

In Table 7, we explore the relationship between distance from the city center and regulatory stringency for all of the housing regulations that we measure and both the first and second principal components. The first four columns show these relationships between regulations and distance from city center by region, while the final column shows the relationship for the U.S. as a whole. Across all regions, we observe that the number of zoning districts decreases robustly with distance from the city center, suggesting simpler zoning structures in more peripheral areas. This pattern is consistent across all regions but is particularly pronounced in the Midwest and South. Some components of allowable density decrease with distance from the center, especially the permission of townhouses and mixed-use development. Interestingly, the allowance of multifamily housing shows a positive correlation with distance in the West and South, contrary to the general expectation of decreasing density with distance. This might reflect the presence of suburban multifamily developments in these regions.

The Northeast stands out with several distinct patterns. Unlike other regions, it shows increasing restrictiveness with distance for several measures. For instance, the longest frontage requirement, highest residential minimum lot size, and mean residential minimum lot size all increase with distance from the city center in the Northeast, while these measures show no significant relationship in other regions. These results highlight the unique regulatory landscape of the Northeast, where bulk regulations and exclusionary zoning practices appear to intensify in suburban and exurban areas, contrary to patterns observed in other regions. These results are also especially surprising in the context of well-developed public transit and highway links in this area, which should, all else equal, facilitate greater development and density even outside of city centers in this region.

Several factors may help to explain this regional variation. The Northeast was the first region of the U.S. to urbanize and industrialize, allowing for greater development before the advent of zoning. Many of its suburbs were established earlier than in other regions, often as affluent enclaves seeking to preserve their character against urban expansion (Fischel, 2015). In addition, the region is characterized by a highly fragmented system of local governments, with many small, independent municipalities. This structure facilitates more localized and potentially more restrictive zoning policies. These suburbs also have a strong home rule tradition of local control over land use decisions. The Northeast's early experience with industrial pollution and urban congestion may have fostered a culture of environmental protection that manifests itself in stricter land use controls, particularly for suburban lot size requirements, which were commonly justified on the basis of preserving natural land. Finally, the region has had particularly strong fights over access to local schools (i.e., school busing (Angrist et al., 2022)) that can increase the use of exclusionary zoning practices to maintain local school districts by limiting access to lower-income and minority households.

In contrast, California, which also has high house prices and where housing regulation is commonly thought to be tight, appears surprisingly to have more nationally typical bulk regulations. We show a map representation of these regulations in Figure 6, which highlights the high minimum residential lot size requirements in the northeast relative to California. In contrast, California's housing market is heavily influenced by state-level regulations such as environmental reviews and an onerous permitting process. We find some evidence of this in Appendix Figure A4, which shows that California has the highest potential waiting time for review of a typical new multi-family building (see Mayer and Somerville (2000) for a link between regulatory uncertainty and reduced construction) and highest prevalence of public hearing requirements for multi-family buildings. These results therefore suggest that while the Northeast and California have high housing costs, they are the result of very different factors that may call for different policy responses. Furthermore, California's housing market faces significant constraints due to its challenging topography (Saiz, 2010), which limits developable land, and the high demand for natural amenities and local jobs. These factors can create a situation in which even relatively less restrictive local bulk regulations can result in binding constraints on housing supply, especially in combination with a challenging permitting process.

To explore the drivers of exclusionary zoning practices in the Northeast further, we follow (Glaeser and Ward, 2009) who connected minimum lot sizes drawn from the Pioneer Institute survey in the Boston metropolitan area to historical variables: particularly historical density and industrial production. The key conclusion from that analysis is the gradient between minimum lot sizes and distance to city center (Boston) was largely explained by the fact that low density areas historically (in 1940) were able to adopt minimum lot size requirements, and thereby stay low density. We are able to expand the scope of this analysis to a much broader set of municipalities across the nation.

In Table 8, we show the key object of interest—regional gradients of minimum lot size against distance to city center—across a range of controls. Our initial specification reveals that the Northeast and (to a lesser extent) the Midwest feature substantially higher minimum lot sizes as distance to city center increases, a relationship which actually strengthens after controlling for a fixed effect for the nearest metropolitan area. However, we are able to partially account for these gradients by controlling for a variety of historical variables, drawn from 1940, which capture density, manufacturing output, and demographic variables in that period. Our preferred specification, in column (7), controls for an interaction of historical municipal-level density with county-level industrial production. This specification allows for the possibility that low-density areas near high industrial production regions adopted more strict minimum lot size rules, which explains the overall gradient. This specification reduces the regional gradient by 40–50%, depending on whether we measure distance as a raw number or a log. We are able to explain an ever larger fraction of the raw gradient in a final specification which controls for a large battery of historical variables.

To shed additional light on this finding, we show in Figure 12 the key mechanisms driving this result. In Panel A of this figure, we show the minimum lot size, at the municipality level, as a function of relative density compared to their MSA-average in 1940. We see a strong pattern that Northeastern municipalities with *less* density than their local average exhibit substantially higher minimum lot sizes, while Northeastern municipalities with higher than average density do not exhibit any unusual pattern for minimum lot sizes. In Panel B of this figure, we show the coefficient of historical density against minimum lot size for a sample of the top 30 metropolitan areas. We see that historical density is very strongly negatively associated with minimum lot sizes in Northeastern and Midwestern areas, but not in the rest of the country. Appendix Figure A11 also shows a non-parameteric relationship of distance to city center and minimum lot sizes to highlight the skewed pattern in the Northeast, and also highlights the high manufacturing share of employment in this region.

Our findings confirm that the relationship that (Glaeser and Ward, 2009) found also broadly applies to industrial centers across the Northeast, and to some extent in the Midwest. However, it does not apply to the West or South regions, where the average level of minimum lot sizes are far lower, and they do not exhibit the same gradients with respect to distance to city center. These more restrictive zoning regulations have reinforced historical density patterns, as shown in Appendix Figure A12, which shows the correlation between 1940 and 2019 housing unit density is highest in the Northeast and weakest in the South. Therefore it appears that the Northeast locked in a particular pattern of spatial development, which is then highly persistent over time.

There are several potential factors which could explain this result. Less dense areas in the vicinity of industrial centers in the Northeast and Midwest could be averse to industrial pollutants; they may have been motivated by social exclusionary factors against industrial workers (who were disproportionately immigrant and non-white, see Cui (2024)); and the spread of industrial production through automobile traffic may have made such areas adopt stricter zoning to thwart direct industrial spillovers, as suggested by Fischel (2015). While a deeper exploration of this pattern is outside of the scope of the current paper, this finding is consistent with several plausible channels connecting industrial spillovers and exclusionary motives, and we formally consider this channel in the context of our model in Section 4.

Our findings on spatial patterns of zoning have important implications for urban economic models. While the general relationship between regulatory intensity and distance from city centers is broadly consistent with standard monocentric city models (i.e., Rossi-Hansberg (2004)), the strong sorting of high-income households into expensive, strictly regulated suburbs is more of a challenge for these frameworks. Traditional models predict that all households prefer central locations to minimize commuting costs, with wealthier residents outbidding others for scarce urban land. However, the high market values of suburban housing and the willingness of affluent residents to accept longer commutes contradict these predictions and represent a recognized challenge to urban economic theory (Glaeser et al., 2008). This spatial pattern is more pronounced in the United States compared to other countries, where it is more typical for wealthy residents to reside in the city center, and appears particularly pronounced in the Northeast. Our results suggest that this sorting is accompanied by regulatory restrictions on the minimum allowable housing size, which truncates the housing size distribution to the left, and results in sorting of higher-income residents to distant suburbs against the typical pattern expected in the monocentric city model. We summarize these patterns as our fourth and final fact:

Fact 4. *Zoning regulations generally follow a monocentric pattern, with value capture tools predominant in city centers and exclusionary zoning practices intensifying with distance from urban cores, a pattern most pronounced in the Northeast where suburban areas employ particularly strict density restrictions.*

4 Discussion and Framework

4.1 Model Framework

Having established several key findings on housing regulatory variation across the United States, we next interpret these empirical findings through a model of inter-municipal competition. In our framework, local governments strategically select housing regulations. A more complete version of the model is discussed in Appendix E.

The model features households sorting across locations with two different regulatory regimes intended to proxy for our first two principal components. The key elements include:

Agents and Locations: The economy contains two types of households differentiated by productivity: high-wage (H) and low-wage (L), earning $w^H > w^L$. Households choose between two distinct zones: city centers (c) and suburbs (s). Each zone offers different amenities (α_c, α_s) and implements different regulatory environments. Housing supply in each zone is somewhat inelastic, such that local rents r_i are increasing in housing demand and population.

Government Policies: Local governments in each zone $i \in \{c, s\}$ impose two types of taxes: a uniform wage tax τ and zone-specific housing taxes t_i, which are intended to proxy for value capture regulations. These correspond to our first principal component of housing regulations, and the interpretation is that regulations such as mandatory inclusionary zoning can be seen as an additional implicit taxes. In principle the proceeds could be either redistributed to other residents (as in the case of affordable housing units), captured as private benefits by local policymakers, or else extracted by the government for the purpose of public goods. We assume these benefits can be extracted without penalty to produce government revenue. With these revenues, they provide public services s_i using labor inputs G_i (compensated at the low-productivity wage), while maximizing net revenue. Additionally, governments can impose minimum housing requirements $\underline{h}_i$, which proxy for exclusionary zoning requirements.

For a household of type j in zone i, utility is given by:

$$U_i^j = \max_{c,h} \ \gamma \log c + (1-\gamma)\log h + \alpha_i - \log(N_i) + \log(s_i).$$

This optimization is subject to the budget constraint:

$$c + h \cdot r_i = (1-\tau) \cdot w^j$$

Local governments maximize revenue:

$$\max_{G_i, t_i, \underline{h}_i} \ [\tau + (1-\gamma)\cdot t_i] \cdot (w^L \cdot N_i^L + w^H \cdot N_i^H) - w^L \cdot \log(G_i)$$

subject to the minimum housing constraint ($h_i^j \geq \underline{h}_i$) and a positive earnings condition for local governments. Production opportunities are equally accessible from both zones, so there are no reasons to sort for increasing wage income.

In spatial equilibrium, households sort across zones until indifferent at the margin, yielding:

$$[\alpha_c - \alpha_s] + \log \frac{s_c}{s_s} = (1-\gamma)\left[\log \frac{r_c}{r_s} + \log \frac{1+t_c}{1+t_s}\right] + \log \frac{N_c}{N_s} \tag{1}$$

i.e., that the benefits of locating in one zone (based on local amenities and government services) equal the costs of rents, housing taxes, and congestion disamenities.

4.2 Optimal Policy Conditions

Exclusionary Zoning: Minimum housing requirements ($\underline{h}_i$) become binding when they increase revenue by attracting high-productivity households:

$$\Delta \underline{h_i} \left(\Delta \eta_i N_i \cdot \left[t_i \frac{\partial r_i}{\partial \eta_i} + \tau \frac{\partial \tilde{w}}{\partial \eta_i} \right] + \Delta N_i \cdot \left[t_i r_i + \tau \tilde{w}_i + N_i t_i \frac{\partial r_i}{\partial N_i} \right] \right) \geq 0.$$

For $\widetilde{N}_i = w_i^L \cdot N_i^L + w_i^H \cdot N_i^H$ as the "effective" population, and similarly $\widetilde{w}_i = w_i^H \cdot \eta_i + w_i^L \cdot (1-\eta_i)$ as the effective wage, and $\eta_i = \frac{N_H}{N^H + N^L}$ as the high-income share of total population.

The intuition behind this condition is that exclusionary zoning practices affect local government revenues through two channels. The first term in this equation captures the gain to local governments from attracting a higher share of high-income workers. These workers are increasingly attracted to areas with minimum lot size requirements because they are subject to lower congestion disamenities. However, the loss in population also reduces local government revenue in the second term.

Value Capture: Housing taxes (t_i) become optimal when tax revenue gains outweigh population losses:

$$(1-\gamma) \cdot \widetilde{N}_i + \frac{\partial \widetilde{N}_i}{\partial t_i} [\tau + (1-\gamma) \cdot t_i] \geq 0.$$

Proposition 1. *Urban Value Capture: City cores c prioritize revenue extraction (t_c).*

In the model, city cores choose a positive housing tax requirement t_c to maximize local revenues, and pick no exclusionary housing requirement. The intuition is that city cores have a sufficiently high population of low-income workers, so the income losses from an exclusionary housing requirement, and the resulting loss of the low-income population, would exceed gains from gaining high-income population share. However, if city cores have sufficiently high amenities relative to suburban areas ($\alpha_c - \alpha_s$), from indifference equation (1) we see this amenity differential can offset higher housing costs, taxes, and congestion disamenities in the city core. This gives urban centers leverage to extract value from housing and residents without depopulating the area.

Proposition 2. *Urban zones with higher amenities (α_i) have more revenue extraction.*

This proposition follows a similar argument as the previous one. The core intuition is that higher amenities increase residents' willingness to pay for housing, generating economic rents which governments can profitably tap into. Again from equation 1 we see that households accept higher housing costs (both due to rent and additionally due to governments value capture taxes) because these are offset by higher amenities. Therefore, the optimal t_i rises with α_i: urban centers with higher amenities can impose even higher taxes without losing residents, because leaving entails losing access to valuable local benefits.

Proposition 3. *Suburban Exclusion: Suburbs s adopt stricter bulk regulations $\underline{h}_s$.*

The key difference between urban cores and suburban areas lies in population composition: when suburban areas have a substantially smaller population relative to urban cores, they are in a position to shift to a completely concentrated tax base consisting only of high-income workers through exclusionary zoning. The trade-off is that such strict bulk regulations limit the total number of households (N_s), which lowers income-based tax revenue. However, this loss in quantity can be offset fiscally by an increase in per-household tax contribution, and also results in lower congestion costs, which results in an additional motive for high-income workers to move to suburban areas.

Proposition 4. *Suburban areas with fewer amenities have weakly stricter bulk regulations.*

In the model, local amenities and exclusionary zoning can work as substitutes. This situation is not relevant if the local government is already at a corner solution and has implemented a minimum house size requirement to exclude all low-income residents. But if a suburban area is indifferent between imposing such a restriction or not, lowering the presence of amenities will encourage the municipality to implement such policies. The intuition is that a low α_s raises the marginal benefit of exclusion (i.e., attracting rich households), because the suburb can offset low taxation in quantities by improving the quality of the tax base.

4.3 Connecting Model Predictions to Empirical Findings

Our first and third propositions relate to the spatial locations of housing regulations, which closely align with where we find them in the data The value capture motive for housing regulation shows up most strongly in the centers of American cities, where demand for living is highest and so the ability for local governments to extract value is correspondingly higher as well. Consistent with this result in the model, we find empirically that areas high in the first principal component are substantially larger, and also appear to be somewhat more centrally located.

By contrast, exclusionary regulations in the model are exclusively a suburban phenomenon, and this is also a finding of our data. We find that bulk regulations and the second principal component are commonly found in suburban areas in the peripheries of those cities across the United States, but are particularly pronounced in the Northeast. These spatial regulatory pattern contributes to the sorting of Americans along the dimensions of age and income. Cities are home to poorer residents that need access to smaller housing, wealthy residents, as well as younger working households. By contrast, suburban areas with exclusionary zoning requirements are more typically home to richer households, despite high commuting costs.

Our results help to connect classic previous theories of zoning. (Fischel, 1987) and (Hamilton, 1975, 1976) argue that housing regulation and property taxes can create efficient public goods provision in the context of (Tiebout, 1956) sorting. By contrast, (Zodrow and Mieszkowski, 1986) argue instead that local taxes are distortionary and function like excise taxes, while more recent scholarship has emphasized the segregation motives of zoning, especially exclusionary zoning (Rothstein, 2017; Cui, 2024). Our two sets of regulatory controls by municipalities spans this prior literature, and helps to explain the circumstances under which housing regulation can appear extractive, and the conditions under which they sustain local public goods investment under exclusionary environments.

Our model also makes additional sets of predictions on the extensive margin of such regulations, which also line up with our empirical findings. Proposition 2 predicts that greater amenities should predict higher value capture, which is an extremely strong pattern in our data as well across amenities measured through different methods (natural amenities, consumption amenities, and productive/job amenities). This close alignment between empirical and model results here helps to further justify our interpretation of the first principal component as being primarily about demand and about extracting value.

The model also helps us interpret our fourth fact concerning the regional nature of exclusionary zoning patterns. In the model, exclusionary zoning can emerge in suburban environments exposed to negative amenities. Our empirical results suggest that the strongest patterns of exclusionary zoning emerged historically among Northeastern (and to some extent Midwestern) municipalities exposed to local manufacturing. While we cannot separate the specific role of industrial pollutants and direct spillovers versus exclusionary motives around the presence of manufacturing workers (many of whom were immigrants or non-white); all plausible mechanisms entail such early industrialization exposing smaller suburbs to some set of negative disamenities. Our model helps therefore reconcile the motives for why such municipalities may pick exclusionary zoning as a response to this disamenity shock.

5 Conclusion

Our paper introduces a new methodology we refer to as generative regulatory measurement, which leverages LLMs to systematically interpret municipal zoning regulations at scale. Our results demonstrate that state-of-the-art LLMs can achieve near-human levels of accuracy in classifying zoning rules from textual documents, with accuracy levels of 96% for binary questions and correlations of 0.87 on average relative to existing hand-classification from the Pioneer Institute. This generative regulatory measurement approach therefore enables the creation of a comprehensive, publicly available, nationwide dataset of municipal zoning regulations. By leveraging LLMs to extract structured information from unstructured textual data, our methodology opens up new avenues for analyzing vast amounts of previously untapped regulatory documents across multiple domains. With further development, this generative regulatory measurement framework can be extended to other categories of housing regulation (for example, to building codes), regulations in other domains, and even regulatory systems across different countries and languages.

By combining this measurement with a theoretical framework, we establish four facts about housing regulation. Municipal zoning regulations are not well-summarized by single dimensional measures of strictness, reflecting the myriad ways municipalities have designed their zoning codes. We focus on first two primary regulatory dimensions in the data: value capture and exclusionary zoning. We find that urban cores, especially those with higher amenities, implement value capture regulations to extract housing surplus in order to fund other public goods. Conversely, suburban municipalities often adopt exclusionary zoning and strict bulk regulations, especially in the Northeast and Midwest, a pattern which we connect to historical exposure to industrial production.

Our findings highlight the broader consequences of housing regulations on socioeconomic sorting and affordable housing. Suburban practices of exclusionary zoning truncate the left part of the housing price distribution, which effectively prices out lower-income and minority households from areas with high-quality public services. Such households sort instead to urban cores, which also feature elevated housing costs due to extractive housing regulations. Our findings therefore underscore the importance of zoning regulations as tools used by municipalities to manage growth, capture value from residential demand, and impact demographic composition.

References

Alonso, William, *Location and land use: Toward a general theory of land rent*, Harvard university press, 1964.

Angrist, Joshua, Guthrie Gray-Lobe, Clemence M Idoux, and Parag A Pathak, "Still worth the trip? School busing effects in Boston and New York," Technical Report, National Bureau of Economic Research 2022.

Anthropic, "Introducing Contextual Retrieval," 2024.

Azamfirei, R, SR Kudchadkar, and J Fackler, "Large language models and the perils of their hallucinations," *Critical Care*, 2023, *27* (1), 1–2.

Babalievsky, F, K Herkenhoff, LE Ohanian, and EC Prescott, "The Sky is Not the Limit: The Macroeconomic Consequences of Commercial Zoning Laws," Technical Report 2021.

Baum-Snow, Nathaniel and Lu Han, "The microgeography of housing supply," *Journal of Political Economy*, 2024, *132* (6), 1897–1946.

Brueckner, Jan K, "Strategic control of growth in a system of cities," *Journal of Public Economics*, 1995, *57* (3), 393–416.

Bryan, Michael, "2020 General Election Voting by US Census Block Group," 2022.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond, "Generative AI at work," Technical Report, National Bureau of Economic Research 2023.

Bybee, J Leland, "The ghost in the machine: Generating beliefs with large language models," Technical Report, Technical report, Working Paper 2023.

Calder-Wang, Sophie and Gi Heung Kim, "Coordinated vs efficient prices: The impact of algorithmic pricing on multifamily rental markets," Technical Report, Working Paper 2023.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," Technical Report, Harvard University, Brown University, Massachusetts Institute of Technology, and U.S. Census Bureau January 2025.

_ , Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst, "Income growth and the distributional effects of urban spatial sorting," *Review of Economic Studies*, 2024, *91* (2), 858–898.

Cui, Tianfang, "Did Race Fence Off The American City? The Great Migration and the Evolution of Exclusionary Zoning," Technical Report, Working Paper 2024.

D'Amico, Leonardo, Edward L Glaeser, Joseph Gyourko, William Kerr, and Giacomo AM Ponzetto, "Why Has Construction Productivity Stagnated? The Role of Land-Use Regulation," Technical Report, Working Paper 2023.

Dell, Melissa, "Deep learning for economists," *Journal of Economic Literature*, 2025, *63* (1), 5–58.

Diamond, Rebecca, "Housing supply elasticity and rent extraction by state and local governments," *American Economic Journal: Economic Policy*, 2017, *9* (1), 74–111.

Duranton, Gilles and Diego Puga, "Urban growth and its aggregate implications," Technical Report, National Bureau of Economic Research 2019.

Eisfeldt, Andrea L, Gregor Schubert, and Miao Ben Zhang, "Generative AI and firm values," Technical Report, National Bureau of Economic Research 2023.

Epple, Dennis and Allan Zelenitz, "The implications of competition among jurisdictions: Does Tiebout need politics?," *Journal of Political Economy*, 1981, *89* (6), 1197–1217.

Fischel, William, *Zoning Rules!: The Economics of Land Use Regulation*, Lincoln Institute of Land Policy, 2015.

Fischel, William A, *The economics of zoning laws: A property rights approach to American land use controls*, JHU Press, 1987.

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90.

Giesecke, Oliver, "State Government Debt Structure," *Available at SSRN*, 2023.

Glaeser, Edward and Joseph Gyourko, "The economic implications of housing supply," *Journal of Economic Perspectives*, 2018, *32* (1), 3–30.

Glaeser, Edward L and Bryce A Ward, "The causes and consequences of land use regulation: Evidence from Greater Boston," *Journal of Urban Economics*, 2009, *65* (3), 265–278.

_ , Joseph Gyourko, and Raven Saks, "Why is Manhattan so expensive? Regulation and the rise in housing prices," *The Journal of Law and Economics*, 2005, *48* (2), 331–369.

_ , Matthew E Kahn, and Jordan Rappaport, "Why do the poor live in cities? The role of public transportation," *Journal of urban Economics*, 2008, *63* (1), 1–24.

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

Gyourko, Joseph, Albert Saiz, and Anita Summers, "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 2008, *45* (3), 693–729.

_ and Raven Molloy, "Regulation and housing supply," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 1289–1337.

_ and Sean McCulloch, "Minimum Lot Size Restrictions: Impacts on Urban Form and House Price at the Border," Technical Report, National Bureau of Economic Research 2023.

_ , Jonathan S Hartley, and Jacob Krimmel, "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index," *Journal of Urban Economics*, 2021, *124*, 103337.

Hamilton, Bruce W, "Zoning and property taxation in a system of local governments," *Urban Studies*, 1975, *12* (2), 205–211.

_ , "Capitalization of intrajurisdictional differences in local tax prices," *The American Economic Review*, 1976, *66* (5), 743–753.

Hansen, Anne Lundgaard and Sophia Kazinnik, "Can ChatGPT Decipher Fedspeak?," *Available at SSRN*, 2023.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun, "Firm-level political risk: Measurement and effects," *The Quarterly Journal of Economics*, 2019, *134* (4), 2135–2202.

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott, "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown," *Journal of Monetary Economics*, 2018, *93*, 89–109.

Hoffman, David A and Yonathan A Arbel, "Generative interpretation," Technical Report, Working Paper 2023.

Horton, John J, "Large language models as simulated economic agents: What can we learn from homo silicus?," Technical Report, National Bureau of Economic Research 2023.

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

Huang, Haifang and Yao Tang, "Residential land use regulation and the US housing price cycle between 2000 and 2009," *Journal of Urban Economics*, 2012, *71* (1), 93–99.

Ihlanfeldt, Keith R, "The effect of land use regulation on housing and land prices," *Journal of Urban Economics*, 2007, *61* (3), 420–435.

Jackson, Kristoffer, "Do land use regulations stifle residential development? Evidence from California cities," *Journal of Urban Economics*, 2016, *91*, 45–56.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, "ChatGPT and corporate policies," Technical Report, Chicago Booth Research Paper 2023.

Kahn, Matthew E, "The environmental impact of suburbanization," *Journal of Policy Analysis and Management*, 2000, *19* (4), 569–586.

Khot, Tushar, Harsh Trivedi, Matthew Finlayson, Yao Fu, Kyle Richardson, Peter Clark, and Ashish Sabharwal, "Decomposed Prompting: A Modular Approach for Solving Complex Tasks," 2023.

Lagakos, David, Stelios Michalopoulos, and Hans-Joachim Voth, "American Life Histories," Working Paper 33373, National Bureau of Economic Research January 2025.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, 2020, *33*, 9459–9474.

Lewis, Paul G and Nicholas J Marantz, "What planners know: Using surveys about local land use regulation to understand housing development," *Journal of the American Planning Association*, 2019, *85* (4), 445–462.

Li, Junyou, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye, "More agents is all you need," Technical Report, Working Paper 2024.

Lopez-Lira, Alejandro and Yuehua Tang, "Can chatgpt forecast stock price movements? return predictability and large language models," *arXiv preprint arXiv:2304.07619*, 2023.

Markley, Scott N., Steven R. Holloway, Taylor J. Hafley et al., "Housing unit and urbanization estimates for the continental U.S. in consistent tract boundaries, 1940–2019," *Scientific Data*, 2022, *9*, 82.

Mayer, Christopher J and C Tsuriel Somerville, "Land use regulation and new construction," *Regional Science and Urban Economics*, 2000, *30* (6), 639–662.

Mills, Edwin S, "An aggregative model of resource allocation in a metropolitan area," *The American Economic Review*, 1967, *57* (2), 197–210.

Mleczko, Matthew and Matthew Desmond, "Using natural language processing to construct a National Zoning and Land Use Database," *Urban Studies*, 2023, *60* (13), 2564–2584.

Modarressi, Ali, Hanieh Deilamsalehy, Franck Dernoncourt, Trung Bui, Ryan A. Rossi, Seunghyun Yoon, and Hinrich Schütze, "NoLiMa: Long-Context Evaluation Beyond Literal Matching," 2025.

Muth, Richard F, "The derived demand for urban residential land," *Urban studies*, 1971, *8* (3), 243–254.

Pierson, Kawika, Michael Hand, and Fred Thompson, "The Government Finance Database: A Common Resource for Quantitative Research in Public Financial Analysis," *PLoS ONE*, 2015, *10* (6), e0130119.

Quigley, John M and Steven Raphael, "Regulation and the high cost of housing in California," *American Economic Review*, 2005, *95* (2), 323–328.

Raymond, Lindsey, "The Market Effects of Algorithms," Technical Report, Working Paper 2023.

Reardon, Sean F., Andrew D. Ho, Benjamin R. Shear, Erin M. Fahle, Demetra Kalogrides, and Joseph Saliba, "Stanford Education Data Archive (Version 5.0)," 2024. Retrieved from https://purl.stanford.edu/cs829jn7849.

Reimers, Nils and Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," 2019.

Romer, Christina D and David H Romer, "A new measure of monetary shocks: Derivation and implications," *American economic review*, 2004, *94* (4), 1055–1084.

Rossi-Hansberg, Esteban, "Optimal urban land use and zoning," *Review of Economic Dynamics*, 2004, *7* (1), 69–106.

Rothstein, Richard, *The Color of Law: A Forgotten History of how our Government Segregated America*, Liveright Publishing, 2017.

Rothwell, Jonathan T., "Racial Enclaves and Density Zoning: The Institutionalized Segregation of Racial Minorities in the United States," *American Law and Economics Review*, 2011, *13* (1), 290–358.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Schwarcz, Daniel, Sam Manning, Patrick Barry, David R. Cleveland, J. J. Prescott, and Beverly Rich, "AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice," March 2025.

Service, USDA Economic Research, "Natural Amenities Scale," 8 2019.

