



FIRST INSIGHTS OF LAND DATA FROM WEB-SCRAPING

Pr. Jean-Charles Bricongne,
Thomas Berthelot,
Youssef Zribi

11/06/2023

Constructible land: why web-scraping may be useful

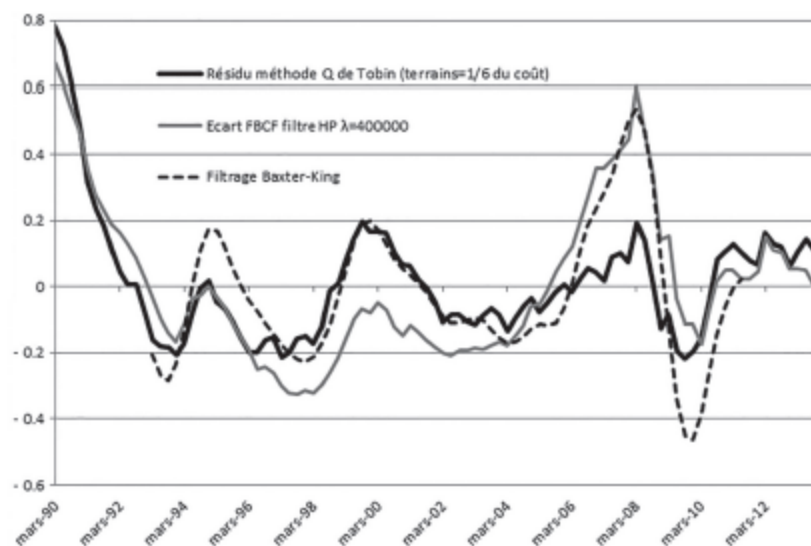
Data on housing: often scarce or not harmonized, even more for constructible land

Posted offers on websites complete official statistics, from the supply side, with more dynamic approach (stock of offers, duration, price revisions...)

Use prices of land for other purposes like the calculation of housing Tobin Q, calculate possible over-investment situations (see for example Bricongne Pontuch, 2017)

an-Charles Bricongne, Peter Pontuch « La crise s'est-elle accompagnée d'un sous-investissement dans l'immobilier résidentiel dans les pays de la zone euro ? », *revue économique* 2017/HS2 (Vol. 68), p. 129-157.
DOI 10.3917/reco.hs02.0129

5. Écarts d'investissement selon différentes méthodes (valeurs normalisées en divisant par la différence entre valeur maximale et valeur minimale)



Sources : Eurostat, Commission européenne, OCDE, BRI, sources nationales et calculs des auteurs.

Websites covered for France:

- Properstar
- Être Proprio
- Terrains
- Les Terrains
- Achat Terrain
- Figaro
- ParuVendu

