TECHNICAL REPORT 2011-3

Value and Rent per ft2 of Non-residential and Non-vacant Parcels at Model Zone Level**

Huiling Zhang*#
Richard Arnott*

November 2011

Abstract: Value and rent per square foot of floor area of parcels whose land use type is neither residential nor vacant were calculated. Those land use types include retail, other commercial, industrial, warehousing, public, transportation and communication. Missing sale values were imputed from SCAG parcel database, while rents were calculated from sale values by multiplying corresponding rent-value ratios, whose relationships with accessibility measures, building ages, and value per ft2 can be obtained from Costar Office database.

** This report was prepared in conjunction with the University of California MRPI Project LA-Plan.
* Department of Economics, University of California, Riverside, CA 92506#
# Please address questions and comments to the primary author, Huiling Zhang, at hzhan017@ucr.edu.

Value and Rent per ft2 of Non-residential and Non-vacant Parcels at Model Zone Level

Every parcel was assigned a three digit land use code by SCAG, but these codes were reclassified into thirteen building types in LA project. And these building types can be further aggregated to three categories in terms of the methodology adopted in calculating rents and values. The three categories are: Vacant Lots, Residential Buildings, and Non-residential Non-vacant Buildings. The first two have been treated in separate technical reports, and this report is committed to the third one.

Non-vacant and non-residential category contains nine building types: Office, Retail, Other Commercial, Public, Warehousing, Industrial, Transportation and Communication, and Mixed. Among them Office Buildings were treated separately as data on both rents and values of a portion of office buildings in South California is available in Costar database. SCAG parcel database has no rent data on any land use type, therefore the rent data in Costar office buildings also contributed to the calculations of rents of other building types in this category.

The general procedure of calculating value per square foot of floor area for a particular building type at aggregated model zone level is virtually the same across all building types. The regression and imputation framework and data processing methods were explained in detail in an earlier technical report on vacant parcels. This technical report

focuses on the problems and corresponding solutions special to the non-residential non-vacant non-office building types.

The biggest challenge is to compute rent for non-office building types given that SCAG parcel database provides no data on rents. Among all building types, office buildings are the only type that has both rent and value data available for some parcels. Therefore the relationship between rent and possible explanatory variables was investigated first in office buildings in Costar database, and then copied to non-office buildings.

The second biggest challenge comes from the imperfection of SCAG parcel database. Last sale price is the most important data, but it suffers severe missing value problem. Among the six counties of interest, only Riverside[1] and Orange have reliable data on last sale value. The best substitute of last sale price in SCAG database is an assessed value generated by county assessors. Assessed values for non-vacant parcels are calculated using price of last sale, income, and the sales price of comparable properties, and might be systematically biased from market value. Therefore last sale price was used whenever it was available, and assessed value multiplied by an adjustment factor was used as pseudo sale price whenever it was not.

Another problem is the high percentage of missing data in floor area[2]. Sale prices must be scaled by floor area before they can be comparable. Missing data on floor area reduces the sample size of regressions and discards lots of otherwise useful information.

[1] From hereby Riverside means Riverside County, and similarly for the other five counties.
[2] Refer to the floor area missing data percentage table provided by Yizhen.

Furthermore, non-random missing of floor area implies the reduced sample may not be representative of the population, and thus causing the revealed relationships misleading and imputed values biased. UCSB team did some ground truth tracking of parcels that have floor area listed as zeros or blanks in several building types and counties, based on which Yizhen Gu estimated floor area for parcels that don't have one. The resulting value per square foot of parcels would be less biased if the estimated floor area is unbiased. In this report results using original floor area and result using estimated floor area are both listed.

The idea and therefore procedures of calculations are virtually the same for all non-residential and non-vacant building types, so retail type was taken as an example hereafter.

