

For What It's Worth: Measuring Land Value in the Era of Big Data and Machine Learning

SGE/ASSA 2022

Gary Cornwall

Bureau of Economic Analysis
Office of the Chief Economist

Jeremy Moulton

University of North Carolina, Chapel Hill
Department of Public Policy

Scott Wentland

Bureau of Economic Analysis
Office of the Chief Economist



***Disclaimer:** Any views expressed here are those of the authors and not necessarily those of the Bureau of Economic Analysis or the U.S. Department of Commerce. The results and opinions are those of the authors and do not reflect the position of Zillow Group.

This work is preliminary and is still being developed. Please contact the authors before you cite this paper for the latest draft/results.



Stocks & Flows of National Accounts: Where's Land?

Table 1. Households and Nonprofit Institutions Serving Households
[Billions of dollars]

	Line	2003	2004	2005		Line	2003	2004	2005
Current account									
Gross value added	1	1,269.2	1,356.5	1,419.6	Long term (mortgages)	76	864.4	981.7	1,107.1
Less: Consumption of fixed capital	2	201.7	235.9	293.5	Insurance technical reserves (unpaid premiums)	77	0.9	1.6	1.0
Equals: Net value added	3	1,067.5	1,120.6	1,126.1	Other accounts payable (trade debt)	78	4.8	2.0	2.0
Compensation paid by households and NPISHs	4	503.5	529.8	552.4	Addendum:				
Wages and salaries	5	436.8	458.9	477.2	Net lending or net borrowing, financial account (39-68)	79	-97.7	-302.4	-629.6
Employers' social contributions	6	66.7	70.9	75.3	Other changes in volume account				
Taxes on production and imports less subsidies	7	128.0	136.6	142.3	Total other volume changes	80	514.9	209.6	150.2
Operating surplus, net	8	436.0	454.1	431.4	Net investment in consumer durable goods	81	205.7	208.2	210.2
Net national income/balance of primary incomes, net	9	8,427.9	8,930.9	9,357.9	Other volume changes	82	199.8	12.5	47.6
Operating surplus, net	10	436.0	454.1	431.4	Less: Statistical discrepancy (37-39-68) ^a	83	-109.4	11.1	107.6
Compensation of employees (received)	11	6,325.4	6,650.3	7,030.3	Revaluation account				
Wages and salaries	12	5,127.7	5,377.1	5,664.8	Nonfinancial assets	84	1,122.3	1,652.8	1,986.4
Employers' social contributions	13	1,197.7	1,273.2	1,365.5	Real estate	85	1,220.5	1,674.8	2,025.8
Property income (received)	14	2,201.1	2,385.6	2,528.4	Consumer durable goods	86	-98.4	-21.7	-39.0
Interest	15	915.4	892.1	946.3	Equipment and software	87	0.2	-0.3	-0.3
Distributed income of corporations	16	1,285.8	1,493.5	1,582.1	Financial assets	88	3,342.8	1,988.4	1,713.3
Dividends	17	422.6	537.1	574.4	Shares and other equity	89	1,912.8	1,251.5	1,323.0
Withdrawals from income of quasi-corporations ¹	18	863.2	956.5	1,007.7	Corporate equities	90	1,078.9	361.2	232.7
Less: Uses of property income (interest paid)	19	534.6	559.1	632.2	Mutual fund shares	91	424.4	276.2	244.8
Net national income/balance of primary incomes, net	20	8,427.9	8,930.9	9,357.9	Equity in noncorporate business	92	409.5	614.1	845.5
Less: Current taxes on income, wealth, etc. (paid)	21	1,001.1	1,049.8	1,203.1	Insurance technical reserves	93	1,429.9	736.8	845.5
Plus: Social benefits (received)	22	1,316.7	1,398.4	1,480.9	Changes in net worth due to nominal holding gains or losses	94	4,465.4	3,994.4	3,798.8
Less: Social contributions (paid)	23	778.6	826.4	880.6	Changes in balance sheet account				
Plus: Other current transfers (received)	24	34.3	28.1	45.7	Change in net worth (32+37+80+94)			3,994.4	3,798.8
Less: Other current transfers (paid)	25	105.7	110.4	93.3	Balance sheet account (end of period)				
Equals: Disposable income, net	26	7,893.5	8,370.9	8,707.5					
Less: Final consumption expenditures	27	7,703.6	8,211.5	8,742.4					
Equals: Net saving	28	189.9	159.3	-34.8					
Capital account									

Balance sheet account (end of period)				
Total assets	96	53,780.0	58,967.7	63,975.4
Nonfinancial assets	97	20,238.8	22,523.9	25,173.1
Real estate	98	16,675.0	18,759.0	21,222.6
Consumer durable goods	99	3,380.3	3,566.8	3,738.0
Equipment and software	100	183.4	198.1	212.5
Financial assets	101	33,541.1	36,443.8	38,802.3
Currency and deposits	102	4,330.3	4,730.5	5,110.9
Currency and transferable deposits	103	286.8	319.0	308.1
Other deposits	104	4,043.5	4,411.5	4,802.8
Foreign deposits	105	52.1	57.5	62.7
Time and savings deposits	106	3,991.3	4,353.9	4,740.1
Securities other than shares	107	2,381.7	2,587.7	2,853.1
Open market paper	108	105.9	136.1	164.2
U.S. savings bonds	109	203.8	204.4	205.1
Treasury securities	110	236.8	358.3	342.1
Agency- and GSE-backed securities ²	111	388.5	435.3	638.8
Municipal securities	112	707.7	740.9	816.0
Corporate and foreign bonds	113	739.0	712.6	687.0
Loans	114	623.2	738.6	740.7
Short term (security credit)	115	475.4	578.3	567.4
Long term (mortgages)	116	147.9	160.2	173.3
Shares and other equity	117	15,058.0	16,187.1	17,310.1
Corporate equities	118	5,613.1	5,715.2	5,491.6
Mutual fund shares	119	3,085.4	3,610.7	4,121.4
Money market fund shares	120	959.8	903.5	957.3
Equity in noncorporate business	121	5,399.6	5,957.7	6,739.9
Insurance technical reserves	122	11,147.9	12,200.0	12,787.4
Net equity in life insurance and pension funds	123	10,685.9	11,697.2	12,259.3
Net equity in life insurance reserves	124	1,013.2	1,060.4	1,082.6
Net equity in pension fund reserves	125	9,672.7	10,636.8	11,176.7
Prepayments of premiums and reserves against claims	126	462.0	502.8	528.1
Net equity in reserves of property-casualty insurance companies	127	252.0	273.0	295.1
Net equity in other life insurance company reserves	128	199.5	211.0	217.3
Net equity in Uniformed Services Retiree Health Care Fund	129	10.5	18.8	15.7
Total liabilities and net worth	130	53,780.0	58,967.7	63,975.4

Land is a nonproduced, nonfinancial asset.

- Implicitly included in nonfinancial assets among household real estate
 - Does not separate out structure vs. land value
- Excludes land from corporate assets and other places on the balance sheet
- Excludes land from government assets

Why are we here?

