Does the S&P 500 Fluctuation Really Correlate With the Fluctuation of Housing Prices in California?

Final Project Econ 335 Summer '16
8/7/16
Prof. Matthew Elmer

By: Julian Erbil

Honor Pledge: I did not give, receive, or use any unauthorized assistance on this project.
Signed: Julian Erbil

Introduction and Statement of Research Question:

If the value S&P 500 is up for a particular year, does that mean housing prices in California are going to rise as well? This is the report's general research question, taking into account several possible causational factors including the value of the S&P 500 and its rises and falls, inflation or deflation, the consumer sentiment index, and motor vehicle and parts dealership's sales: each measured monthly from January 2010 to December 2015. This gives 72 pieces of historical data for each category, insinuating normality in the case of statistical analysis according to the given guidelines. The goal with this OLS regression is to find a correlational situation potentially between the dependent variable, the median house price in California, and hopefully any to all of the independent variables.

Formulation of the Model

The question is, whether changes in the dependent variable median housing prices in California, is correlated with changes in primarily the S&P 500, alongside the other independent variables mentioned in the introduction above. The S&P 500, as more thoroughly described above, is an index following 500 of the biggest market capitalization stocks of well known public domestic companies. Inflation is a measure of how much prices increase and how much the value of the dollar depreciates nominally, and deflation is how much prices decrease and how much dollar appreciates nominally, both according to the Consumer Price Index and the defined basket of goods. The Consumer Sentiment Index, created by a professor at the University of Michigan, is normalized at a value of 100 points from 1964, and exists to indicate consumer optimism or lack thereof, forecasts of the economy based on consumer's outlook, and to gauge current and future spending behavior. Motor vehicle and parts dealership's sales is simply the sales of domestic dealerships that sell motor vehicles and parts, and that data is measured in $1,000,000s.

The mathematical formula found by running an OLS regression with heteroskedastic-robust standard errors is CAExistingSingleFamilyHome= 135075 + 177.542(SP500) - 15550.3(InflationDeflati) -1135.14(ConsumerSentimen) +.991325(MotrVhclPartDea).