Section I describes the data and explains three basic techniques intensively used in most imputations. Section II outlines the logic flows of imputing rents and values, both for counties with data on last sale price and counties that are without. Section III explains the regression framework and presents results for estimating the relationship between rent-value ratio and some shared explanatory variables by two databases. Section IV presents computed average value and rent per square foot per model zone, and section V concludes.

1. Data and Basic Techniques

1.1 Data

SCAG parcel database is the primary data source in this sub-project. It provides data on a parcel's geospatial features (x-y coordinates, lot size, floor area, city, and model zone code), building features (current land use code, planned land use code, year built), and measures of value (assessed total value of land and improvement, last sale price and date). Those variables come from different data files, and some data preparation including matching and spatial join are necessary before analysis[3].

Costar database has data on both rents and values for a sample of office buildings in Los Angeles Metropolitan Area. However, the sample is not a representative one. It includes only Class A and Class B, which are of higher quality than average. Meanwhile, SCAG parcel database does not provide information on building class.[4] It was for these reasons that Costar values (rents) per square foot was not used directly to impute values (rents) per square foot to office buildings in the SCAG parcel database. In other words, Costar database is useful in investigating the relationship between rent-value ratio and other explanatory variables, not in investigating the relationship between values (rents) per square foot and explanatory variables.

Variables in Costar database that were used in this sub-project include a building's geospatial features (x-y coordinates, rentable floor area, county), building features (year built), and measures of value (**annual rent at 2008**, last sale price and date).

[3] Regular problems occurred in data processing and corresponding solutions can be found in the technical report on vacant land evaluation.

[4] in principle it does provide some information but the land use classification is inconsistent, with a building perhaps being coded according to its specific class and perhaps due to its generic class

The fact that the two databases share lots of variables allows the revealed relationship between rent and other variables in one dataset to be applied to the other.

1.2 Basic techniques

1.2.1 Adjust sales price to year 2000

Comparing sales prices in different sales years directly is not reasonable. Inflation makes a dollar less valuable, and a positive demand shock increases the market prices of all properties. This subproject tried to give all properties a value in a common year 2000 to exclude year fixed effects. Therefore all sales prices have to be adjusted to the default year whenever they are to be used.

Adjusting sale prices to year 2000 always contains two steps:

- Step 1: Run a regression on natural logarithm of sales price per square foot of floor area against sales year dummies and other explanatory variables. Set year 2000 as the default.

  ln(vsq) = alpha*(control variables) + beta* i.saleyear
- Step 2: Sales value in year 2000 is then equal to last sale price divided by the exponential of the regression coefficient on corresponding last sale year dummy.

  vsq2000 = vsq / exp(b_saleyr)

1.2.2 Impute missing sales prices

For counties that have last sale prices data available for some parcels, sales price in year 2000 can be estimated after running a regression on available last sale prices. The regression framework is the same as the one used in adjusting sales prices to year 2000[5].

- Step 1: Run a regression on the natural logarithm of value per square foot of floor area of all parcels that have data available on last sale price, last sale date. The explanatory variables are four kinds of accessibility measures, building age, last sale year dummies, land use code dummies, and city dummies.
- Step 2: Use the set of regression coefficients to predict value in default year 2000 for parcels that have missing data on either last sale value or last sale year.

1.2.3 Discount rents from year 2008 to year 2000

Similarly for rents in Costar database, they are annual rents for year 2008. Bill-Wheaton's table provides a TW Rent Index which can be used to discount rents in year 2008 to year 2000. San Bernardino and Imperial are assumed to have the same index as Riverside, and Ventura has the same index as Orange.

TW Rent Index ($/SF)	Los Angeles	Orange County	Riverside
2000.1	$21.38	$25.76	$16.15
2008.1	$27.94	$29.54	$22.12

[5] A larger R-squared usually indicates a better curve fitting and a better prediction, so all variables in the dataset that might be correlated with the dependent variable were included as explanatory variables in the regression framework for imputations.