- **Research question:** Can we estimate what (private) U.S. land value is worth?
 - Real estate appears in non-produced, non-financial assets but there is no breakout for **land**.
 - **Our contribution:** building aggregate land value estimates from the “ground up” using fine microdata (“Big Data”) and methodological approaches that separate land from structure value – beginning with single family residential (SFR) land...so far
 - **Data:** Zillow ZTRAX microdata
 - **Methods:**
 - Hedonic OLS regression approach vs. machine learning approach; linear extrapolation of structure value
 - Data driven clusters using ML approach → better land value estimate?
- **What do we find?**
 - Data driven clusters with GBT has [substantially] better price prediction.
 - Land values are [on average] higher than traditional hedonic approach.

Conceptually, how should we value land?

- Vacant land transactions?
 - Few sales & selection bias → less desirable land developed last...
- Market transactions of properties with structures & land
 - a linear and additive model where the selling price of a property V , the value of the structure $p_s S$, and the value of the plot of land $p_L L$ can be written as:
$$V = p_s S + p_L L$$
 - where S is the size (and quality) of the structure, L is the land area and location, and p_s and p_L are the prices of a unit of S and L , respectively





Zillow

Save

Share

More

\$499,900

4 bd

3 ba

3,561 sqft

903 Kings Mill Rd, Chapel Hill, NC 27517

For sale Zestimate: \$523,664

Est. payment: \$2,660/mo

Get pre-qualified

Contact Buyer's Agent

Overview

Days on Zillow

143

Views

2,455

Saves

80

Check out fun upcoming neighborhood event: <https://youtu.be/VTKZePoCA5A> ! Own a piece of North Carolina history with this sweet house in James Taylor's childhood neighborhood! Probably LESS than "half a mile down to Morgan Creek," right next to Botanical Gardens, can even ride a bike to UNC. MASSIVE INCOME POTENTIAL with large, exterior-entry apartment with updated kitchenette-- perfect for professional students, as an in-law suite, or just extra space for the young or young-hearted. Otherwise, house lives like a one-level palace in the trees. Recently-improved landscaping!

What I love about this home

Great home in an incredible neighborhood

Facts and features

Type: Single Family

Year built: 1962

Heating: Forced air

Cooling: Central

Parking: On street

Lot: 0.92 acres

Price/sqft: \$140

Interior features

BEDROOMS

Beds: 4

HEATING AND COOLING

Heating: Forced air

Heating: Gas

Cooling: Central

APPLIANCES

Appliances included: Dishwasher

FLOORING

Floor size: 3,561 sqft

Flooring: Carpet, Hardwood, Tile

OTHER INTERIOR FEATURES

Fireplace

Ceiling Fan

Room count: 11

Spaces and amenities

SIZE

Unit count: 1

SPACES

Jetted Tub

Construction

TYPE AND STYLE

Structure type: Other

Single Family

MATERIALS

Roof type: Shake/Shingle

Exterior material: Brick

Foundation type: Masonry

Skylight

DATES

Last remodel year: 1963

Built in 1962

OTHER CONSTRUCTION FEATURES

Stories: 2

Exterior features

PATIO

Deck

Porch

Patio

LOT

Lot: 0.92 acres

OTHER EXTERIOR FEATURES

Parcel #: 9788807899

Home value

~~\$523,664~~
Zestimate

Zestimate range \$497,000 - \$571,000

Last 30-day change -\$982 (-0.2 %)

One-year forecast \$541,259 (+3.4 %)

Tax history

YEAR	PROPERTY TAXES	TAX ASSESSMENT
2018	\$8,330 --	\$506,600 (+1.4%)
2017	\$8,330 (+2.3%)	\$499,700 (+2.1%)
2016	\$8,144 --	\$489,316 --
2015	\$8,144 --	\$489,316 --
2014	-- --	\$489,316 --

Price history

DATE	EVENT	PRICE
2/2/2019	Back on market	\$499,900 --
1/26/2019	Pending sale	\$499,900 --
11/7/2018	Price change	\$499,900 (-2.9%)
10/19/2018	Listed for sale	\$515,000 (-0.8%)
9/16/2018	Listing removed	\$519,000 --
7/11/2018	Price change	\$519,000 (-5.6%)
6/17/2018	Price change	\$550,000 (-2.7%)
5/16/2018	Listed for sale	\$565,000 (+25.6%)
8/5/2016	Sold	\$450,000 (-5.3%)
6/22/2016	Listing removed	\$475,000 --
6/1/2016	Price change	\$475,000 (-5%)
4/7/2016	Listed for sale	\$500,000 --

[See less price history](#)

Zestimate history

— This home --
- - - Chapel Hill --



Hedonic Approach (Baseline)

- Use market transaction data and structure characteristics to decouple structure and land value
 - By year and state, regress the transaction prices on a detailed set of covariates (square footage, bedrooms, bathrooms, number of stories, age, etc.) and acreage, along with location fixed effects and interactions

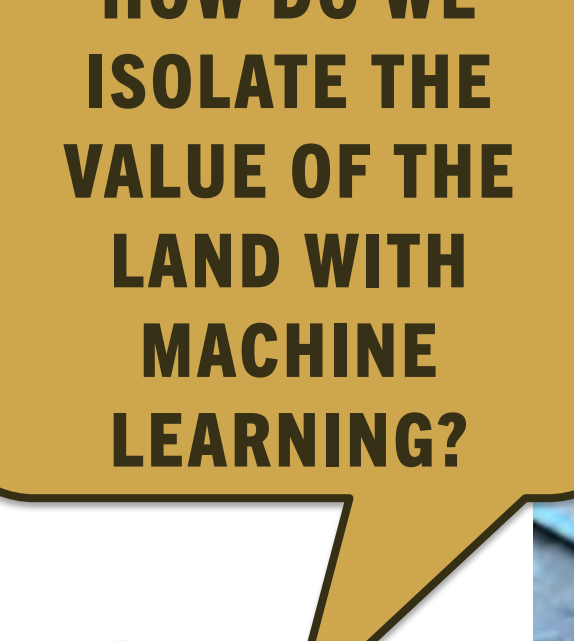
$$\text{Property Sale Price}_{ijt} = \alpha + \sum \beta X_i + \gamma \text{LOCATION}_j + \sum \delta \text{sq.ft.}_i * \text{LOCATION}_j + \sum \phi \text{acreage}_i * \text{LOCATION}_j + \rho \text{TIME}_t + \varepsilon$$

- Use these coefficients to generate individual level price and land estimates

$$\text{Residential Land Value}_{ijt} = \alpha + \gamma \text{LOCATION}_j + \sum \phi \log(\text{acreage}_i) * \text{LOCATION}_j + \rho \text{TIME}_t$$

- Construct weighted average estimates of any geography (national, regional, state, county) aggregating up from millions of individual observations for nearly the universe of properties in the tax assessment data
 - Can run separately: dense urban, urban, suburban (SFH), rural, commercial, industrial, ag