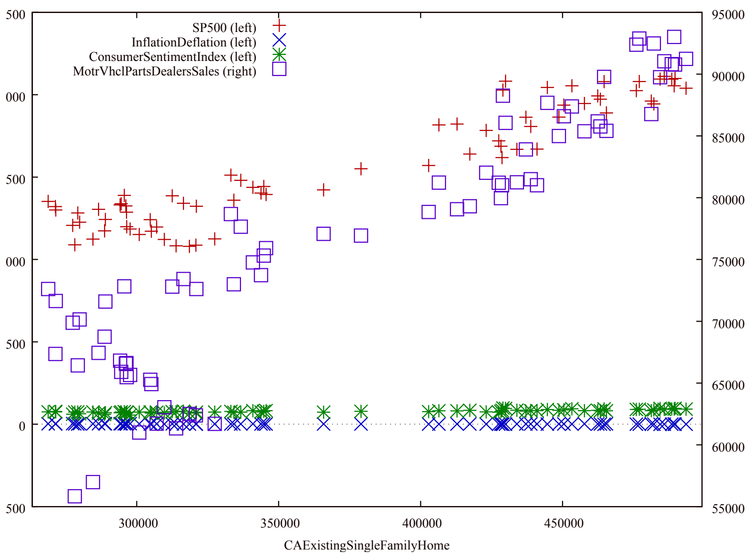
Data Description

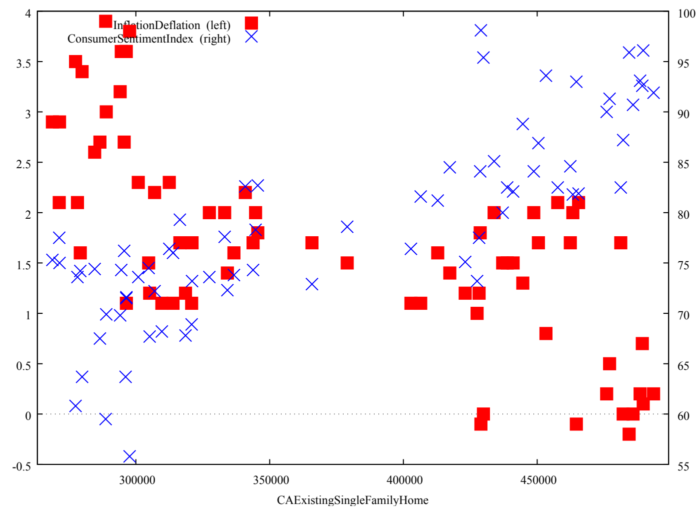
Summary Statistics	**Mean**	**Median**	**Min**	**Max**	**Std. Dev.**
CA Existing Single Family Home Average Median Sales Price	$374210	$345160	$268810	$493510	$76799
S&P500 Values	$1570.4	$1461.9	$1079.8	$2111.9	$347.41
Inflation/Deflation	1.6861%	1.7%	-.2%	3.9%	1.0194%
Consumer Sentiment Index	78.668 points	77.5 points	55.8 points	98.1 points	9.3088 points
Motor Vehicles and Parts Dealer's Sales. ($ In Millions)	$76518	$77006	$55839	$93020	$10161

I found my data on websites that are listed in the Bibliography below. The housing prices dependent variable California Existing Single Family Home is the numeric average median existing single-family home sales price per month, the average being of all counties in California. Single-family homes are the most commonly sold real estate, as they represent the average family's household purchase. The S&P 500 data is the value of the fluctuating Standard and Poor's Index, representing to many what America's stock market behaves like. Inflation is inflation: measuring price increases and the devaluation of money, deflation the opposite. Consumer sentiment index is just that: the sentiment of consumers towards spending, the sentiment on the economy in general, and forecasts based upon the current sentiment. The sales of dealerships is pretty straightforward; it is the rough total of reported sales from domestic dealerships combined. Each data set has 72 values, for the 12 months of year 0 starting January 2010 until December 1, 2015.

Looking at these summary statistics, one measure continuously stands out: the standard deviation. In my opinion, the standard deviation in each category is pretty large. However, remember that this is a measure over 6 years of data. When looking at the minimum, maximum, and deviation, it seems to suggest that the economy experienced much volatility, but in the end, growth as a whole, as all the medians are above the minimum's and close to the averages. However, certain theoretical economic aspects must be taken into account when viewing this table. Housing values almost always rise in value over time. This is due to a number of things that can be discussed in another report, but the key here is that the max and minimum almost represent the last and first data value for housing prices. Thus, regardless of correlations, housing prices rise on average, making the minimum and max data somewhat useless. The same idea applies with the S&P 500, but not as dependably. The stock market on average has gone up historically in value, however it has many more dips and sideways trends than the housing market. In general, the max and minimum should not be interpreted for housing prices nor the S&P 500 data. Yet, again, it appears to look like the economy, particularly California's, experienced volatility and growth.

On the scatterplot shown on the next page, a couple of observations can be made. Like I said previously, it appears that the S&P500 generally went up in value over time as the housing prices also went up. From roughly 2010-2011, there seems to be quite a bit of volatility when housing prices took on lower values. Outliers seem to potentially exist within the dealer's sales. As the consumer sentiment index and inflation does not really show on this graph as a result of vastly varying sizes of data values, I made another graph with just inflation and CSI, pictured on the right.