But if the sole objective is to discount sales prices to a default year, then R-squared is not important as long as there is no omitted variable bias. So the regression framework requires only sale year dummies and variables that are correlated with both sale year dummies and dependent variable.

But in this sub-project, the number of variables is not that large, therefore one regression framework is used in imputations and in discounting sales prices.

Table 1: Bill-Wheaton's table of TW Rent Index

2. Procedure Outline

For Riverside and Orange, the procedure to calculate value and rent per square foot of floor area can be summarized in three steps:

- Step 1: In SCAG retail parcel database, impute a sale value for parcels that don't have one, and adjust all values to year 2000.
- Step 2: In Costar office database, run a regression on rent-value ratio of office buildings against a bunch of explanatory variables, and keep the regression coefficients.
- Step 3: In SCAG parcel database, predict rents for all retail parcels using the obtained regression coefficients in part two.

For Los Angeles, San Bernardino, Ventura, and Imperial, last sale price data is not available. But the ratio of average last sale price to average assessed value for parcels in a specific building type in Riverside can be calculated, and thus multiplying the assessed value by this ratio gives an estimate of sales price for any parcel in San Bernardino and Imperial. The procedure has three steps, and only SCAG parcel database was used:

- Step 1: Adjust last sale price to year 2000 for retail parcels in Riverside.
- Step 2: For each building type, calculate average sales price in year 2000 and average assessed value in year 2007. Then obtain ratios.
- Step 3: For each building type in San Bernardino and Imperial, multiply assessed values by the corresponding ratio to get estimated sales prices in year 2000.

Similarly, the ratios can be calculated for Orange County, and then applied to Los Angeles and Ventura County.

3. Rent-value ratio

3.1 Framework

Costar database has rents and value data available, but SCAG parcel database has no data on rents for any parcel of any building type. A regression on rent-to-value ratio against variables that are shared by both databases would provide an estimation of rents for non-office parcels in SCAG parcel database. And the estimation can be a good one if the revealed relationship for office parcels in Costar database well describes the relationship for non-office parcels. In this sub-project, this is assumed:

Assumption 1: The relationship between rent-to-value ratio and accessibility variables, value per unit of floor area, and building age is the same for Costar office buildings and non-residential non-office parcels in SCAG database.

The relationship expressed in a regression equation is:

$$ln(rv) = \beta 1 * ln(vsq_{2000}) + \beta 2 * fsub + \beta 3 * cbd + \beta 4 * fwy + \beta 5 * ocean + \beta 6 * age + constant + \varepsilon$$

Where *rv* is rent-to-value ratio in year 2000, and vsq_{2000} means value per square foot of rentable floor area adjusted to year 2000. Fsub, cbd, fwy, and ocean are the four accessibility measures, which mean the distance to the nearest subcenter, cbd, neareast freeway, nearest coastline respectively.

Rent-value ratio is believed to be negatively correlated with a building's age, as a new building depreciates more quickly than an old building. Similarly a higher quality building depreciates slower than a low quality one; therefore rent-value ratio is believed to be negatively correlated with quality of the building. Here value per square foot of rentable floor area acts as a proxy for quality, and it is in natural logarithm form. If a dollar increase in per unit price of an ordinary office building in rural Riverside leads to some x percentage decrease in its rent-value ratio, then one would not expect a dollar increase in per unit price of a luxurious office building in downtown Orange would make its rent-value ratio decrease by the same percentage. More reasonably, one would expect the latter to be higher than the former. That's why a log-log model in terms of value per square foot of floor area is preferred to a log model.

3.2 Regression coefficients

Costar database is of high quality, but it is not perfect. The database unit is one office building. Sometimes several office buildings in a business park were sold in a package, and all those buildings were recorded in the database to have a last sale price equal to the price for the whole package. If there are five buildings in the package, then value per square foot calculated for those five building is biased upwards 4 times. This is not a negligible problem given the usual large scale of a package sale.