Total number of lands scraped (after cleaning): **122,226 Lands**

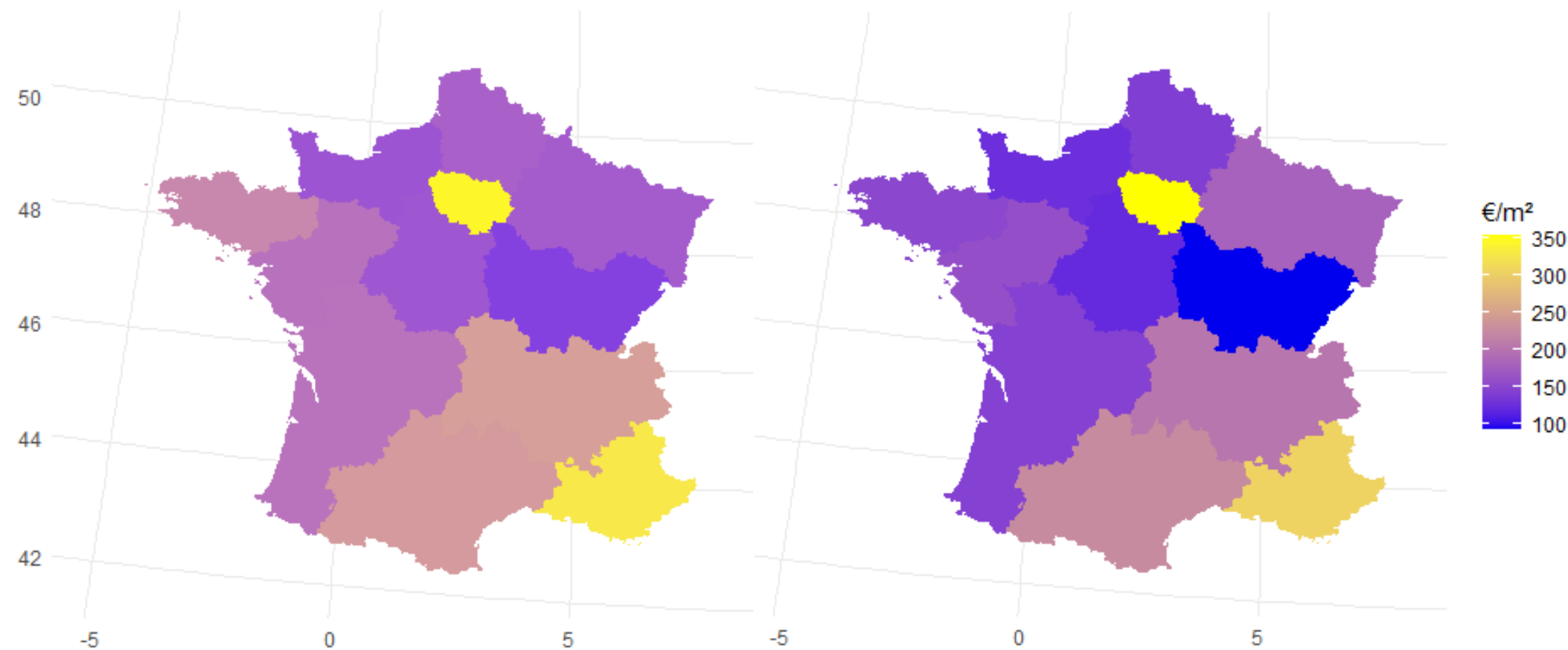
Comparing our data to available data: the EPTB (1/3)

- Data on the prices and characteristics of the land, the construction of the house, the applicant
- Perimeter: housing permits to households to build a house in non dense sectors
- Data collected over the year and the exploitation is annual
- Granularity: Region
- Exhaustive in theory but questions may not be answered

Our data:

- Data collected monthly or even more frequently on the price and characteristics of the land
- Perimeter: constructible land for sale, whatever the buyer, the sector and the construction
- Available in real time right after being scraped (posted offers and not transactions)
- Spatial granularity: Region, Department, City (GPS coordinates available for some websites such as Properstar)

Comparing our data to available data: the EPTB (2/3)



