On the market: The longitudinal relationship between housing market valuation and job market postings in California

1. Introduction

1.1 Question

What is the temporal relationship between job postings and housing prices in California, -- specifically we will investigate the strength and effect of the lags between these two factors, and also looking for its dependence on job categorization and education level, which we think are two most important demographic data that will influence the characteristics of the job market.

1.2 Significance of Our Question

California is the largest economy in the United States, and is a major provider of jobs in the technology and healthcare sector (Table 7). However with rapidly increasing housing prices in California nowadays, even people who work in highly paid industries feel great pressure living there. Will this eventually stop the growth of California's economy considering companies there will have to pay much higher salaries than other parts of US to cover employee's living expense and thus there would be less job openings from these companies? This question will be satisfyingly answered with the analysis of data of housing prices and job openings within a certain time period.

Job postings data are easily available online and could be regularly scraped from the web using ordinary methods. These postings could be used as an important source of data for predicting market trends, including real estate values (including predicting the values of unlisted properties). Because the temporal ordering of causal influence is unclear, it is important to assess the time specificity of the relationship between job growth and real estate values.

1.3 Data Sets Utilized

Four of the provided datasets were used for the analysis:

1. jobs
2. real_estate
3. education
4. geographic

Because of hardware constraints, special measures were taken to reduce the size of the analysis dataset. We'll specifically focus on the data from the most recent 5 years which can reflect the situation nowadays most accurately.

2. Non-Technical Executive Summary

Our main finding in the cross-sectional analysis was that for every $100,000 increase in home value over the period of 2011-2015 in California, the number of job postings increased by about 2 per thousand persons (Table 1). This effect was robust to adjustment for variables that could potentially confound the relationship (Table 2).

Additionally, the effect was very significant, suggesting that home values could be a consistent predictor of job posting activity. Interestingly, education was not strongly associated with job postings, suggesting that the expectation for migration of new hires might heavily influence how positions are filled (this could also reflect high competition for jobs.)

3. Technical Executive Summary

3.1 Data Selection and Manipulation

Special measures were taken to reduce the size of the dataset to make the analysis more feasible for a highly limited time period. The Job dataset was first filtered to only jobs in California, and variables of interest. This dataset was then merged with similarly subsetted data sets on home value and education.

The education dataset was summarized as one variable - years of education - by calculating a weighted average of all of the years of education variable. This was done by assigning discrete values to each category, e.g. 1st grade is one year while a Bachelor's Degree is 14 years. This greatly simplifies our analysis of education factor while at the same time captures the most intrinsic property of education--length.

The home values dataset was transposed so that each row represented a time-place combination (replacing the original wide format of the dataset). This yielded an analysis dataset where each observation represented a unique time-place-job category combination. There were 128,607 total jobs posted for California across 174 cities. The study period of 2011-2015 was chosen based on the date range of data from the education dataset, so there were 60 time periods representing 1 month each.

Finally, the **geographic** data set was chosen for the creation of heat maps.

3.2 Methodology

Linear models were used as our starting point, as in most social-economic problem, to assess the cross-sectional association between home value and number of jobs. Specifically, a Blundell-Bond longitudinal model was used to assess the strength of lags between variables at different time points. The relationship was also assessed using graphical methods (Table 3).

4. Appendix

Table 1. Crude linear model of number of job postings per capita as a function of home value.

	Dependent variable:
	jobspercapit
hundredthousand	0.002^{***}
	(0.00004)
Constant	0.002^{***}
	(0.0002)
Observations	92,102
Log Likelihood	223,479.900
Akaike Inf. Crit.	-446,955.800
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2. Model of home value adjusting for education (in years) of the city.