Inflation seems to have an inverse relation with housing prices, and the consumer sentiment seems to have a positive relationship with the Y variable. What is interesting is the lack of data points in the middle. Remember, both of these graphs are not chronological; they are simply based on the given data. That does not imply the graphs are not helpful, but that one should interpret them as plain scatterplots. The gap in the middle on both graphs would insinuate that a long period of average housing prices around $375,000 was uncommon, and that basically prices jumped quickly from around $350,000 to over $400,000. Thus, the lack of data simply implies a lack of values at the respective numbers, not a lack of a relationship. Another interesting point is that the data plots for each X variable almost seem exact opposites of each other, if one were to flip the inflation data on top of the CSI data, the shape would look very similar. The relationship between CSI and inflation seems to be that, as inflation "deflates," the consumer sentiment rises, which makes sense because people generally feel more confident in their purchasing power when the dollar is worth more, which exists in falling inflation/deflationary periods. I am glad I included the inflation/deflation variable in this project as it poses as a control variable for the other monetary X variables.

Empirical Results

Using the first OLS regression ran, a regular multiple regression model, I found the following results. With a constant of $135075, significant at 90 and 95%, that is the conditional mean of the dependent variable when all X values are 0. The model gives the following interpretations: if the S&P500 increases by 1 point, the average median housing price in California will increase by $177.54. If the economy inflates by 1%, the a.m.h.p. (average median housing price) will decrease by $15,550.30. If consumer sentiment goes up by 1 point on the index, the a.m.h.p. will decrease by $1,135.14. Finally, for each $1000 (data is in millions) that is sold by domestic motor vehicles/parts dealers, the a.h.m.p. will increase by about $990 (.99 x $1000). The standard error is also representing about $992, a pretty high standard error, which correlates with the insignificance of the variable. Each coefficient's effect is "on average" of course.

Both the inflation variable and the S&P500 variable are significant to the model at all conventional confidence levels! This makes sense because inflation has to do with all things monetary, and the fact that S&P500 also should be correlating somewhat with income, as many hold a portion of their savings in the stock market. Neither of the other two variables were significant, although the CSI variable looks like it could be significant at 10% if given more data, since it's p-value is .1007. The R-squared apparently explains approximately 90.81% of the variation in the average median house price in California, which may be an issue right up front. The Standard Error of the Regression is $23,966.92, which seems very reasonable.

For the second regression, I decided to include an interaction variable of S&P500 x InflationDeflation, and see what results a nonlinear model would predict. The S&P500 magnitude drops about 30 points from the original. The negative effect of the inflation variable increased significantly by about $40,000. The interaction variable has significance at all conventional levels, reinforcing the significance of the change in the inflation effect's magnitude, being a positive corrector itself. A ~$6000 decrease in the SER occurred, relating to the small increase in R-squared. The interaction has the interpretation where the S&P500 effect is now dependent on the effect of InflationDeflation's value. One big difference is InflationDelfation's standard error, drastically decreasing from the original model.

In the third regression, I decided to exclude S&P500, as a result of the previous model having a huge R-squared. I did not want to exclude the inflation variable as it accounts for controlling the other monetary variables as well in terms of real values. Interestingly enough, the inflation was of course still significant at all levels, but the amount of significance decreased by quite a bit, showing a much higher p value, however, the magnitude only decreased in size by about $1200. The most crazy occurrence is that the coefficient became negative, and by a large amount. As a result of the dealer's sales becoming extremely* significant at every conventional level, and the magnitude increasing about 5 times over to ~5.7, it makes sense that the coefficient is trying to correct, as the sales increase quite a bit over the 6 years measured, and the change numerically is in thousands. R-squared did not decreased as much, explaining now approximately 87.26% of the variation. The SER only went up by about $50. I like this regression the best out of the three because it involves the importance of car/car part sales to the economy. One issue is the fact that the coefficient change so much, and the Sales variable becomes significant as a result of excluding the S&P500 may suggest omitted variable bias.

