In this memo, I present preliminary descriptive findings about two possible large-sample measures of land use regulatory intensity and discuss their differences/similarities:

1. Wharton Residential Land Use Regulatory Index (WRLURI)
   1. See [Gyourko, Hartley, and Krimmel (2021)](https://www.sciencedirect.com/science/article/pii/S009411902100019X)
2. LLM Generative Regulatory Measure
   1. See [Bartik, Gupta, and Milo (2025)](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4627587)

**Differences**

These are fundamentally different approaches to building a comprehensive dataset of land use regulatory intensity. WRLURI conducts a survey (their survey instrument can be found [here](https://real-faculty.wharton.upenn.edu/wp-content/uploads/~gyourko/WRLURI/GyourkoHartleyKrimmel_NBERw26573appendix.pdf)), whereas the LLM Generative Regulatory Measure process administrative documents and legal information to directly estimate regulatory intensity. Unsurprisingly, these approaches yield very different stylized facts about regulatory intensity in the United States.

*Coverage*

First, the LLM measure has a much larger sample size. At the CBSA-level, the Wharton measure has data for 55 CBSAs, while the LLM measure has data for 713 CBSAs. 50 of those CBSAs are identified in both datasets. We can do a quick check for whether selection into the sample is biased by comparing the LLM-estimated overall regulatory intensity in the 50 CBSAs present in both datasets and the 663 CBSAs identified only in the LLM dataset:

Chart

AI-generated content may be incorrect.

This suggests there is an *upwards* bias for selection into the Wharton sample; thus, if we went only off the Wharton sample, we would over-estimate (relative to the LLM sample) regulatory stringency. Next, we can create a basic heatmap showing LLM-measured regulatory intensity across CBSAs in the country:

Map, scatter chart

AI-generated content may be incorrect.

*Interpretation of Index Magnitude*

The Wharton Index is composed of different types of questions; many of these questions, particularly regarding the involvement of local political actors, require respondents to place their answers on a simple 1-5 scale (from ‘no involvement’ to ‘very involved’). Further, questions about the specific zoning characteristics are ranked from a 1-6 scale (from ‘Land-use is not zoned’ to ‘Far less than demanded’). This is contrast to the LLM-generated measure, for which all questions are either continuous or binary.

Thus, the Wharton Index’s magnitude has no inherent value; what matters is a locality’s index relative to other localities. For the sake of consistency, here, I standardize all measures to the municipality’s z-score for that particular index.

*Measure Similarity*

First, we can compare the overall indices measures; by comparing z-scores, we’re essentially checking how similarly localities are placed in the distribution of regulatory intensity.

Chart, scatter chart

AI-generated content may be incorrect.

There is a positive association, although clearly there are series discrepancies between the measures; they are far from perfectly correlated. The LLM index has two principal comments: 1) Value Capture and 2) Exclusionary Zoning.

Chart, scatter chart

AI-generated content may be incorrect.

Chart, scatter chart

AI-generated content may be incorrect.

This is rather surprising to me; these results suggest that the Wharton Index has almost no correlation to the LLM regulations’ estimates of the (relative) intensity of zoning.

Nonetheless, both exhibit the about the same correlation between regulatory intensity and housing affordability:

Scatter chart

AI-generated content may be incorrect.

Chart, scatter chart

AI-generated content may be incorrect.