Is housing data responsible for the market recovery?

VEE Regression Analysis
Student Project
Fall 2014

Conor Hoekstra

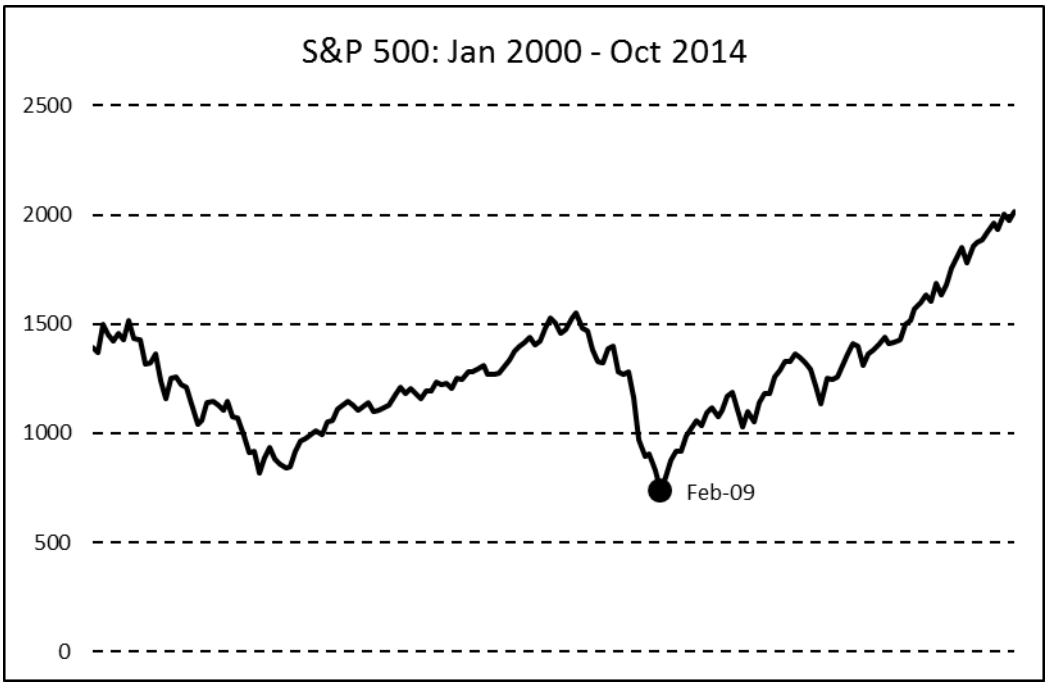
Introduction

The question attempting to be answered in this regression analysis project is whether or not housing data is responsible for the market recovery since the stock market crash of 2008. The idea for this project came from something I heard Warren Buffett say once in a Bloomberg interview. Buffett was asked whether or not the market was in full recovery mode and he responded by saying he thought the market was recovering but the main driver was housing data and that the market wouldn't be fully recovered until the housing data reached pre-2008 levels. So this leaves us with the question: is housing data responsible for the market recovery?

Analysis

The first thing that needs to be done is to determine when the market recovery began. This is done by identifying the nadir of the 2008 market crash. Below in Figure 1 is a chart of the S&P 500, the index that was chosen to represent the market for this project, from the year 2000 to October 2014, the data for which the most recent economic numbers (i.e. the housing data) is available. It is shown in the graph that the lowest point that the S&P 500 reached during the crash was in February of 2009. Therefore, the data that we will use will be the S&P 500 and housing data from February 2009 to October 2014.

Figure 1



Now that the period of time has been identified, the housing data must be selected. The most reported on economic housing data that I know of is Existing Home Sales. The S&P 500 will be the response variable and the Existing Home Sales data will be the dependent variable. Below in Figure 2 to are the results of a simple linear regression.