Shanks, Brendan, "Land Use Regulations and Housing Development," 2021. manuscript.

Singla, Shikhar, "Regulatory costs and market power," Technical Report 2023.

Song, Jaehee, "The effects of residential zoning in US housing markets," Technical Report, Working Paper 2021.

Stacy, C, C Davis, YS Freemark, L Lo, G MacDonald, V Zheng, and R Pendall, "Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?," *Urban Studies*, 2023. 00420980231159500.

Tetlock, Paul C, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 2007, *62* (3), 1139–1168.

Tiebout, Charles M, "A pure theory of local expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

Turner, Matthew A, Andrew Haughwout, and Wilbert Van Der Klaauw, "Land use regulation and welfare," *Econometrica*, 2014, *82* (4), 1341–1403.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,

Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 2017, *30.*

Wei, Jason, Nguyen Karina, Hyung Won Chung, Yunxin Joy Jiao, Spencer Papay, Amelia Glaese, John Schulman, and William Fedus, "Measuring short-form factuality in large language models," 2024.

Wu, Tongshuang, Michael Terry, and Carrie Jun Cai, "AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts," in "Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems" CHI '22 Association for Computing Machinery New York, NY, USA 2022.

Yang, Stephen, "Predictive Patentomics: Forecasting Innovation Success and Valuation with ChatGPT," Technical Report, Working Paper 2023.

Zhang, Zhuosheng, Aston Zhang, Mu Li, and Alex Smola, "Automatic Chain of Thought Prompting in Large Language Models," 2022.

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong et al., "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

Zodrow, George R and Peter Mieszkowski, "Pigou, Tiebout, property taxation, and the underprovision of local public goods," *Journal of Urban Economics*, 1986, *19* (3), 356–370.

Tables

Table 1: Sample Coverage

Panel A: Sample and Local Government Coverage Metrics

Coverage Metrics:	National	Northeast	Midwest	South	West
Total Munis	19,488	2,101	8,481	6,587	2,319
% of Munis in Sample	25	32	19	22	48
Total Townships	16,213	4,111	12,102	0	0
% of Townships in Sample	6	23	0	-	-
Total Pop. (Millions)	331	57	69	127	77
% of Pop. Under Local Gov.	76	100	95	55	78

Panel B: % of Pop. Under Local Gov. Covered By Sample

Ordinance Aggregator:	National	Northeast	Midwest	South	West
American Legal Publishing	11	15	15	6	8
Municode	23	1	19	54	12
Ordinance.com	30	52	12	1	60
Total	63	68	46	61	80

Notes: This table highlights our sample coverage across region and data source. For local governments available in multiple datasets, we prioritize using Ordinance.com and then Municode and reflect that in the population count. We also adjust for geographical overlap between townships and municipalities in tallying population by using census block level population data and corresponding shape files. We use population estimates from the 2022 Census of Governments for municipality population, and 2022 State-Level Census Population Data for census region and national population.
Links to data sources are American Legal Publishing, Municode, and Ordinance.com.

Table 2: Performance Validation for LLM-Generated Zoning Classifications

Panel A: Continuous Questions

Question	RSE	Correlation
How many zoning districts, including overlays, are in the municipality?	0.06	0.98
What is the longest frontage requirement for single family residential development in any district?	1.16	0.70
What is the lowest residential minimum lot size?	0.16	0.92
Cumulative Average	0.46	0.87
Cumulative Median	0.31	0.90

Panel B: Binary Questions

Question	RSE	% Accuracy
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	0.00	100%
Are apartments above commercial (mixed use) allowed in any district?	0.07	96%
Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?	0.08	96%
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	0.30	90%
Does zoning include any provisions for housing that is restricted by age?	0.14	96%
Are accessory or in-law apartments allowed (by right or special permit) in any district?	0.09	96%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	0.00	100%
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	0.00	100%
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	0.00	100%
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	0.33	90%
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	0.14	96%
Cumulative Average	0.11	96%
Cumulative Median	0.09	96%

Notes: This Table highlights our model accuracy by comparing the accuracy of our generated regulatory dataset against a hand-classification by the Pioneer Institute on a hold-out sample of 30 municipalities. In Panel A, we measure accuracy using a Relative Squared Error (RSE) that compare the model's results to the naive model that guesses the sample mean. The correlation column is the correlation between the model answer and the Pioneer Institute answer. We calculate performance metrics and sample means (for RSE) only on the set of question municipality pairs that GPT-4 Turbo does not say "I don't know." We winsorize data from our models at the 1% level but do not winsorize data from the Pioneer Institute. The Cumulative Average and Cumulative Median are calculated across questions giving equal weight to each question. In Panel B, for Relative Squared Error (RSE) we compare each model's results to the naive model that guesses the sample mode. The accuracy column is calculated as the percent of municipalities where the model matches the adjusted Pioneer Institute answer for each question. We drop any errors where the answer is considered ambiguous. For details on adjustments to the Pioneer data, see Figure 3.

Table 3: Decomposition of Model Accuracy

Panel A: Continuous Questions

		Correlation		I Don't Know Rate	
Model	Specification	Mean	Std. Dev.	Mean	Std. Dev.
RAG Strategies (+ No Prompting)					
Gemini Flash 1.5	No RAG	0.74	0.01	38.7%	2.3
Gemini Flash 1.5	+ Basic RAG	0.66	0.03	32.7%	1.7
Gemini Flash 1.5	+ Full RAG	0.84	0.02	12.9%	0.6
Prompting Strategies (+ Full RAG)					
GPT-4 Turbo	No Prompting	0.72	0.05	2.7%	0.6
GPT-4 Turbo	+ Prompt Engineering	0.78	0.05	2.9%	1.3
GPT-4 Turbo	+ Background Info	0.73	0.11	2.2%	1.4
GPT-4 Turbo	+ Prompt Chaining	0.87	-	4.4%	-
Model Selection					
GPT-3.5 Turbo	Full RAG + Full Prompting	0.86	0.03	3.8%	1.0
Gemini Flash 1.5	Full RAG + Full Prompting	0.81	0.01	9.6%	1.7
GPT-4 Turbo	Full RAG + Full Prompting	0.87	-	4.4%	-

Panel B: Binary Questions

		Accuracy		I Don't Know Rate	
Model	Specification	Mean	Std. Dev.	Mean	Std. Dev.
RAG Strategies (+ No Prompting)					
Gemini Flash 1.5	No RAG	81.0%	1.24	21.1%	1.0
Gemini Flash 1.5	+ Basic RAG	84.5%	1.18	22.0%	1.7
Gemini Flash 1.5	+ Full RAG	86.2%	0.87	18.6%	0.8
Prompting Strategies (+ Full RAG)					
GPT-4 Turbo	No Prompting	83.5%	1.4	14.2%	5.3
GPT-4 Turbo	+ Prompt Engineering	86.0%	0.8	10.7%	1.4
GPT-4 Turbo	+ Background Info	93.9%	1.0	1.1%	0.2
GPT-4 Turbo	+ Prompt Chaining	96.7%	-	0.0%	-
Model Selection					
GPT-3.5 Turbo	Full RAG + Full Prompting	83.9%	0.6	3.1%	0.7
Gemini Flash 1.5	Full RAG + Full Prompting	89.1%	1.2	4.1%	0.4
GPT-4 Turbo	Full RAG + Full Prompting	96.7%	-	0.0%	-

Notes: This Table reports the accuracy of LLM-generated regulatory classifications against hand-classified data from the Pioneer Institute. We focus on a leave-out sample of 30 municipalities drawn from the Pioneer Institute, and show accuracy measures across a range of strategies and LLMs. To quantify accuracy, we calculate the mean and standard deviation across five runs of each specification. See Appendix Section B for more details on sources of variation across runs. We manually verified disagreements between our headline specification (GPT-4 Turbo Full RAG + Full Prompting) and the Pioneer dataset, so we only run this specification once. See Figure 3 for more details on error adjustments. All accuracy statistics are calculated on the disagreement adjusted testing sample.

Table 4: National Sample Question Means

Panel A: Continuous Questions

Question	National			Income Tercile			Urban/Rural		
	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
How many zoning districts, including overlays, are in the municipality?	14	19	5,471	13	14	14	10	16	13
What is the longest frontage requirement for single family residential development in any district?	92	69	5,213	74	83	113	93	97	79
What is the highest residential minimum lot size? (Thousand Square Feet)	52	45	5,424	36	46	70	60	57	31
What is the lowest residential minimum lot size? (Thousand Square Feet)	10	6	5,440	7	9	13	12	10	8
How many mandatory steps are involved in the approval process for a typical new multi-family building?	4.5	4.3	5,791	4.4	4.4	4.7	4.5	4.5	4.5
For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?	3.1	3.0	5,759	3.2	3.1	3.1	3.1	3.2	3.1
What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?	218	211	5,109	195	222	233	200	222	226

Panel B: Binary Questions

Question	National			Income Tercile			Urban/Rural		
	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	95	99	5,703	99	98	90	95	96	92
Are apartments above commercial (mixed use) allowed in any district?	63	71	5,717	64	67	57	55	66	61
Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?	14	20	5,766	13	14	14	10	15	13
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	80	89	5,795	80	83	80	65	84	83
Does zoning include any provisions for housing that is restricted by age?	45	60	5,129	34	42	57	24	50	52
Are accessory or in-law apartments allowed (by right or special permit) in any district?	33	39	5,781	27	33	37	23	41	22
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	9	10	5,797	8	8	10	5	10	8
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	80	80	5,679	79	81	80	69	86	73
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	24	50	5,540	10	20	40	9	28	27
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	17	18	5,803	11	17	22	10	19	16
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	10	7	4,617	4	8	16	7	12	7
Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?	7	10	5,784	1	4	16	2	7	10
Are there townwide requirements for public hearings on any type of multi-family residential projects?	30	32	5,709	23	30	37	27	30	32

Notes: This table reports the averages of sample questions from our generated national regulatory dataset across a range of demographic associates. We define the count (sample size) as the number of municipalities where the model does not say "I don't know" as the answer. The "Weight" column weights each municipality by its population in the 2022 census of governments. We designate Urban/Rural using the percent overlap of the 2022 shape file for the municipality with the 2020 shape file for urban areas. Specifically, we define Urban as a municipality being 100% in an urban area, Mix as a municipality being partially in an urban area, and Rural as a municipality being 0% in an urban area. We use median income from the 2021 Five-Year American Community Survey (`B19013_001E`). For continuous questions we upper winsorize at the 1% level for frontage, minimum lot sizes, and maximum potential review waiting time.

Table 5: Loadings on Principal Components

	First	Second	Third	Fourth	Fifth
Affordable Incentive	0.42	0.10	-0.37	-0.10	0.04
Affordable Mandate	0.32	0.11	-0.46	-0.17	0.06
Age-Restricted Provisions	0.31	0.00	-0.18	0.15	-0.02
Zoning District Count	0.30	-0.20	-0.00	0.14	-0.07
Wetlands Restricted in Lot Size Calc	0.23	0.20	0.06	0.21	0.11
Permit Cap Or Phasing	0.22	0.03	0.05	-0.22	-0.24
Highest Res Min Lot Size	0.19	0.37	0.22	0.15	-0.02
Longest Frontage Requirement	0.17	0.40	0.21	0.15	0.03
Public Hearing Requirements	0.15	0.11	-0.13	-0.29	-0.09
Max Review Waiting Time	0.12	0.03	0.06	-0.39	0.41
Lowest Res Min Lot Size	0.05	0.47	0.19	0.05	-0.06
Distinct Approval Bodies	0.02	0.02	0.43	-0.31	-0.08
Mandatory Approval Steps	0.01	0.10	0.17	-0.52	0.39
No Conversion to Multifamily	-0.09	0.00	-0.19	-0.38	-0.48
No Flexible Zoning By Right	-0.13	0.04	-0.23	0.11	0.53
Multifamily Not Allowed	-0.14	0.35	-0.08	-0.10	-0.13
No Mixed-Use Buildings	-0.21	0.29	-0.20	-0.04	-0.21
Townhouses Not Allowed	-0.23	0.33	-0.14	0.04	0.03
Accessory Apartments Banned	-0.30	-0.02	-0.06	-0.10	0.06
No Flexible Zoning By Permit	-0.31	0.20	-0.31	0.05	0.10
Variance Explained	0.13	0.11	0.06	0.06	0.05

Notes: This table reports loadings between the first five principal components of our regulatory dataset and specific regulatory questions. We upper winsorize at the 1% level the values for highest residential minimum lot size, lowest residential minimum lot size, longest frontage requirement, and maxaximum review waiting time. We transform the highest residential min lot size variable into a dummy for whether it is above one acre. Missing data, where the model output "I don't know," were imputed with k-nearest neighbors. Prior to performing principal component analysis, all variables were normalized into z-scores. Additionally, each variable was expressed in terms of its expected univariate association with stricter zoning policies, such that more positive values indicate a greater degree of restrictiveness. For example, the variable representing the allowance of multi-family housing was inverted, so that a more positive value indicates that multi-family housing is not permitted, while a more negative value suggests that it is not.

Table 6: Economic and Social Correlates of Housing Regulatory Dimensions

	First PC		Second PC	
	No FE	Metro FE	No FE	Metro FE
Fundamental Amenities (County Level)				
Combined Amenity Index	0.56***		-0.08***	
	(0.02)		(0.03)	
– Natural Amenities Index	0.27***		0.01	
	(0.02)		(0.02)	
– Retail Establishments Index	0.46***		-0.03	
	(0.02)		(0.02)	
– Log Patents Per Capita	0.37***		-0.01	
	(0.02)		(0.02)	
– Log Employment Density	0.49***		-0.13***	
	(0.02)		(0.03)	
Socioeconomic Characteristics (Local Government Level)				
Socioeconomic Index	0.22***	0.10***	0.35***	0.24***
	(0.01)	(0.03)	(0.01)	(0.03)
– Share Mid to High Income	0.20***	0.12***	0.24***	0.18***
	(0.01)	(0.03)	(0.01)	(0.03)
– White Share	-0.03***	-0.03	0.22***	0.22***
	(0.01)	(0.03)	(0.01)	(0.03)
– Share Households Over 35	0.06***	-0.07**	0.38***	0.26***
	(0.01)	(0.03)	(0.01)	(0.03)
– Non-Poverty Rate	0.21***	0.09***	0.26***	0.15***
	(0.01)	(0.02)	(0.01)	(0.02)
– College Degree Share	0.27***	0.16***	0.27***	0.13***
	(0.02)	(0.03)	(0.01)	(0.03)
Government Services (Local Government Level)				
Government Services Index	0.24***	0.12***	0.36***	0.16***
	(0.02)	(0.04)	(0.02)	(0.03)
– Math Test Scores	0.23***	0.16***	0.29***	0.16***
	(0.01)	(0.03)	(0.01)	(0.02)
– Local Revenue Per Student	0.14***	0.01	0.32***	0.17***
	(0.03)	(0.07)	(0.06)	(0.06)
– Total Revenue Per Capita	0.11***	0.04	0.08***	-0.04***
	(0.02)	(0.03)	(0.01)	(0.01)
Housing Density (Local Government Level)				
Housing Density Index	0.05***	0.04	-0.37***	-0.36***
	(0.01)	(0.05)	(0.02)	(0.04)
– Housing Unit Density	0.01	-0.07*	-0.17***	-0.24***
	(0.01)	(0.04)	(0.03)	(0.06)
– Share Structures with 2 or More Units	0.13***	0.10***	-0.32***	-0.35***
	(0.01)	(0.04)	(0.01)	(0.03)
– Share Rental Units	-0.02	0.04	-0.41***	-0.33***
	(0.01)	(0.04)	(0.01)	(0.03)
Geographic Attributes (Local Government Level)				
Geographic Attributes Index	-0.29***	-0.28***	-0.16***	0.12***
	(0.01)	(0.03)	(0.01)	(0.02)
– Log Land Area	0.47***	0.47***	0.07***	-0.05
	(0.01)	(0.03)	(0.01)	(0.05)
– Log Neighbors within 25 Miles	0.17***	0.29***	0.21***	-0.06*
	(0.01)	(0.06)	(0.01)	(0.03)
– Log Miles to Metro Center	-0.14***	-0.07	-0.03**	0.10***
	(0.01)	(0.04)	(0.01)	(0.03)
Political Characteristics (Local Government Level)				
Percent Democrat	0.27***	0.17***	0.02*	-0.22***
	(0.01)	(0.04)	(0.01)	(0.04)

Notes: This Table regresses our two key principal components against a variety of county-level covariates (first panel) as well as municipal-level variables (remaining panels). All variables are standardized (Z-scores). Metro FE columns include state FE for municipalities not within 100 miles of a metro center. Significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors (parentheses) are clustered at metro/state level when including fixed effects. County regressions use population weights. The Combined Amenity Index is the first PC of Natural Amenities (climate, topography, water features), Patents per Capita (2000–2015), and Retail Establishments per capita. Category indices (e.g., Government Services) are first PCs of their components. See Appendix Table A1 for variable definitions and Figure A7 and Table A8 for detailed components.

Table 7: Housing Regulation Intensity and Distance from Metropolitan Centers

Independent Variable	US Census Region				All Regions
	West	Northeast	Midwest	South	
First Principal Component (Value Capture)	-0.10	0.09	-0.23***	-0.16***	-0.07
	(0.06)	(0.06)	(0.08)	(0.05)	(0.04)
Second Principal Component (Exclusionary Zoning)	0.06	0.22***	0.06**	0.03	0.10***
	(0.08)	(0.04)	(0.03)	(0.04)	(0.03)
Accessory Apartments Allowed	-0.11*	0.13***	-0.10*	-0.07**	-0.02
	(0.07)	(0.04)	(0.06)	(0.03)	(0.03)
Flexible Zoning By Right	-0.03	0.06	-0.07	-0.02	-0.01
	(0.05)	(0.04)	(0.05)	(0.04)	(0.02)
Flexible Zoning By Permit	0.01	0.10*	-0.09**	-0.03	0.01
	(0.04)	(0.06)	(0.04)	(0.04)	(0.03)
Affordable Incentive	-0.04	0.00	-0.21**	-0.17***	-0.07***
	(0.07)	(0.05)	(0.09)	(0.04)	(0.03)
Affordable Mandate	-0.01	-0.01	-0.07	-0.03	-0.01
	(0.04)	(0.02)	(0.06)	(0.04)	(0.01)
Zoning District Count	-0.18**	-0.06	-0.24***	-0.22***	-0.17***
	(0.07)	(0.07)	(0.08)	(0.05)	(0.04)
Permit Cap Or Phasing	-0.02	0.03	-0.07***	-0.00	-0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Wetlands Restricted in Lot Size Calc	-0.03	0.14***	-0.03	-0.03	0.04
	(0.07)	(0.04)	(0.05)	(0.05)	(0.03)
Longest Frontage Requirement	0.03	0.18***	-0.03	-0.03	0.06
	(0.06)	(0.03)	(0.07)	(0.04)	(0.03)
Highest Res Min Lot Size	0.00	0.19***	0.01	-0.00	0.06**
	(0.05)	(0.03)	(0.06)	(0.04)	(0.03)
Lowest Res Min Lot Size	0.04	0.20***	0.04	-0.02	0.09***
	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)
Mandatory Approval Steps	-0.03	0.02	0.03	0.01	0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)
Distinct Approval Bodies	0.02	0.03	0.02	0.02	0.02
	(0.04)	(0.02)	(0.05)	(0.03)	(0.02)
Public Hearing Requirements	0.00	0.11***	-0.01	-0.05	0.02
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Max Review Waiting Time	-0.09***	0.05	-0.09**	-0.00	-0.03
	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)
Multifamily Allowed	0.13	-0.04	0.07	0.08**	0.03
	(0.10)	(0.05)	(0.06)	(0.04)	(0.03)
Mixed-Use Buildings	-0.05	-0.03	-0.08***	-0.12***	-0.07***
	(0.05)	(0.04)	(0.03)	(0.04)	(0.02)
Conversion To Multifamily	-0.04	-0.05	0.08***	0.02	0.00
	(0.04)	(0.04)	(0.02)	(0.04)	(0.02)
Townhouses Allowed	-0.06	-0.10*	-0.11***	-0.03	-0.08***
	(0.05)	(0.06)	(0.04)	(0.04)	(0.02)
Age-Restricted Provisions	-0.07	-0.03	-0.13**	-0.11**	-0.08***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.03)
Metro Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: This Table shows a regression of distance to city center against a variety of regulatory measures. We subset to within 100 miles of the center of a metropolitan area which accounts for 3,605 observations in our sample. The dependent variable is log distance to metro center. A positive coefficient indicates that the variable increases with log distance from the metro center and a negative coefficient means that the variable decreases with log distance from the metro center. See Appendix Table A9 for full definitions of zoning questions. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We cluster standard errors at the metro level.

Table 8: Historical Determinants of Minimum Lot Size Gradients

Panel A: Raw Distance

Regional Slopes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Midwest	85.19***	79.37***	28.19	93.13***	72.31***	83.45***	12.90	82.36**
	(24.01)	(23.05)	(21.97)	(31.32)	(27.86)	(24.79)	(23.24)	(33.02)
Northeast	345.2***	358.7***	205.6***	321.8***	330.6***	351.1***	220.9***	154.5***
	(44.42)	(42.69)	(44.12)	(47.35)	(43.67)	(45.39)	(42.63)	(51.48)
South	6.283	2.912	-7.213	-1.022	10.15	19.91	-11.41	2.300
	(26.41)	(26.14)	(27.52)	(34.01)	(29.14)	(30.93)	(25.97)	(40.64)
West	-5.160	-1.264	-10.50	9.998	5.697	0.226	-31.51**	7.588
	(10.45)	(12.95)	(12.07)	(13.44)	(14.46)	(12.82)	(14.07)	(15.30)
Controls								
Nearest Metro		✓						✓
Density × Nearest Metro			✓					✓
Foreign Born × Nearest Metro				✓				✓
Owner-occupied × Nearest Metro					✓			✓
Manufacturing × Nearest Metro						✓		✓
Manufacturing × Density							✓	✓
N	3,353	3,344	3,344	3,344	3,344	3,344	3,344	3,344
R-squared	0.136	0.248	0.329	0.263	0.267	0.266	0.288	0.371

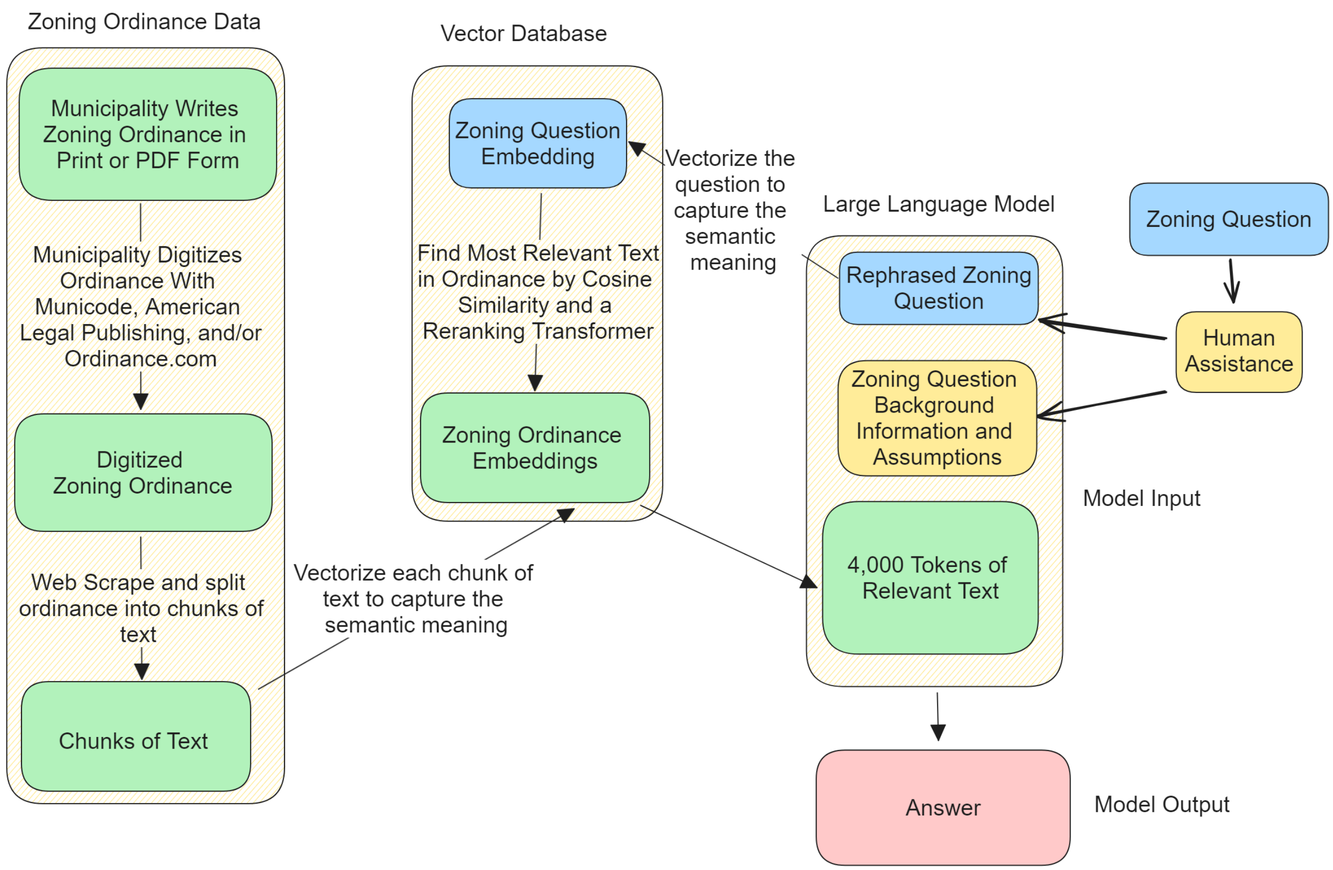
Panel B: Log Distance

Regional Slopes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Midwest	71.49	271.2	-177.1	541.7**	105.7	310.1	-275.2	148.1
	(183.3)	(183.1)	(178.2)	(238.1)	(220.0)	(209.5)	(201.4)	(285.6)
Northeast	5,087***	6,033***	1,936***	5,316***	5,763***	6,406***	3,087***	1,291**
	(655.3)	(631.1)	(708.9)	(676.6)	(658.1)	(681.8)	(655.4)	(641.0)
South	-317.8	-251.7	-332.2	-418.8	-298.0	10.49	-380.2	-432.6
	(322.7)	(313.3)	(313.7)	(351.7)	(327.4)	(354.9)	(310.2)	(362.9)
West	169.0	252.5	214.3	317.3*	380.2*	244.1	-9.487	238.6
	(175.0)	(193.9)	(193.4)	(184.7)	(218.4)	(197.8)	(210.2)	(220.3)
Controls								
Nearest Metro		✓						✓
Density × Nearest Metro			✓					✓
Foreign Born × Nearest Metro				✓				✓
Owner-occupied × Nearest Metro					✓			✓
Manufacturing × Nearest Metro						✓		✓
Manufacturing × Density							✓	✓
N	4,479	4,467	4,467	4,467	4,467	4,467	4,467	4,467
R-squared	0.131	0.244	0.323	0.262	0.263	0.259	0.285	0.362

Notes: This table examines the regional variation in minimum lot size gradients relative to distance from metropolitan centers. The dependent variable is the lowest residential minimum lot size requirement (in square feet). Panel A presents results using linear distance measures, and restricts to municipalities within 50 miles of a metro center while Panel B uses a log of distance for municipalities within 100 miles of a metro center. Column (1) shows the baseline regional relationships and subsequent columns adding various controls. Column (2) additionally controls for a fixed effect for the nearest metropolitan area. Column (3) controls for an interaction of a nearest metro fixed effect with historical density in the municipality from 1940. Column (4) controls for the foreign-born share, measured in 1940 at the municipality level, interacted with a nearest metro fixed effect. Column (6) controls for manufacturing share of employment in 1940 at the county level interacted with a nearest metro fixed effect. Column (7) controls for manufacturing worker share at the county level interacted with historical density at the municipal level, with both variables measured in 1940. Column (8) includes all previous controls. Standard errors are shown in parentheses and are clustered at the municipality level. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figures