	Dependent variable:
	jobspercapit
hundredthousand	0.002^{***}
	(0.0001)
aved	0.0001
	(0.0001)
Constant	0.0004
	(0.001)
Observations	92,102
Log Likelihood	223,481.000
Akaike Inf. Crit.	-446,956.100
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3. results from Blundell-Bond longitudinal model (jobs as outcome)

```
Number of instruments =    1.7e+03              Wald chi2(14)     =   14788.66
                                                Prob > chi2       =     0.0000
One-step results
```

stdnjobs	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stdnjobs						
L1.	.3277366	.0109788	29.85	0.000	.3062184	.3492547
L2.	.2630164	.0117927	22.30	0.000	.2399031	.2861296
L3.	-.0054594	.0145114	-0.38	0.707	-.0339012	.0229825
L4.	.0626614	.0154252	4.06	0.000	.0324285	.0928942
L5.	.0271586	.0152408	1.78	0.075	-.0027129	.05703
L6.	-.0377441	.0150774	-2.50	0.012	-.0672953	-.0081929
L7.	.0284109	.0155298	1.83	0.067	-.002027	.0588489
L8.	.0580048	.0152738	3.80	0.000	.0280687	.0879408
L9.	-.0236378	.0156853	-1.51	0.132	-.0543803	.0071048
L10.	.0811042	.0157507	5.15	0.000	.0502333	.1119751
L11.	.1494017	.0151963	9.83	0.000	.1196175	.1791859
L12.	.0251746	.0143754	1.75	0.080	-.0030006	.0533499
stdhome	.0452091	.0128058	3.53	0.000	.0201101	.0703081
aved	.0234494	.0175912	1.33	0.183	-.0110287	.0579276
_cons	-.2832405	.2312384	-1.22	0.221	-.7364595	.1699785

```
Instruments for differenced equation
        GMM-type: L(2/.).stdnjobs
        Standard: D.stdhome
Instruments for level equation
        GMM-type: LD.stdnjobs
        Standard: _cons
```

Table 4. results from Blundel-Bond longitudinal model (home value as outcome)

```
Number of instruments =   1.7e+03                Wald chi2(14)      =   6.22e+07
                                                 Prob > chi2        =     0.0000
One-step results
```

stdhome	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stdhome						
L1.	2.58461	.0134416	192.28	0.000	2.558265	2.610955
L2.	-2.990139	.0373831	-79.99	0.000	-3.063408	-2.916869
L3.	2.36676	.0545093	43.42	0.000	2.259924	2.473596
L4.	-1.306335	.0624406	-20.92	0.000	-1.428716	-1.183953
L5.	.1375123	.0637907	2.16	0.031	.0124848	.2625398
L6.	.726359	.0627944	11.57	0.000	.6032842	.8494338
L7.	-1.093506	.0636132	-17.19	0.000	-1.218186	-.9688269
L8.	.9791137	.0654367	14.96	0.000	.8508601	1.107367
L9.	-.572128	.0638587	-8.96	0.000	-.6972887	-.4469673
L10.	.2053921	.0548336	3.75	0.000	.0979202	.312864
L11.	-.0332714	.0365656	-0.91	0.363	-.1049387	.0383959
L12.	-.0071945	.0128079	-0.56	0.574	-.0322975	.0179084
stdnjobs	.0003726	.000299	1.25	0.213	-.0002133	.0009586
aved	.0068919	.000413	16.69	0.000	.0060823	.0077014
_cons	-.0870539	.0053989	-16.12	0.000	-.0976356	-.0764723

```
Instruments for differenced equation
        GMM-type: L(2/.).stdhome
        Standard: D.stdnjobs
Instruments for level equation
        GMM-type: LD.stdhome
        Standard: _cons
```