It is difficult to calculate the sale price for a building involved in a package sale if the building's share in the package price is unknown. However, Costar database provides some variables that help identify whether a building is involved in a package sale.

Therefore only buildings whose last sale was a stand-alone one were included in the regression sample. And a stand-alone sale was "defined" as having a last sale price and date that is not shared by any other building.

Table 2 below lists the set of regression coefficients for stand-alone sample (column 2) and for all buildings (column 3). Both have high R-squared, and sign of coefficients on explanatory variables are as expected, but the stand-alone sample coefficients were chosen to be applied to non-office parcels' rent calculations.

Also notice that Stand-alone sales constitutes more than 97% of all sales records, so the sample size reduction is still decent after excluding sold-in-package buildings.

Table 2: Comparing the regression coefficients of one regression framework under two samples. The regression framework is the one for estimating rent-value ratio, and one sample includes all office buildings that have a sales price, while the other sample includes stand-alone sales data only.

	(1): All data lg(rv)	(2): stand-alone sales data only Lg(rv)
$lg(vsq_{2000})$	-0.944***	-0.940***
	(0.000)	(0.000)
fsub	-0.00198	-0.00222
	(0.277)	(0.248)
cbd	-0.00202***	-0.00187***
	(0.000)	(0.000)
ocean	-0.00854***	-0.00858***
	(0.000)	(0.000)
fwy	-0.0268*	-0.0222
	(0.023)	(0.068)

age	-0.00319***	-0.00255***
	(0.000)	(0.000)
_cons	3.018***	2.964***
	(0.000)	(0.000)
N	1776	1671
R^2	0.957	0.966

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4. Imputed value and rent per square foot of floor area

Yizhen Gu estimated floor area for parcels whose floor area is listed as missing or blank. The estimated floor area together with originally available floor area data in SCAG parcel database is called updated floor area hereafter. Value per square foot calculated using updated floor area data is around 10 to 100 times larger than that calculated from original floor area data.

Several facts may shed some light on this difference: First, A parcel with a higher sales price is more likely to have missing floor area. Second, the floor area ratio (FAR) for parcels with a floor area is higher for parcels without one (whose FAR is then calculated from updated floor area).

One possible explanation is that parcels with missing floor area are indeed different from parcels that have floor area data available. They may have higher quality floor area, so their true value per square foot is higher. Another explanation is that parcels that sold in a package are more likely to have missing floor area, and their value per square foot is biased upward severely even if the estimated floor area is true. But it is hard to justify this case. Another explanation is simply that the estimations are systematically biased upwards. A ground truth tracking would cast light on this problem, but for now there are

no dominant explanations and therefore no plausible way to improve the floor area estimations.

Model Zone (ID)	Model Zone (Name)	vsq_mz ($/sqft)	rsq_mz ($/sqft)
1	Downtown Los Angeles	83.68	14.07
2	Westside	108.01	18.08
3	Glendale	106.16	17.92
4	East Los Angeles	76.04	13.83
5	Maywood	68.03	12.80
6	Florence	64.40	11.72
7	Baldwin Hills	87.74	14.11
8	Beverly Hills	175.34	27.76
9	El Segundo	92.06	13.56
10	Santa Monica	149.09	22.10
11	Marina del Rey	174.09	26.95
12	Westwod	178.89	27.49
13	East Santa Monica Mtns	123.48	19.78
14	Reseda - van Nuys	94.79	15.53
15	East van Nuys	81.91	14.08
16	Burbank	100.97	16.61
17	Pasadena	100.22	17.61
18	East Pasadena	93.28	15.73
19	Rosemead	97.32	17.15
20	Pico Rivera	83.74	11.42
21	South Gate	78.69	13.41
22	West Compton	65.16	10.70
23	Torrance	104.49	15.33
24	Palos Verdes	96.13	17.03
25	Carson	84.06	12.34
26	Long Beach	95.76	15.38
27	Signal Hill	89.48	13.78
28	Compton	59.63	9.54
29	Hawaiian Gardens	100.89	14.76
30	Cerritos	65.47	9.18
31	Norwalk	70.15	11.06
32	Industry	71.58	11.08
33	Diamond Bar	69.46	10.92
34	North El Monte	75.13	12.99
35	West Covina	66.43	11.60