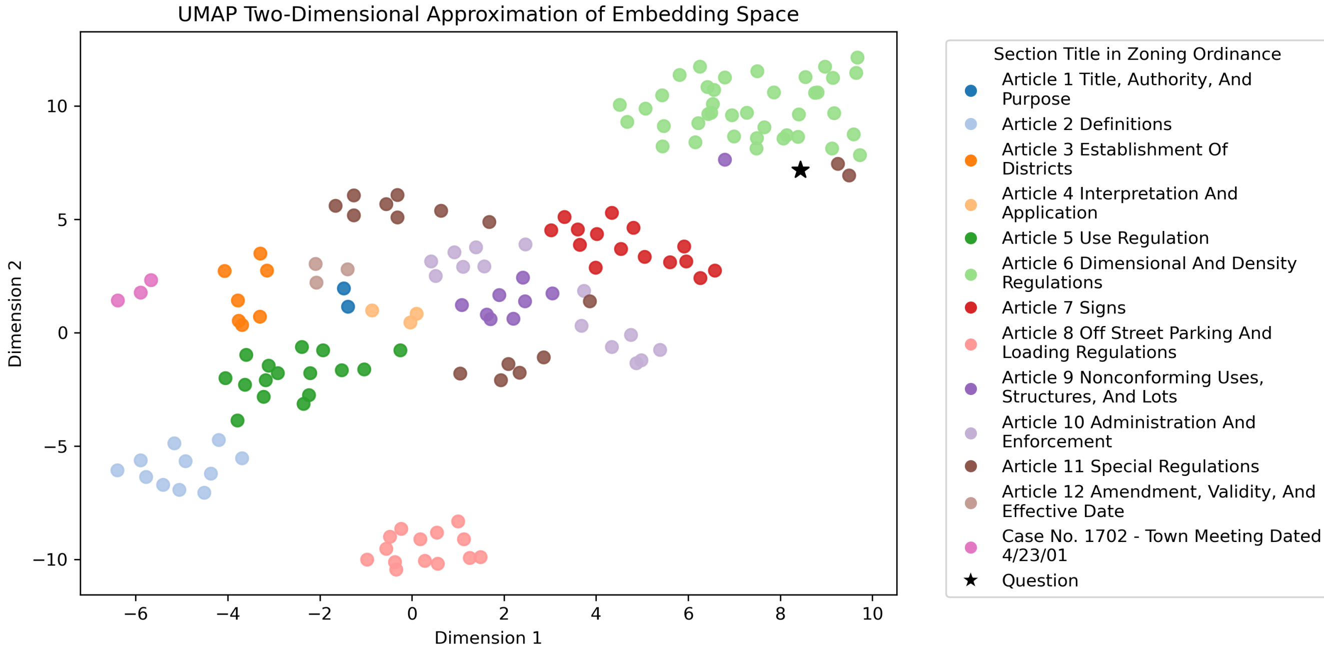
Figure 1: Model Overview



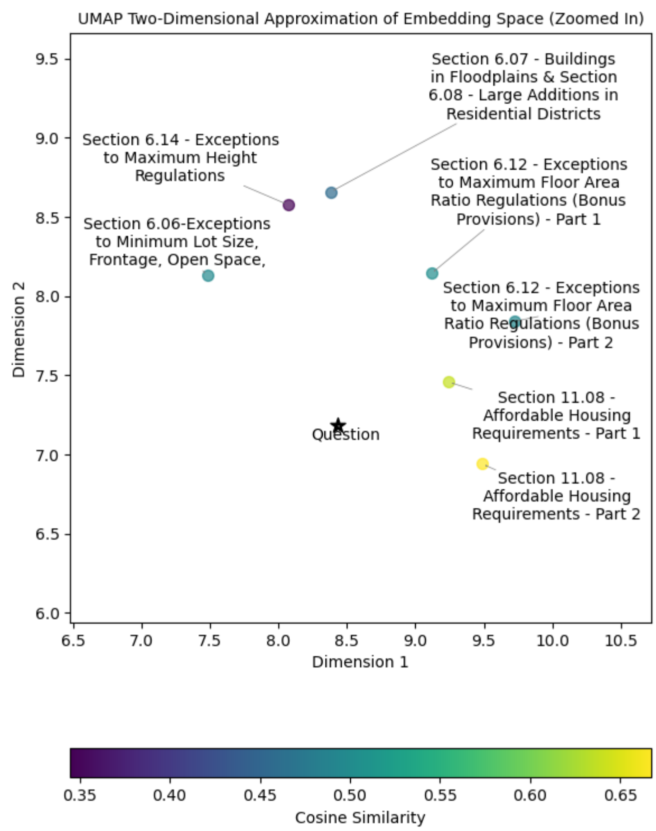
Notes: This Figure illustrates our overall process for Generative Regulatory Measurement. We download each section within a zoning ordinance separately. We split up sections that are longer than one thousand tokens into chunks of at most one thousand tokens. We also combine adjacent sections of less than 50 tokens. So, each section of text varies in length but is between 50 and one thousand tokens. We vectorize each chunk of text using OpenAI embeddings models (link). Specifically, we use the "text-embedding-3-large" algorithm. Sometimes digital aggregators leave tables in image form, especially the aggregator Ordinance.com. So that the model can still read the table, we transcribe images of tables using Amazon Textract. We elicit an open-ended response to each question and then use function calling to parse out a structured answer (i.e., to ascertain whether an answer is "Yes," "No," or "I don't know" to a binary question). Question background information and model assumptions are based on a combination of the "Issue Overview" and the "Research Coding" sections for each question from the Pioneer study as well as from trial and error in the training sample of municipalities. Rephrased zoning questions came entirely from trial and error on the training sample. Ordinances from digital aggregators (Municode, American Legal Publishing, and Ordinance.com) are either entirely about zoning, partially about zoning (i.e., have one or more sections about zoning), or not about zoning at all. We filter out ordinances not at all about zoning by searching for key phrases, table headers, and zoning district names (i.e., R-1 for the first residential zoning district). See Appendix Section C for further details on question background information and assumptions as well as system prompts.

Figure 2: Embeddings of Municipal Code Sections

Panel A: Full Embedding Space

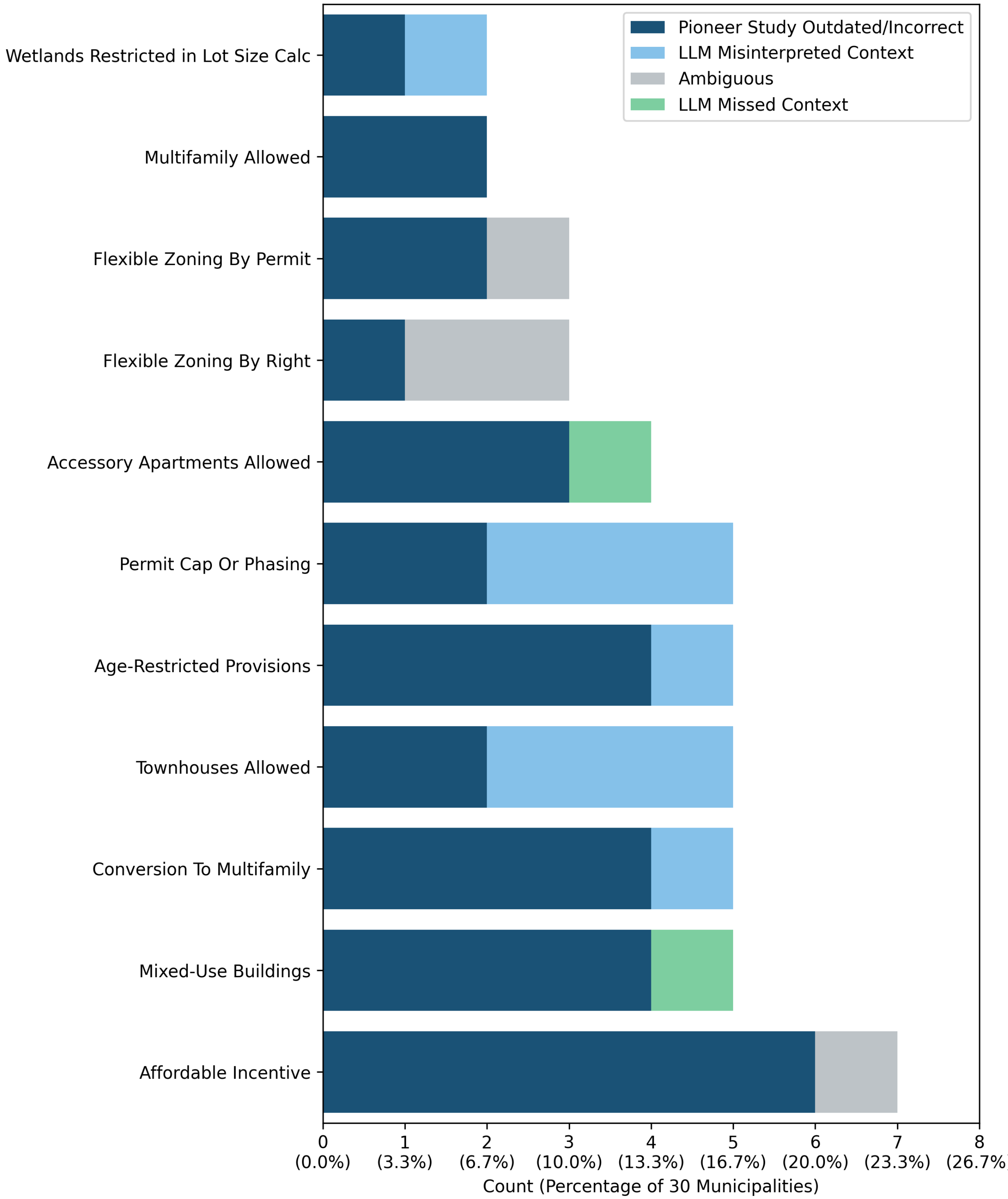


Panel B: Zoomed View of Similar Sections



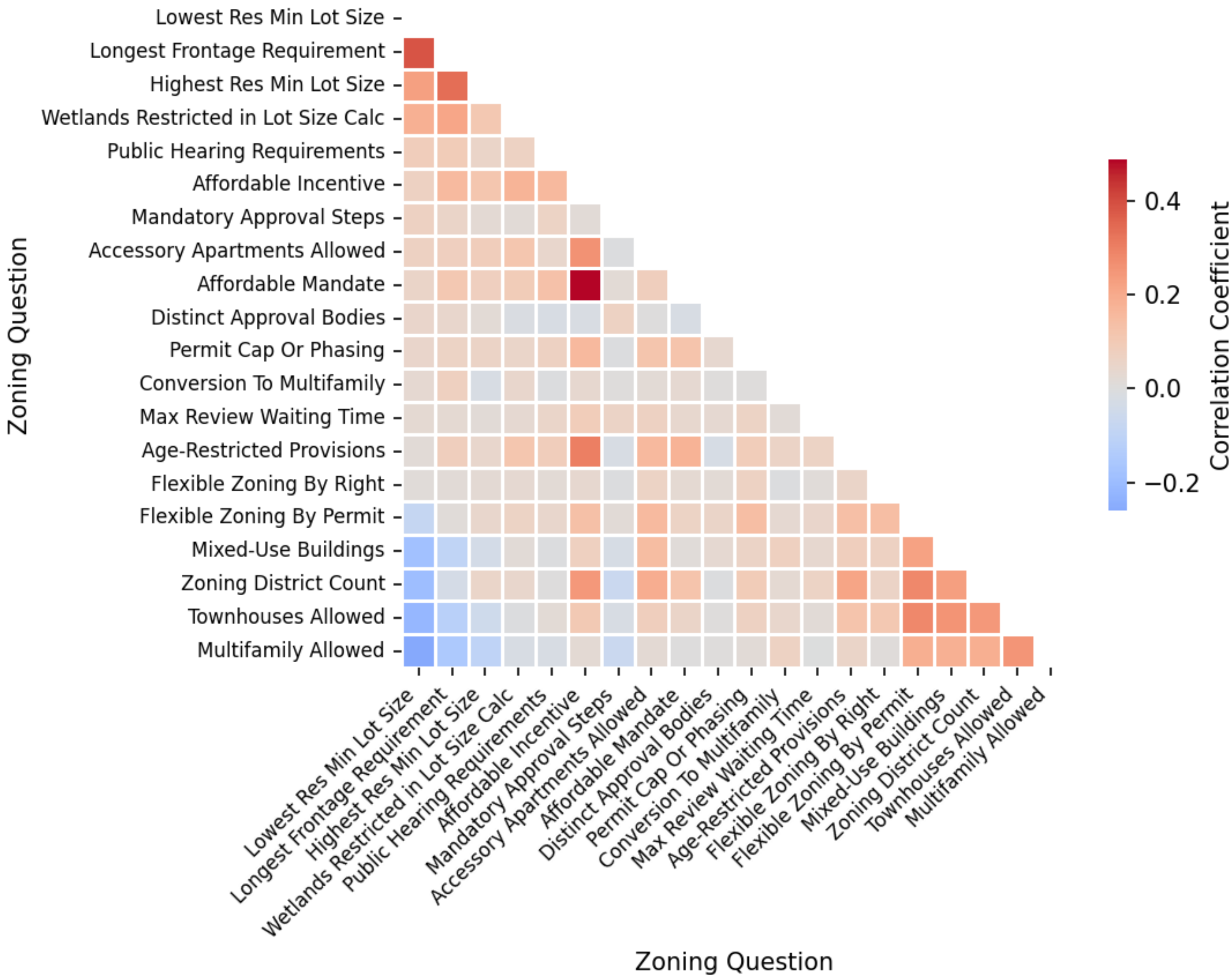
Notes: This Figure shows the two-dimensional projection of embeddings from sections of municipal codes for Arlington, Massachusets. Panel A shows the full embedding space, while Panel B provides a zoomed-in view of closely related sections. The embeddings are created using OpenAI's text-embedding-3-large model and visualized using UMAP dimensionality reduction.

Figure 3: Sources of Discrepancy Between LLM and Human Classifications of Zoning Regulations



Notes: This Figure reports reasons for disagreement between our generated model data from GPT-4 Turbo and the original version of the Pioneer Institute data for binary questions. We first ran GPT-4 Turbo on the testing sample of 30 randomly selected municipalities that were included in the Pioneer Institute's study but were not used to train our model. We then identified the binary questions where the model responses disagreed with the Pioneer study. A law student reviewed each of these disagreements individually to determine the reason for the discrepancy, classifying them into the categories shown in the chart. When measuring the performance of the model we adjust for disagreements where the Pioneer study was outdated/incorrect and also drop ambiguous cases.

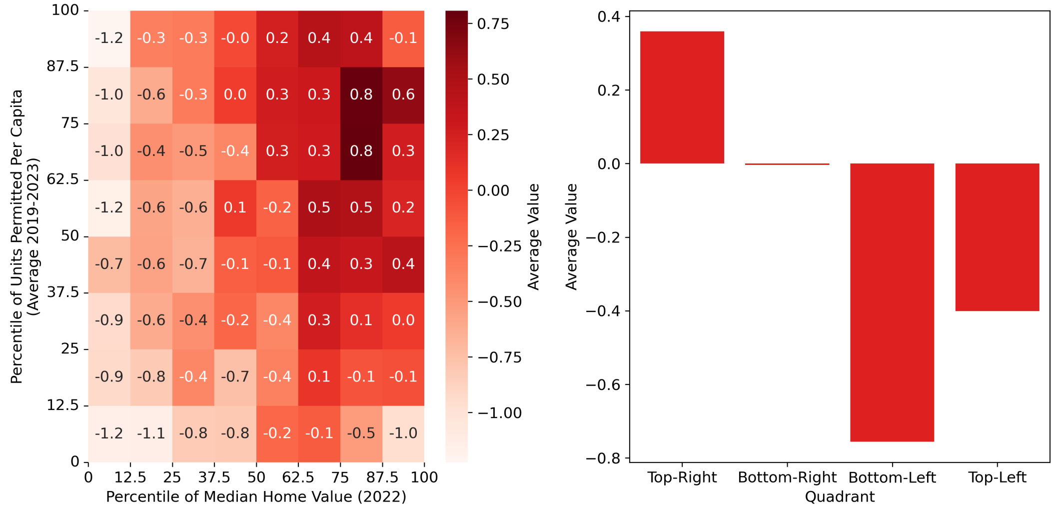
Figure 4: Pairwise Correlations Between Zoning Questions



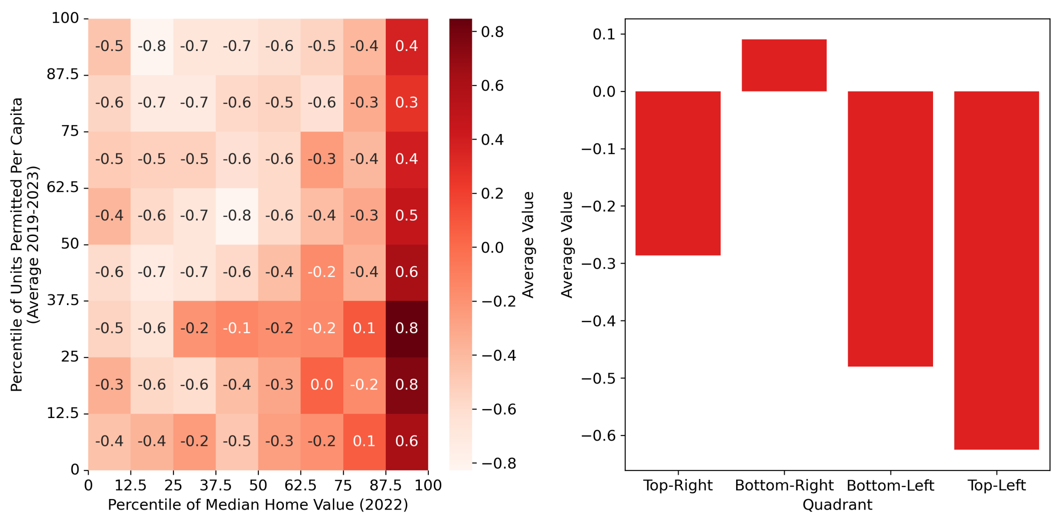
Notes: This heatmap illustrates the pairwise correlations between various zoning regulations across U.S. municipalities. Each cell represents the correlation coefficient between two zoning measures, with color intensity indicating the strength and direction of the relationship. Darker red indicates stronger positive correlations, while darker blue represents stronger negative correlations. White or light-colored cells suggest weak or no correlation. See Appendix Table A9 for full question names. See Table 4 footnote for details on sample construction.

Figure 5: Regulatory Dimensions and Housing Market Dynamics

Panel A: First Principal Component (Value Capture)



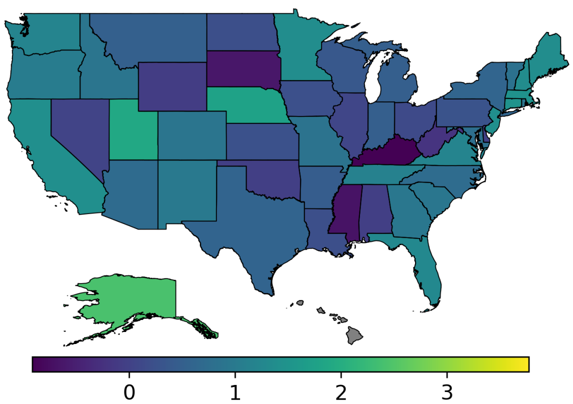
Panel B: Second Principal Component (Exclusionary Zoning)



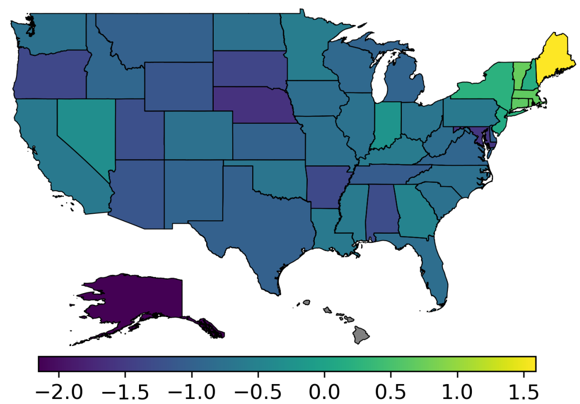
Notes: This Figure illustrates the relationship between two principal components of housing regulation and key housing market outcomes. Panel A shows the first principal component, associated with value capture. Panel B displays the second principal component, which corresponds to exclusionary zoning practices. The x-axis in both panels represents the percentile of median housing value, drawn from the 2022 ACS, while the y-axis measures the percentile of units permitted per capita (averaged from 2019–2023). Darker colors illustrate a larger correlation in the heatmap between each regulatory principal component and each coordinate of prices and building. The right figures show the overall correlation between each quadrant of the space of housing prices and quantities and the regulatory outcome. See Table A1 for details on variable definitions.

Figure 6: Nationwide Maps of Population-Weighted Averages

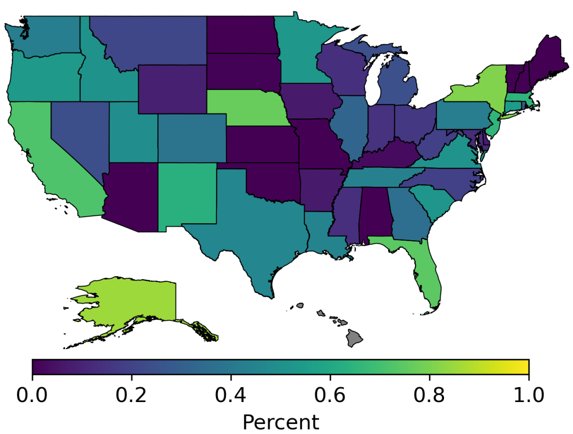
Panel A: First PC



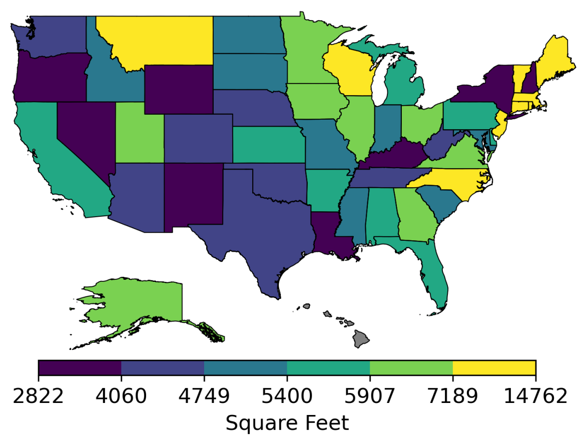
Panel B: Second PC



Panel C: Affordable Incentives/Mandates



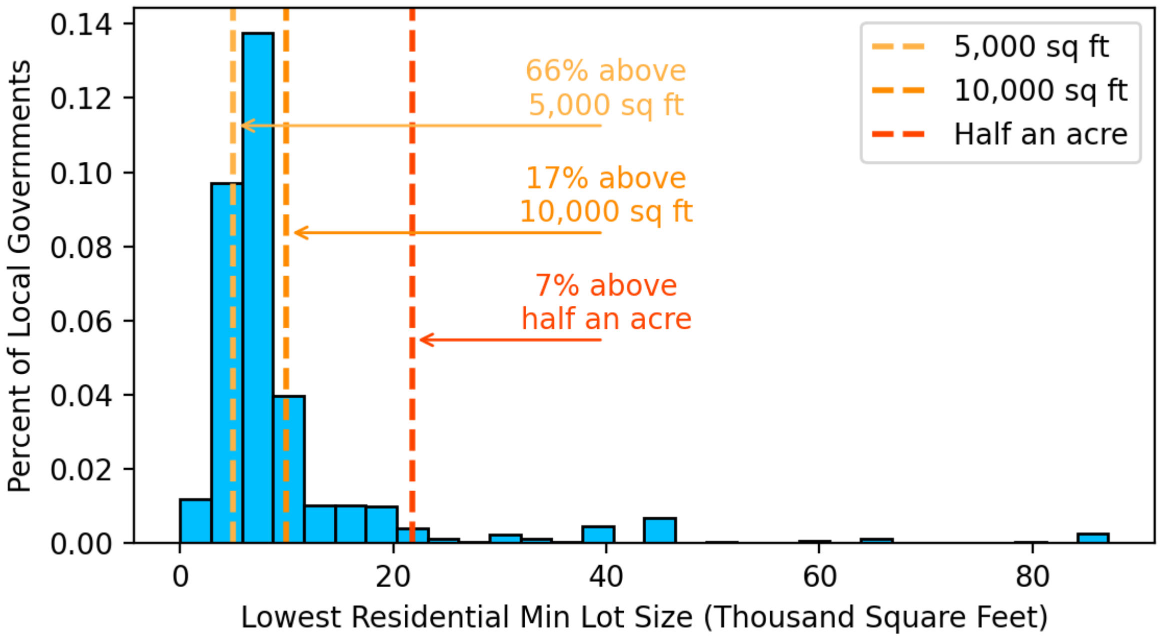
Panel D: Lowest Residential Minimum Lot Size



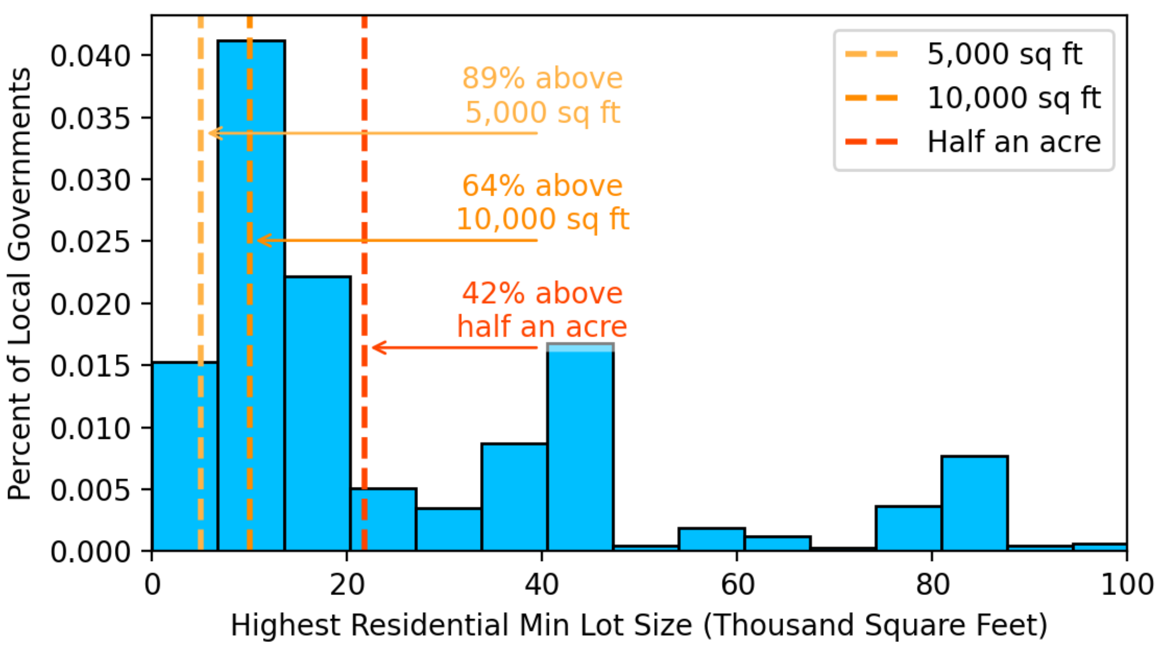
Notes: This Figure plots regulatory variables by state, weighted by local municipality population. We use the 2022 ACS Population as the population weight. Hawaii is grey because only one municipality (Honolulu) is in the dataset. For county level maps see Appendix Figure A6.

