# Do Stock Market Values Affect Housing Values

Question

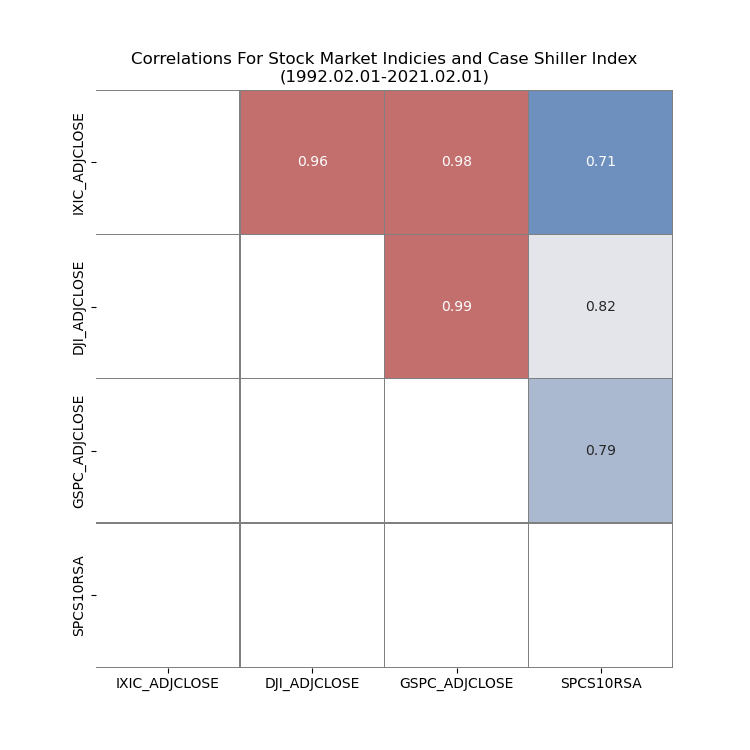
Do stock market values affect housing values?

Steps

Our first step is to determine if there are any correlations between stock market indices and home values. We choose three major indices Down Jones Industrial Average (^DJI), the S&P 500 (^GSPC), and the NASDAQ (^IXIC). These indices were chosen due to the slightly different composition of each index in terms of companies that make up that index. The Dow is made up primarily of large manufacturing companies in different sectors. The S&P500 has a wider representation of companies across Tech, Manufacturing, Consumer Goods, Financials and Retail. The NASDAQ is tech heavy containing software, and hardware producers in the tech sector. Housing values were represented by the Case-Shiller 10 City Composite Index (SPCS10RSA). This Case-Shiller index is a composite of the 10 largest metro areas in the United States. We also have the Case-Shiller index for United States (CSUSHPISA).

Using this data, our first step is to determine if any correlations exist between the stock market indices and home values in the 10-city composite. We will also validate that there is a strong correlation between the Case-Shiller 10 City Composite and the Case Shiller National Index. If so, we will proceed with our analysis using only the 10 City Composite.

Table 1: Correlation Values for Major Stock Indices and Home Values



We can see that there is a strong correlation between the stock indices and home values represented by the Case-Shiller 10 city composite index. We also observe, as expected, that the Case-Shiller 10 City Composite Index may be used to represent housing data as a whole based on the strong correlation between the national and 10 City indices. It’s important to consider that these correlations include the housing bubble from the late 90’s and early 200’s let’s take a look at that with normalized data.

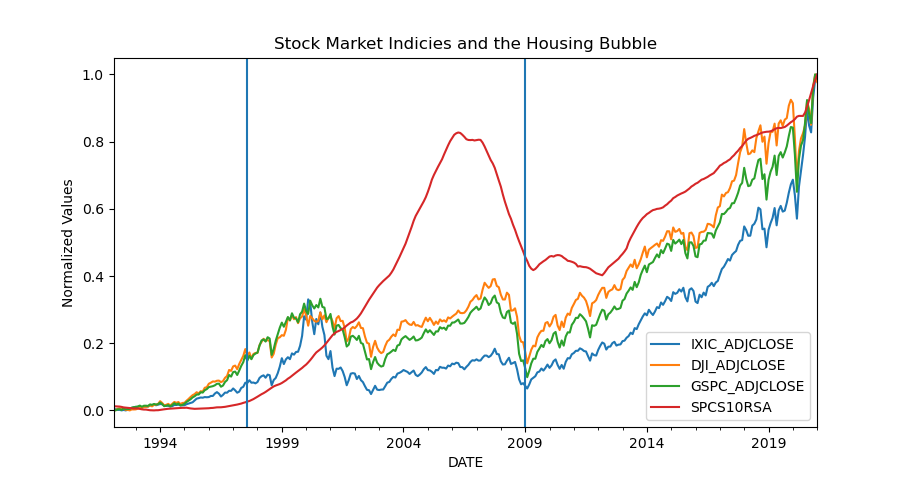


Figure 1 Housing Bubble

Figure 1. Illustrates the Housing Bubble with the estimated start of August 1st 1997 and ending on Jan 1st 2009. We now want to explore the correlations in each of these three time periods. To understand how the housing bubble impacted the correlations.