36	Glendora	64.76	11.71
37	La Verne - Azusa	61.46	10.21
38	Altadena	88.62	16.17
39	North Hills - Sylmar	81.88	14.34
40	Chatsworth	79.27	13.61
41	Calabasas	113.54	16.54
42	Malibu - Point Dume	330.58	51.95
43	Agoura Hills	87.14	13.70
44	Lake Los Angeles	65.60	25.95
45	Lancaster - Palmdale	63.65	17.57
46	Santa Clarita	97.63	19.06
47	Ventura North County	191.31	34.12
48	Thousand Oaks	216.62	41.52
49	Oxnard - Camarillo	162.57	26.03
50	Seal Beach - Los Alamitos	179.72	27.40
51	Cypress	372.67	59.00
52	South Buena Park	377.29	57.20
53	Buena Park - La Habra	96.18	15.31
54	Placentia	48.99	7.80
55	Yorba Linda	99.32	16.82
56	Huntington Beach	205.83	31.71
57	Garden Grove	128.58	20.65
58	Anaheim	129.31	21.18
59	North Tustin	121.02	19.09
60	Costa Mesa	121.22	18.63
61	Santa Ana	110.13	18.07
62	Tustin	98.54	13.87
63	Newport Coast	207.03	32.36
64	Irvine	91.84	13.16
65	East Orange County	138.37	25.47
66	San Juan Capistrano	168.48	27.75
67	Montclair - Chino	65.48	10.29
68	Ontario	26.12	4.70
69	Rancho Cucamonga	40.48	6.55
70	Upland	63.81	10.44
71	Fontana	37.77	6.53
72	Colton	29.13	5.29
73	San Bernardino	19.07	3.66
74	Redlands - Highland	28.26	5.01
75	Crestline	35.15	9.88
76	Victorville	33.67	9.80
77	Lucerne Valley	0.71	0.28
78	San Bernardino	34.20	22.24

	Mountains		
79	Northwest Mojave	19.75	176.96
80	Northeast Mojave	17.85	8.16
81	Corona	31.52	5.70
82	East Riverside	38.49	7.74
83	Indio	27.30	5.86
84	East Mojave	20.97	6.14
85	Lake Elsinore	26.52	5.01
86	Riverside	41.01	8.84
87	Moreno Valley	28.23	6.58
88	Perris	29.46	6.15
89	Banning	24.31	5.77
90	Hemet	27.97	6.08
91	Temecula	33.96	5.58
92	Palm Springs	28.22	6.64
93	La Quinta	30.68	6.11
94	Cathedral City	24.66	5.10
95	Palm Desert	25.30	4.97
96	Imperial Valley[6]	282.92	57.58
97	El Centro	180.10	53.79

Table 3: Value and rent per square foot of floor area for **retail** parcels at aggregated model zone level in year 2000. Note that the floor area used were all original data in SCAG parcel database.

5. Conclusions

The regression framework for estimating rent-value ratio yields rather good statistical results, and the signs of the explanatory variables are consistent with expectations. However the framework seems to convey the idea that rents are determined by values, which is not necessarily the case. Either the regression framework needs to be adjusted or the interpretation of the regression framework needs to be improved.

[6] The rents and values per square foot of floor area for Imperial County, including model zone 95 and 96, are severely overestimated, as an underestimation of floor area in Imperial County. Note that floor area data in Imperial County is completely unavailable in SCAG database, and it was estimated by UCSB team. Once an updated floor area is received, the values and rents for the two model zones in this table will be adjusted accordingly.