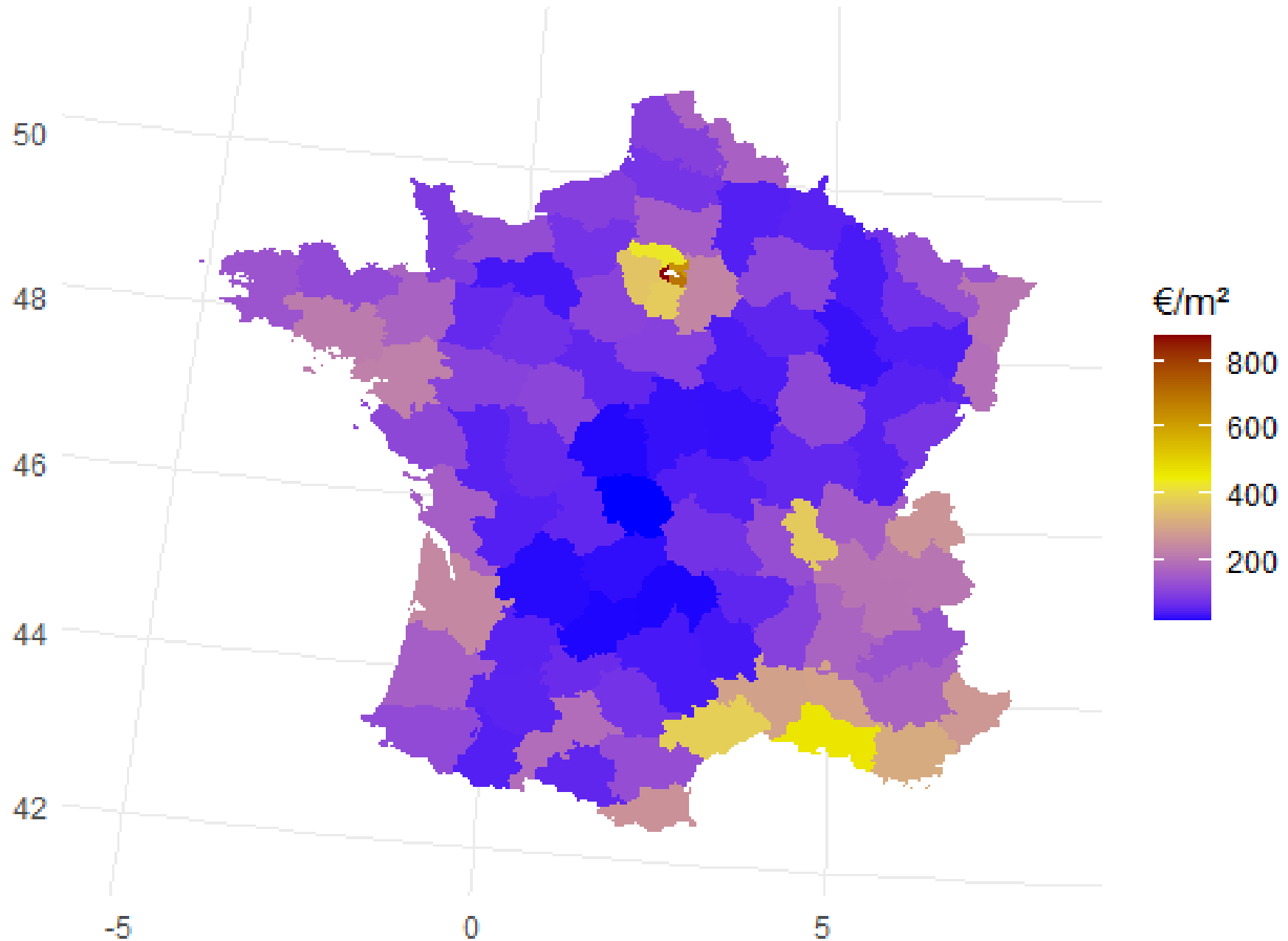
Mean Price/m² (€) from web-scraping data

Mean Price/m² (€) from the EPTB survey for Q3.2021

Comparing our data to available data: the EPTB ^(3/3)

Statistic	Variable	WebScr	EPTB2021
Mean	Price (€)	111231	85292
Mean	Area (m ²)	1271	948
Mean	Price/m ² (€)	159	94
Median	Price (€)	76900	-
Median	Area (m ²)	719	-
Median	Price/m ² (€)	106	-