OLS Mult. Regression #1	Coefficient	SE	T ratio	P-value
Constant	135075	61622.2	2.192	.0319
S&P 500	177.542	33.5273	5.295	.00000142
Inflation/Deflation	-15550.3	3960.6	-3.926	.0002
Consumer Sentiment	-1135.14	682.056	-1.664	.1007
Motor Vehicles and Parts Dealer's Sales	.991325	1.00488	.9865	.3274
R^2	.908097			
SE Regression	23966.92			

OLS Regr. #2 Interaction included S&P500*InflDelf	Coefficient	SE	T ratio	P-value
Constant	184076	56295.7	3.27	.0017
S&P500	144.436	35.3703	4.084	.0001
InflationDeflation	-54997.7	11693.3	-4.703	.0000135
CSI	-1012.11	670.824	-1.509	.1361
Motor Vehicles and Parts Dealer's Sales	.937828	.991737	.9456	.3478
Interaction Variable S&P500*InflDelf	26.3472	7.58045	3.476	.0009
R^2	91.9967%~92%			
SER	22534.49			

OLS Regr. #3 W/out S&P500	Coefficient	SE	T ratio	P-value
Constant	-79471.2	42463.5	-1.872	.0656
InflationDeflation	-13921.6	4509.51	-3.087	.0029
CSI	517.96	687.86	.753	.454
Motor Vehicles and Parts Dealer's Sales	5.70342	.592964	9.618	2.61e-14*
R^2	87.2656%			
SER	28004.03			

Summary and Discussion

The relationship between the independent variables and the dependent variable I was testing is obviously one of interest, as looking at history can explain the present. In my first regression, the significant variables were the constant, S&P500, and InflationDeflation. That is the trend in most of the models I ran, with the exception of the third model I ran in which Motor Vehicles and Parts Dealer's Sales came into play as the most significant variable in the model. However, as stated before, I unfortunately think this regression is inaccurate, as a result of the major coefficient change and inclusion of the Motor variable into definite significance. The reason for this significant alteration is the exclusion of the S&P500 variable, indicating major omitted variable bias, as the variable was a very significant one. Inflation has the highest magnitude of any every effect in each model, and has the lowest p value in the two models with the S&P500. This makes sense because inflation affects everything with a price tag or measured monetary value, so it represents a control variable in a sort. Another interesting factor I noticed in the first regression was the fact that Consumer Sentiment Index was very close to being significant at the 90% level, being at a.1007 p-value. I think if I had used more data, it may have been significant.

One error in my study was the fact that using all the variables I did accounted for about 87-92% of all the variation in the average median single-family housing price in California. That would lead me to believe that my studies were somewhat "kitchen sink" regressions, as the factors I put in were all pretty much important, big-data measurements of the economy, all of which are pretty much correlated with housing. This also explains the Standard Error of the Regression, which at first I noticed was pretty reasonable, staying around $20-30 thousand, but the tight SER may be a result of having accounted for many of the larger factors that make up the economy. In hindsight, I may have tried to use smaller factors, such as a particular index, or a particular mutual fund, instead of basing the study off a huge portion of the market.

My next major question would be why dealership sales didn't play a bigger role in the housing market. My guess would be that most people would skimp on cars if they have to in order to buy a house for their family. I would like to track inflation and how it plays a role on car sales, and then compare that to this study's data. With my original regression, I don't believe there was much omitted variable bias, as the S&P500 has all kinds of companies in it, each affected by different commodity prices, and company sales correlate to how retail sales are doing in general. The consumer sentiment index necessarily should be correlated with income levels on average, as well as inflation accounting for adjustments in pricing levels for all other services and goods. Thus, the internal validity is well accounted for in the study. The only exception I could see occurring is simultaneous causality. A trend in housing prices could cause a trend in the S&P, the opposite being what my report shows.

My hypothesis about the relationship between housing and the S&P was correct, however it was a sort of obvious guess. What I did not expect was both the CSI and Motor Vehicles being insignificant, the CSI consistently. The one time I found Dealer's Sales to be significant was when I took out S&P, but the CSI was still insignificant to the housing prices. However, what I did find was more than interesting, which is the *major* significance of the S&P500 movement and the huge effect just a 1% inflation or deflation in the price level has on the single family housing market in California.