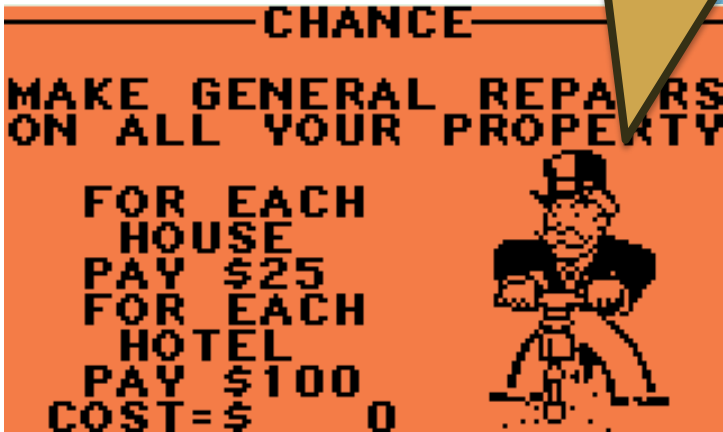


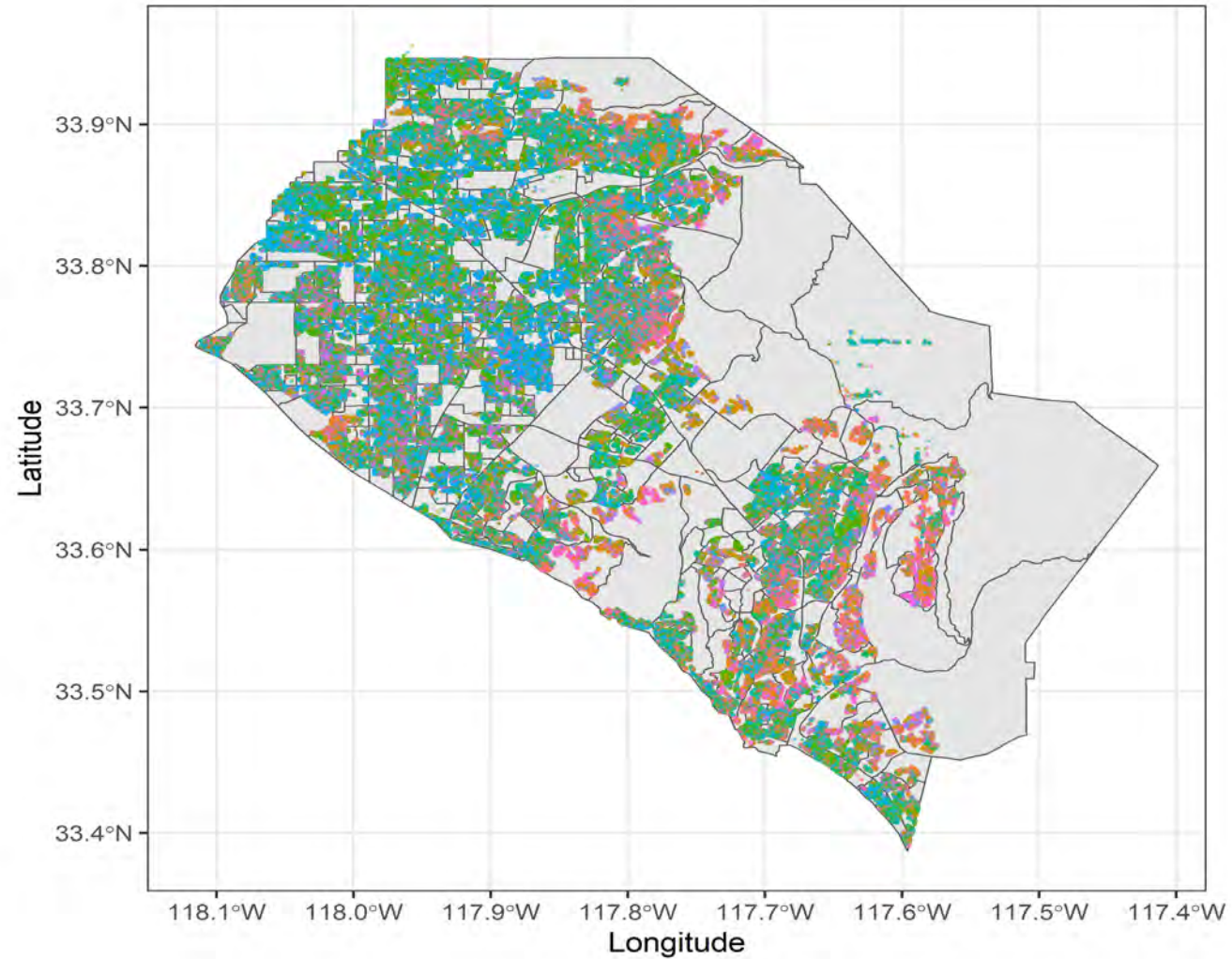
**HOW DO WE
ISOLATE THE
VALUE OF THE
LAND WITH
MACHINE
LEARNING?**



By clustering over location and common elements we load within cluster price variation on square footage.

This allows us to back out the structure value!





Machine Learning Approach

Table 1: Within Cluster Standard Deviations (California Assessment Data)

	Minimum	First Quartile	Median	Mean	Third Quartile	Max
Generated Clusters: Square Feet	206.9	328.8	451.1	440.7	526.4	812.2
Generated Clusters: Acreage	0.052	0.178	0.242	0.256	0.340	0.548
Generated Clusters: Bedrooms	0.000	0.035	0.245	0.264	0.425	0.857
Generated Clusters: Bathrooms	0.000	0.104	0.212	0.237	0.370	0.840
Census Block Group: Square Feet	0.000	346.6	441.1	474.4	578.5	1908.5
Census Block Group: Acreage	0.000	0.030	0.057	0.132	0.156	1.311
Census Block Group: Bedrooms	0.000	0.605	0.707	0.709	0.807	2.121
Census Block Group: Bathrooms	0.000	0.458	0.571	0.575	0.683	2.121

Machine Learning Approach

- Assign time invariant clusters based on like characteristics of property and location.
- Predict price using Gradient Boosted Trees

$$\textit{Property Sale Price} = f(\textit{latitude}, \textit{longitude}, \textit{sqft}, \textit{acreage}, \textit{year built}, \textit{cluster})$$

