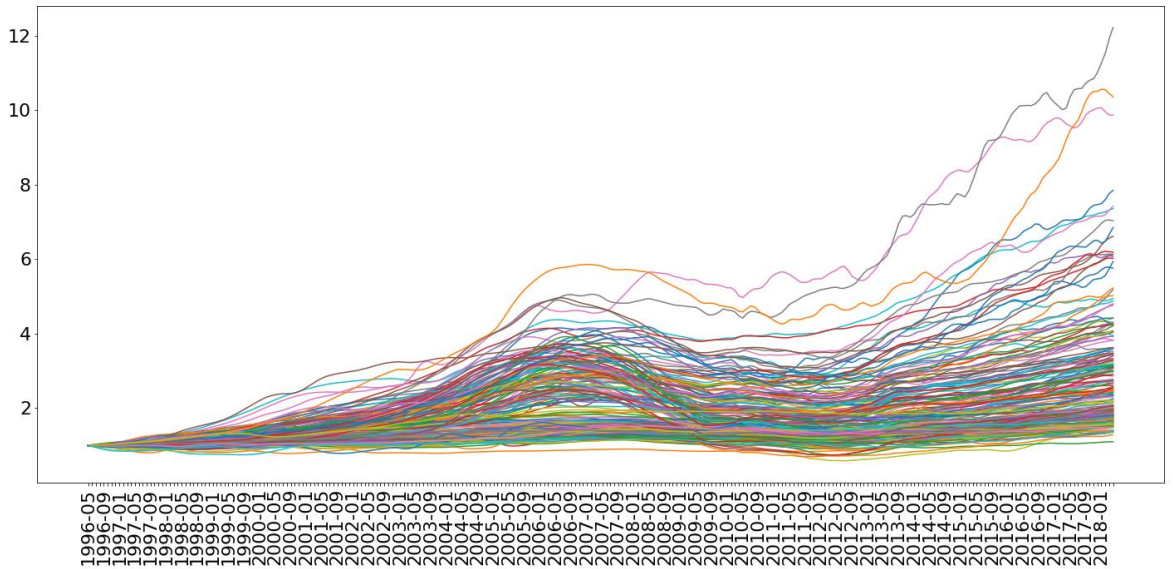
Zillow Housing Data Analysis

Akshay Ghalsasi

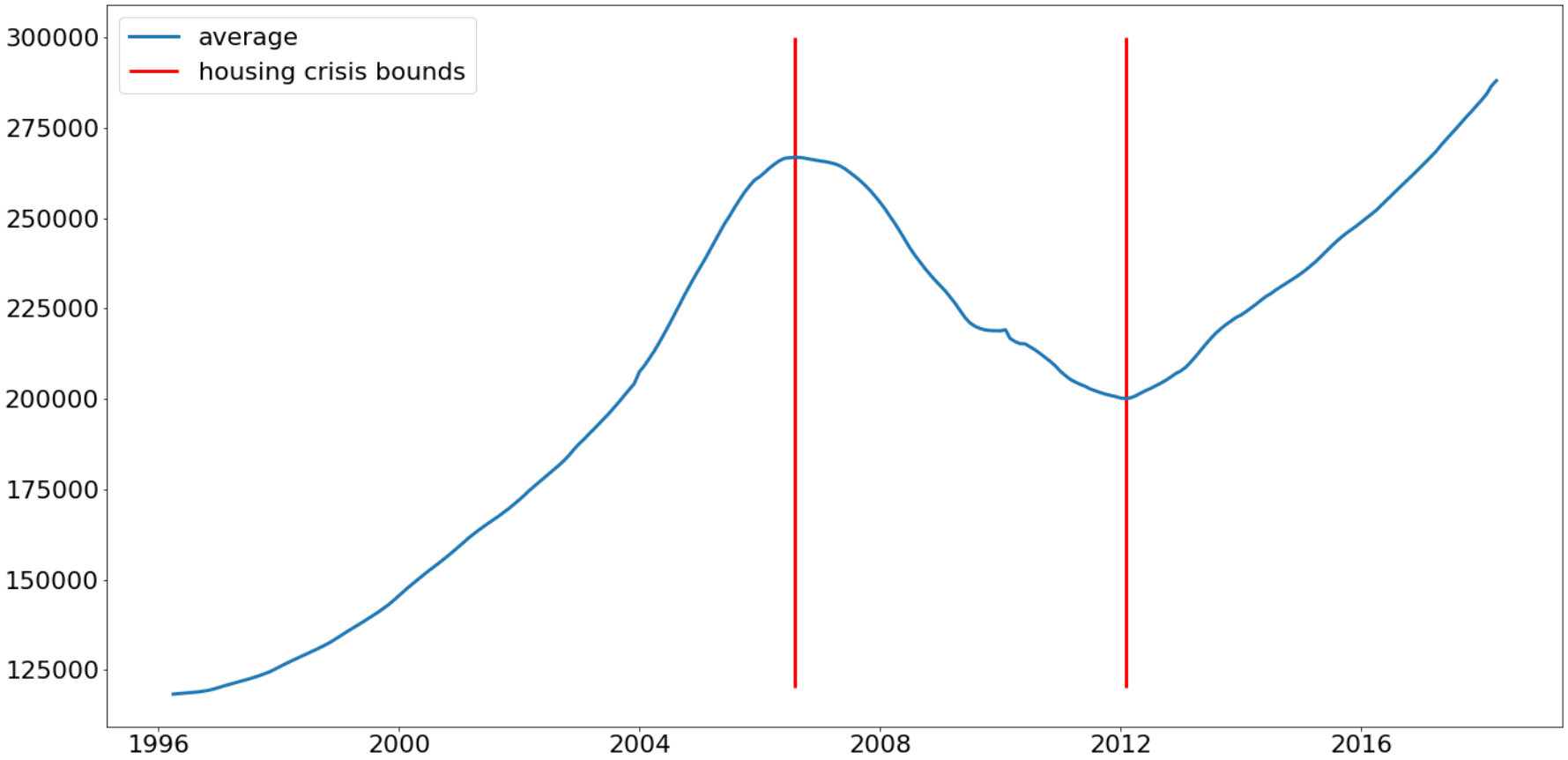
Module 4 Project

Data

RegionName	City	State	Metro	CountyName	SizeRank	1996-04	1996-05	1996-06	1996-07	...
60657	Chicago	IL	Chicago	Cook	1	334200.0	335400.0	336500.0	337600.0	...
75070	McKinney	TX	Dallas-Fort Worth	Collin	2	235700.0	236900.0	236700.0	235400.0	...
77494	Katy	TX	Houston	Harris	3	210400.0	212200.0	212200.0	210700.0	...
60614	Chicago	IL	Chicago	Cook	4	498100.0	500900.0	503100.0	504600.0	...
79936	El Paso	TX	El Paso	El Paso	5	77300.0	77300.0	77300.0	77300.0	...
77084	Houston	TX	Houston	Harris	6	95000.0	95200.0	95400.0	95700.0	...
10467	New York	NY	New York	Bronx	7	152900.0	152700.0	152600.0	152400.0	...
60640	Chicago	IL	Chicago	Cook	8	216500.0	216700.0	216900.0	217000.0	...
77449	Katy	TX	Houston	Harris	9	95400.0	95600.0	95800.0	96100.0	...
94109	San Francisco	CA	San Francisco	San Francisco	10	766000.0	771100.0	776500.0	781900.0	...