Web-scraping data at the "French-department" level



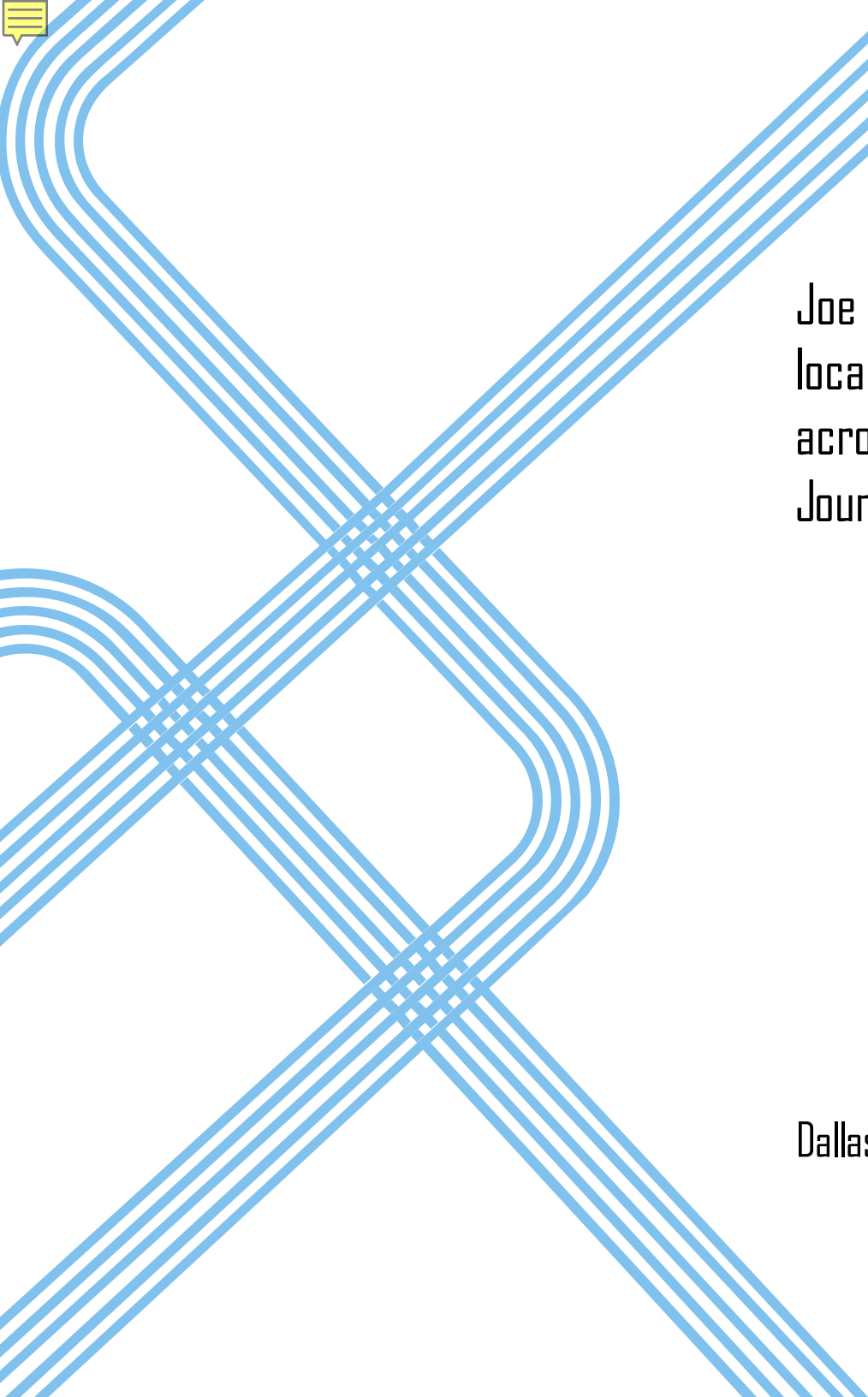
Average price per square meter in "French-departments"

OTHER COVERED COUNTRIES:

Country	Website	Number of lands
USA	LandWatch	32412
Ireland	Daft.ie	2956
New-Zealand	realestate.nz	2565

Descriptive statistics for scraped land prices in Ireland by region

Region	Number of lands	Mean Price per m ²	Mean Surface
Dublin	17	449,63	4054,00
Cork	162	91,64	3264,71
Galway	228	29,83	4462,90
Kildare	22	155,93	16860,68
Meath	44	30,22	3905,22
Limerick	62	31,66	5279,19
Tipperary	35	26,22	4524,39
Donegal	162	19,82	4983,38
Wexford	62	23,73	5299,42
Kerry	87	38,86	6281,93
Wicklow	32	137,13	6841,72
Louth	17	24,01	5158,55
Mayo	216	17,36	11456,73
Clare	107	27,68	3673,94
Waterford	25	45,95	5610,56
Kilkenny	37	35,23	5677,63
Westmeath	53	29,98	5047,88
Laois	30	28,51	4749,66
Offaly	31	22,74	4234,84
Cavan	51	14,50	7600,95
Roscommon	48	24,56	3799,83
Sligo	41	23,47	6266,71
Monaghan	10	14,75	4552,71
Carlow	10	26,19	2893,50
Longford	26	16,27	3888,10
Leitrim	10	17,59	4795,52


