REPORT

BERKELEY INSTITUTE FOR YOUNG AMERICANS

Benchmarking the Zillow Transaction and Assessment Dataset

February, 2023

Max Zhang and Peter Amerkhanian

Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

ZTRAX

The Zillow Transaction and Assessment Dataset (ZTRAX) is a comprehensive, national real estate database made available free of charge to U.S. academic, nonprofit and government researchers.[1] ZTRAX is largely made up of administrative data collected from public agencies that assess real estate. Given the scale of the ZTRAX data and variance in record keeping practices across localities, researchers must exercise care in working with and creating analysis that involve the data. In order to verify the accuracy of the data that we extract from ZTRAX, we draw heavily on work previously completed by the Places Lab at Boston University that documents missingness and error within ZTRAX (Nolte et al., 2021). We also create a series of benchmarks that compare statistics obtained from ZTRAX against various public data sets that allow for examining equivalent geographic subsets of housing data.

The California State Board of Equalization

Our principal data source for benchmarking ZTRAX comes from the California State Board of Equalization (BOE), which maintains a variety of public datasets with property tax statistics disaggregated by county. By aggregating ZTRAX up to the county level, we can directly compare its data with that of the BOE. Table 1 shows the degree to which ZTRAX and the BOE data align when aggregated at the state level, with all measures showing less than a 5% error.

Table 1: State-Level Alignment between ZTRAX and BOE

Data Set	Parcel Count	Non-missing Parcels	Non-missing Total Assessed Value	Avg. Assessed Value	Total Taxes Paid	Avg. Tax Rate
ZTRAX	12,814,271	12,209,357	$5.893tn	$482,685	$68.074bn	1.155%
BOE	unknown	11,775,114	$5.656tn	$480,378	$66.490bn	1.151%
ZTRAX / BOE	N/A	1.037	1.042	1.005	1.024	1.004

Source: ZTRAX - ZTA dataset, 2017 and California Board of Equalization public data

[1] *Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.*

Parcels, Assessed Value, and Taxation

The following table details the county-by-county alignment of records between ZTRAX and the BOE. Each figure expresses a ratio, $\frac{ZTRAX}{BOE}$, meaning that the closer the figure is to 1.0, the more aligned ZTRAX and BOE data are for that statistic in that county. For example, in the table below, we can see that for Alameda County, the total number of parcels (non-missing) in ZTRAX is 98.8% of the parcel count expressed in the BOE data, while for total taxes paid, the amount in ZTRAX is 110% of the total amount of taxes paid in Alameda County listed in the BOE.

Table 2: County-Level Alignment between ZTRAX and BOE

	Ratio of ZTRAX to State Board of Equalization Data				
County	**Non-missing Parcels**	**Total Assessed Value**	**Avg. Assessed Value**	**Total Taxes Paid**	**Avg. Tax Rate***
ALAMEDA	0.988	0.998	1.01	1.1	1.118
AMADOR	1.009	1.014	1.005	0.72	0.735
BUTTE	0.982	1.007	1.026	0.771	0.783
COLUSA	1.006	1.001	0.995	0.628	0.868
CONTRA COSTA	1.01	1.026	1.016	1.154	1.123
EL DORADO	1.032	1.005	0.974	1.066	1.062
FRESNO	1.031	1.004	0.974	0.894	0.944
HUMBOLDT	0.906	1.004	1.107	0.693	0.701
IMPERIAL	1.144	1.011	0.883	0.58	0.577
INYO	0.76	0.917	1.207	0.819	0.921
KERN	0.949	1.004	1.058	0.911	0.895
KINGS	0.971	1.029	1.06	0.903	0.931
LASSEN	1.016	0.997	0.981	0.939	1.012
LOS ANGELES	1.009	0.931	0.923	1.046	1.128
MARIN	0.966	0.994	1.028	1.209	1.208
MARIPOSA	15.378	0.957	0.062	0.932	1.003
MERCED	0.997	0.976	0.98	0.954	1.013
MODOC	1.007	1.046	1.04	1.392	1.446
MONO	1.019	1.007	0.988	0.942	0.958
MONTEREY	0.966	0.986	1.021	0.736	0.749
NAPA	0.968	1.002	1.035	0.836	0.847
NEVADA	1.01	0.999	0.99	1.11	1.112
ORANGE	0.946	0.964	1.019	1.052	1.101
PLACER	1.07	0.998	0.933	1.135	1.139
RIVERSIDE	1.067	1.02	0.956	1.113	1.105
SACRAMENTO	0.973	0.991	1.019	1.009	1.005
SAN BENITO	1.016	1.005	0.989	0.973	1.018
SAN BERNARDINO	1.018	0.977	0.96	1.024	1.069
SAN DIEGO	0.986	0.997	1.012	1.042	1.037
SAN FRANCISCO	0.972	0.956	0.984	0.861	0.894
SAN JOAQUIN	0.999	1.007	1.008	1.084	1.135