Figure 7: Minimum Lot Size Distribution

Panel A: Lowest Residential Minimum Lot Size



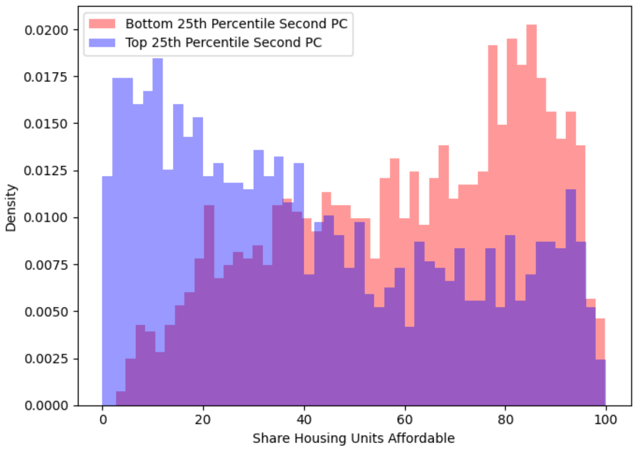
Panel B: Highest Residential Minimum Lot Size



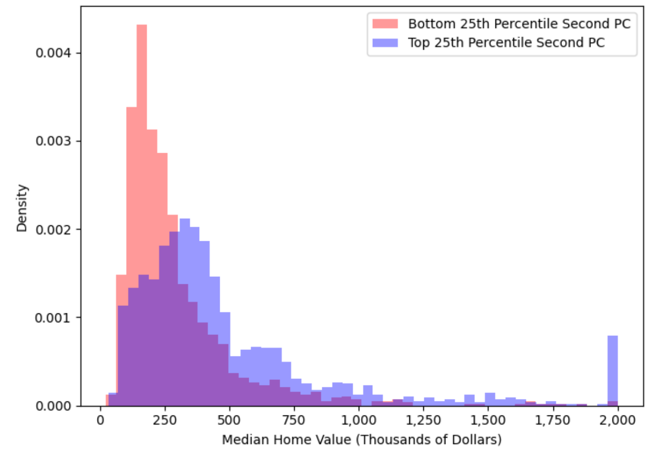
Notes: This histogram shows the distribution of minimum lot size requirements across local governments. The x-axis represents lot size in square feet, with key thresholds labeled. The y-axis shows the percentage of local governments falling into each lot size category. Vertical lines mark important thresholds: 5,000 sq ft (common suburban lot size), 10,000 sq ft (quarter-acre), and 21,780 sq ft (half-acre). We measure minimum lot sizes across all residential districts, and show the lowest such requirement within a municipality in Panel A, and the highest such requirement in Panel B. The x-axis in Panel B stops at 100 thousand square feet, though 12% of local governments have highest residential minimum lot sizes above this level.

Figure 8: Second Principal Component and Housing Affordability

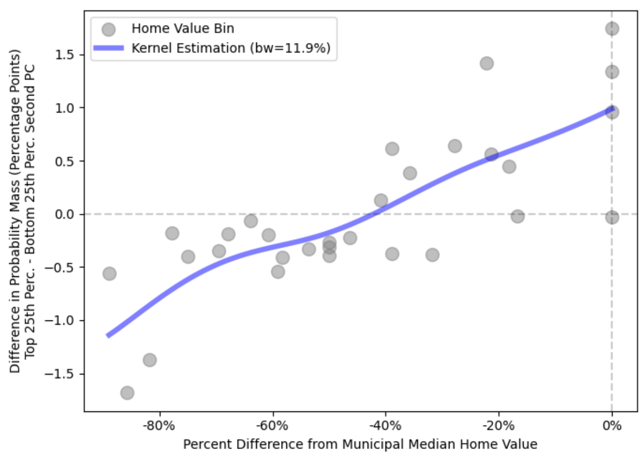
Panel A: Share Units Affordable



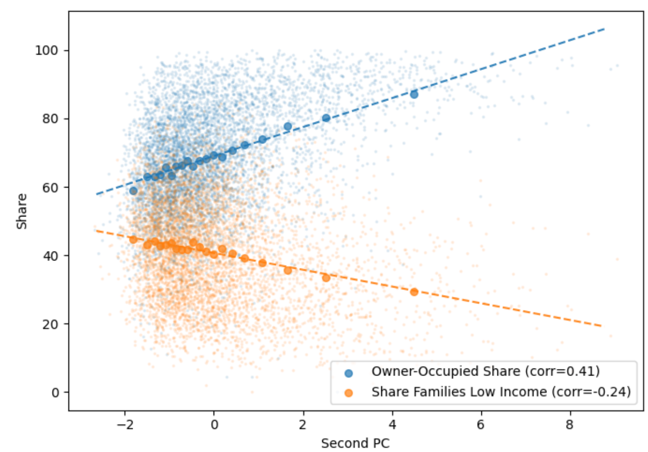
Panel B: Median Home Value



Panel C: Difference in Probability For Left Tail of Home Value

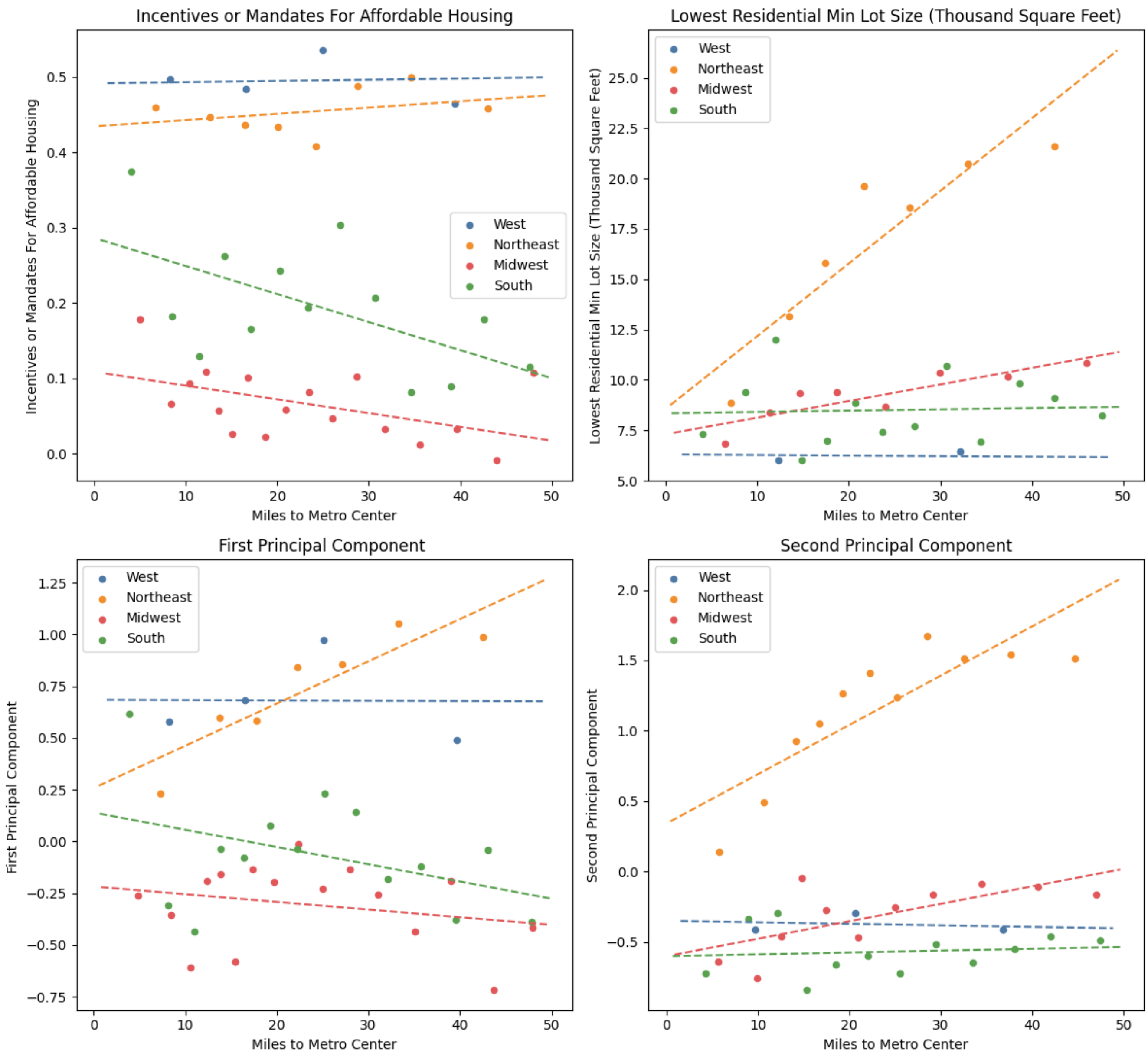


Panel D: Owner-Occupied Share and Low-Income Family Share by Second PC



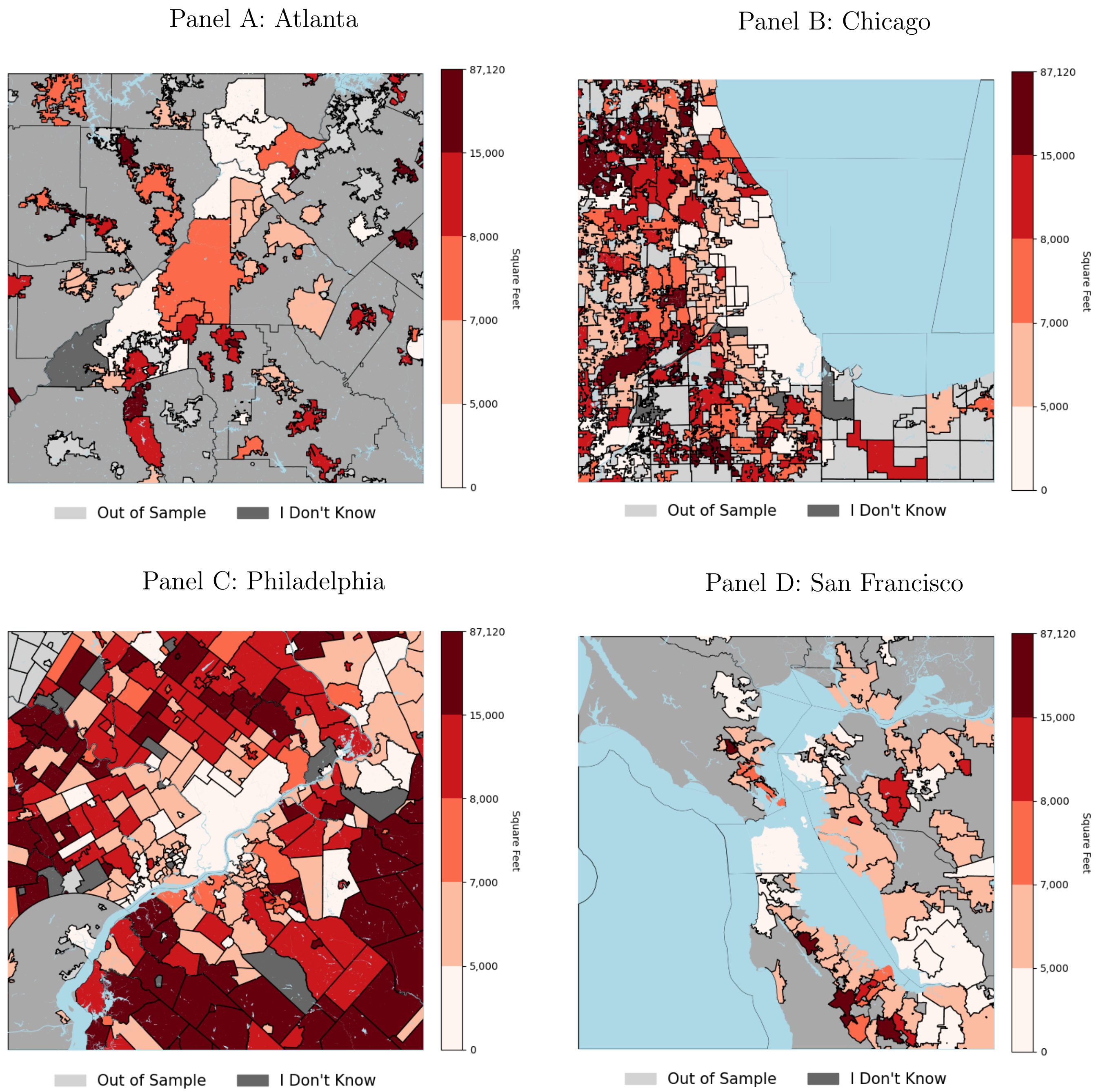
Notes: This paper highlights the associations of the second principal component against different measures of housing affordability. Panel A splits the sample into the top and bottom 25th percentile on the second PC, ad plots the distribution of affordable housing units in each sample. Rental units are considered affordable if monthly rent does not exceed 30% of the monthly median income; owner-occupied units are affordable if their value is less than three times the annual median income. Panel B shows the distribution of median house values across municipalities across the top and bottom 25th percentiles. In Panel C, we compare the distribution of home values between local governments with high and low second principal component (PC) scores for owner occupied units from the 2022 American Community Survey (ACS). We focus on local governments with median home values between \$200k and \$500k, where the most overlap occurs in the second PC median home value histogram (Panel B). For each ACS home value range:\$200k-–\$249k, \$250k-–\$299k, \$300k—\$399k, and \$400k—\$499k—we filter the dataset to include only local governments with median home values in that range. Within each filtered subset, we divide local governments into the top and bottom 25th percentiles of the second PC. We then calculate the percentage point difference in the probability mass of housing units for each ACS home value bin for high and low second PC local governments. To simplify the analysis, we collapse all home values below \$100k into a single category with a midpoint of \$50k. The x-axis measures the percentage difference between the bucket's median home value midpoint (e.g., \$225k for the \$200k—\$249k bucket) and the midpoint of each ACS home value bin, while the y-axis represents the difference in probability mass between high and low second PC municipalities. Finally, a kernel regression, with bandwidth determined using Silverman's rule of thumb, is fitted to the plotted points to highlight the underlying trend. Panel D shows the relationship between different levels of the second PC and the municipal owner-occupied share (correlation of 0.41) as well as the share of families that are low-income (income below 80% of the state median income, correlation of -0.24).

Figure 9: Spatial Variation of Zoning Regulations Relative to Metropolitan Centers



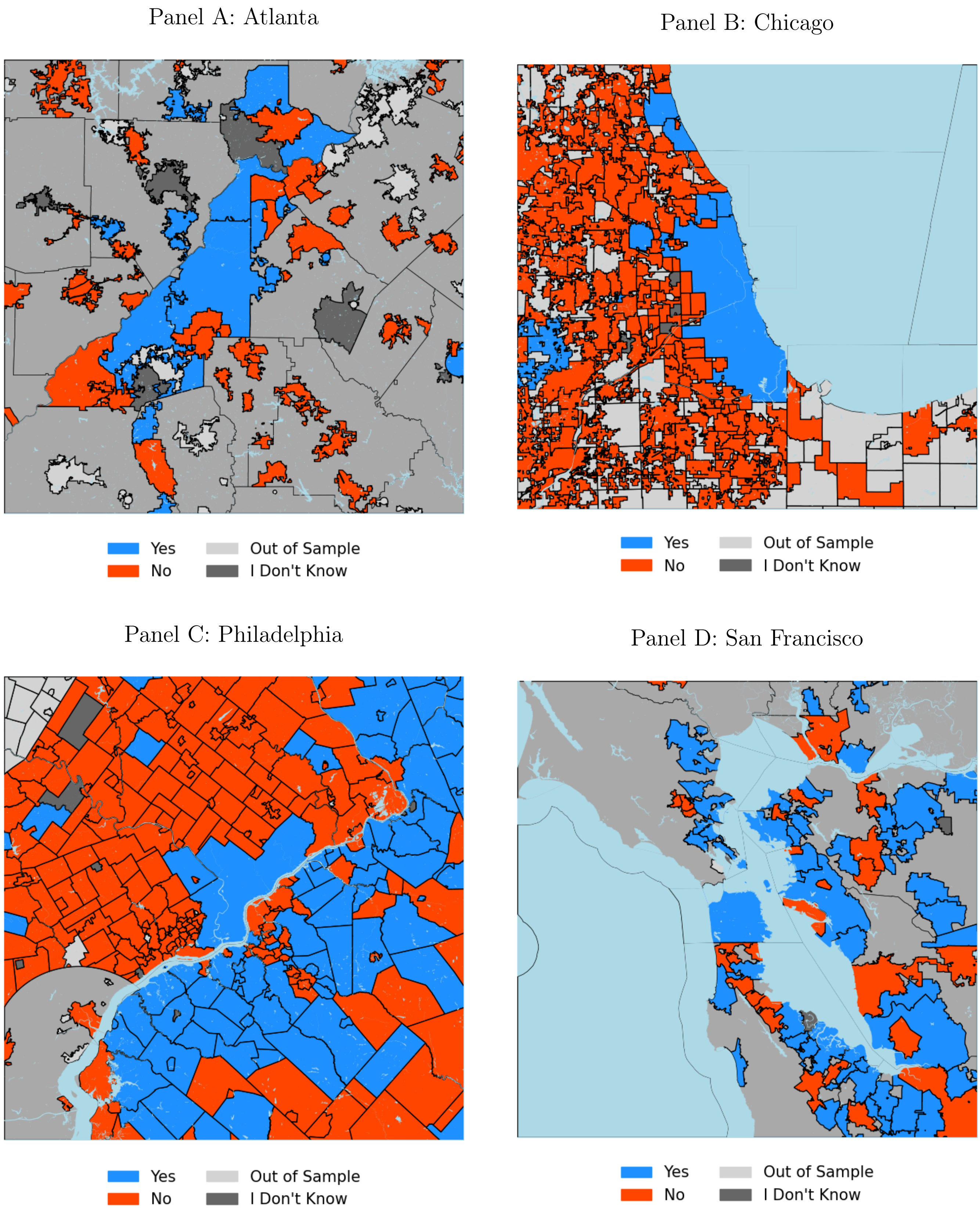
Notes: Note: We plot regulatory variables at the local government level based on the distance from the center of the respective metro, defined as city hall. We show whether a city has an affordable housing incentive or mandate; the minimum lot size; the first principal component of housing regulation (value capture), and the second principal component of housing regulations (exclusionary zoning).

Figure 10: Lowest Residential Minimum Lot Sizes For Select Metro Areas



Notes: This Figure plots the lowest local government wide minimum lot size for each local government in the Atlanta (Panel A), Chicago (Panel B), Philadelphia (Panel C), and the San Francisco Bay Area (Panel D) metro areas. Within each map we plot all Census-designated places as well as county subdivisions that represent local governments. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files. Each map shows roughly a 100km × 100km square area. Non-Incorporated areas are shown in light grey, and areas for which the model reported "I don't know" are shown in dark gray.

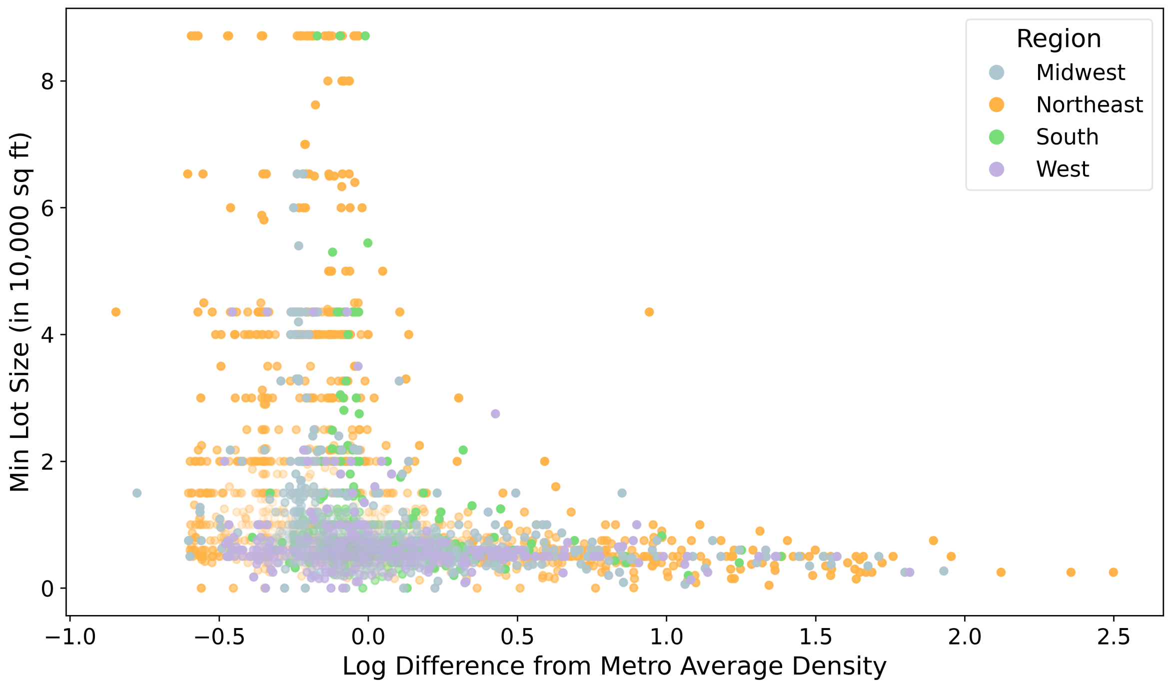
Figure 11: Affordable Housing Incentives or Mandates For Select Metro Areas



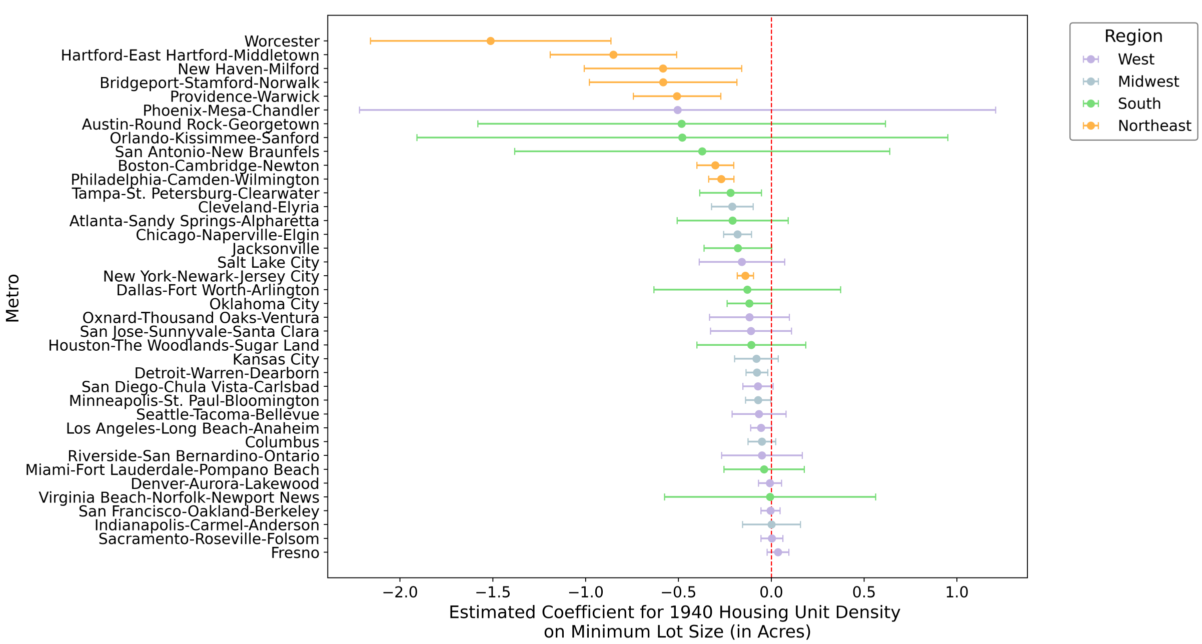
Notes: This Figure plots the presence of affordable housing incentives for each local government in the Atlanta (Panel A), Chicago (Panel B), Philadelphia (Panel C), and the San Francisco Bay Area (Panel D) metro areas. Areas with incentives or mandates are plotted in blue; areas without such mandates are shown in red. Within each map we plot all Census-designated places as well as county subdivisions that represent local governments. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files. Each map shows roughly a 100km $\times$ 100km square area. Non-Incorporated areas are shown in light grey, and areas for which the model reported "I don't know" are shown in dark gray.

Figure 12: Minimum Lot Sizes and Historical Density

Panel A: Deviation from Metro Historical Density and Minimum Lot Size



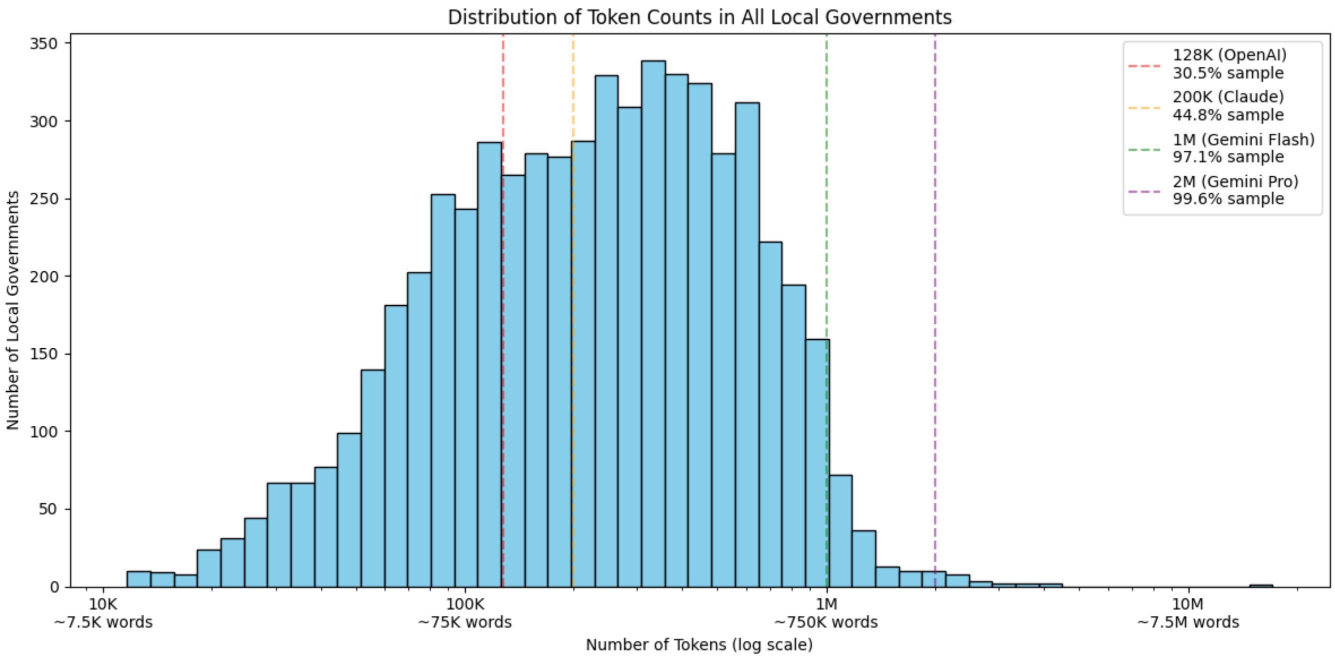
Panel B: OLS Regression of Min Lot Size against Historical Housing Density



Notes: Panel A of this Figure examines the relationship between historical population density and contemporary minimum lot size requirements. Each point represents a municipality, with the x-axis showing the deviation of each municipality's density in 1940 from the closest metropolitan area's average historical density (standardized). The y-axis showing minimum lot size requirements in thousands of square feet. Panel B presents an OLS regression analysis of modern minimum lot size requirements against 1940 housing unit density. The unit of observation is municipality, and we estimate coefficients at the MSA level. In both figures, different regions are indicated through colors.