Table 5. Distribution of home values in California in 2011-2015

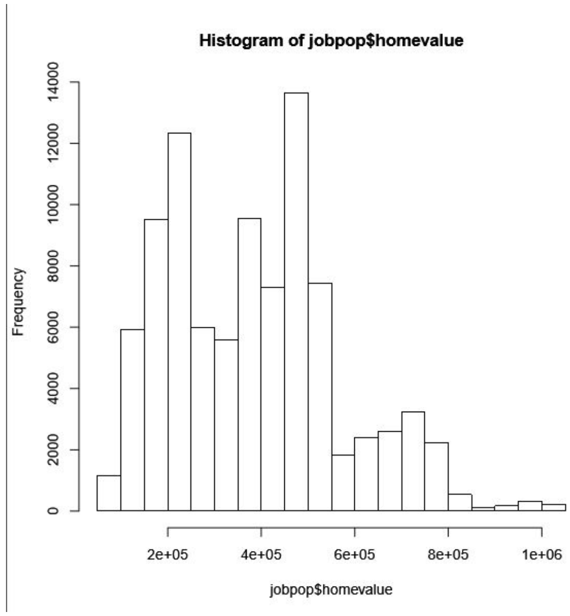
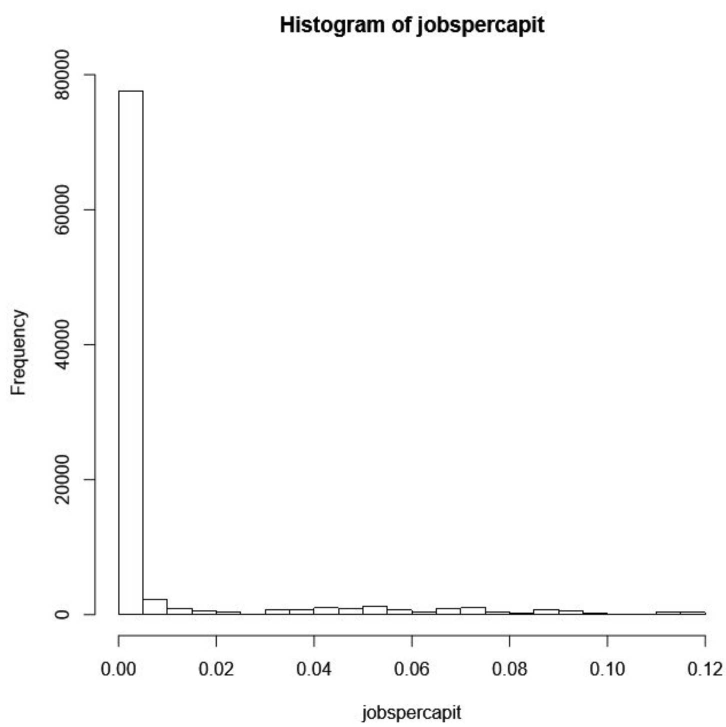


Table 6. Distribution of job per capita in California in 2011-2015



<u>Table 7.</u> Jobs postings in California in 2011-2015

	Job Category	Frequency
1	Missing	9376
2	Account Management	578
3	Accounting & Finance	2615
4	Admin & Clerical	591
5	Administrative Assistant	279
6	Advertising	83
7	Aerospace, Aviation & Defense	75
8	Agriculture, Forestry & Fishing	10
9	Architecture	10
10	Arts, Media & Publishing	437
11	Auditing	56
12	Automotive	465
13	Bank Teller	100
14	Banking & Financial Services	603
15	Banquet, Catering & Events	69
16	Biological Sciences	168
17	Bookkeeping	72
18	Business Development	1113
19	Channel Sales	73
20	Child Care	20
21	CNAs, Aides, MAs, Home Health	308
22	Computers & Hardware	47

23	Concierge & Guest Services	193
24	Construction & Skilled Trade	735
25	Consultants & Freelance Opportunities	268
26	Credit	18
27	Customer Service	1773
28	Data Entry	47
29	Database Administrator	4
30	Direct Sales	196
31	Education & Training	325
32	Engineering & Architecture	380
33	Engineers	1
34	Executive Assistant	59
35	Executive Management	35
36	Finance Management	52
37	Financial Services	145
38	General Management & Business	1897
39	Government	40
40	Green	358
41	Health & Medical	6244
42	Healthcare Management & Finance	300
43	Healthcare Office & Finance	24
44	Healthcare Support Services	4
45	Healthcare Technologists & Technicians	46
46	Hospitality & Travel	286