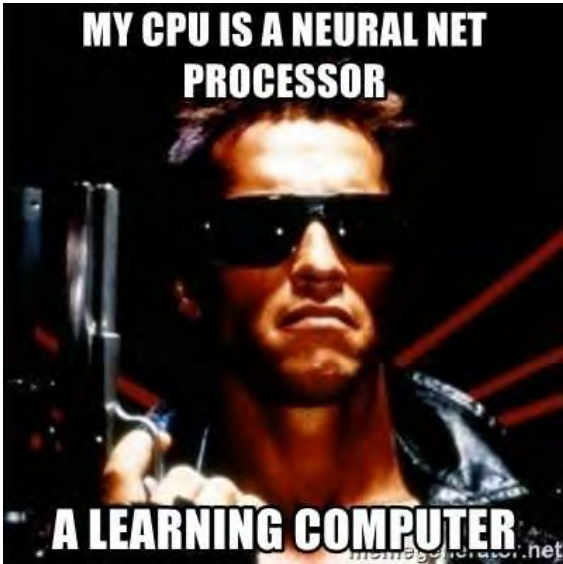
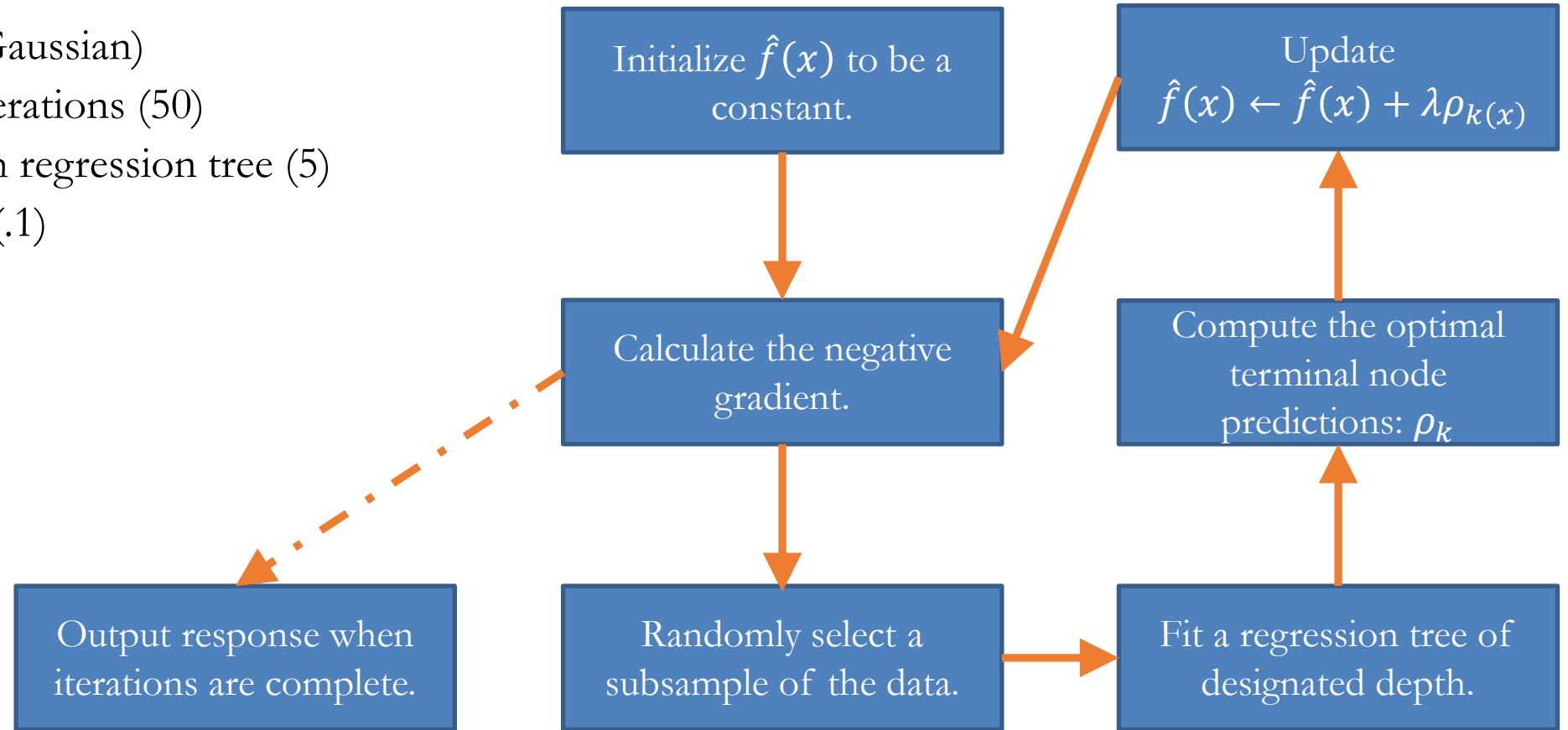
- Estimate new price conditional upon marginal change in structure square footage and extrapolate back to zero square feet.
- Construct an estimate of the land value:

$$lv_{i,t} = \tilde{P}_{i,t|sqft} - \theta(\tilde{P}_{i,t|sqft+} - \tilde{P}_{i,t|sqft})$$