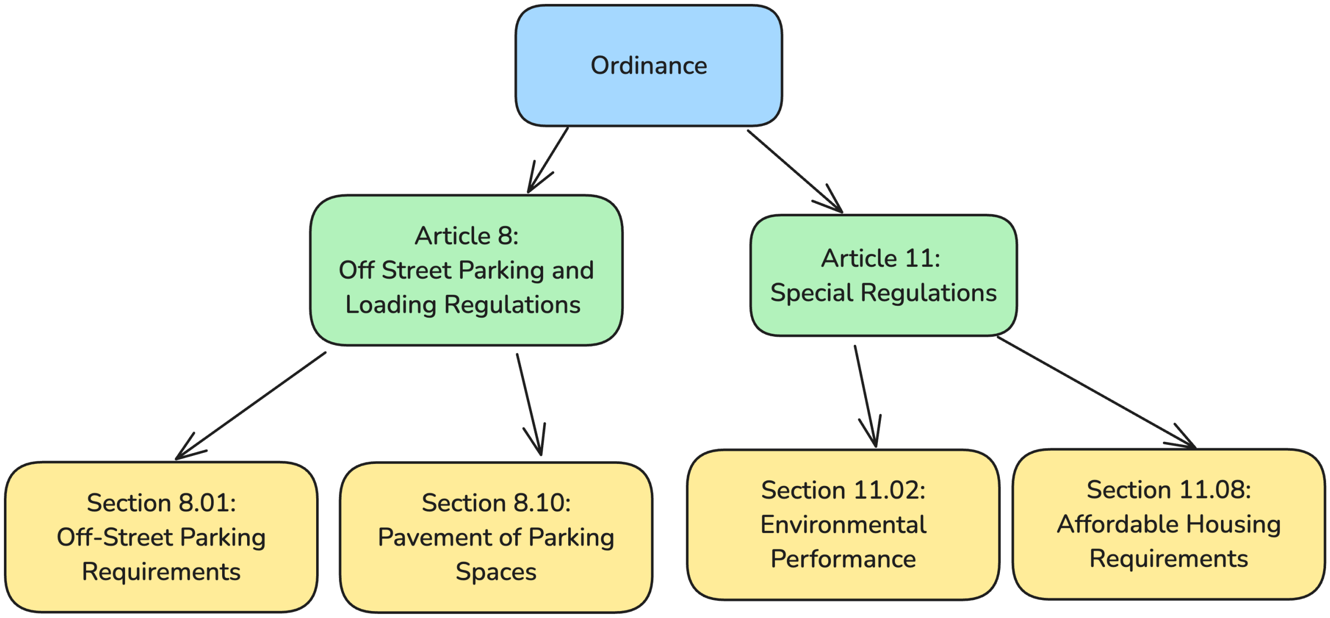
A Appendix: Additional Graphs and Tables

Figure A1: Distribution of Ordinance Lengths



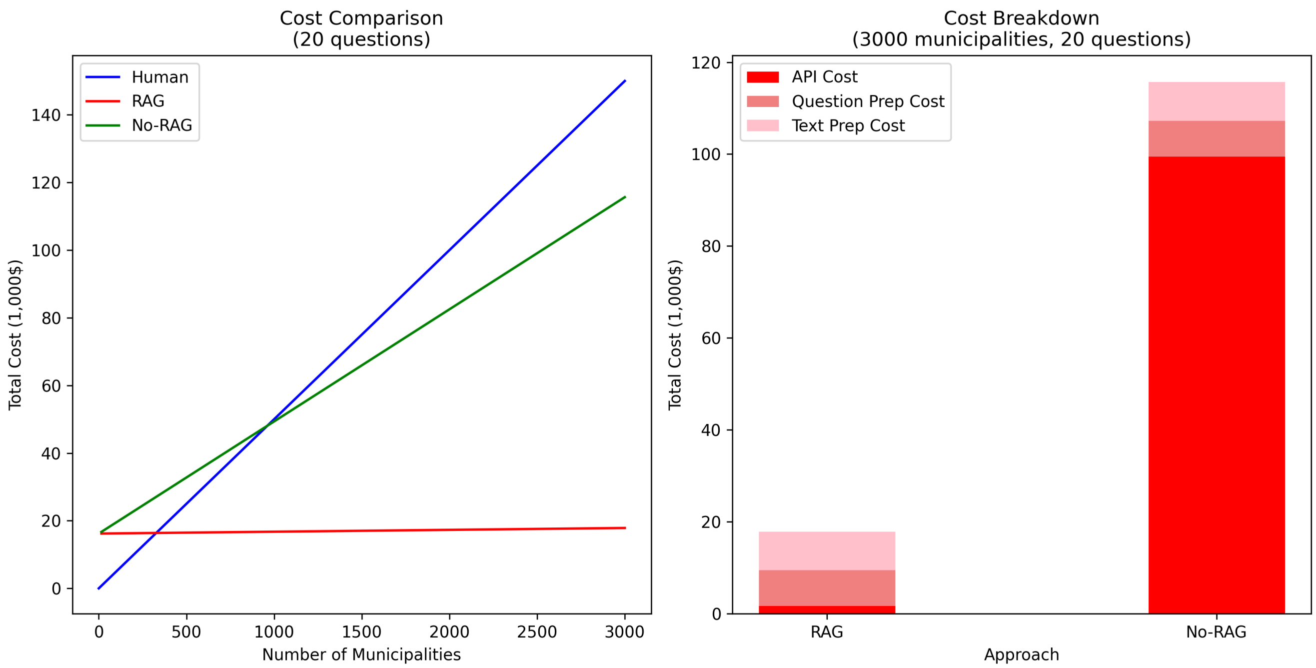
Notes: This figure shows a distribution of token counts for municipal zoning ordinances in our sample. Vertical dashed lines represent maximum context length for various popular Large Language Models in number of tokens. We use the OpenAI tokenizer used in GPT-4 Turbo (cl100k_base), though tokenizers for Claude and Gemini produce similar numbers.

Figure A2: Splitting Up Ordinance Into Chunks Using Hierarchical Structure: Arlington, MA Example



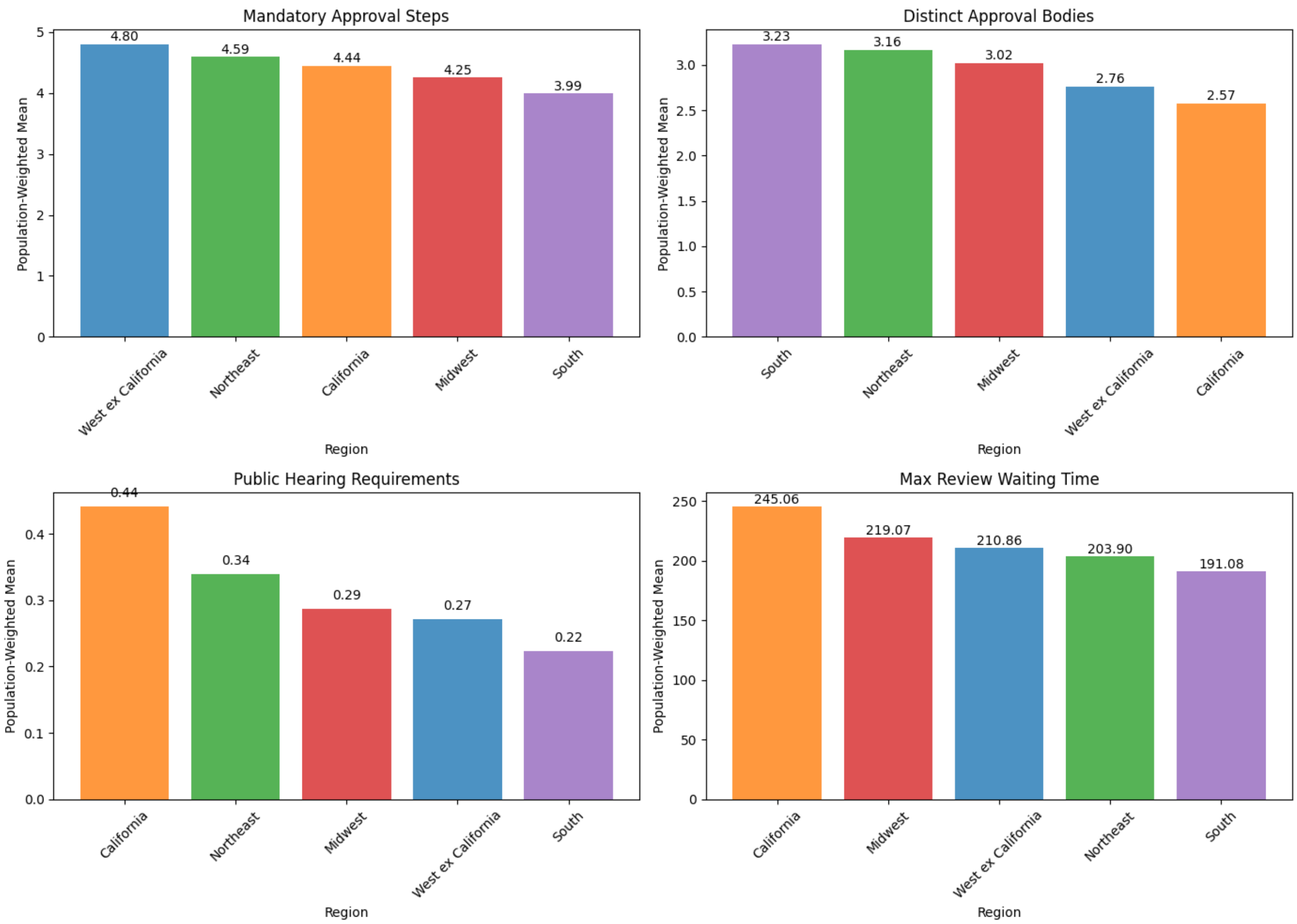
Notes: This Figure shows how we split ordinances into chunks of text using the hierarchical structure of the zoning ordinance. We first split the ordinance into chunks based on the sections and subsections of the document. Then we combine subsections together if they are less than 50 tokens in length and split them up if they are more than 1000 tokens in length.

Figure A3: Comparison of Human vs. LLM Cost



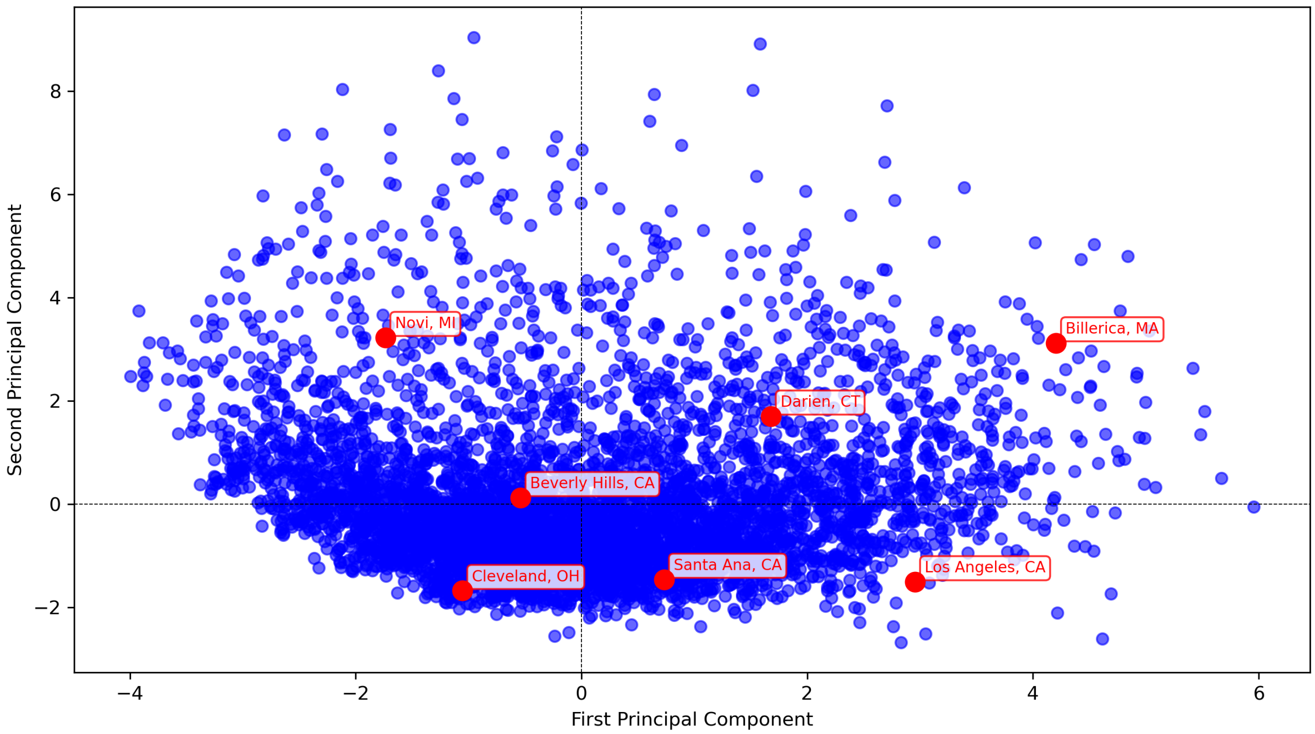
Notes: This figure compares the cost of human versus LLM-based analysis of zoning regulations across different numbers of municipalities. We assume: (1) Lawyers take an average of five minutes per question-municipality pair, compensated at $50/hour; (2) For the RAG approach, each LLM API call involves 4,000 input tokens and 500 output tokens, with costs of $5 and $15 per million tokens respectively; (3) Question preparation costs $390/question, including lawyer time for answering 60 municipalities (30 for training, 30 for testing) and 10 hours of human labor at $140 to train the model; (4) Scraping, cleaning, and embedding each municipality's ordinance costs $2.80. For the No-RAG approach, input tokens average 330,000 per question-municipality pair based on mean ordinance length. All costs are assumed to scale linearly with the number of municipalities and questions, shown here for 20 questions.

Figure A4: Measuring Housing Process Variation



Notes:. This Figure plots population weighted averages of four LLM-generated process regulations at the census region level.

Figure A5: Relationship between First and Second Principal Components

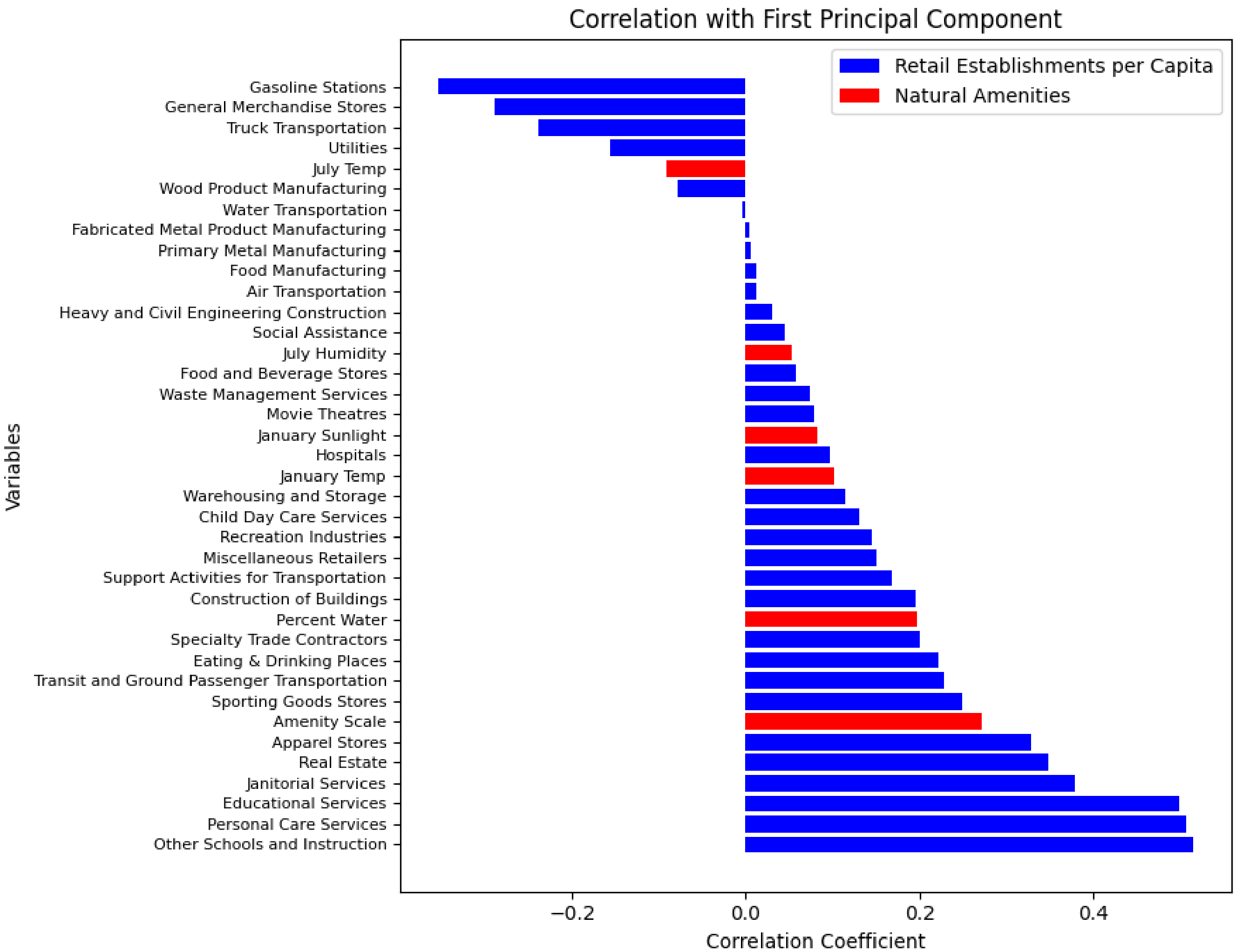


Notes: This scatter plot illustrates the relationship between the first two principal components of our zoning regulation analysis across U.S. municipalities. The x-axis represents the first principal component (PC1), which we interpret as a measure of value capture. The y-axis shows the second principal component (PC2), which corresponds to exclusionary zoning practices. Each point represents a municipality. Municipalities in the upper right quadrant tend to have both value capture and more exclusionary practices, while those in the lower left capture less value and exclude less.

Figure A6: County Maps of Zoning Measures

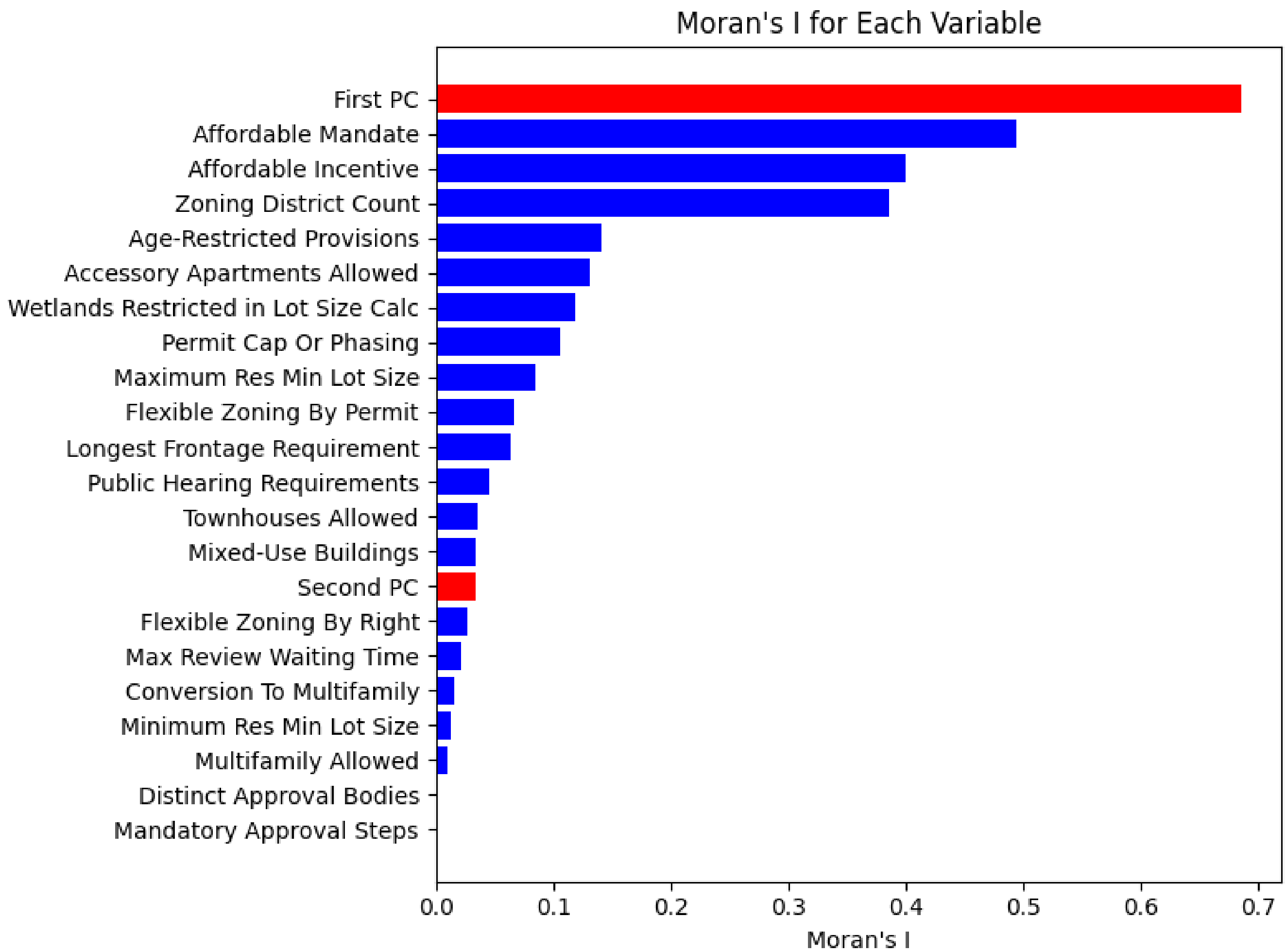
Notes: This Figure shows county-level maps of four key zoning measures across the United States. Panels A–B show the first principal component (PC1) of zoning regulations, interpreted as value capture. Panels C–D display the second principal component (PC2), associated with exclusionary zoning practices. Panels E–F illustrate the prevalence of affordable housing incentives or mandates. Panels G–H depict the minimum residential lot size requirements. For each measure, we present both unweighted averages (left column) and population-weighted averages (right column). Darker colors indicate higher values or greater prevalence of each measure. We use the 2022 ACS Population as the population weight. Hawaii is grey because only one municipality (Honolulu) is in the dataset. For population weighted state maps see Figure 6.

Figure A7: Granular County Amenities Correlations with Principal Components



Notes: This Figure shows the correlation of the first principal component of regulatory variables against a range of amenities. Retail establishment data are taken from the U.S. Census Bureau's County Business Patterns (CBP) 2022 dataset. Correlations are calculated at the county level, with the number of establishments for each industry normalized to per capita measures using county population estimates. Principal Component indices are population-weighted averages of municipality level data, aggregated to the county level. Industries are classified using 2017 NAICS codes. Natural amenity data comes from USDA county natural amenities dataset.

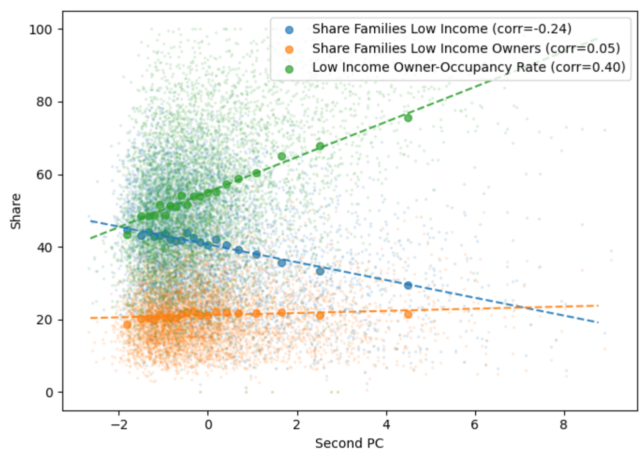
Figure A8: Spatial Autocorrelation of Zoning Regulations and Indices



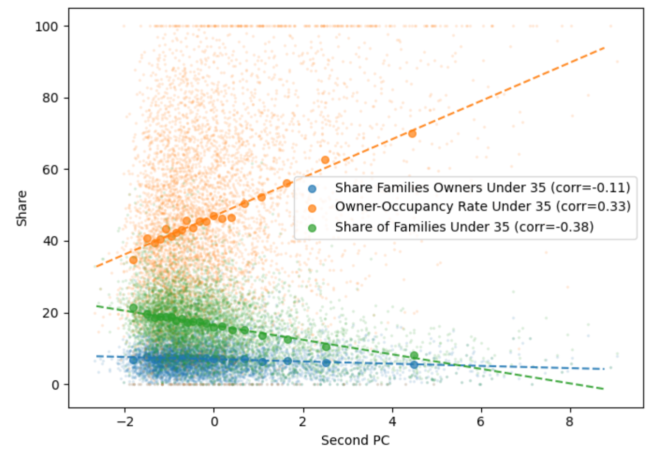
Notes: This figure displays Moran's I spatial autocorrelation statistics for various zoning regulations and indices. Moran's I measures the degree to which similar values cluster spatially, with values closer to 1 indicating stronger positive spatial autocorrelation.

Figure A9: Correlations Between Second Principal Component and Housing Outcomes

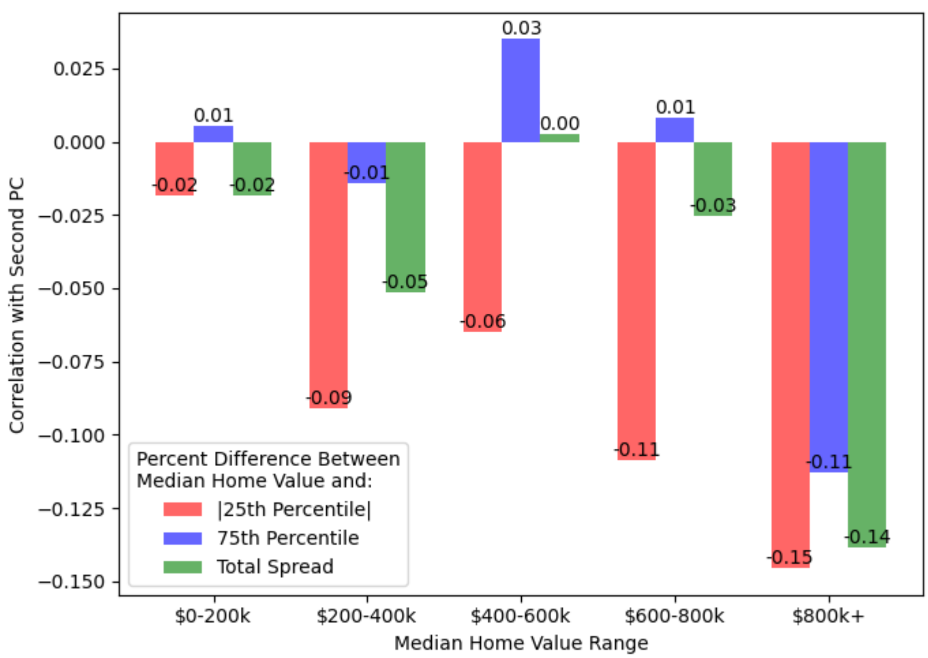
Panel A: Low-Income Housing and Ownership



Panel B: Housing and Ownership Among Younger Households

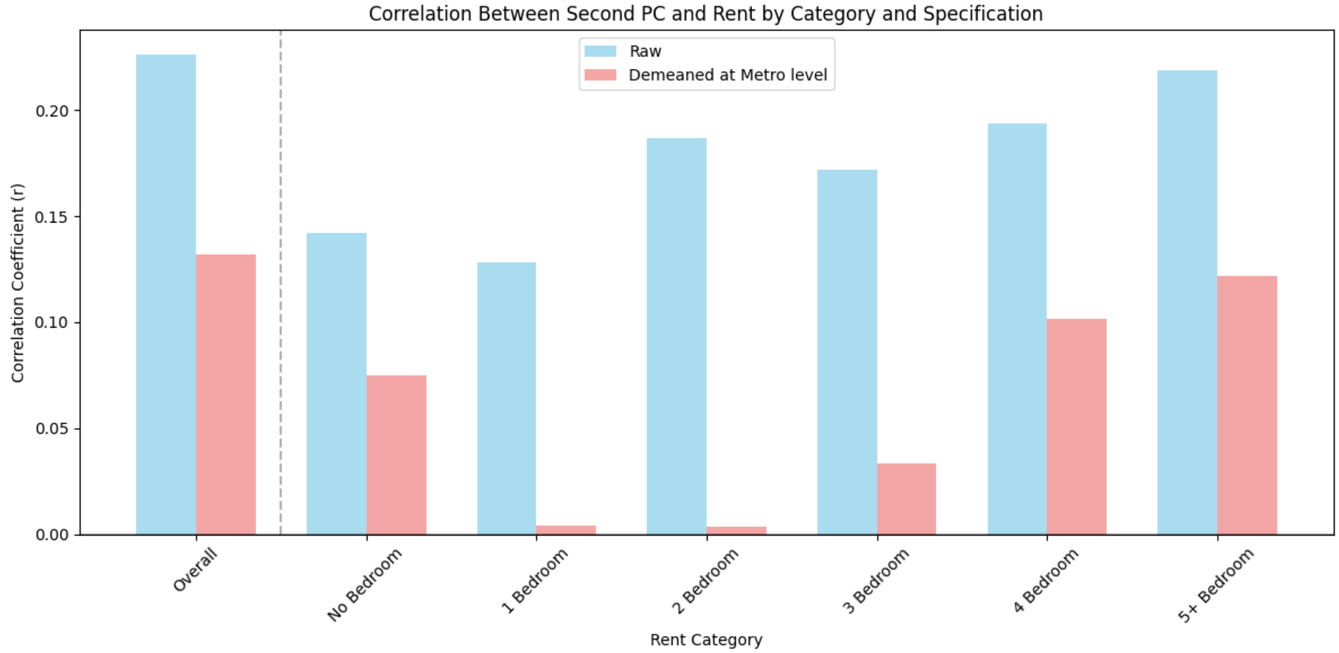


Panel C: Correlation of Second PC with Home Value Dispersion



Notes: This Figure shows the relationship between the second principal component and other key housing indicators. Panel A plots the relationship between the second PC and the share of low-income families, the share of low-income homeowners, and the low-income owner-occupancy rate. Panel B examines similar trends for younger households, plotting the share of family owners under 35, the owner-occupancy rate among those under 35, and the overall share of families under 35. Panel C presents correlations between the second PC and measures of home value dispersion across different median home value bins. The correlation with the 25th percentile home value reflects the absolute percent difference from the median, so a negative correlation indicates that higher-second PC municipalities have a tighter lower tail, with the 25th percentile closer to the median. Similarly, a negative correlation with the 75th percentile means that the upper tail is also compressed. The interquartile range correlation captures the combined effect of these shifts, with negative values indicating an overall tighter home value distribution in higher-second PC municipalities.