Figure 2

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.66187
R Square	0.43807
Adjusted R Square	0.42969
Standard Error	243.37752
Observations	69

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3093867.64	3093867.64	52.23	0.00
Residual	67	3968585.21	59232.62		
Total	68	7062452.84			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-794.7657	302.2284	-2.6297	0.0106	-1398.0161	-191.5153	-1398.0161	-191.5153
X Variable 1	2.3773	0.3289	7.2272	0.0000	1.7207	3.0338	1.7207	3.0338

The first thing that can been seen is that there is definitely a relationship between Existing Home Sales and the S&P 500. The adjusted R-squared is equal to 0.43 which means that although not significantly strong, the relationship is definitely there. Furthermore, the standard error of the explanatory variable is small relative to the coefficient. This is also another indication that explanatory variable does a good job explaining the change in the response variable.

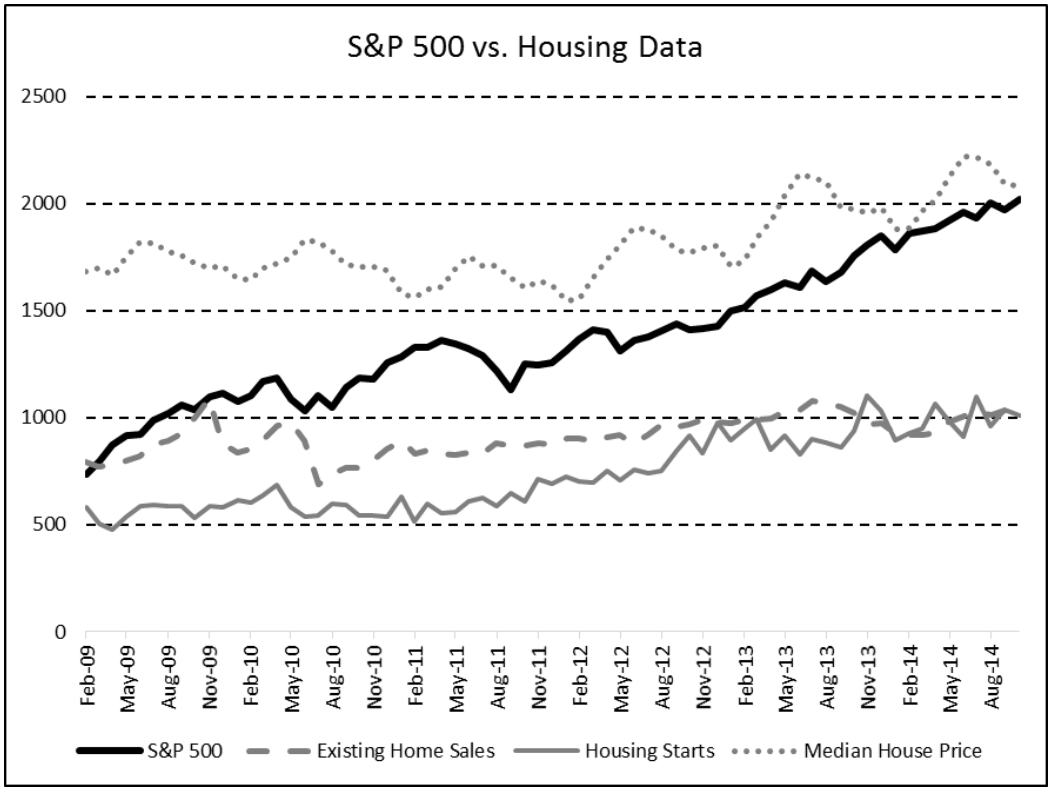
However, the standard error of the intercept is quite large. As only one source of data was looked at (i.e. one explanatory variable), this is expected. If other sources of data are regressed the results are sure to improve. Therefore, two new explanatory variables were added. These were:

- Housing starts
- Median house prices

This in total gives of three explanatory variables (including the original Existing Home Sales). The three explanatory variables are graphed against the S&P 500 in Figure 3 on the next page. It should be noted that the three explanatory variables were scaled down so they would fit on the same graph.