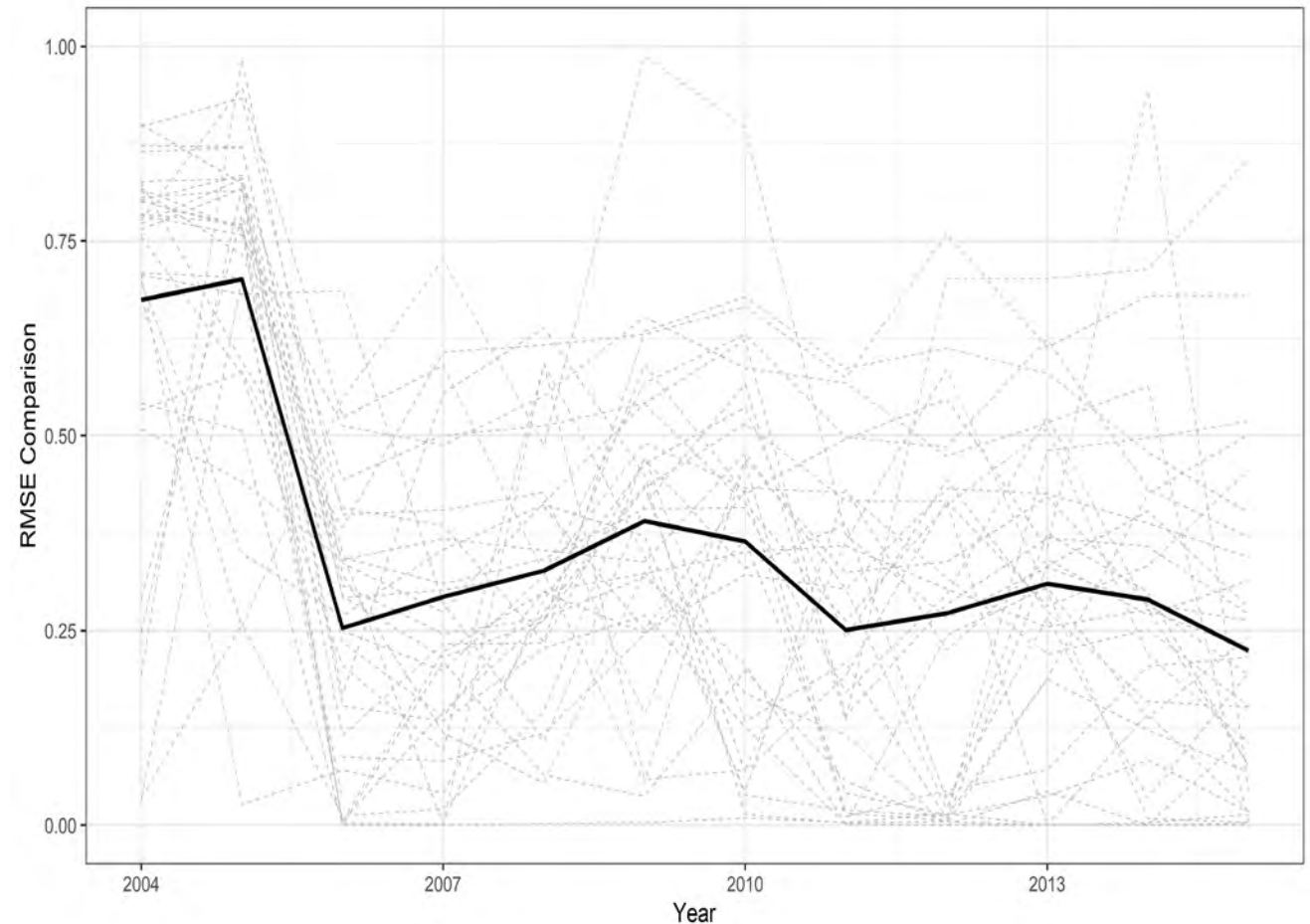
Machine Learning (ML) Approach

- Choose:
 - A loss function (Gaussian)
 - The number of iterations (50)
 - The depth of each regression tree (5)
 - The learning rate (.1)

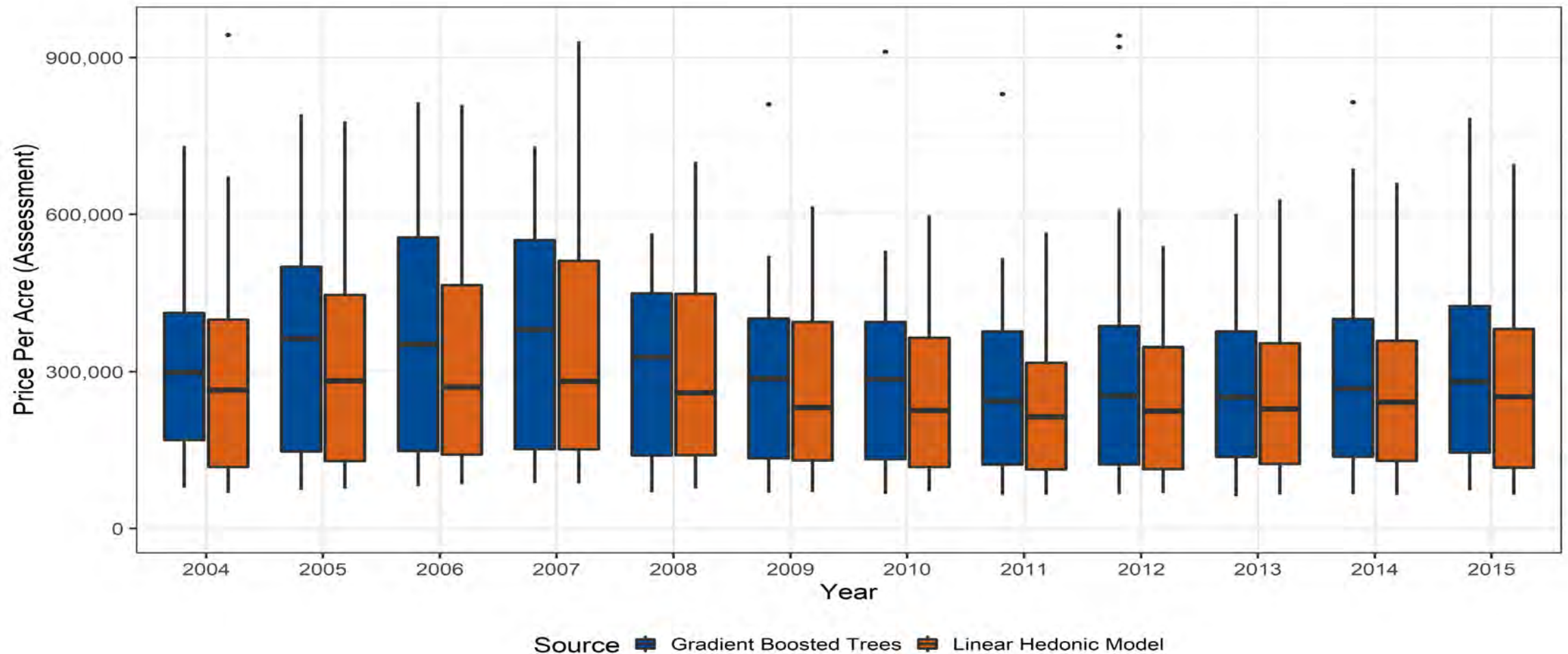


Better Predictions, Better Land Values?

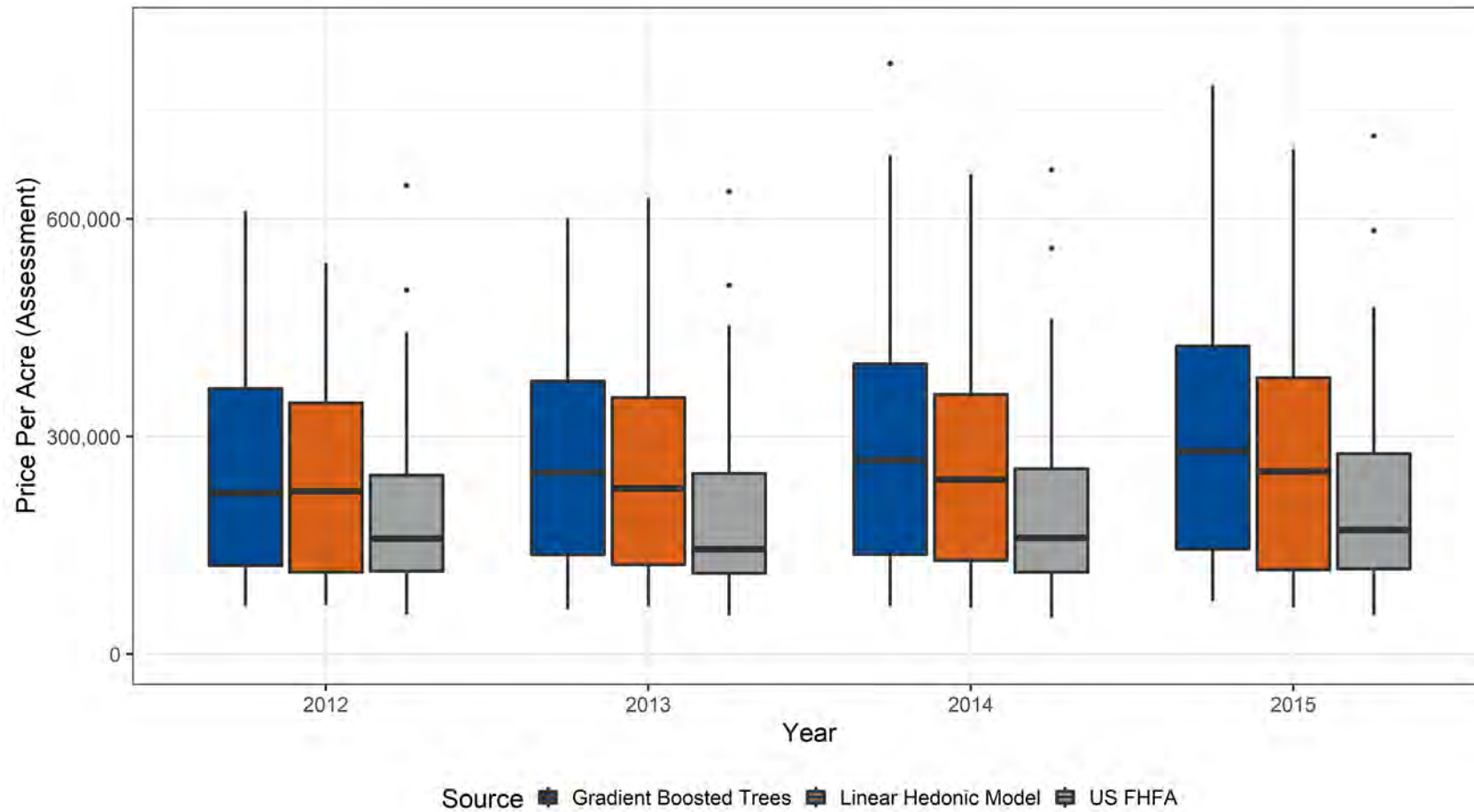
- Calculating the land value is an exercise in decomposing predictions.
- We can evaluate, how well our models predict on price:
 - $RMSE_{m,t} = \sqrt{N^{-1} \sum_{i=1}^N (p_{i,t} - \tilde{p}_{i,t})^2}$
 - $e_t = RMSE_{GBT,t} / RMSE_{LH,t}$
 - If $e_t = 1$ then same loss from each model.
 - If $e_t > 1$ then hedonic approach provides better predictions of price.
 - If $e_t < 1$ then gradient boosted trees is better at prediction.