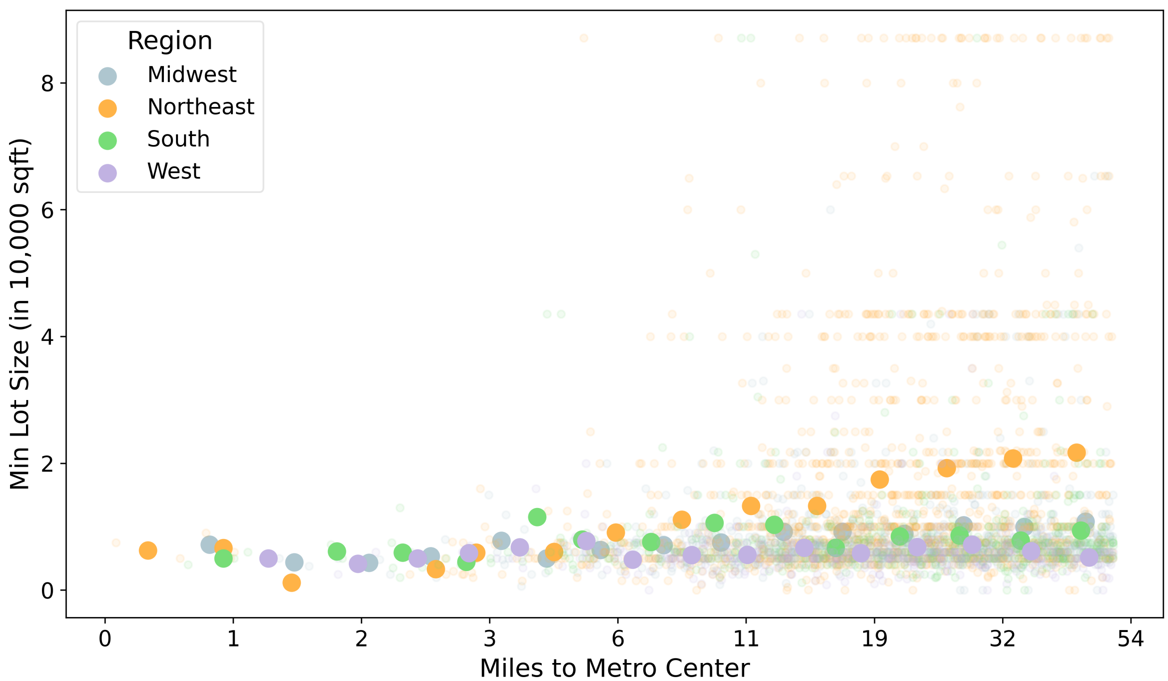
Figure A10: Second PC and Rent by Number of Bedrooms



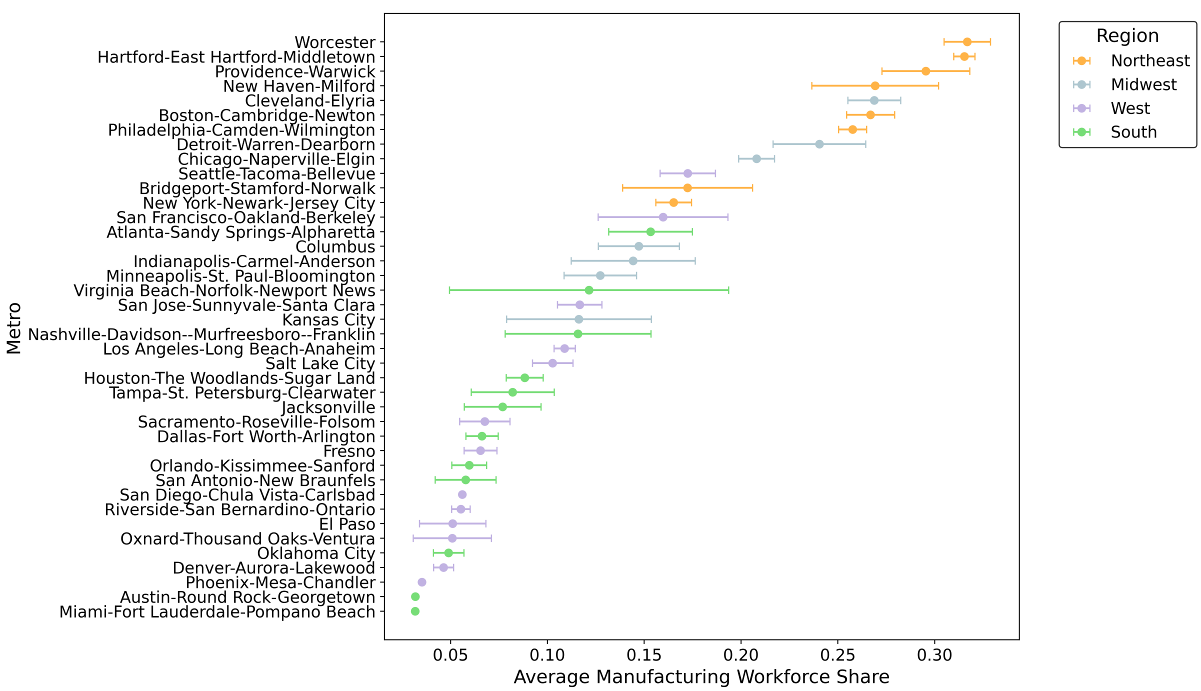
Notes: This Figure plots the correlation between various rent measures and the second principal component of our housing regulatory dataset both with raw values (blue) and demeaned at the MSA level. Rent measures come from the 2022 ACS. "Overall" is the median gross rent value (B25064_001E), while other categories are median gross rents conditioned on number of bedrooms in the housing unit (i.e. B25031_003E for median gross rent for one-bedroom apartments).

Figure A11: Minimum Lot Sizes and Historical Variation

Panel A: Binscatter of Min Lot Size Against Distance

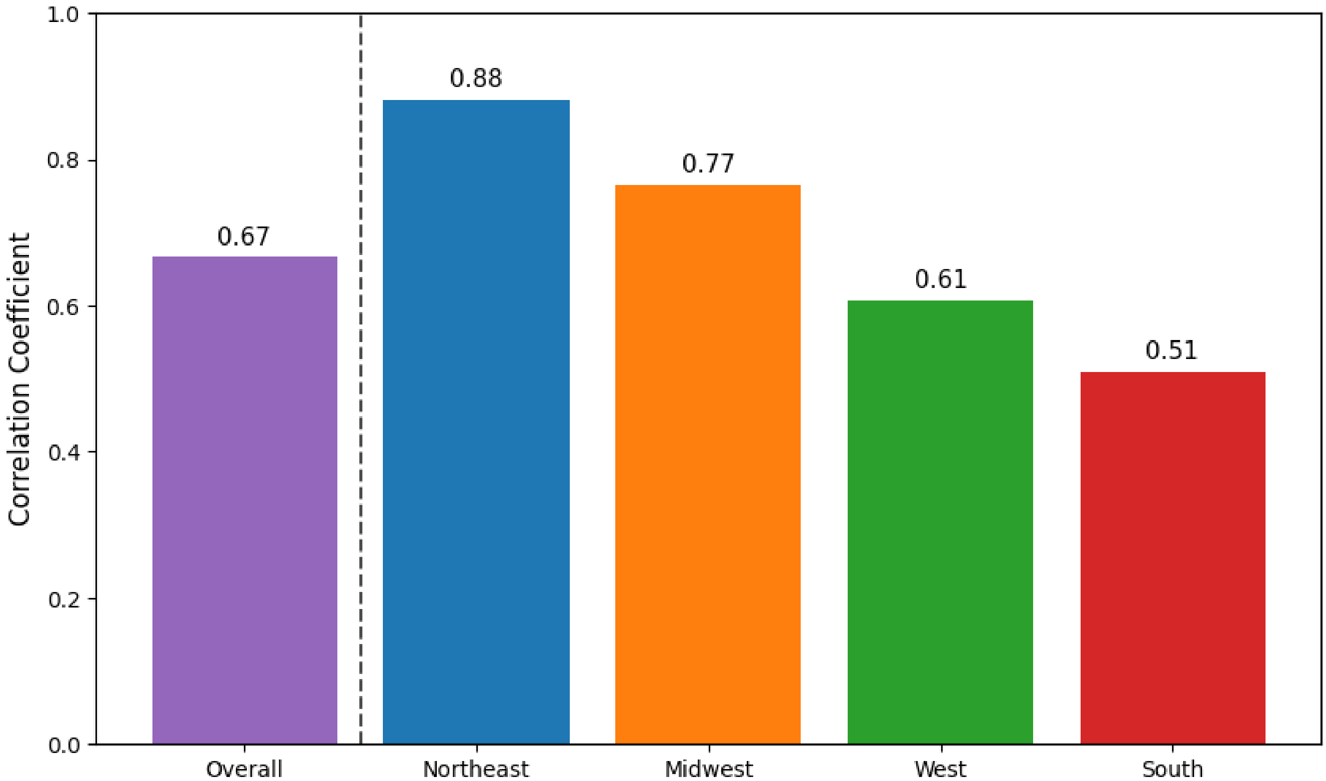


Panel B: Manufacturing Workforce Share



Notes: Panel A presents a binscatter plot showing how minimum lot size requirements (y-axis, in thousand square feet) vary with distance from metropolitan centers (x-axis, in miles shown on a log scale) across different U.S. regions. Panel B maps the historical manufacturing workforce share across U.S. counties. Both plots break out different regions.

Figure A12: Correlation Between 1940 and 2019 Log Housing Unity Density



Notes: This Figure plots the correlation between log housing density within municipality between 1940 and 2019. We use 1940 and 2019 housing units per square mile for all 2010 census tracts from Markley et al. (2022).

Table A1: Description of Variables Used in the Study

Variable	Source	Definition
Auto Commute Share	2022 ACS	Percentage of commuters using either cars, trucks, or vans as their primary commute method.
Born in Same State Share	2022 ACS	Ratio of those born in the state (B05002_003E) to those all those with information on location of birth (B05002_001E).
College Degree Share	2022 ACS	The percentage of the population aged 25 and over with a bachelor's degree or higher (B15003_022E, B15003_023E, B15003_024E, B15003_025E / B15003_001E).
Foreign Born Share	2022 ACS	The percentage of the population that is foreign-born (B05002_013E / B05002_001E).
Housing Unit Density	2022 ACS	The number of housing units in a local government divided by the area from its shape file.
Median Gross Rent	2022 ACS	The median gross rent for rental units (B25064_001E).
Median Home Value	2022 ACS	The median value of owner-occupied housing units (B25077_001E).
Non-Poverty Rate	2022 ACS	Theratio of the population with income in past 12 months below poverty line (B17001_002E) over the population with poverty status (B17001_001E). We subtract this ratio from 100% to get the non-poverty rate.
Owner-Occupied Share	2022 ACS	The percentage of housing units that are owner-occupied (B25003_002E / B25003_001E).
Share Families Low Income	2022 ACS	Share of families with income below 80% of the state median income. The distribution of incomes in a local government comes from ACS income buckets by tenure (i.e. B25118_003E for number of owner-occupied households making below $5,000).
Share Households Over 35	2022 ACS	We sum the number of rental and owner occupied units with a households under the age 35 (i.e. B25007_003E) and divide by the total number of units with age of householder information (B25007_001E). We subtract this ratio from 100% to get the share of households over 35.

Continued on next page

Table A1 continued from previous page

Variable	Source	Definition
Share Mid to High Income	2022 ACS	Share of families that are not low income, see "Share Families Low Income" definition.
Share Population 65 and Over	2022 ACS	The percentage of the population aged 65 and over (B01001_020E to B01001_025E and B01001_044E to B01001_049E / B01001_001E).
Share Population Under 18	2022 ACS	The percentage of the population under 18 years old (B01001_003E to B01001_006E and B01001_027E to B01001_030E / B01001_001E).
Share Rental Units	2022 ACS	100% minus the owner-occupied share, see owner occupied share definition.
Share Structures Built Before 1970	2022 ACS	The percentage of housing structures built before 1970 (B25034_008E, B25034_009E, B25034_010E, B25034_011E / B25034_001E).
Share Structures with 2 or More Units	2022 ACS	The percentage of housing structures with 2 or more units (B25024_004E to B25024_009E / B25024_001E).
Share Units Affordable	2022 ACS	The percentage of housing units affordable to households earning the state median income. This measure combines rental and owner-occupied housing affordability, determined using the state median income. Rental units are affordable if the monthly rent does not exceed 30% of the monthly median household income, and owner-occupied units are affordable if their value is less than three times the annual median household income. The total number of affordable rental and owner-occupied units is summed and divided by the total number of housing units to determine the share of units that are affordable.
Share with Commute Over 30 Minutes	2022 ACS	The percentage of workers with a commute time over 30 minutes (B08303_008E to B08303_013E / B08303_001E).
Vacancy Rate	2022 ACS	The percentage of vacant housing units (B25002_003E / B25002_001E).
White Share	2022 ACS	The percentage of the population identifying as White (B02001_002E / B02001_001E).

Continued on next page

Table A1 continued from previous page

Variable	Source	Definition
Local Revenue Per Student	2022 Annual Survey of School System Finances	The local revenue for a school district divided by the enrollment of that school district. We spatially merge school districts into the Census of Gov- ernments. Local revenue for school districts includes property taxes directly raised by the school and transfers from local governments for subordinate school districts.
Log Land Area	2022 Census Shapefiles	The area in acres of a local government with a log transform.
Miles to Metro Center	2022 Census Shapefiles	The number of miles from the centroid of a local government's shape file to the center of a metropolitan area.
Log Neighbors within 25 Miles	2022 Census of Governments	The number of other local governments within 25 miles of a local government's border.
Units Permitted Per Capita	BPS	The number of housing units permitted per capita averaged over 2019-2023.
Log Employment Density	CBP	Log of total employment at the county level normalized by county land area from the Census Gazetteer files.
1940 Foreign Born Share	NHGIS	Ratio of foreign born males (BXY003) and females (BXY004) to both native and foreign born population.
1940 Manufacturing Workforce Share	NHGIS	Ratio of the number of annual average wage earner in manufacturing (BW001) to the sum of employed males (BW9001) and females. (BW9002)
1940 Owner Occupied Share	NHGIS	Ratio of owner occupied dwellings (BYM001) to the sum of owner occupied dwellings and tenant occupied dwellings (BYM002).
Log Patents Per Capita	USPTO	Domestic patents granted between 2000–2015, normalized by each county's average population during that period.
Percent Democrat	(Bryan, 2022)	The share of votes that are Democrat in 2020.
Year of Incorporation	(Goodman, 2023)	The year a municipality was incorporated. Not available for townships.
Property Tax Rate	(Pierson et al., 2015)	The property tax rate is calculated as the total property tax revenue (Property_Tax_2017) divided by the aggregate home value from the 2017 ACS. This excludes property taxes raised from independent school districts.

Continued on next page

Table A1 continued from previous page

Variable	Source	Definition
Total Revenue Per Capita	(Pierson et al., 2015)	Total local government revenue (Total_Revenue) normalized by population.
Math Learning Rate	(Reardon et al., 2024)	The slope of the increase in math test scores from 3rd to 8th grade pooled across years (2008-2019) (cs_mn_grd_mth_ol).
Math Test Scores	(Reardon et al., 2024)	The average math test score pooled across grades (3rd-8th) and years (2008-2019) (cs_mn_avg_mth_ol).
Natural Amenities Index	(Service, 2019)	Calculated by USDA as the first principal component of various climate, topography, and water area features.
Opportunity Index	(Chetty et al., 2025)	The kid family rank, a measure of economic mobility. This data is merged from census tracts to local governments using Geocorr.
1940 Housing Unit Density	(Markley et al., 2022)	Estimated 1940 housing units per square mile in 2010 tract borders merged to local governments with geocorr.

Notes: USPTO is the United States Patent and Trademark Office. NHGIS is the National Historical Geographic Information System. CBP is the County Business Patterns dataset. BPS is the Building Permits Survey dataset. 2022 ACS refers to the 2022 American Community Survey dataset.

Table A2: Question Level Accuracy Contributions From Prompting

Panel A: Binary Questions

Question	Accuracy (%)			
	No Prompting	Prompt Engineering	Background Info	Prompt Chaining
Accessory Apartments Allowed	86.1%	99.0%	96.7%	96.7%
	(3.9)	(2.1)	(0.0)	(-)
Flexible Zoning By Right	100.0%	100.0%	100.0%	100.0%
	(0.0)	(0.0)	(0.0)	(-)
Flexible Zoning By Permit	96.0%	94.8%	95.8%	100.0%
	(0.3)	(3.4)	(2.9)	(-)
Affordable Incentive	81.4%	86.6%	99.3%	100.0%
	(2.5)	(1.9)	(1.5)	(-)
Permit Cap Or Phasing	83.3%	67.1%	91.3%	90.0%
	(0.0)	(5.2)	(1.8)	(-)
Wetlands Restricted in Lot Size Calc	76.0%	90.6%	92.7%	96.7%
	(1.5)	(5.6)	(1.5)	(-)
Multifamily Allowed	93.0%	94.7%	95.3%	100.0%
	(2.3)	(1.8)	(3.0)	(-)
Mixed-Use Buildings	79.7%	91.4%	92.6%	96.7%
	(7.5)	(2.9)	(1.5)	(-)
Conversion To Multifamily	59.5%	60.8%	88.0%	96.7%
	(1.3)	(4.1)	(3.8)	(-)
Townhouses Allowed	73.6%	68.1%	82.6%	90.0%
	(6.2)	(1.5)	(2.8)	(-)
Age-Restricted Provisions	89.5%	93.4%	98.6%	96.7%
	(5.1)	(2.9)	(1.9)	(-)

Panel B: Continuous Questions

Question	Correlation			
	No Prompting	Prompt Engineering	Background Info	Prompt Chaining
Zoning District Count	0.88	0.95	0.94	0.98
	(0.03)	(0.03)	(0.02)	(-)
Longest Frontage Requirement	0.50	0.53	0.46	0.70
	(0.10)	(0.12)	(0.07)	(-)
Lowest Res Min Lot Size	0.80	0.87	0.79	0.92
	(0.11)	(0.04)	(0.29)	(-)

Notes: This Table shows how different prompting strategies affect accuracy for specific zoning questions. Accuracy is measured on a validation sample of 30 municipalities from the Pioneer Institute which were not used during prompt development. Performance metrics are reported for binary questions (percent accurate, Panel A) and continuous ones (correlation, Panel B). Standard errors calculated across five model runs are shown in parentheses. Each row corresponds to a different question, and each column associates with a different LLM prompting strategy.

Table A3: Confusion Matrix For Binary Performance Results

Question	True Positive	False Positive	True Negative	False Negative	True Positive Rate
Multifamily Allowed	28	0	2	0	1.00
Mixed-Use Buildings	15	0	14	1	0.94
Conversion to Multifamily	12	1	17	0	1.00
Townhouses Allowed	18	1	9	2	0.90
Age-Restricted Provisions	22	0	7	1	0.96
Accessory Apartments Allowed	18	0	11	1	0.95
Flexible Zoning by Right	1	1	27	0	1.00
Flexible Zoning by Permit	26	0	3	0	1.00
Affordable Housing	22	0	7	0	1.00
Permit Cap or Phasing	8	2	19	1	0.89
Wetlands Restricted in Lot-Size Calc	23	1	6	0	1.00
Total	193	6	122	6	0.97

Notes: This Table shows a confusion matrix of model accuracy for binary variables. True Positive refers to an outcome where the model correctly predicts the positive class. False Positive is an outcome where the model incorrectly predicts the positive class. True Negative denotes an outcome where the model correctly predicts the negative class. False Negative represents an outcome where the model incorrectly predicts the negative class. The true positive rate (also known as sensitivity or recall) is the proportion of actual positive cases correctly identified by the model. The false positive rate (also known as the false alarm rate or fall-out) is the proportion of actual negative cases incorrectly identified as positive by the model. Precision (also known as positive predictive value) is the proportion of positive identifications that are actually correct. See Panel B footnote of Table 2 for details about the sample.

Table A4: Further Manual Validation

Panel A: Process Questions on Nationwide Random Sample of 30 Municipalities

Question Description	Correct	Ambiguous	Incorrect	Correct (Non-Ambiguous %)
Public Hearing Requirements	80%	10%	10%	89%
Max Review Waiting Time	87%	13%	0%	100%
Mandatory Approval Steps	93%	7%	0%	100%
Distinct Approval Bodies	80%	13%	7%	92%

Panel B: Bulk Questions on Random Sample of 30 California Municipalities

Question Description	Correct	Ambiguous	Incorrect	Correct (Non-Ambiguous %)
Lowest of Residential Min Lot Sizes	83%	6%	10%	89%
Longest Frontage Requirement	80%	10%	10%	89%

In Panel A we randomly drew 30 municipalities from our national dataset and had a law student review each answer. For Panel B, we drew a random sample of 30 municipalities from California and had a law student review each answer. Answers that are ambiguous include situations where the bulk regulations depend on whether a lot is interior or corner and the model chose one of the cases that could be correct. The column "Correct (Non-Ambiguous %)" calculates the percentage of correct responses among non-ambiguous cases. See Table A9 for full detailed versions of questions.

Table A5: Housing Regulation and Housing Supply Elasticity

Dependent Variable:	First Principal Component		Second Principal Component	
New Housing Unit Elasticity	-0.43***	-0.17**	-0.74***	-0.10
	(0.05)	(0.09)	(0.05)	(0.10)
Share Land Developed (2001)	-0.87***	-0.28	-2.23***	-0.90***
	(0.11)	(0.23)	(0.12)	(0.23)
Squared Share Land Developed (2001)	0.38***	-0.05	1.26***	0.41***
	(0.08)	(0.15)	(0.08)	(0.15)
Share Land Flat Plains	0.01	0.03	0.19***	0.12*
	(0.03)	(0.05)	(0.03)	(0.07)
Log Miles to Metro Center	-0.10***	-0.08***	-0.15***	-0.09***
	(0.02)	(0.03)	(0.02)	(0.03)
Intercept	0.14***	-0.49***	0.06***	1.45***
	(0.02)	(0.07)	(0.02)	(0.06)
R-squared	0.06	0.20	0.17	0.40
N	3890	3890	3890	3890
Metro Fixed Effects	No	Yes	No	Yes

Notes: This specification has as the dependent variable the first regulatory principal component (value capture, first two columns) and the second regulatory principal component (exclusionary zoning, second two columns). We regress these variables against a range of variables relating to new housing production and land availability. All variables are normalized to z-scores for the regression. Housing elasticity controls follow Baum-Snow and Han (2024) and include fraction of land developed in 2001, squared fraction of land developed in 2001, and the fraction of land with a flat topography. Metro fixed effects include state fixed effects for municipalities not within 100 miles of a metro center. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. Standard errors are clustered at the metro level when using metro fixed effects and are robust otherwise.

Table A6: Predicting Housing Market Outcomes With Zoning Regulation

Panel A: Individual Housing Regulations

Dependent Variable:	Median Home Value			Total Building Permits			Median Gross Rent			Housing Unit Den	
	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO
Lowest Res Min Lot Size	50	0.02	0.12***	32		0.01	66	0.03	0.08***	100	-0.13
Multifamily Not Allowed	99	0.17	0.24***	18		0.00	100	0.14	0.16***	49	-0.04
Affordable Mandate	62	0.01	0.05***	100		-0.00	79	0.06	0.09***	58	-0.01
Highest Res Min Lot Size	51	0.01	0.09***	11		0.02	49	0.02	0.07***	96	-0.11
Affordable Incentive	62		0.01	35		0.02	61	0.03	0.08***	59	
Townhouses Not Allowed	100	0.02	0.12***	9		-0.03**	56	-0.02	-0.01	28	-0.04
Longest Frontage Requirement	47		0.06***	37		0.01	52	0.01	0.06***	50	-0.12
No Conversion to Multifamily	49		0.02*	23		0.02	64		0.02	36	-0.03
Age-Restricted Provisions	67		-0.07***	18		0.01	58	0.00	0.02	28	-0.00
Zoning District Count	56		-0.04***	16	0.00	0.06***	58	0.06	0.06***	40	0.03
Max Review Waiting Time	30		-0.00	47		0.02	42		0.01	48	-0.01
No Mixed-Use Buildings	59		0.07***	29		-0.02	45	0.03	0.04***	33	-0.06
Public Hearing Requirements	57		-0.04***	12		0.04**	56	-0.00	-0.01	24	0.00
Wetlands Restricted in Lot Size Calc	32		0.04***	17		0.01	44	0.02	0.04***	33	-0.06
Mandatory Approval Steps	43		0.03**	19		0.03*	44	0.01	0.03**	29	
Flexible Zoning By Permit	33		0.08***	12		-0.05***	37		-0.01	49	0.09
Permit Cap Or Phasing	19		-0.04***	26		0.03**	37		0.01	33	-0.03
Distinct Approval Bodies	37		0.02	17		0.01	29	0.01	0.03**	36	-0.01
Accessory Apartments Banned	22		-0.03***	21		-0.05***	31	-0.02	-0.05***	27	0.06
No Flexible Zoning By RIght	15		-0.00	13		0.00	35		-0.01	19	0.01

Panel B: Housing Regulations Indices

Dependent Variable:	Median Home Value			Total Building Permits			Median Gross Rent			Housing Unit Density			Share Hou
	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost
Overall Index	82	-0.05	-0.01	73	0.02	0.03**	91	-0.06	0.00	100		0.01	91
First PC	91		0.02	92		-0.02	96	-0.02	0.00	80		0.02	91
Second PC	100	0.05	0.01	100		0.02	100	0.07	0.03*	76		0.01	100

Notes: This specification reports the results of three sets of regressions: a random forest specification, LASSO, and bivariate regressions of our generated regulatory variables against four dependent variables. Median house value and median gross rent are drawn from the 2022 ACS. Total building permits are defined as the number of housing units permitted divided by the population of the local government, averaged over 2019–2023, from the Census Building Permits Survey. Housing unit density is the number of housing units in a local government in the 2022 ACS divided by the area from its shape file. The share of affordable housing units is defined as the percentage of housing units affordable to someone earning the state median income. Rental units are considered affordable if the monthly rent does not exceed 30% of the monthly median household income, and owner-occupied units are affordable if their value is less than three times the annual median household income. For bivariate regressions, stars indicate statistical significance: *** p<0.01, ** p<0.05, * p<0.1. LASSO coefficients are shown where selected, with blank cells indicating variables not retained in the model. XGBoost scores represent “gain” importance from a random forests specification, measuring the average gain of splits using each feature, normalized so the most important feature for each dependent variable has a score of 100, with others scaled relatively. Higher scores indicate greater importance in the model’s predictions. All variables are first demeaned at the metro level, or for municipalities not within 100 miles of a metro center at the state level, then transformed into z-scores (mean=0, std=1). We use imputed regulations from our PCA analysis for LASSO and Bivariate regressions when the LLM reports “I don’t know,” see footnote of Table 5 for further details. We express variables so that a more positive value is associated with stricter zoning regulations, i.e. we transform the question of whether multi-family housing is allowed to whether it is not allowed. We allow missing data for XGBoost letting the algorithm both impute and predict. The Overall index in Panel B is a sum of normalized individual housing regulations for a municipality.