Bibliography

"Historical Inflation Rates: 1914-2016." *US Inflation Calculator*. COIN News, 24 July 2008. Web. 02 Aug. 2016. <http://www.usinflationcalculator.com/inflation/historical-inflation-rates/>.

"Login." *Login*. California Association of Realtors, n.d. Web. 02 Aug. 2016. <http://www.car.org/marketdata/data/ahds/?redirectFrom=login>.

"Motor Vehicle and Parts Dealers." *Annual and Retail Trade*. Census.gov, n.d. Web. 02 Aug. 2016. <https://www.census.gov/retail/marts/www/adv44100.txt>.

"Surveys of Consumers - Data." *Surveys of Consumers - Data*. University of Michigan, Jan. 2015. Web. 02 Aug. 2016. <https://data.sca.isr.umich.edu/data-archive/mine.php>.

"S&P 500 Historical Prices by Month." *S&P 500 Historical Prices by Month*. Multpl, n.d. Web. 02 Aug. 2016. <http://www.multpl.com/s-p-500-historical-prices/table/by-month>.

Note to reader/Professor: I would have included the outputs from my other two regressions, but I saved the data and commands, exited Gretl, and reopened it and find that it did not show me the same data results as it did originally. This leads me to believe that I did not save correctly, which was unfortunate, because I cannot seem to get the same outcomes as I did on my original regressions/statistics. Luckily, this is what I saved before I accidentally exited Gretl. This shows me that I should always save every piece of work before exiting anything.

```
                              Mean           Median          Minimum
Maximum
SP500                       1570.4           1461.9           1079.8
2111.9
MotrVhclPartsDea~           76518.           77006.           55839.
93020.
InflationDeflati~           1.6861           1.7000         -0.20000
3.9000
ConsumerSentimen~           78.668           77.500           55.800
98.100
CAExistingSingle~      3.7421e+05       3.4516e+05       2.6881e+05
4.9351e+05

                         Std. Dev.             C.V.         Skewness    Ex.
kurtosis
SP500                       347.41          0.22123          0.22286
-1.4474
MotrVhclPartsDea~           10161.          0.13279        -0.080301
-1.1048
InflationDeflati~           1.0194          0.60461          0.18642
-0.32020
ConsumerSentimen~           9.3088          0.11833         0.068038
-0.33661
CAExistingSingle~           76799.          0.20523          0.15452
-1.5873

                          5% perc.        95% perc.         IQ range
Missing obs.
SP500                       1088.5           2094.4           658.84
0
MotrVhclPartsDea~           61216.           92416.           18734.
0
InflationDeflati~        -0.035000           3.6000           1.0750
0
ConsumerSentimen~           62.685           95.575           11.200
0
CAExistingSingle~       2.7536e+05       4.8876e+05       1.4923e+05
0
```

```
Model 2: OLS, using observations 1-72
Dependent variable: CAExistingSingleFamilyHome
Heteroskedasticity-robust standard errors, variant HC1

                      coefficient   std. error   t-ratio  p-value
  ----------------------------------------------------------------
  const              135075        61622.2        2.192   0.0319   **
  SP500                 177.542       33.5273     5.295   1.42e-06
***
  InflationDeflati~  -15550.3       3960.60     -3.926   0.0002
***
  ConsumerSentimen~   -1135.14       682.056    -1.664   0.1007
  MotrVhclPartsDea~       0.991325     1.00488   0.9865  0.3274

Mean dependent var   374214.2   S.D. dependent var   76799.07
Sum squared resid    3.85e+10   S.E. of regression   23966.92
R-squared            0.908097   Adjusted R-squared   0.902610
F(4, 67)             205.6249   P-value(F)           7.62e-37
Log-likelihood      -825.6515   Akaike criterion     1661.303
Schwarz criterion    1672.686   Hannan-Quinn         1665.835
```