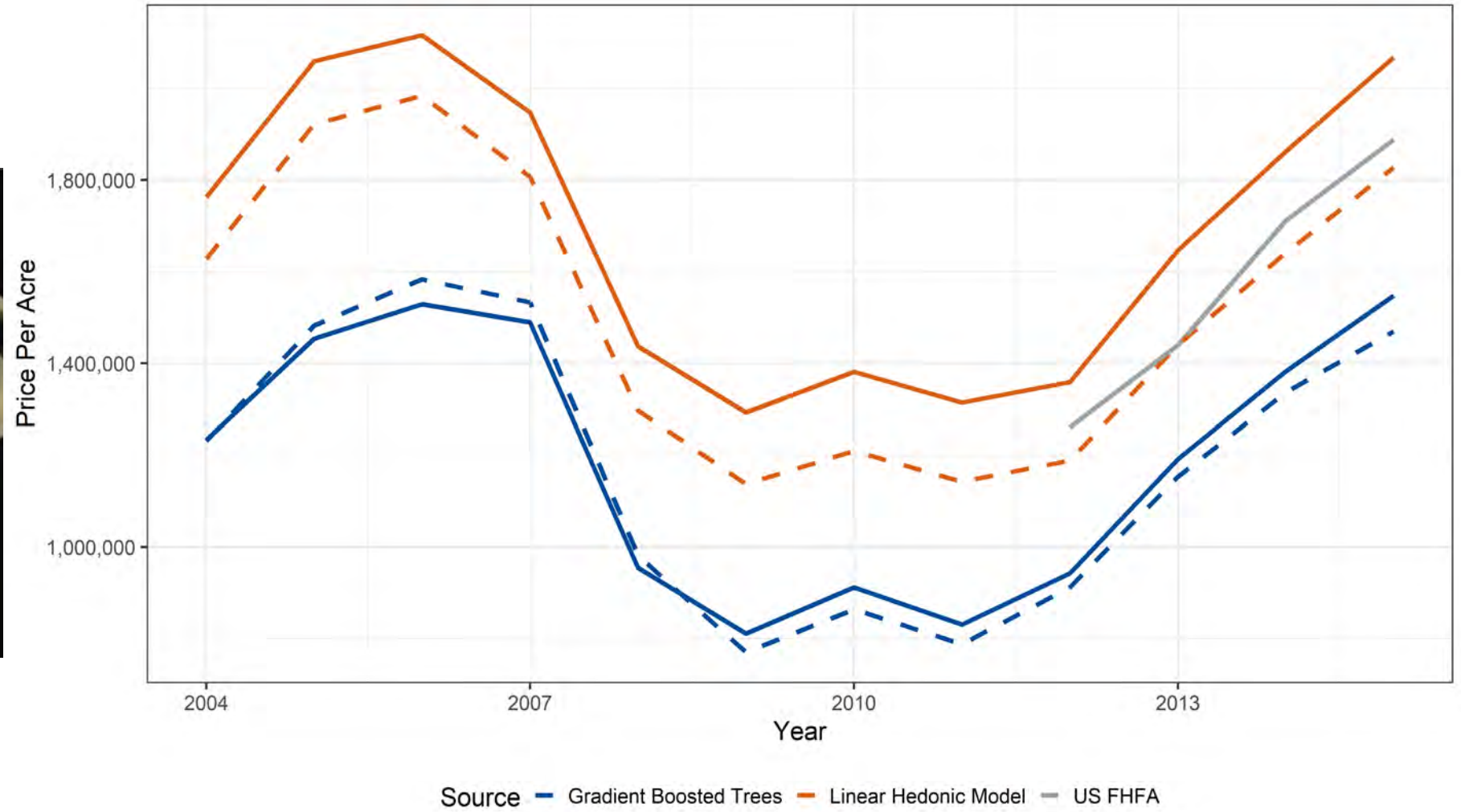
The Value of Land



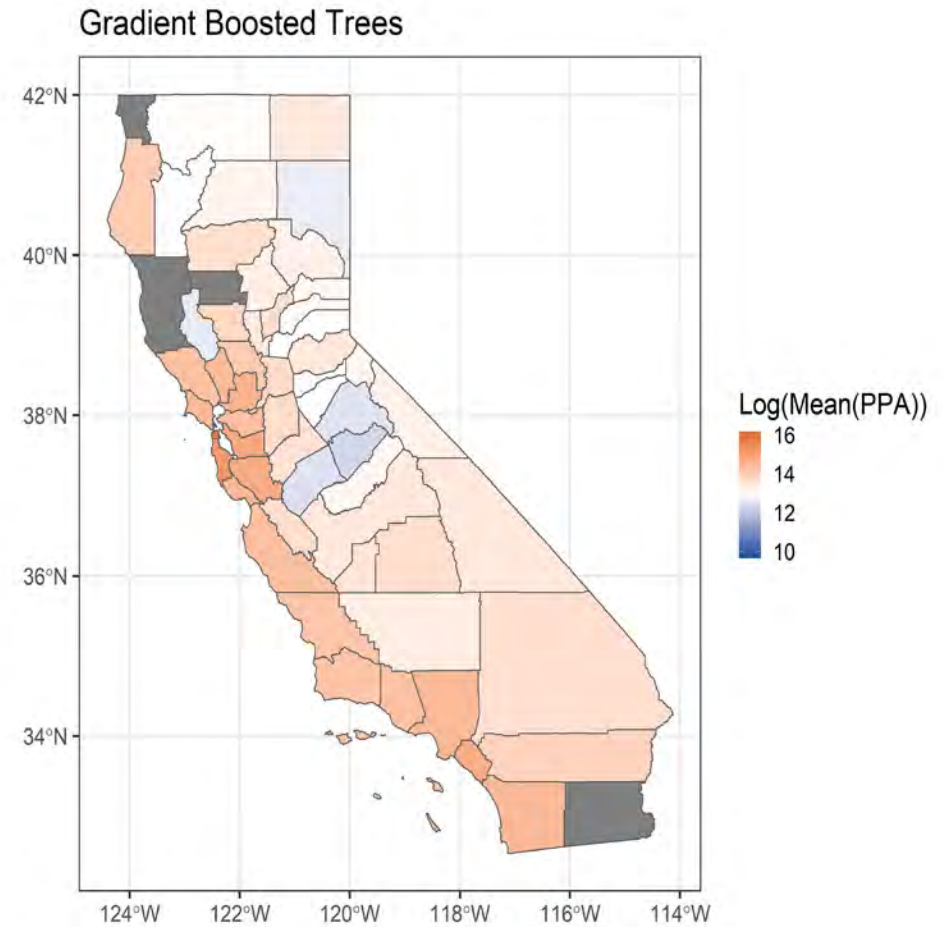
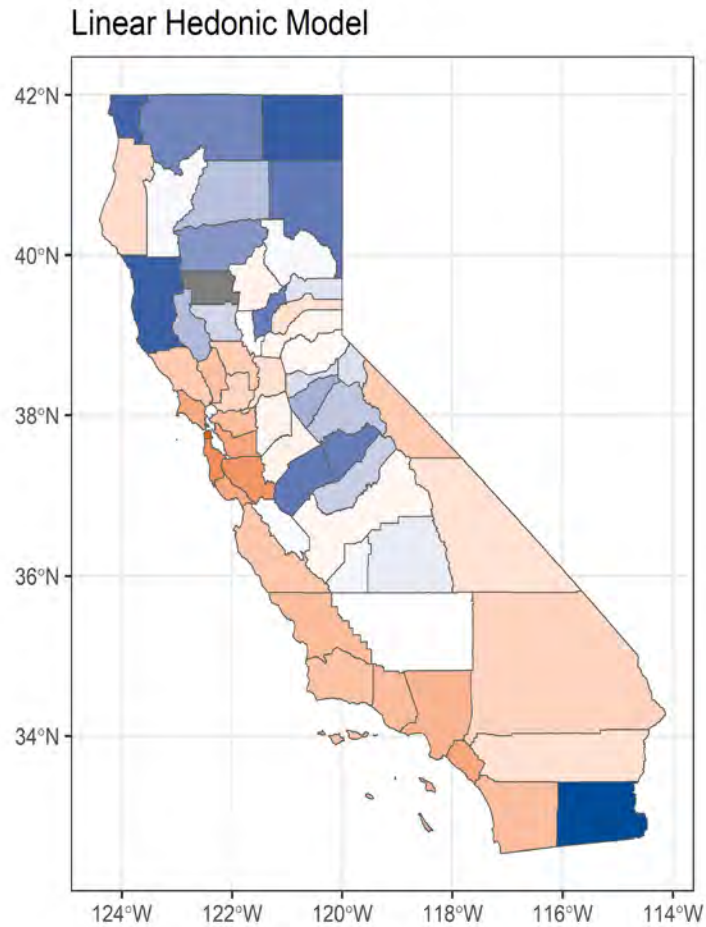