Table A7: Comparative Model Performance in Predicting Housing Outcomes

LHS Variable	Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Median Home Value	Lasso	1.07	1.11	0.73	1.05	0.71	0.73	0.71
	OLS	1.06	1.11	0.72	1.05	0.70	0.72	0.70
	XGBoost	1.00	1.12	0.74	0.96	0.66	0.77	0.67
Total Building Permits	Lasso	1.07	1.07	1.06	1.07	1.07	1.07	1.07
	OLS	1.06	1.06	1.06	1.06	1.06	1.06	1.06
	XGBoost	1.32	1.31	1.31	1.32	1.31	1.31	1.31
Median Gross Rent	Lasso	1.08	1.10	0.82	1.07	0.81	0.82	0.81
	OLS	1.08	1.10	0.81	1.07	0.81	0.81	0.81
	XGBoost	1.06	1.08	0.80	1.04	0.76	0.80	0.77
Housing Unit Density	Lasso	1.08	1.16	1.13	1.08	1.07	1.13	1.07
	OLS	1.08	1.16	1.13	1.08	1.07	1.13	1.07
	XGBoost	0.98	1.13	1.10	0.98	0.98	1.08	0.98
Share Housing Units Affordable	Lasso	0.99	1.02	0.70	0.97	0.69	0.70	0.69
	OLS	0.99	1.02	0.69	0.97	0.68	0.69	0.68
	XGBoost	0.99	1.03	0.63	0.95	0.53	0.62	0.52
Variables Included	Regulations	Yes	No	No	Yes	Yes	No	Yes
	Land	No	Yes	No	Yes	No	Yes	Yes
	Income	No	No	Yes	No	Yes	Yes	Yes

Notes: These specification report the RMSE of models predicting the housing market outcomes in Table A6. We include housing regulation measures in columns 1, 4, 5, and 7. We also include land availability controls (the share of land that is flat plains) in columns 2, 4, 6, and 7. We include median household income from the 2022 ACS as a control in specifications 3, 5, 6, and 7. All variables are first demeaned at the metro level, or when not available at the state level, then transformed into z-scores (mean=0, std=1). We use imputed regulations when the LLM reports “I don’t know” to avoid different imputation values across different controls, see footnote of Table 5 for further details.

Table A8: Other Associates of Principal Components

	First PC		Second PC	
	No FE	Metro FE	No FE	Metro FE
Year of Incorporation	0.01 (0.01)	-0.06*** (0.02)	0.10*** (0.01)	0.13*** (0.02)
Property Tax Rate	0.06*** (0.01)	-0.03 (0.02)	0.07*** (0.02)	-0.12*** (0.02)
Vacancy Rate	-0.15*** (0.01)	-0.15*** (0.03)	0.04*** (0.01)	0.06*** (0.02)
Share with Commute Over 30 Minutes	0.13*** (0.01)	-0.04 (0.03)	0.24*** (0.01)	0.12*** (0.03)
Auto Commute Share	-0.18*** (0.02)	-0.05 (0.05)	-0.16*** (0.01)	-0.03 (0.03)
Share Population Under 18	-0.07*** (0.01)	0.00 (0.02)	-0.11*** (0.01)	-0.02 (0.02)
Share Population 65 and Over	-0.07*** (0.01)	-0.10*** (0.02)	0.16*** (0.01)	0.14*** (0.02)
Median Home Value	0.18*** (0.02)	-0.04 (0.07)	0.28*** (0.02)	0.27*** (0.03)
Median Gross Rent	0.29*** (0.02)	0.13*** (0.05)	0.22*** (0.01)	0.15*** (0.03)
Foreign Born Share	0.16*** (0.02)	0.07* (0.04)	-0.04*** (0.01)	-0.17*** (0.03)
Units Permitted Per Capita	0.09*** (0.02)	0.06*** (0.02)	-0.03* (0.01)	0.00 (0.01)
Share Structures Built Before 1970	-0.19*** (0.01)	-0.28*** (0.04)	0.06*** (0.01)	-0.09*** (0.03)
Math Learning Rate	0.10*** (0.01)	0.08*** (0.02)	0.11*** (0.01)	0.08*** (0.02)
Share Units Affordable	-0.32*** (0.01)	-0.16*** (0.03)	-0.29*** (0.01)	-0.20*** (0.04)
Born in Same State Share	-0.19*** (0.01)	-0.17*** (0.03)	0.04*** (0.01)	0.05* (0.03)
Opportunity Index	0.08*** (0.01)	0.00 (0.03)	0.24*** (0.01)	0.16*** (0.03)

Notes: This Table reports a regression of our two principal components against additional covariates. For variable definitions, see Appendix Table A1. All right-hand side variables are measured as Z-scores. Fixed effects are for metros with State FE for municipalities not within 100 miles of a metro center. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We cluster standard errors at the metro/state level when including metro fixed effects and use robust standard errors otherwise. County level regressions use population weighted regulations. Results for other covariates can be found in Table 6.

Table A9: Mapping of Full Pioneer Institute Study Questions to Short Names

Full Question	Short Question
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	Multifamily Allowed
Are apartments above commercial (mixed use) allowed in any district?	Mixed-Use Buildings
Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?	Conversion To Multifamily
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	Townhouses Allowed
Does zoning include any provisions for housing that is restricted by age?	Age-Restricted Provisions
Are accessory or in-law apartments allowed (by right or special permit) in any district?	Accessory Apartments Allowed
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	Flexible Zoning By Right
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	Flexible Zoning By Permit
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	Affordable Incentive
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	Permit Cap Or Phasing
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	Wetlands Restricted in Lot Size Calc
What is the lowest residential minimum lot size?	Lowest Res Min Lot Size
What is the mean residential minimum lot size?	Mean Res Min Lot Size
What is the highest residential minimum lot size?	Highest Res Min Lot Size
How many zoning districts, including overlays, are in the municipality?	Zoning District Count
What is the longest frontage requirement for single family residential development in any district?	Longest Frontage Requirement
Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?	Affordable Mandate
How many mandatory steps are involved in the approval process for a typical new multi-family building?	Mandatory Approval Steps
For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?	Distinct Approval Bodies
Are there townwide requirements for public hearings on any type of multi-family residential projects?	Public Hearing Requirements
What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?	Max Review Waiting Time

Notes: This Table shows the mapping between the full regulatory questions and the short question name use throughout the paper. See Appendix Section C for further details on questions.

B Appendix: LLM Replicability

LLMs responses are not fully deterministic and so the results of this study cannot be completely replicated for several reasons. First, LLMs sample tokens from a probability distribution leading to variation in responses to the same query across iterations. Second, the GPT-4 class of models that we use in this study follow a Mixture-of-Exports (MoE) architecture (see here and here for further details). This means that the specific expert that an LLM query gets routed to varies across API calls depending on supply/demand of experts. In turn, this implies that the underlying probability distribution that tokens are sampled from may change from one query call to the next, depending on the availability of experts. Moreover, OpenAI's models are closed source so there may exist other sources of randomness across API calls that we cannot explain.

In this section we quantify how deterministic LLM queries are in our use case and how ensembling many query calls may help mitigate the issue. In general, we query the LLM to respond with a detailed answer, (i.e. "Think step by step") followed by a structured output (i.e. "Yes" or "No"). Randomness in output for LLMs leads to a high frequency of variation in open-ended responses, but many of these differences do not change the overall meaning of a response, for example just swapping synonyms. However, we do find some variation in the structured output of LLM responses in our use case.

One potential way to mitigate non-determinism is to request multiple responses from the LLM and then aggregate the answers by majority rule, an ensemble approach. Previous research has also found that ensemble methods can greatly improve LLM performance (Li et al., 2024).

With the OpenAI API there are two ways to ensemble API calls. The first way is to request multiple chat completions for a given query (by setting the API parameter n > 1). This effectively samples the distribution of tokens several times. This method is also cost effective because OpenAI only charges the user once for the input tokens regardless of how many iterations of output tokens are requested. However, this approach fails to sample from the distribution of potential experts or other potential sources of variation, for example the hardware of the server in which the LLM was run. A more costly approach is to separately query the LLM for each of the ensemble queries paying for both the input and output tokens used in each call. This second approach more broadly samples from the various sources of randomness for an LLM response.

We measure replicability with two measures in this analysis. Both measures average pairwise matching rates. We compare the final structured answer from a given LLM query across multiple API calls for all pairs (n choose 2), and take the ratio of the number of pairs that match to the total number of pairs. We call this ratio a consistency score. We measure both internal consistency, scores from requesting multiple chat completions for a given query, and external consistency, scores from comparing separate API calls.

We confirm that lower temperatures create more deterministic responses, even after ensembling. In Table B1, we compare temperatures of 0, 0.5, and 1 (lower temperatures should mean more deterministic responses) as well as whether including a random seed makes responses more deterministic. We use a random sample of 30 municipalities from our national sample and use two questions, a binary one (whether there are permits caps or project phasing) and a continuous one (how many districts there are). For each specification we run the model five times, each time requesting 10 chat completions. We measure internal consistency scores within a model run, and external consistency across aggregated majority rule answers from each model run. We find that lower temperature models are more internally consistent, especially for the continuous question, and have a lower variance of internal consistency. After aggregating responses, we find that external consistency scores are fairly similar for the binary question, but still higher for the continuous one and with lower variance. We also do not find evidence that including a random seed makes responses more deterministic.

We next show in Figure B1 that external consistency grows with ensemble size, especially for the continuous question. This suggests that answers begin to stabilize at larger ensemble sizes, though not fully. We still find that at least five percent of pairwise comparisons do not match even with ensemble sizes of 10 and a temperature of 0. We also find that the zero temperature specification already begins at a fairly high level of external consistency even with an ensemble size of 1.

Next, we ask how informative the internal consistency score is for predicting the external consistency score. In Figure B2 we find that higher temperature model internal consistency scores are highly predictive of external consistency and that this effect grows with ensemble size. However, lower temperature models internal consistency scores are not very informative for external consistency. This suggests that within model run variation for low temperature models may not reflect the same source of randomness as across model runs. For example, the variation in which expert the query is routed to may be more important than the within expert sampling distribution for predicting external consistency.

We next explore whether performance increases with ensemble size. For this analysis we return to our testing sample of 30 Massachusetts municipalities where we have a clean dataset to compare answers to. We use temperatures of 0 and 1 and request 10 chat completions from each model run. We do not find evidence of increased performance with larger ensemble sizes. In Figure B3 we find a fairly persistent outperformance of the 0 temperature model for the binary question and of the 1 temperature model for the continuous question.

We next ask whether internal consistency scores are helpful for predicting accuracy of questions. If the internal consistency score is highly predictive of the external consistency score then the measure can be used as a model confidence measure. We find in Table B2 that the internal consistency score is somewhat helpful for predicting accuracy with the temperature 1 model but not helpful for the temperature 0 model. With a larger ensemble size of 100, (Wei et al., 2024) found that answer frequency within an ensemble was highly predictive of accuracy, especially for more advanced reasoning models.

We suggest researchers consider one of two specifications. First, a low temperature single shot approach. This approach is cost effective by only requesting one chat completion, has a high degree of external consistency, and is straightforward to explain. If a researcher wants to use an ensemble approach then we suggest using a high temperature model with an ensemble size of at least five. High temperature ensemble models have a high degree of external consistency and have informative internal consistency scores for both predicting external consistency and accuracy. We choose to use the first approach in this paper for the higher external consistency, cost savings, and for simplicity.

Table B1: Internal and External Consistency Varying Temperature and Seed

Question	Temperature	Seed	External Consistency Mean	External Consistency Variance	Internal Consistency Mean	Internal Consistency Variance
Continuous: Number of Districts	0	No	0.91	0.05	0.9	0.05
		Yes	0.89	0.04	0.9	0.04
	0.5	No	0.83	0.09	0.68	0.1
		Yes	0.8	0.11	0.67	0.1
	1	No	0.75	0.11	0.61	0.1
		Yes	0.78	0.11	0.64	0.1
Binary: Whether Permit Caps or Phasing	0	No	0.93	0.03	0.96	0.02
		Yes	0.92	0.03	0.95	0.02
	0.5	No	0.93	0.03	0.88	0.04
		Yes	0.9	0.05	0.89	0.04
	1	No	0.91	0.04	0.88	0.04
		Yes	0.92	0.04	0.86	0.04

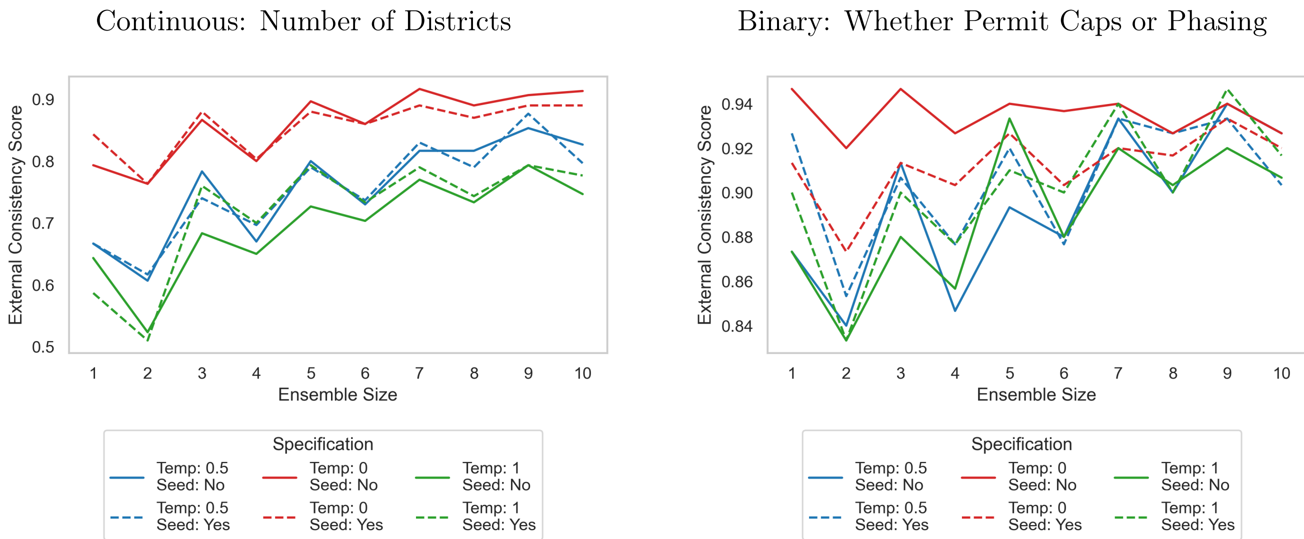
Notes: This Table compares the internal and external consistency of LLM responses across different temperature settings and with/without a random seed. "Temperature" refers to the randomness in the LLM's output (0 being most deterministic, 1 being most random). "Seed" indicates whether a random seed was used for replicability. "External Consistency" measures agreement across separate API calls, while "Internal Consistency" measures agreement within a single API call requesting multiple completions. Results are shown for two types of questions: a continuous question about the number of zoning districts, and a binary question about permit caps or phasing. Mean values closer to 1 indicate higher consistency. Lower variance indicates more stable results across trials.

Table B2: Regression of Internal Consistency Score on Whether Correct

	Temp = 0		Temp = 1	
	(1)	(2)	(3)	(4)
Consistency Score		0.2603		0.4175*
		(0.3614)		(0.2097)
Question FE	Yes	Yes	Yes	Yes
R-squared	0.1383	0.1461	0.2188	0.2696
N	60	60	60	60

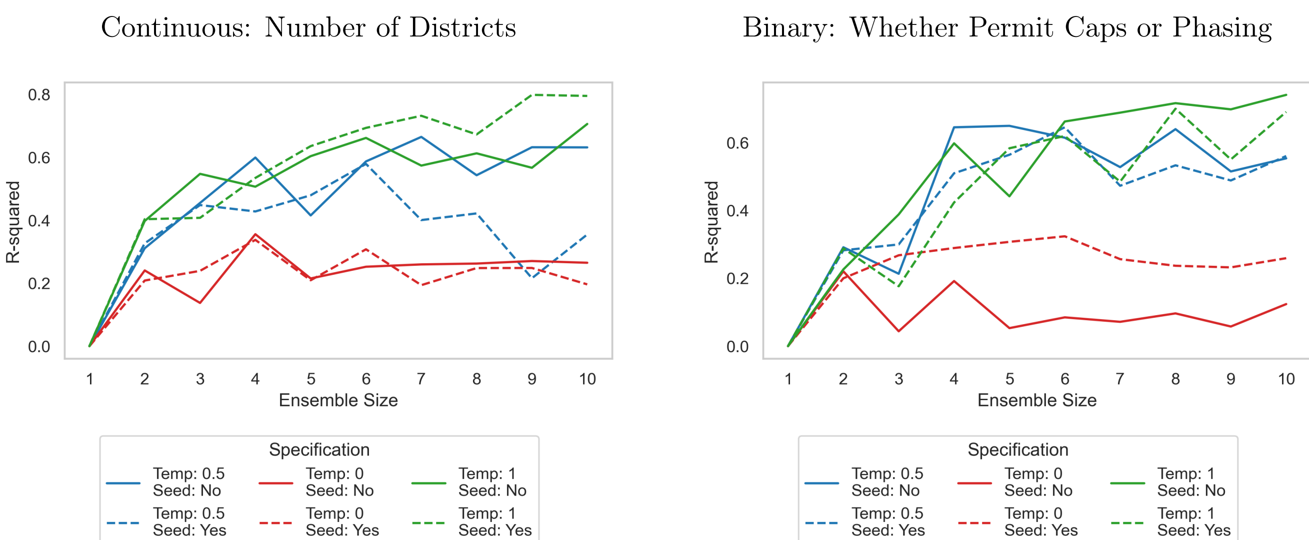
Notes: This table presents regression results examining the relationship between the internal consistency score of LLM responses and their accuracy. The dependent variable is a binary indicator for whether the LLM's response is correct. Results are shown for two temperature settings: 0 (most deterministic) and 1 (most random). Columns 1 and 3 include only question fixed effects, while columns 2 and 4 add the consistency score as an explanatory variable. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B1: External Consistency vs. Ensemble Size



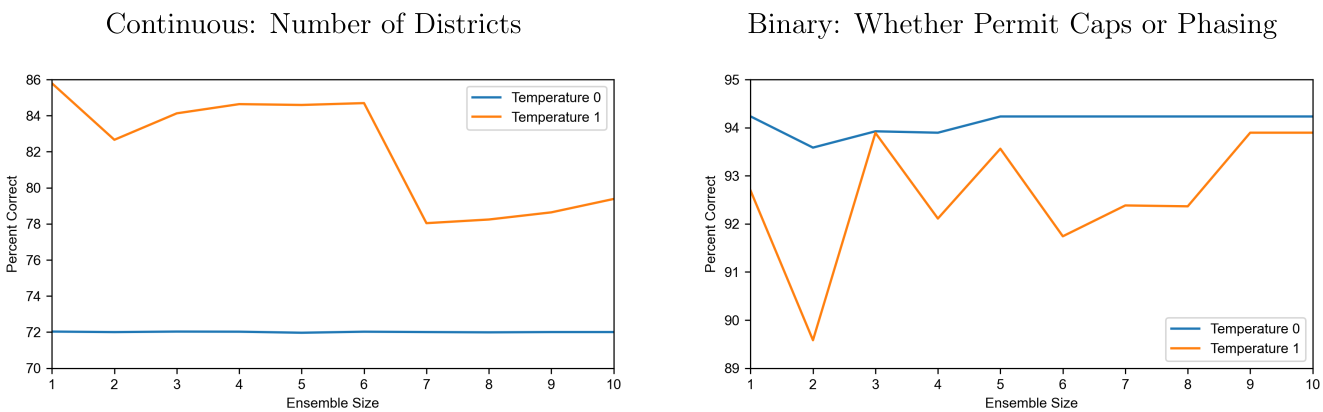
Notes: This Figure shows how external consistency of LLM responses varies with ensemble size for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size (number of model runs aggregated), while the y-axis shows the external consistency score. Different lines represent various temperature settings (0, 0.5, 1) and whether a random seed was used.

Figure B2: Relationship Between Internal and External Consistency vs. Ensemble Size



Notes: This figure demonstrates the relationship between internal and external consistency of LLM responses as a function of ensemble size for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size, while the y-axis shows the R-squared value, indicating how well internal consistency predicts external consistency. Different lines represent various temperature settings (0, 0.5, 1) and whether a random seed was used.

Figure B3: Ensemble Size vs. Percent Correct



Notes: This Figure illustrates the relationship between ensemble size and accuracy of LLM responses for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size (number of model runs aggregated), while the y-axis shows the percent of correct responses. Two temperature settings are compared: 0 (blue line, most deterministic) and 1 (orange line, most random). For the continuous question, accuracy is measured as the percentage of responses within a certain tolerance of the true value, while for the binary question, we use the percentage of correct classifications.

C Appendix: Question Details

This appendix provides detailed information about each question used in the study. Each question is presented with its original phrasing by the Pioneer Institute, the text that we embed for the question, background information and assumptions, question type, and the rephrased question that the language model sees. For some questions, we also include a value that triggers double-checking if the model's answer does not match it, along with the rephrased question used for double-checking and the keywords used to build context during the double-checking process. Additionally, certain questions involve subtasks, which are described in detail.

System Prompts for Each Question Type

We use a system prompt to guide the LLM in how to respond. Part of the system prompt includes details on how to structure the response, so we vary the exact system prompt by qustion type.

Numerical Questions: *"You are a municipal zoning ordinance expert. Use the following context which follows 'Context: ' from a municipal ordinance about zoning laws to answer the question which follows 'Question: '. You think step by step and justify each step with explanations and evidence from the context. At the end of your argument, you explicitly state your answer in the format of 'ANSWER: ' followed by a number or 'I DON'T KNOW'."*

Binary Questions: *"You are a municipal zoning ordinance expert. You use the following context which follows 'Context: ' from a municipal ordinance to answer the question which follows 'Question: '. You first review the background information on the question following 'Background Information on Question:' and treat it as additional instructions. You assume that the context includes all of the relevant legal information for the question. You review the context thoroughly for evidence to answer the question. When you cannot find any relevant information in the context, you realize that the town does not have relevant laws for the question and you reference the question background for how to handle this situation. You think step by step and justify each step with explanations and evidence from the context. At the end of your argument, you review what the answer should be and then explicitly state your answer in the format of 'ANSWER: ' and then one of 'YES', 'NO', or 'I DON'T KNOW'."*

Lot Size Questions: *"You are a municipal zoning ordinance expert. Use the following context which follows 'Context: ' from a municipal ordinance about zoning laws to answer the question which follows 'Question: '. Refer to the question background section for detailed instructions on how to answer the question. You think step by step and justify each step with explanations and evidence from the context. At the end of your answer, you say 'ANSWER:' and then reply with a CSV format with a column for 'District Name', 'Minimum Lot Size', 'Unit', and perhaps more depending on the question background. Ensure that you only include one row per district."*

Question 4

Question Phrased by Pioneer: Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?

Question Text That We Embed: Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?

Question Background and Assumptions: Multi-family housing comes in a wide variety of forms and sizes. The ways municipalities define and categorize "multi-family" housing varies widely, as do the use-regulations that govern multi-family housing development. This study includes as "multi-family" any building with three or more dwelling units. Multi-family dwelling units can be rental or condominium. They can be in a freestanding residential building or part of a mixed-use building, new construction or conversion of a preexisting building. Zoning documents usually specify what kinds of buildings qualify for conversion to multi-family housing: single family houses, two family houses, mills, schools, churches, municipal buildings or other types of facilities. Freestanding new "Multi-family" housing is defined as any building with three or more dwelling units, excluding townhouses, unless a municipality includes townhouses in its broader definition of multi-family housing and effectively permits only townhouses as such. Assisted living facilities, congregate care homes, dormitories, and lodging houses are not considered multi-family housing. If the zoning laws allow for conversion to multi-family housing, but do not comment on whether new multi-family housing is allowed, then the answer is 'YES'. Most towns allow a form of multi-family housing.

Question Type: Binary

Rephrased Question the LLM Sees: Is multi-family housing allowed at all in any district or overlay? If multi-family housing is allowed by special permission in any district or overlay then that counts allowed.

Question 5

Question Phrased by Pioneer: Are apartments above commercial (mixed use) allowed in any district?

Question Text That We Embed: Are apartments above commercial (mixed use) allowed in any district?

Question Background and Assumptions: Zoning bylaws and ordinances in various municipalities often contain provisions for combining residential dwellings with commercial uses such as retail or office spaces, creating mixed-use developments. While some zoning regulations explicitly allow multi-family housing and retail to coexist within the same district, they may not clarify whether these uses can share the same building, leaving this to be determined in practice. Certain municipalities explicitly permit "combined dwelling/retail" configurations in their use regulation tables, sometimes noting that any uses allowed within the same district can occupy the same building. Additionally, detailed provisions for mixed-use are facilitated through special zoning arrangements like overlay districts (e.g., mixed use district, downtown overlay, or planned unit development) or conversion projects, such as transforming former mills to accommodate both retail and housing. However, it's important to note that some references to "mixed use" may actually pertain to commercial and industrial combinations, excluding residential components. If you cannot find any reference to residential and commercial uses in the same building within the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is a combination of commercial and residential uses in the same building or structure allowed in any zoning district?

Question 6

Question Phrased by Pioneer: Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?

Question Text That We Embed: Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?

Question Background and Assumptions: The development of multifamily housing through the conversion of existing buildings encompasses two primary approaches: transforming single-family or two-family houses into structures with at least three units, and repurposing non-residential buildings, such as mills, other industrial buildings, schools, and municipal buildings, for multi-family residential use. This is different from the ability to construct new multi-family housing. The conversion of non-residential structures often occurs through designated overlay districts, like Mill Conversion Overlay Districts, or within industrial zones, whereas the conversion of houses to accommodate more units typically takes place in residential or business districts. The question does not count the conversion of single-family homes into two-family dwellings as allowing conversion to multi-family dwellings because multi-family is defined as having at least three units. If the conversion requires a special permit then we consider that as allowing conversion. Assisted living facilities, congregate care homes, dormitories, and lodging houses are not considered multi-family housing. The allowance of multi-family housing does not imply the allowance of the conversion to multi-family housing. You must search for an explicit statement allowing the conversion to multi-family housing from another type of structure. If you do not find any mention of conversions in the context then you assume the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: In any district, is the conversion to multi-family explictly allowed under any scope?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: In any district, is the conversion to multi-family explictly allowed under any scope?

Keywords We Use to Build Context When Double Checking in Order of Importance:

'conver'

Question 8

Question Phrased by Pioneer: Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?

Question Text That We Embed: Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?

Question Background and Assumptions: The question asks whether some form of attached housing is allowed in the municipality. Common forms of attached housing are single-family attached homes, townhouses, rowhouses, and zero lot line dwelling units. Attached housing is often allowed through special zoning provisions, such as overlay districts or use provisions tailored for cluster developments, Planned Unit Developments (PUD), or communities for active adults aged 55 and over. Remember that accessory apartments to a single-family home or the ability to attach one unit to a single-family home do not count as attached housing. Duplexes also do not count as attached housing. A form of attached housing may be listed as a type of single-family or multi-family housing. However, the allowance of single-family or multi-family housing does not imply the allowance of attached housing. This context raises the question of whether any type of attached housing are allowed either as their own category of housing or explicitly as a type of single family or multi-family housing. If you do not find any mention of a type of attached housing in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is some form of attached housing allowed in any district of the town?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: Is some form of attached housing allowed in any district of the town?

Keywords We Use to Build Context When Double Checking in Order of Importance:

'town house', 'town houses', 'townhouse', 'townhouses', 'attached dwelling', 'attached dwellings', 'row house', 'row houses', 'rowhouse', 'rowhouses', 'attached single family', 'attached unit', 'attached units', and 'attached'

Question 9

Question Phrased by Pioneer: Does zoning include any provisions for housing that is restricted by age?

Question Text That We Embed: Does zoning include any provisions for housing that is restricted by age?