County	Ratio of ZTRAX to State Board of Equalization Data				
	Non-missing Parcels	Total Assessed Value	Avg. Assessed Value	Total Taxes Paid	Avg. Tax Rate*
SAN LUIS OBISPO	0.963	0.998	1.036	0.962	1.018
SAN MATEO	0.989	0.988	0.999	1.047	1.083
SANTA BARBARA	0.938	0.996	1.061	1.003	0.996
SANTA CLARA	1.05	1.629	1.552	0.992	0.609
SANTA CRUZ	0.977	0.991	1.014	1.099	1.101
SHASTA	0.892	1.004	1.125	0.932	0.969
SIERRA	0.699	0.973	1.391	1.216	1.248
SISKIYOU	0.844	1.014	1.202	0.551	0.575
SOLANO	0.947	0.983	1.038	1.069	1.064
SONOMA	1.022	1.007	0.985	0.742	0.738
STANISLAUS	0.979	0.996	1.017	1.048	1.061
SUTTER	1.003	1.003	0.999	1.028	1.087
TEHAMA	1.026	1.029	1.003	0.937	0.948
TUOLUMNE	1.006	1.012	1.007	0.751	0.755
VENTURA	0.965	0.994	1.03	1.024	1.038
YOLO	0.979	0.982	1.003	1.156	1.191
YUBA	0.953	0.949	0.996	0.864	0.887

Source: ZTRAX - ZTA dataset, 2017 and California Board of Equalization public data

**Note: Average Tax Rate is calculated as the aggregate total property taxes paid in a county divided by the total assessed property value in the same county.*

While Table 1 shows promising figures at the state level, Table 2 makes clear that, county-to-county, there is significant variance in accuracy. While most counties are clustered around a 1.0 ratio for ZTRAX/BOE, it seems that the accuracy in our state-wide figures may be in part due to some counties offsetting others, including the fact that Santa Clara County is a severe overcount in ZTRAX, while Kern County is a severe undercount.

Transfers

We can see in Table 3 that property transaction figures in ZTRAX align fairly well with those in BOE at the aggregate. However, benchmarking measurements beyond sale volume, such as sale prices, proves to be a complicated task. While property assessment records are uniform in what they measure – assessed value – transaction records do not simply reflect sale prices. Many properties are transferred for prices less than "full consideration" as a result of intra-family transfers, gifts, sales for an agreed upon price that is lower than market value, etc. In order to isolate only those transfers that we believe are market value transactions, we use detailed documentation that the PLACES Lab produced, outlining how to isolate a subset of transfers that reflect true market value transactions, though only at "medium" or "high" confidence (Nolte et al., 2021). We follow their guide, filtering out various types of transfers unlikely to be at full market value, and, in order to balance both accuracy and sample size considerations, we arrive at a sample for which we can have "medium" confidence in it reflecting all true market value transactions. The key steps of our filtering process are as follows:

1. Filter down to data with document type codes corresponding to documents associated with market value sales as identified by the PLACES Lab, i.e. “Agreement of Sale,” “Administrator’s Deed,”, “Assumption/Deed Agreement,” etc.
2. Filter down to data with data class codes that correspond to market value sales. This includes deed transfers and deeds with concurrent mortgages.
3. Filter down to data with sales price codes that indicate market value sale, i.e. “Full Consideration,” “Price from recorded affidavit,” “Sale price computed from transfer tax,” etc.
4. Drop all sale records with a sales price less than $1,001.

Table 3: State-Level Alignment between ZTRAX and BOE Transfer Data

Data Set	Total Transfers
ZTRAX	1,281,778
BOE	1,343,543
ZTRAX / BOE	0.954

Source: ZTRAX - ZTT 2015, 2016 and California Board of Equalization public data

County Realtors’ Associations

Following our filtering of transfers down to only those we believe reflect market value transactions, we benchmark our final counts for one sample county, San Mateo, with public real estate transaction data, in this case data from The San Mateo County Association of Realtors (2022).

San Mateo County Realtors

The results of this benchmarking exercise are shown in Table 4. We find that the market value transactions that we obtained following the filtering process set forth by the PLACES Lab gets us sale counts and mean sale prices that are generally very similar to the figures published by the San Mateo County Association of Realtors. Given these results, we have further confidence in the filtering process set forth by the PLACES Lab even beyond San Mateo, as we assume that if the data are accurate according to this county’s realtors’ records, they would be accurate in other counties as well.

Table 4: Alignment between ZTRAX and San Mateo County Realtor’s Data

Date	ZTRAX Sale Count	ZTRAX Sale Mean	SMR Sale Count	SMR Sale Mean	Sale Count Ratio	Sale Mean Ratio
2017-09	368	1718539	354	1739517	0.988	1.04
2017-08	418	1595575	381	1739015	0.918	1.097
2017-07	385	1884086	358	1891341	0.996	1.075
2017-06	510	1832713	482	1782672	1.028	1.058
2017-05	520	1836049	437	1778136	1.033	1.19
2017-04	411	1770670	348	1876376	0.944	1.181
2017-03	388	1575543	328	1558630	1.011	1.183
2017-02	276	1676875	190	1610361	1.041	1.453
2017-01	196	1586783	189	1452117	1.093	1.037

Source: ZTRAX - ZTT 2017 and San Mateo County Realtor’s Public Data

Citations

California State Board of Equalization. (2022). *State and County Assessed Property Taxes: Assessed Property Values (Table 5) and Local Roll Value and Statistics* [Data set]. Retrieved from https://www.boe.ca.gov/dataportal/catalog.htm?category=Property%20Taxes.

Nolte, C., Boyle, K. J., Chaudhry, A. M., Clapp, C. M., Guignet, D., Hennighausen, H., Kushner, I., Liao, Y., Mamun, S., Pollack, A., Richardson, J., Sundquist, S., Swedberg, K., & Uhl, J. H. (2021). *Studying the Impacts of Environmental Amenities and Hazards with Nationwide Property Data: Best Data Practices for Interpretable and Reproducible Analyses*. *No. 2021-013*. https://dx.doi.org/10.2139/ssrn.3900806.

San Mateo County Association of Realtors. (2022). *Market Data* [Data set]. Retrieved from https://www.samcar.org/member-resources/market-data/.