Comparing Land Values

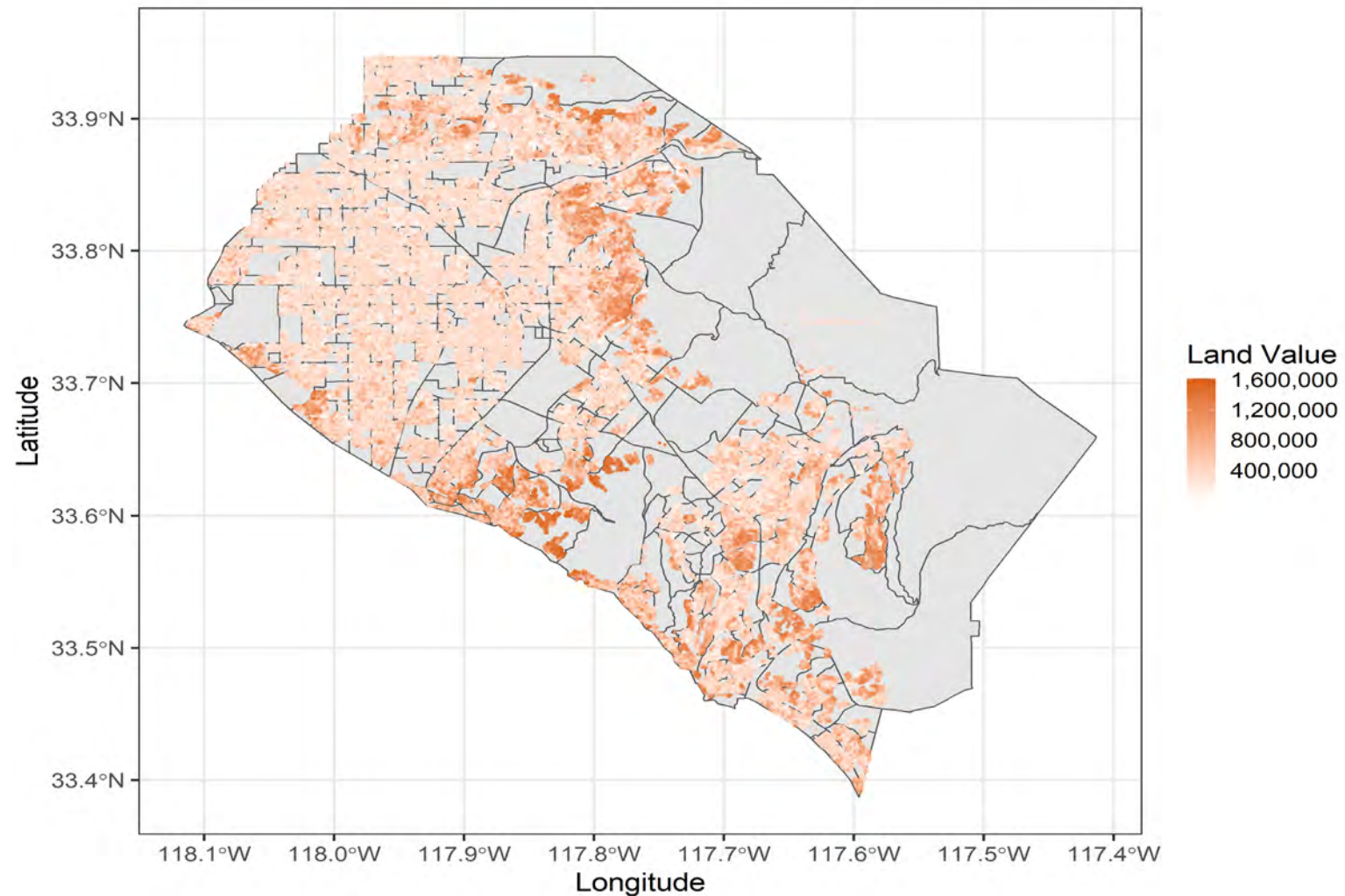


How expensive is land in California?