Question Background and Assumptions: Many zoning bylaws/ordinances include provisions for housing that is deed restricted to occupants 55 (or another age) and older. Some of the provisions are for developments that are entirely age-restricted, while other provisions are incentives, often density bonuses, to include age-restricted units within an unrestricted development, such as cluster or multi-family. The restricted developments are called active adult housing, adult retirement village, senior village, planned retirement community, or something similar.

The answer should be Yes if any provisions exist for age-restricted single-family, townhouse, duplex, multi-family or accessory apartments. Provisions can be in the form of an age-restricted overlay, cluster development, density bonus for age-restricted units, or other zoning requirements or incentives for age-restricted housing.

Question Type: Binary

Rephrased Question the LLM Sees: Does zoning include any provisions for housing that is restricted by age?

Question 11

Question Phrased by Pioneer: Are accessory or in-law apartments allowed (by right or special permit) in any district?

Question Text That We Embed: Are accessory or in-law apartments allowed (by right or special permit) in any district?

Question Background and Assumptions: Accessory dwellings are separate housing units typically created in surplus or specially added space in owner-occupied single-family homes. Accessory dwellings can also be attached to the primary dwelling or be situated on the same lot (for example in a carriage house or small cottage.) An accessory dwelling typically has its own kitchen and bathroom facilities, not shared with the principal residence. Many zoning bylaws/ordinances call the dwellings "in-law apartments" or "family apartments" and restrict their occupancy to relatives of the homeowner - "related by blood, marriage or adoption." Some of these also allow domestic employees, caregivers, elderly people or people with low incomes to live in the units. Some municipalities allow the apartment by right if a family member will occupy the accessory apartment, but require a special permit otherwise. If you cannot find any reference to accessory apartments in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Are accessory or in-law apartments allowed in any district? If they are allowed by special permit in any district then we count that as allowed.

Question 13

Question Phrased by Pioneer: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?

Question Text That We Embed: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?

Question Background and Assumptions: Flexible zoning, encompassing terms like open space residential design, cluster, planned unit development, or conservation subdivision, provides municipalities with a more adaptable approach to zoning beyond the traditional "as-of-right" options. This methodology allows developers to bypass the stringent requirements of standard zoning, such as specific lot sizes and setback mandates, and enables the incorporation of various residential unit types like townhouses, duplexes, and multi-family homes that might not be allowed under conventional zoning regulations. The question only considers provisions that are primarily for residential uses. Most municipalities require special permits for cluster/flexible development.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is cluster development allowed explictly by right in any district? Question 2: Is open space residential design allowed explictly by right in any district? Question 3: Is any type of flexible zoning other than cluster development and open space residential design allowed explictly by right in any district?

Question 14

Question Phrased by Pioneer: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?

Question Text That We Embed: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?

Question Background and Assumptions: Flexible zoning, encompassing terms like open space residential design, cluster, planned unit development, or conservation subdivision, provides municipalities with a more adaptable approach to zoning beyond the traditional "as-of-right" options. This methodology allows developers to bypass the stringent requirements of standard zoning, such as specific lot sizes and setback mandates, and enables the incorporation of various residential unit types like townhouses, duplexes, and multi-family homes that might not be allowed under conventional zoning regulations. The question only considers provisions that are primarily for residential uses. Most municipalities require special permits for cluster/flexible development so if you find suggestive evidence that the municipality allows cluster/flexible development by special permit then you assume that the answer is 'YES'.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is cluster development allowed in any district, including by special permit? Question 2: Is open space residential design allowed in any district, including by special permit? Question 3: Is any type of flexible zoning other than cluster development and open space residential design allowed in any district, including by special permit?

Question 17

Question Phrased by Pioneer: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question Text That We Embed: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question Background and Assumptions: Inclusionary zoning requires or encourages developers to include affordable dwelling units within new developments of market rate homes. Some municipalities call it "incentive zoning" - when provision of affordable units is voluntary. The affordable units are typically located on site, but some municipalities also allow off-site development under certain circumstances. Often, payments may be made to a trust fund in lieu of building housing. Housing designated as "affordable" must be restricted by deed or covenant, usually for a period of 30 or more years, to residents with low or moderate incomes. The deed restrictions also limit sales prices and rents as the units are vacated, sold or leased to new tenants.

Do not include provisions for entirely affordable, subsidized housing development by public or non-profit corporations. Also do not include provisions under "rate of development" headings that exempt affordable units from project phasing and growth caps.

Question Type: Binary

Rephrased Question the LLM Sees: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question 20

Question Phrased by Pioneer: Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?

Question Text That We Embed: Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?

Question Background and Assumptions: Some municipalities enact town-wide caps limiting the number of units that can come on line annually or biannually. The number of permits is often set at the average in the previous years. Note that this question asks only about town-wide caps and does not consider caps exclusive to a specific district in the town. Some municipalities require phased growth for individual developments (also known as development scheduling or buildout scheduling) - a technique that allows for the gradual buildout of approved subdivisions over a number of years. Note that we only consider project phasing when it is required and not when it is optional. Project phasing is usually triggered by a minimum number of units in the project, so small subdivisions can be constructed in one year. Some phasing provisions are only triggered at the town-wide level once a threshold number of units have been permitted. Most of the "rate of development" provisions include an expiration or "sun set" date (some that have expired have been updated and re-adopted). Many include a "point system" where points are awarded for provision of community goods such as open space or affordable units, and projects with more points are given priority for permits. If you do not find any information in the context about a town-wide annual or biannual cap or about project phasing then you assume the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is there a town-wide annual or biannual cap on residential permits issued Question 2: Is project phasing required?

Question 21

Question Phrased by Pioneer: Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?

Question Text That We Embed: How is lot area defined and how is the lot size calculated?

Question Background and Assumptions: Remember to first review your research so far on how a lot size is calculated and defined. If you have already found a restriction on including wetlands, sloped land, or easements in your prior research then the answer is 'YES'.

Some municipalities require that the minimum lot size requirement be met by a percentage of land that does not include wetland resource areas, steeply sloped land or easements. A subset of those municipalities requires that the buildable area be contiguous on the lot – called "contiguous buildable area" or "contiguous upland area." Upland area is non-wetland area. It is much more common for municipalities to restrict the use of wetlands areas in meeting lot size requirements than sloped land or easements.

Note that this question only asks about whether there are restrictions on calculating the lot size. It does not ask about whether there are restrictions to buildable area or whether there are any restrictions in wetland areas.

If you do not find any restrictions for lot size calculations in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Detail how lot area is defined and how a lot size is calculated. Then, answer the question of are there restrictions on counting wetlands, uplands, or sloped land in lot area/lot size calculation?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?

Keywords We Use to Build Context When Double Checking in Order of Importance: 'wetland', 'upland', 'sloped land', and 'easement'

Question 28

Question Phrased by Pioneer: What is the minimum lot size for single-family homes in each residential district?

Question Text That We Embed: What is the minimum lot size for single-family homes in each residential district?

Question Background and Assumptions: When compiling a list of minimum lot sizes for districts that permit single-family housing, prioritize clarity by selecting the specific minimum lot size for single-family homes within each district. If multiple options exist, choose the most common standard size, excluding sizes for historic properties or special cases. Report sizes in square feet over acres unless only acre measurements are available. Only include districts with a defined minimum lot size or those adhering to a town-wide minimum if no district-specific size is established. Finalize the data in a CSV format with columns for 'District Name', 'Min Lot Size', 'Unit', and 'Estate', ensuring a straightforward, single entry for each district that reflects the standard requirement for single-family homes.

Question Type: Lot Size

Rephrased Question the LLM Sees: What is the minimum lot size for single-family homes in each residential district?

Subtask:

- Subtask Question That Gets Embedded: Find the name of each district that allows single-family housing

- Rephrased Subtask Question the LLM Sees: Find the name of each district that allows single-family housing

- Additional Subtask Instructions: Please list out the name of each residential district in the town that primarily consist of detached single-family housing. If you cannot find any districts that explictly allow single-family detached housing then just assume that any residential districts allow single-family detached housing. Respond with a detailed answer followed by a CSV format with the name of the district in the first column and whether a district has the label 'Estate' in the second column as a True/False statement. Use the column headers of 'District Name' and 'Whether Estate District'.

- How The Subtask Results Are Described to the LLM Afterwards: Your previous work finding which districts to find minimum lot sizes for and whether they are estate districts

Question 2

Question Phrased by Pioneer: How many zoning districts, including overlays, are in the municipality?

Question Text That We Embed: How many zoning districts, including overlays, are in the municipality?

Question Type: Numerical

Rephrased Question the LLM Sees: How many zoning districts and overlays are in the municipality?

Question 22

Question Phrased by Pioneer: What is the longest frontage requirement for single family residential development in any district?

Question Text That We Embed: What is the longest frontage requirement for single family residential development in any district?

Question Type: Numerical

Rephrased Question the LLM Sees: What is the longest frontage requirement for single family residential development in any district?

Subtask:

- Subtask Question That Gets Embedded: Find the name of each single-family residential district

- Rephrased Subtask Question the LLM Sees: Find the name of each single-family residential district
- Additional Subtask Instructions: Please list the names of each single-family residential district. Only include districts that are primarily residential. Usually, this means districts that start with the letter R like R1. If there is only one residential district that permits single-family zoning then just name that one district. If you are unsure whether a residential district permits single-family zoning then assume that it does, but ensure that the district is primarily residential. An agricultural (A) or industrial (I) district would not be included for example.
- How The Subtask Results Are Described to the LLM Afterwards: Only consider the frontage requirements in the following districts

Question 17w

Question Text That We Embed: Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?

Question Background and Assumptions: Zoning codes may require developers to include affordable housing in market-rate residential projects, but the applicability of these requirements can vary. Some inclusionary policies apply broadly to all residential development, while others are tied to optional zoning designations, incentive programs, or specific areas.

To determine if a zoning code contains a mandatory inclusionary requirement, look for clear language stating that all or most market-rate residential projects must provide affordable units as a standard condition of approval under normal zoning rules. The requirement should not be limited to projects that opt into a special zoning designation, participate in an incentive program, or are located in a particular overlay zone.

Focus on whether the code unambiguously requires all or most market-rate residential development to include affordable housing under the generally applicable rules. Do not select "YES" if affordable housing is only mandatory in narrow, specialized situations. The mere presence of affordable housing provisions is not sufficient if they are elective or only apply in atypical circumstances. If the affordable housing requirements are not clearly universally applicable, the likely answer is "NO".

Question Type: Binary

Rephrased Question the LLM Sees: Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?

Question 30

Question Text That We Embed: How many mandatory steps are involved in the approval process for a typical new multi-family building?

Question Background and Assumptions: The approval process for constructing a new multi-family building typically involves multiple mandatory steps, each representing a distinct interaction or requirement that a developer must fulfill before construction can begin. Focus on identifying only the core, pre-construction approval steps that are required for all multi-family building projects, from initial application submission to final permit issuance. Each required interaction with a distinct city department or agency should be counted as a separate step, but be careful not to artificially separate closely related actions within a single process. For example, applying for and obtaining a building permit should be considered one step, not two. Be cautious not to include optional or discretionary steps, post-approval activities such as inspections during construction or certificate of occupancy issuance, steps that are only required in specific circumstances or for certain types of properties, or internal processes within departments that don't require direct developer interaction. When analyzing the ordinances, pay close attention to language indicating whether a step is mandatory (e.g., "shall", "must", "is required") versus optional or conditional (e.g., "may", "at the discretion of", "if applicable"). The goal is to identify the minimum number of distinct, mandatory steps that every multi-family building project must go through in the approval process, avoiding redundancy and over-segmentation of closely related actions.

Question Type: Numerical

Rephrased Question the LLM Sees: How many mandatory steps are involved in the approval process for a typical new multi-family building?

Question 31

Question Text That We Embed: For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?

Question Background and Assumptions: When answering this question, focus on the approval process for a typical new multi-family building project as described in the provided ordinance sections. Only count distinct governing bodies or agencies whose approval is explicitly required by the ordinances for all multi-family building projects, including those allowed "by right" under existing zoning. To be counted, an entity must have clear, independent approval authority that is mandatory for the project to proceed. This approval must be specifically for the multi-family project itself. Look for unambiguous language indicating required, independent approval steps. Distinguish between actual approval authority and advisory roles; entities that only review or provide input should not be counted. Consider roles like the Planning Board, Board of Health, Building Commissioner, and special permit granting authorities, but include them only if their approval is explicitly required and independent. For coordinated review processes, determine whether they represent multiple independent approvals or a single approval incorporating multiple inputs. Provide your answer as a number, followed by a brief explanation of which entities you counted and why. Cite relevant ordinance sections, explaining why each approval is considered independent and mandatory, and how it relates specifically to the multi-family project.

Question Type: Numerical

Rephrased Question the LLM Sees: For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?

Question 32

Question Text That We Embed: Are there townwide requirements for public hearings on any type of multi-family residential projects?

Question Background and Assumptions: When answering this question, examine the zoning ordinances and bylaws for any townwide requirements that mandate public hearings or formal public input processes for multi-family residential developments. Focus on requirements that apply across all zones within the town. Answer YES if public hearings are required for any subset of multi-family projects, even if not all multi-family projects require hearings. For instance, if larger projects require public hearings while smaller ones don't, the answer should still be YES. Requirements specific to certain zones do not count towards a YES answer. Answer NO only if there are no townwide public hearing requirements for multi-family developments of any size or type, or if such requirements only apply in specific zones. Be sure to cite relevant ordinance sections that support your conclusion. The goal is to determine whether there is any mandated opportunity for public input on new multi-family housing developments on a townwide basis, even if this only applies to certain categories of multi-family projects.

Question Type: Binary

Rephrased Question the LLM Sees: Are there townwide requirements for public hearings on any type of multi-family residential projects?

Subtask:

- Subtask Question That Gets Embedded: Do any types of multi-family housing projects require a special permit in this jurisdiction? If so, under what conditions?
- Rephrased Subtask Question the LLM Sees: What is the typical approval process for new multi-family building projects in this jurisdiction? Please describe any required permits, reviews, or other procedures that are standard for multi-family developments.
- Additional Subtask Instructions: Do any types of multi-family housing projects require a special permit in this jurisdiction? If so, under what conditions?
- How The Subtask Results Are Described to the LLM Afterwards: Special Permit Require-

ments for Multi-Family Housing Developments

Question 34

Question Text That We Embed: What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?

Question Background and Assumptions: The review process for constructing a new multi-family building involves several stages, each of which may have a specific waiting period. The total waiting time includes the mandatory review periods as well as any discretionary days that can be added by the governing bodies or agencies. Each agency or department that a developer must interact with, such as city government departments like fire, police, sanitation, building, and planning, has its own review timeline. Additionally, discretionary days that may be required for public hearings, environmental reviews, or other procedural requirements must also be added to the total count of government review days.

Question Type: Numerical

Rephrased Question the LLM Sees: What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?

D Appendix: Comparison With Wharton Study

The Wharton Residential Land Use Regulation Index (WRLURI) is a survey-based measure of local housing market regulations. In this section we compare our study to the 2018 version of the index (Gyourko et al., 2021). The survey was administered through the International City Managers Association (ICMA) to its 10,949 member municipalities at the time, receiving 2,825 responses for a response rate of 25.8%. In Table D1, we compare our dataset's sample coverage with that of the Wharton index. Our analysis reveals superior coverage across the distribution of various demographic and housing characteristic variables. We especially have greater coverage for low population, low white share, high college share, and high median home value local governments.

In Panel A of Table D2, we compare question level responses focusing on two areas where our question bases overlap: affordable housing mandates and minimum lot sizes. Our measure of affordable housing shows a substantial correlation of 0.38 with the Wharton measure. For minimum lot size requirements, we find smaller but meaningful correlations ranging from 0.18 to 0.37 across different size categories. We next compare the accuracy of each study for the highest minimum lot size by benchmarking responses against independently sourced geographic data from Massachusetts GIS in Panel B. Our approach achieves an average accuracy rate of 77%, substantially higher than Wharton's 50% accuracy rate when compared against the same ground-truth data.

In Table D3, we examine correlations between our principal components and the Wharton index. We find positive but somewhat low correlations ranging from 0.11 to 0.33. Besides differences in question-level accuracy, our study exclusively measures regulation, while WRLURI takes a broader approach. Of its 71 sub-questions, only 56.3% directly measure zoning regulations. The remaining questions cover important aspects of the housing production process, such as approval processes and timelines in practice (22.5%), but also questions less relevant to a strictly regulatory index, like housing market outcomes and market conditions (i.e., whether single-family housing supply meets demand). This broader scope is reflected in the PCA weights in the Wharton index. The highest weights are assigned to process-based measures, such as court involvement, state political involvement, and local political pressure indices, while direct regulatory measures such as supply

Table D1: Sample Coverage Comparison With Wharton Index

	Variable Distribution Percentiles					
Variable	Bottom 5%	5-25th	25-50th	50-75th	75-95th	Top 5%
Population						
Wharton Sample (%)	0.0	0.0	0.0	5.1	25.1	31.6
Our Sample (%)	0.3	0.7	3.8	13.7	40.5	65.3
Difference (p.p.)	0.3	0.7	3.8	8.7	15.4	33.7
Owner-Occupied Share						
Wharton Sample (%)	12.2	13.8	8.8	6.1	3.2	0.1
Our Sample (%)	30.9	29.2	17.7	10.8	5.9	0.2
Difference (p.p.)	18.8	15.5	8.9	4.7	2.7	0.1
65+ Population Share						
Wharton Sample (%)	2.7	9.8	10.5	8.6	4.8	1.4
Our Sample (%)	7.7	21.8	20.5	15.3	10.1	6.0
Difference (p.p.)	5.0	11.9	10.0	6.7	5.3	4.6
Under 18 Share						
Wharton Sample (%)	0.4	5.8	10.3	10.5	7.4	1.2
Our Sample (%)	2.6	12.1	19.2	21.3	15.4	4.5
Difference (p.p.)	2.2	6.3	9.0	10.9	8.0	3.3
White Share						
Wharton Sample (%)	11.2	15.1	9.8	6.0	1.5	0.0
Our Sample (%)	34.5	35.8	18.7	8.0	1.7	0.4
Difference (p.p.)	23.3	20.7	8.9	2.1	0.2	0.3
College Share						
Wharton Sample (%)	0.2	2.3	6.0	8.9	14.3	16.9
Our Sample (%)	1.4	6.9	11.6	15.6	27.0	46.8
Difference (p.p.)	1.1	4.6	5.6	6.8	12.7	30.0
Median Home Value						
Wharton Sample (%)	0.5	3.0	6.5	9.1	13.8	23.3
Our Sample (%)	1.7	6.7	12.1	14.1	27.9	71.4
Difference (p.p.)	1.1	3.7	5.5	5.1	14.1	48.1

restrictions and density restrictions indices receive the smallest loadings.[18]

[18] The factor loadings from the principal component analysis used to construct the WRLURI2018 are: Court Involvement Index (0.42), State Political Involvement Index (0.41), Local Political Pressure Index (0.40), Environmental Index (0.28), Approval Delay Index (0.28), Local Project Approval Index (0.29), Local Zoning Approval Index (0.27), Open Space Index (0.24), Affordable Housing Index (0.27), Local Assembly Index (0.17), Supply Restrictions Index (0.12), and Density Restrictions Index (0.09).

Table D2: Correlation With Wharton Index

Panel A: Averages and Correlation For Wharton Questions

Question		Wharton Average	Our Average	Correlation
Affordable Housing		0.20	0.06	0.38
Minimum Lot Size	Less than 1/2 acre	0.50	0.49	0.37
	1/2 to 1 acre	0.17	0.13	0.18
	1 to under 2 acres	0.12	0.17	0.26
	2 acres or more	0.22	0.16	0.27

Panel B: Comparison to Massachusetts GIS Minimum Lot Sizes

	Massachusetts GIS Data	Less than 1/2 acre	1/2 to 1 acre	1 to under 2 acres	2 acres or more	Average Accuracy
Our Model	Less than 1/2 acre	6	2	0	1	
	1/2 to 1 acre	1	12	1	1	
	1 to under 2 acres	0	1	23	2	
	2 acres or more	1	0	4	8	
	Percent Correct	75%	80%	82%	67%	77%
Wharton	Less than 1/2 acre	3	3	1	0	
	1/2 to 1 acre	4	6	7	0	
	1 to under 2 acres	0	4	13	2	
	2 acres or more	1	2	7	10	
	Percent Correct	38%	40%	46%	83%	50%

Notes: The sample overlap between this study and Gyourko et al. (2021) is 1,171 municipalities. We drop municipalities that do not have any minimum lot size requirements. The Affordable Housing questions refers only to affordable housing mandates, not incentives, and the minimum lot size questions refers only to residential districts. Massachusetts GIS data comes from MassGIS (Bureau of Geographic Information). To find the highest residential min lot size from MassGIS we first drop districts with missing/no min lot size info and then drop districts not flagged as single family zoned.

Table D3: Index Level Wharton Correlations

	Wharton Index	PC 1	PC 2	Overall Index
Wharton Index	1.00	0.33	0.11	0.22
PC 1	0.33	1.00	0.07	0.36
PC 2	0.11	0.07	1.00	0.74
Overall Index	0.22	0.36	0.74	1.00

The Overall Index sums the z-scores of all zoning questions.

E Appendix: Economic Model of Housing Regulation and Municipal Zoning

This appendix provides a detailed technical exposition of a model used to analyze municipal zoning regulations, household location choices, and government policy. The model combines a spatial equilibrium framework with non-cooperative local government optimization to explain the key facts around housing production, choice of regulations, and socio-economic sorting which we observe in our key facts.

E1 Model Setup

Agents and Locations: To capture the key socio-economic differences among agents, we define households as of two types: low productivity (L) and high productivity (H), earning wages w^L and w^H, respectively). These agents decide to live in one of two zones, reflecting the key spatial differences within urban areas as either a city core c and a suburb s, which are differentiated by distinct amenities (α_c, α_s) and regulatory environments. $\eta_i = \frac{N_i^H}{N_i}$ is therefore the share of high productivity workers in each zone. Denote $\widetilde{w}_i = w_i^H \cdot \eta_i + w_i^L \cdot (1 - \eta_i)$ as the effective wage, and $\widetilde{N}_i = w_i^L \cdot N_i^L + w_i^H \cdot N_i^H$ as the effective population.

Government Structure: Local governments within each zone $i \in \{c, s\}$ levy a homogeneous wage tax τ and differentiated housing tax t_i. Governments produce public services s_i using labor G_i, paid for by low productivity workers, and maximize net revenue.

Housing Market: The housing market clears through spatial sorting of workers. Housing rents r_i in each region are upward sloping and depend on local population density:

$$r_i = \bar{r}_i + \beta_i \cdot \log(N_i^L \cdot h_i^L + N_i^H \cdot h_i^H)$$

where $\bar{r}_i$ is the baseline rent fixed in each region, $\beta_i > 0$ governs the supply elasticity, and N_i^J is the population of type j in zone i.

E2 Household Problem

Households maximize utility by choosing consumption (c), housing (h), and location (choice of i). The utility function for a household of type j in zone i is:

$$U_i^j = \max_{c,h} \gamma \log c + (1-\gamma) \log h + \alpha_i - \log(N_i) + \log(s_i)$$

subject to the budget constraint:

$$c + h \cdot r_i = (1-\tau) \cdot w^j.$$

Utility in this framework is decreasing in local population (N_i) due to congestion, and increasing in public services s_i. Housing demand in this framework therefore satisfies:

$$h_i = \frac{(1-\gamma)(1-\tau)w_j}{r_i}.$$

E3 Government Problem

Local governments maximize net revenue:

$$\max_{G_i, t_i, \underline{h}_i} [\tau + (1-\gamma) \cdot t_i] \cdot (w^L \cdot N_i^L + w^H \cdot N_i^H) - w^L \cdot \log(G_i),$$

subject to a production function on public services:

$$s_i = \gamma_i \exp(G_i),$$

a requirement on minimum housing consumption, which corresponds to exclusionary zoning:

$$h_i^j \geq \underline{h}_i,$$

a population constraint:

$$N_i^L + N_i^H = N_i$$

and a balanced budget requirement:

$$N_i \cdot [t_i \cdot r_i + \tau \cdot \widetilde{w}_i(\eta_i)] \geq w^L \cdot \log(G_i).$$

The two key government regulatory policies are the choice of housing taxes t_i, which corresponds to value capture or the first principal component, and $\underline{h_i}$, a minimum housing requirement, which corresponds to exclusionary zoning and the second principal component. Governments in both zones choose regulations in a non-cooperative way independent of the other jurisdiction.

E4 Equilibrium and Sorting

The key condition for market clearing is that total population satisfies

$$N_c + N_s = N$$

for each household type.

In spatial equilibrium, households sort across zones until they are indifferent, resulting in the following condition:

$$[\alpha_c - \alpha_s] + \log \frac{s_c}{s_s} = (1 - \gamma) \left[\log \frac{r_c}{rs} + \log \frac{1 + t_c}{1 + t_s} \right] + \log \frac{N_c}{N_s}. \tag{2}$$

In other words, the relative benefit of being in one zone instead of the other (based on local baseline amenity and government services) is equal to the additional costs (taxes and congestion costs). We can also substitute in the government production function and rewrite this as:

$$[\alpha_c - \alpha_s] + [\log \gamma_c - \log \gamma_s] + [G_c - G_s] = (1 - \gamma) \left[\log \frac{r_c}{rs} + \log \frac{1 + t_c}{1 + t_s} \right] + \log \frac{N_c}{N_s}.$$

Exclusionary Zoning ($\underline{h_i}$) This regulatory choice is binding when raising $\underline{h_i}$) raises revenue via the sorting of high productivity households:

$$\Delta \underline{h_i} \left(\Delta \eta_i N_i \cdot \left[t_i \frac{\partial r_i}{\partial \eta_i} + \tau \frac{\partial \tilde{w}}{\partial \eta_i} \right] + \Delta N_i \cdot \left[t_i r_i + \tau \tilde{w}_i + N_i t_i \frac{\partial r_i}{\partial N_i} \right] \right) \geq 0.$$

Value Capture (t_i) This regulatory choice is optimal when increased tax revenue offsets population loss:

$$(1-\gamma)\cdot \widetilde{N}_i + \frac{\partial \widetilde{N}_i}{\partial t_i}\left[\tau + (1-\gamma)\cdot t_i\right] \geq 0.$$

E5 Existence and Uniqueness of Solution

Solution for Value Capture To ensure the existence of an equilibrium, we start from the assumption that the government's objective function is non-negative, as governments always have the alternative of providing no public services, thereby incurring no cost. Hence, the optimal government service provision (G_i) must satisfy the first-order conditions rather than merely binding the budget constraint in equilibrium.

Under symmetric conditions, spatial equilibrium in equation (2) requires that the ratio of government services between the city core and the suburb ($\log(s_c^*/s_s^*)$) equals the ratio of their productivity parameters ($\log(\gamma_c/\gamma_s)$. This reduces equilibrium determination to solving the household indifference condition for population distribution (N_c^*), given that the left-hand side of equation (2) is constant and the right-hand side increases strictly with city core population (N_c). This yields a unique solution for N_c^*, implying unique equilibrium allocations for government services G_c^* and G_s^*.

Specifically, the unique equilibrium holds under the condition $d(RHS)/dN_c \cdot N_c^* \geq 1$ and $d(RHS)/dN_c \cdot N_s^* \geq 1$. If this condition fails, the equilibrium does not exist. However, because $d(RHS)/dN_c$ consistently exceeds $1/N_c + 1/N_s$, a unique equilibrium generally exists.

Solution for Exclusionary Zoning When examining existence under exclusionary zoning conditions, the equilibrium structure changes. Exclusionary zoning mandates imply all low-type households leave the zone that enforces such regulations. The household indifference condition thus applies solely to high-type households, as low-type households always prefer the non-exclusionary zone. Being constrained by exclusionary regulations imposes utility costs on low-type households, effectively excluding them from suburbs. In this scenario, equilibrium again depends on solving the household indifference condition (equation (2)) specifically for high-type households, confirming the existence of at most one internal equilibrium solution for the share of high-type households (η^*)). If such an internal equilibrium does not exist—–meaning the maximum utility difference between zones remains insufficient to attract high-type households—the outcome defaults to a corner solution: the city core is fully populated by low-type workers, and the suburb exclusively by high-type workers.