We can see that all of the three explanatory variables increased from the period of January 2000 to October 2014. Furthermore, we can see a seasonal trend in the Median House Price data series. The median house prices always seems to bottom in January / February, increase up until the summer time and then decrease back down to another local bottom at the start of the next year. This does make sense as more home buyers tend to purchase in the summer time than in the winter time. This may make this data series a bad explanatory variable as we do not see the same seasonal pattern in the S&P 500.

Figure 3



Below are the results of the multiple regression.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9008
R Square	0.8115
Adjusted R Square	0.8028
Standard Error	143.1144
Observations	69

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	5731140.17	1910380.06	93.27	0.00
Residual	65	1331312.68	20481.73		
Total	68	7062452.84			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-291.7853	248.5814	-1.1738	0.2448	-788.2367	204.6661	-788.2367	204.6661
X Variable 1	0.0279	0.2837	0.0983	0.9220	-0.5387	0.5945	-0.5387	0.5945
X Variable 2	1.3434	0.1634	8.2236	0.0000	1.0171	1.6696	1.0171	1.6696
X Variable 3	0.3618	0.1498	2.4146	0.0186	0.0626	0.6611	0.0626	0.6611

Note that the explanatory variables are as follows:

- X Variable 1 = Existing Home Sales

- X Variable 2 = Housing Starts
- X Variable 3 = Median House Price

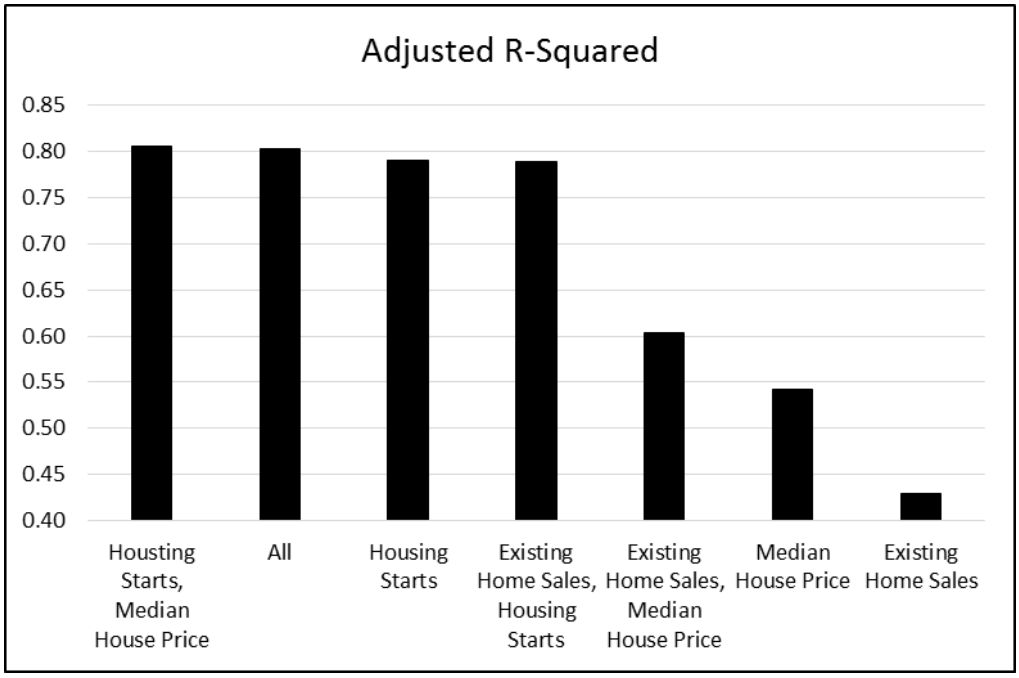
As we can see from the results, the adjusted R-squared has almost doubled to 0.80. This is a very strong relationship. As we take a look at the results of the explanatory variables, we can see that Existing Home Sales, the original data series we compared, does not add much value next to Housing Starts and Median House Price as the coefficient value is close zero and the standard error is quite large. This will probably mean that a regression of just Housing Starts and Median House Price will have a larger adjusted R-squared. It can also be seen that Housing Starts is almost four times more important to the linear relationship than Median House Price.