47	Hotel Housekeeping & Maintenance	8
48	Hotel Management	1
49	HR Benefits & Compensation	84
50	HR Management	100
51	Human Resources	443
52	Inside Sales	283
53	Insurance	116
54	Intern / New Graduate	513
55	Internet	32
56	Inventory	134
57	IT Operations	162
58	Job Fairs	126
59	Lab Technician	133
60	Law Enforcement & Security	1318
61	Legal	45
62	Library	9
63	Life, Physical, and Social Science	14
64	Logistics	246
65	Maintenance & Repair	1847
66	Management Consulting	71
67	Manufacturing & Operations	1059
68	Marketing	896
69	Medical & Dental Practitioners	13
70	Medical Records & Health IT	74
71	Mortgage & Loan	54

72	Network Administrator	216
73	Nonprofit & Volunteer	5
74	Nursing	5884
75	Office Manager	795
76	Oil, Gas & Utilities	216
77	Operations	442
78	Other Healthcare	257
79	Pharmacy	21
80	Physician	8
81	Plant Management	144
82	Product Marketing	202
83	Public Relations	25
84	Publishing	186
85	Purchasing	192
86	Radiology & Imaging	286
87	Real Estate	3
88	Receptionist	161
89	Recruiting	146
90	Research	33
91	Restaurant & Food Service	6455
92	Retail	14888
93	Sales	549
94	Sales & Business Development	6414
95	Sales Engineers	72
96	Sales Rep	1278

97	Salon/Spa/Fitness	243
98	Science, Pharmaceutical & Biotech	410
99	Shipping/Receiving	135
100	Social Services & Mental Health	218
101	Software Architecture	158
102	Software Development	1017
103	Software, Gaming & Web Developers	9
104	Supply Chain & Logistics	449
105	System Admininistrator	118
106	Teaching	51
107	Tech Management	1596
108	Tech Quality Assurance	2055
109	Technical Design	1267
110	Technical Support	400
111	Technology	1784
112	Telecommunications	28
113	Therapy & Rehab	559
114	Trading	23
115	Training & Instructor	680
116	Transportation	2748
117	Travel & Tourism	6
118	TV, Film & Video	6
119	UI Design	34
120	Veterinary & Animal Health	3
121	Warehouse Management	29

122	Warehousing	406
123	Web Design	73
124	Web Development	261
125	Writing & Editing	99
		92102

Figure 1. scatterplot of home value (x-axis) versus number of jobs (y-axis)