General trends

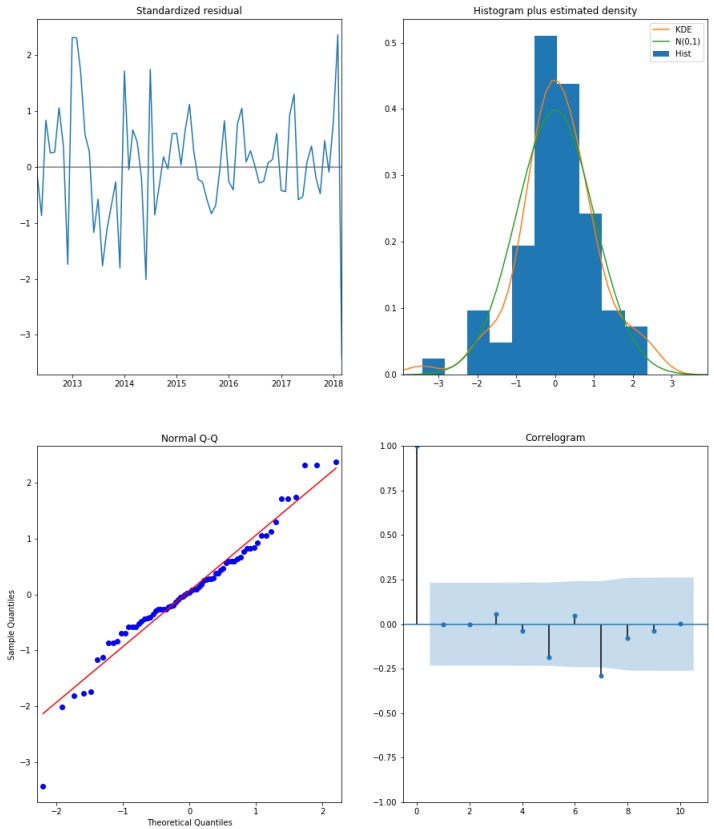


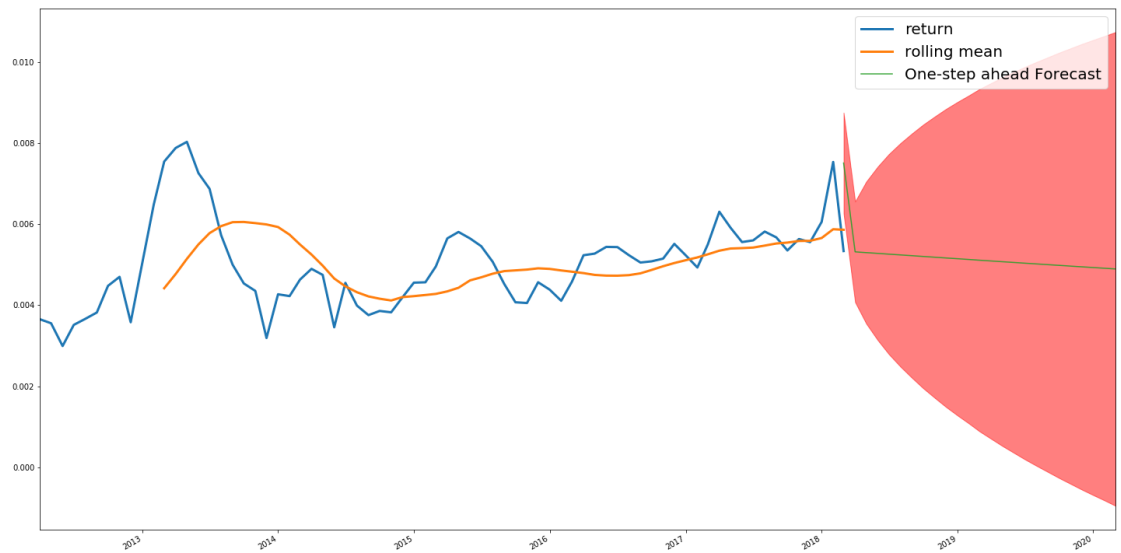
Naïve predictions

- Calculate monthly “ROI” after 2012 for 1000 zipcodes

RegionName	City	State	roi_monthly
85008	Phoenix	AZ	0.015298
89115	Las Vegas	NV	0.014483
89110	Las Vegas	NV	0.014195
33033	Homestead	FL	0.014098
11216	New York	NY	0.014027

Risk free return





ARIMA Analysis predictions

(p,d,q) = ([0,3],[0,1],[0,3]) PDQ = ([0,2],[0,1],[0,2])

Mean

zip	City	mean_ret	low_ret	high_ret	order	seasonal_order	sharpe_ratio
97301	Salem	1.482112	0.867844	2.499384	(0, 1, 0)	(0, 0, 0, 12)	0.210677
28205	Charlotte	1.478319	1.043637	2.083222	(1, 1, 2)	(1, 0, 1, 12)	0.326990
85008	Phoenix	1.476268	0.772465	2.771658	(2, 1, 0)	(0, 0, 0, 12)	0.169010
98012	Bothell	1.462114	1.079082	1.973469	(0, 1, 3)	(1, 0, 1, 12)	0.361958
7306	Jersey City	1.433791	0.940986	2.169114	(0, 1, 3)	(2, 0, 0, 12)	0.240534

Sharpe Ratio

zip	City	mean_ret	low_ret	high_ret	order	seasonal_order	sharpe_ratio
37129	Murfreesboro	1.306796	1.126410	1.514709	(0, 1, 3)	(0, 0, 1, 12)	0.433715
34698	Dunedin	1.367636	1.105605	1.688692	(3, 0, 0)	(2, 0, 1, 12)	0.393169
37130	Murfreesboro	1.300189	1.107280	1.525099	(0, 1, 3)	(0, 0, 0, 12)	0.387260
98012	Bothell	1.462114	1.079082	1.973469	(0, 1, 3)	(1, 0, 1, 12)	0.361958
28205	Charlotte	1.478319	1.043637	2.083222	(1, 1, 2)	(1, 0, 1, 12)	0.326990

Conclusion and Future Work

- We have reommended the 5 best zipcodes based on average expected returns and also on the sharpe ratio
- We have checked the stability of our predictions
- Need to speed up analysis and extend to all zipcodes Can be done by knowing what SARIMA solutions are favored by first 1000 zipcodes
- Need better analysis on whether another recession is in the offing
- Need better metric to account for zipcodes which do better than others during recession