Below in Figure 4 are tabulated adjusted R-squared results for each of the regression combinations. These results are then also graphed in Figure 5.

Figure 4

Explanatory Variables	Adjusted R-Squared
Housing Starts, Median House Price	0.8058
All	0.8028
Housing Starts	0.7904
Existing Home Sales, Housing Starts	0.7884
Existing Home Sales, Median House Price	0.6037
Median House Price	0.5426
Existing Home Sales	0.4297

Figure 5



We can see that Housing Starts and Median House price maximizes the adjusted R-squared. Furthermore, Median House Price only marginally increases the adjusted R-squared by 0.0153. This means that Housing Starts is responsible for almost the entirety of the correlation between the optimal combination of the explanatory variables and the S&P 500.

Conclusion

We can conclude that of the housing data analyzed, Housing Starts is the mostly strongly correlated with the S&P 500 (adjusted R-squared is 0.79) and within the optimal combination of explanatory variables, it is responsible for almost the entirety of the correlation. However, correlation does not always equal causality. We cannot be certain whether or not Housing Starts is responsible for the market recovery, but we can definitely conclude that it played at least some, if not a major part.

Appendix A: S&P 500 and Housing Data

Date	S&P 500	Existing Home Sales	Housing Starts	Median House Price
Feb-09	735	794	582	1682
Mar-09	798	772	505	1699
Apr-09	873	780	478	1666
May-09	919	800	540	1747
Jun-09	919	820	585	1818
Jul-09	987	874	594	1813
Aug-09	1021	890	586	1773
Sep-09	1057	924	585	1761
Oct-09	1036	1004	534	1722
Nov-09	1096	1088	588	1700
Dec-09	1115	880	581	1707
Jan-10	1074	838	614	1649
Feb-10	1104	854	604	1646
Mar-10	1169	898	636	1696
Apr-10	1187	964	687	1723
May-10	1089	976	583	1746
Jun-10	1031	890	536	1829
Jul-10	1102	690	546	1821
Aug-10	1049	736	599	1773
Sep-10	1141	768	594	1714
Oct-10	1183	766	543	1706
Nov-10	1181	804	545	1702
Dec-10	1258	854	539	1688
Jan-11	1286	890	630	1579
Feb-11	1327	830	517	1561
Mar-11	1326	848	600	1598
Apr-11	1364	830	554	1611
May-11	1345	826	561	1693
Jun-11	1321	838	608	1756
Jul-11	1292	830	623	1712
Aug-11	1219	882	585	1712
Sep-11	1131	868	650	1653
Oct-11	1253	870	610	1608
Nov-11	1247	880	711	1640
Dec-11	1258	874	694	1622
Jan-12	1312	902	723	1546

Feb-12	1366	904	704	1556
Mar-12	1408	892	695	1648
Apr-12	1398	906	753	1737
May-12	1310	918	708	1803
Jun-12	1362	882	757	1888
Jul-12	1379	920	740	1878
Aug-12	1407	968	754	1849
Sep-12	1441	956	847	1783
Oct-12	1412	966	915	1769
Nov-12	1416	992	833	1794
Dec-12	1426	980	976	1802
Jan-13	1498	974	896	1706
Feb-13	1515	990	951	1732
Mar-13	1569	992	994	1840
Apr-13	1598	998	848	1918
May-13	1631	1030	915	2031
Jun-13	1606	1032	831	2140
Jul-13	1686	1076	898	2124
Aug-13	1633	1066	885	2097
Sep-13	1682	1052	863	1985
Oct-13	1757	1026	936	1975
Nov-13	1806	966	1105	1955
Dec-13	1848	974	1034	1977
Jan-14	1783	924	897	1879
Feb-14	1859	920	928	1883
Mar-14	1872	918	950	1967
Apr-14	1884	932	1063	2015
May-14	1924	982	984	2120
Jun-14	1960	1006	909	2220
Jul-14	1931	1028	1098	2216
Aug-14	2003	1010	963	2184
Sep-14	1972	1036	1038	2091
Oct-14	2018	1052	1009	2083