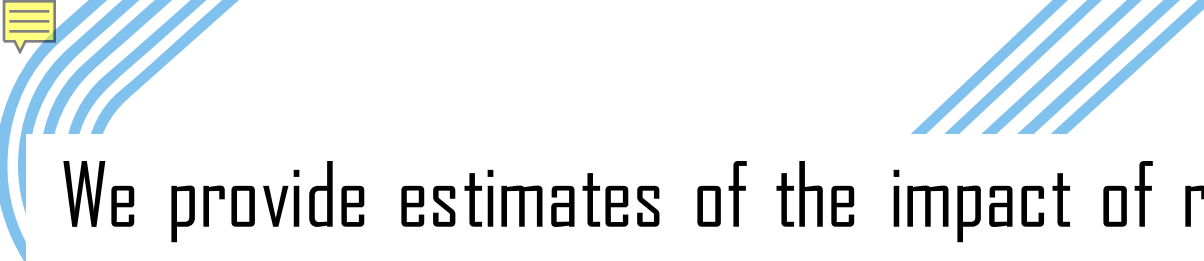


Joe Gyourko and Jacob Krimmel (2021), The impact of local residential land use restrictions on land values across and within single family housing markets, *Journal of Urban Economics*, 126.


Introduction of the zoning tax index

Dallas : 1.298 with scraped data

Dallas : 1.66 from their paper



We provide estimates of the impact of restrictive residential land use environments on the price of land across major American housing markets. Using micro data on vacant land purchased to develop single family housing, we implement a new empirical strategy for estimating so-called 'zoning taxes' –the amount by which land prices are bid up due to supply side regulations. Our results are broadly consistent with previous findings that zoning taxes are especially burdensome in large coastal markets. [...] Finally, we show that our zoning tax estimates are strongly positively correlated with a new measure of local housing market supply constraint (the Wharton Residential Land Use Regulatory Index of 2018) [...]



The price of a house [$P(H)$] can be defined as the sum of physical construction costs (CC) and the price of land [$P(L)$].

$$P(H) = CC + P(L) = CC + qA + Z. \quad (1)$$

Moreover, the value of land can be conceived as being made up of two components. One is the price an existing homeowner places on having an extra square foot of lot (q) times the amount of acreage (A) on which the house sits— qA . This is the value of land on the intensive margin. Market prices of land could exceed qA if additional value is generated by binding supply restrictions. [Glaeser and Gyourko \(2003, 2018\)](#) call that increment the ‘zoning tax’ or Z . Thus, $P(L) = qA + Z$ in [Eq. \(1\)](#); if $Z = 0$ so that there is no binding regulation creating artificial, policy-induced scarcity value, then $P(L) = qA$, with extensive and intensive margin land values being identical.

In this paper, we use micro observations on the actual prices paid for vacant land bought explicitly for the purpose of building single-family homes. In these data, $P(L)$ still is the extensive margin value of land, but now it is the product of the number of houses the buyer intends to build on the land (N), times the difference between what it can sell those houses for [$P(H)$] and what it costs to build those homes (CC). Thus,

$$P(L) = N * [P(H) - CC]. \quad (2)$$

Substituting in from (1) yields

$$P(L) = N * [CC + qA + Z - CC] = N * [qA + Z] \text{ or } P(L)/N = qA + Z. \quad (3)$$

The price of land paid per expected housing unit equals the sum of the intensive margin value and the zoning tax. If $P(L)/N = qA$, the zoning tax per home is zero. As argued above, profitable arbitrage ensures



COMPARING WITH
THE AMERICAN
ENTREPRISE
INSTITUTE

Region/ ppsm	Official stats	Web-scraping data
DALLAS	13,97\$	13,57\$
Greenville	65,85\$	63,60\$

Next steps

Cover more countries

Finish the calculations of zoning tax for the US and extend it to other countries, have an additional indicator on supply constraints that would be quite comparable across countries

Make the data (at least partly aggregated) public