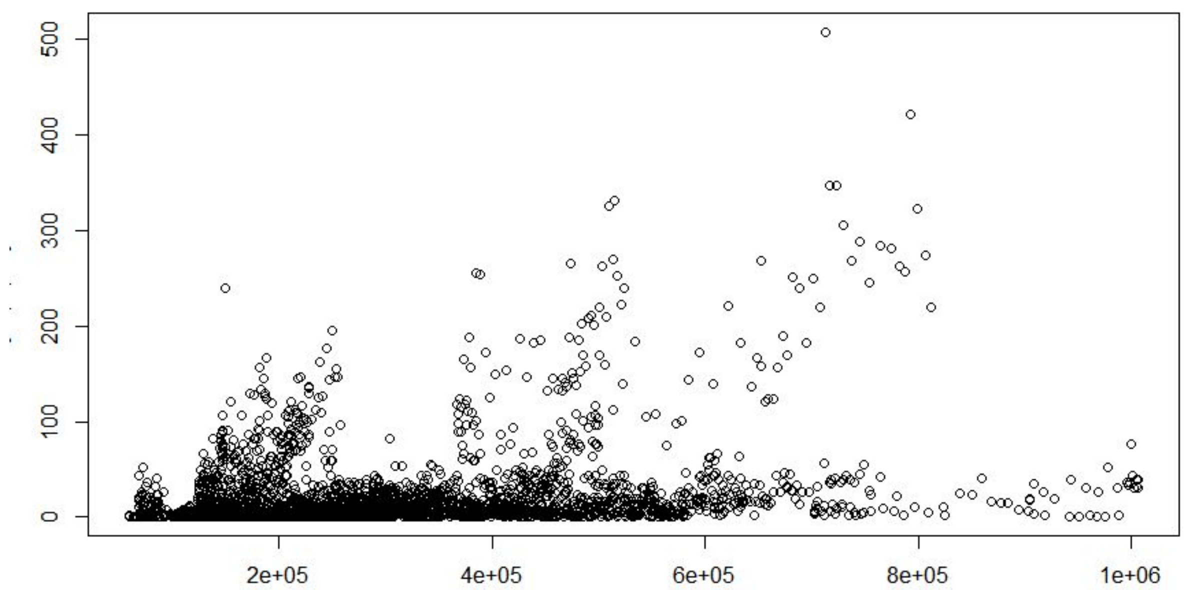
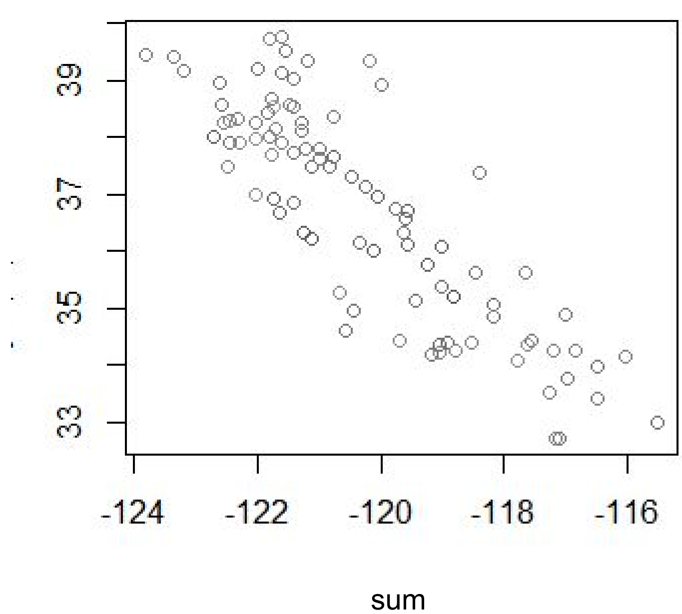


Figure 2. Spatial distribution of jobs in California.



```
import excel "C:\Users\anpas\Documents\Datathon\jobpop.xlsx", sheet("jobpop") firstrow
sum
hist stdhome
hist stdnjobs
xtset city dates
clear
import excel "C:\Users\anpas\Documents\Datathon\jobpop.xlsx", sheet("jobpop") firstrow
xtset city1 dates
gen city12=city1+100
xtset city12 dates
clear
import excel "C:\Users\anpas\Documents\Datathon\jobpop.xlsx", sheet("jobpop") firstrow
xtset city1 dates
xtdpdsys stdnjobs stdhome aved lags(12)
xtdpdsys stdnjobs stdhome aved, lags(12)
xtdpdsys stdnjobs stdhome aved, lags(58)
xtdpdsys stdhome stdnjobs aved, lags(12)
```

```
Random-effects GLS regression                   Number of obs     =      6,672
Group variable: city1                           Number of groups  =        218

R-sq:                                           Obs per group:
     within  = 0.3396                                         min =          1
     between = 0.0409                                         avg =       30.6
     overall = 0.0559                                         max =         54

                                                Wald chi2(7)      =    3323.37
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     stdhome |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    stdnjobs |   .1543129   .0147242    10.48   0.000      .125454    .1831719
       lagnj |   .1134393   .0158489     7.16   0.000      .082376    .1445026
      lagnj2 |   .0840182   .0164294     5.11   0.000     .0518172    .1162191
      lagnj3 |   .0862672   .0163889     5.26   0.000     .0541455     .118389
      lagnj4 |   .0677852   .0170486     3.98   0.000     .0343706    .1011998
      lagnj5 |   .0954284   .0165187     5.78   0.000     .0630523    .1278045
      lagnj6 |    .147972   .0158858     9.31   0.000     .1168364    .1791077
       _cons |   .2599669   .0611714     4.25   0.000     .1400732    .3798607
-------------+----------------------------------------------------------------
     sigma_u |  .88784904
     sigma_e |  .23109708
         rho |  .93654863   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```