Price-Per-Acre By County (2013): California





What About Land Leverage?

Land in California is expensive!

- $LL_t = N^{-1} \sum_{i=1}^N \tilde{v}_{i,t} / \tilde{p}_{i,t}$
- Between 60-70% of purchase price is land alone.
- This varies across state
 - 65% (7.9) for GBT
 - 50% (7.3) for LH



Next Steps

- Expand these methods to all land types across all regions to the US and update data (next slide from prior research)
 - Wentland, S.A., Ancona, Z.H., Bagstad, K.J., Boyd, J., Hass, J.L., Gindelsky, M. and Moulton, J.G., 2020. Accounting for land in the United States: Integrating physical land cover, land use, and monetary valuation. *Ecosystem Services*, 46, p.101178.
- Explore alternative data sources to fill in holes in the Zillow data
- Continue to refine ML Approach
 - Data driven clusters in hedonic model rather than location fixed effects?
 - Three-year window in ML method?



Thank You

gary.cornwall@bea.gov
moulton@email.unc.edu
scott.wentland@bea.gov



Acreage (National Land Use Database), Total Value, and Average (Nominal) Price Per Acre by Census Division

	NLUD 2010 Total Acreage (000s)	2002 – 2006		2007 - 2011		2012 – 2016		
		Total Value (\$Billions)	Average Price Per Acre (\$)	Total Value (\$Billions)	Average Price Per Acre (\$)	Total Value (\$Billions)	Average Price Per Acre (\$)	
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
West Region	Pacific							
	Dense Urban Residential	237	1,098	4,637,323	642	2,713,550	1,032	4,359,191
	Urban Residential	2,415	2,284	945,988	1,360	563,433	2,182	903,476
	Suburban Residential	1,629	1,870	1,147,403	1,116	684,984	1,799	1,103,923
	Rural Residential	9,893	310	31,356	186	18,789	248	25,045
	Commercial	611	376	615,376	295	483,490	495	810,437
	Industrial	261	107	408,685	84	319,963	163	624,334
	Agricultural	78,480	485	6,183	404	5,145	578	7,370
	Mountain							
	Dense Urban Residential	81	214	2,635,320	110	1,353,508	176	2,161,364
	Urban Residential	1,383	786	568,303	675	487,965	792	572,334
	Suburban Residential	1,263	837	662,197	444	351,817	728	576,506
	Rural Residential	7,587	353	46,481	188	24,744	250	32,937
	Commercial	521	379	727,706	256	491,118	343	658,781
	Industrial	212	69	324,449	45	212,820	67	317,188
	Agricultural	218,751	1,605	7,336	934	4,270	1,386	6,335
Midwest Region	West North Central							
	Dense Urban Residential	49	76	1,554,171	54	1,091,550	55	1,124,492
	Urban Residential	1,377	650	471,700	481	349,110	485	351,834
	Suburban Residential	1,246	275	220,419	201	160,961	221	177,411
	Rural Residential	11,073	73	6,552	56	5,078	49	4,447
	Commercial	510	140	273,910	103	202,146	112	218,827
	Industrial	268	45	168,742	41	151,358	40	150,164
	Agricultural	269,990	944	3,495	1,069	3,959	1,325	4,907
	East North Central							
	Dense Urban Residential	148	256	1,729,655	110	746,892	156	1,058,420
	Urban Residential	2,872	596	207,561	390	135,654	624	217,215
	Suburban Residential	2,640	675	255,521	376	142,554	489	185,308
	Rural Residential	24,793	216	8,694	148	5,987	177	7,124
	Commercial	715	177	248,105	114	160,012	146	204,469
	Industrial	441	63	142,343	40	91,775	57	128,183
	Agricultural	95,720	359	3,753	358	3,740	426	4,455

West South Central							
Dense Urban Residential	98	44	448,445	54	546,679	64	645,533
Urban Residential	2,066	364	175,942	454	219,766	427	206,656
Suburban Residential	2,318	243	105,035	258	111,168	206	89,017
Rural Residential	22,875	136	5,960	143	6,247	127	5,556
Commercial	809	132	163,381	129	159,722	134	165,989
Industrial	388	33	86,269	28	72,823	13	32,987
Agricultural	207,344	292	1,410	331	1,597	474	2,288
East South Central							
Dense Urban Residential	32	43	1,353,853	34	1,054,261	40	1,256,018
Urban Residential	797	170	213,266	178	223,257	148	185,475
Suburban Residential	1,810	148	81,754	123	67,906	155	85,543
Rural Residential	29,328	90	3,078	83	2,817	93	3,174
Commercial	412	40	97,521	39	94,097	48	117,321
Industrial	240	14	56,951	13	55,907	15	63,405
Agricultural	64,973	168	2,581	140	2,149	158	2,431
South Atlantic							
Dense Urban Residential	210	619	2,949,178	357	1,699,788	485	2,308,539
Urban Residential	3,049	1,109	363,812	685	224,568	842	276,219
Suburban Residential	5,116	1,293	252,681	769	150,258	1,032	201,709
Rural Residential	44,969	491	10,913	323	7,193	376	8,362
Commercial	886	285	321,464	220	248,248	285	322,219
Industrial	375	71	188,071	46	122,101	63	167,630
Agricultural	67,551	262	3,880	180	2,671	212	3,140
Middle Atlantic							
Dense Urban Residential	232	999	4,312,964	889	3,837,826	1,046	4,518,215
Urban Residential	1,462	1,342	917,921	1,324	906,084	1,408	963,542
Suburban Residential	2,171	859	395,603	688	317,066	726	334,150
Rural Residential	19,415	208	10,735	170	8,773	183	9,435
Commercial	311	133	426,483	118	379,014	143	459,362
Industrial	151	27	176,781	25	164,142	31	208,065
Agricultural	21,632	52	2,417	58	2,695	62	2,884
New England							
Dense Urban Residential	61	277	4,510,260	261	4,251,129	308	5,016,572
Urban Residential	669	522	780,697	336	502,186	454	678,871
Suburban Residential	1,176	360	306,002	268	227,625	315	267,880
Rural Residential	10,836	211	19,430	155	14,316	184	16,990
Commercial	196	64	328,157	48	242,726	62	315,970
Industrial	90	18	201,702	12	138,261	18	199,450
Agricultural	15,761	155	9,848	114	7,202	157	9,960
U.S. National Totals							
	1,264,975	26,592	21,022	19,333	15,283	25,095	19,838