Data and Assumptions

Monthly stock market index data was obtained through the yahoo finance API using the python api https://pypi.org/project/yfinance/ Three CSV files were created. These files were named djii\_1m.csv, nasdaq\_1m.csv, and sp500\_1m.csv. Each of these files contained the following columns:

Table Stock Market Data Format

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Open** | **High** | **Low** | **Close** | **Adj Close** | **Volume** |
| YYYY-MM-DD | Float | Float | Float | Float | Float | Int |

Where **Date** is the year month and day of the stock market readings in four-digit year, two-digit month and two-digit day format. **Close** is the closing price of the market is the actual cash value of the last transacted price before the market closes. **Adj Close** is the closing price adjusted by various corporate actions which include but are not limited to stock splits, dividends, and rights offerings. For the purposes of examining market index trends of long periods of time. We chose to use the adjusted closing price. For the purposes of correlations, we can see that we may use either Close or Adj Close to compare these values by looking at the correlations between Closing price and Adjusted Closing price for each index.

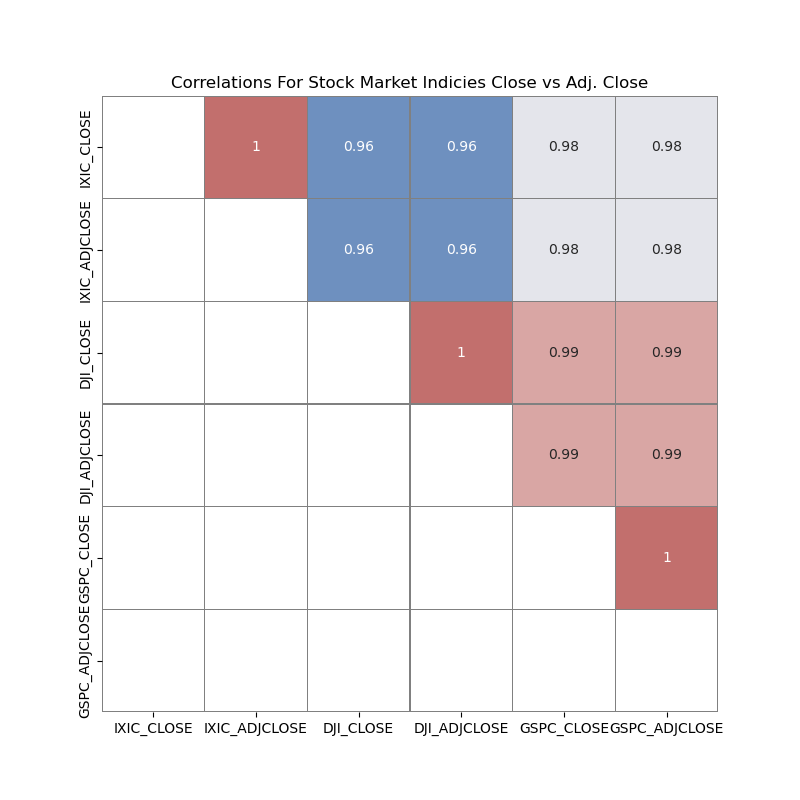


Figure . Indices Closing vs. Adjusted Closing Correlations

Yahoo Finance data for Down Jones Industrial Averages (DJI) data starts on Feb 01 1999 and are reported monthly on the first day of each month through Jan 01, 2021. The two other datasets Nasdaq (IXIC) and Standard and Poor’s 500 (GSPC) are reported starting Jan 01, 1987 and are reported on the first day of the month through Jan 01, 2021. We can see that either adjusted closing or closing could be used to represent an index based on an r value of 1 for each index close with its adjusted close.

The Index data was validated with the following procedure.

Home Price Index data is obtained from the Federal Reserve bank of St. Louis FRED database. [Federal Reserve Economic Data | FRED | St. Louis Fed](https://fred.stlouisfed.org/). We choose to use the Case-Shiller 10 City Composite Home Price Index

Table . Example FRED data Case-Shiller 10 City Housing Index Data Format

|  |  |
| --- | --- |
| **DATE** | **SPCS10RSA** |
| YYYY-MM-DD | float |

This is an index of home prices in the 10 largest US metro areas. The index data is set to be 100 on Jan 2000 and is Seasonally Adjusted. This data is reported on the first day of the month starting in Jan 01, 1987 through Jan 01, 2021. The following cities are represented by the 10-city composite, Boston, MA; Chicago, IL; Denver, CO; Las Vegas, NV; Los Angeles, CA; South Florida; New York, NY, San Diego, CA; San Francisco, CA; and Washington DC.

The Fred and index data were validated with the following procedure.

1. Create a Pandas Dataframe from the .csv data.
2. For Fred data missing values are represented with a “.”. We will need to be mindful of this. This will impact two things.
   1. When the data is checked for missing values with isnull we won’t find any
   2. We should convert those (and do) to np.nan when reading them in.
3. Take a look at the head of the data. We expect two columns for all Fred Data sets. These columns are a DATE and value column. The DATE should be in YYYY-MM-DD format and the value will look like a floating-point number.
4. We check the values for the VALUE column are type float64 for Volume in index data that value should be int64
5. Next a DataFrame is created from the .csv file. Using the DATE column as an Index. We check that the index created is “DatetimeIndex” and that the data type for the entries in the index are datetime64[ns].
6. Count the Columns, make sure we only have one for fred data and six for index data. Date will now be an index.
7. Run describe on the data and make sure that it makes sense. Specifically counts, min and max.

Once these individual data sources are in individual frames that have been validated. They are merged into a single data frame using the DATE index. This “complete” frame will be used to perform analysis in the project.

Appendix

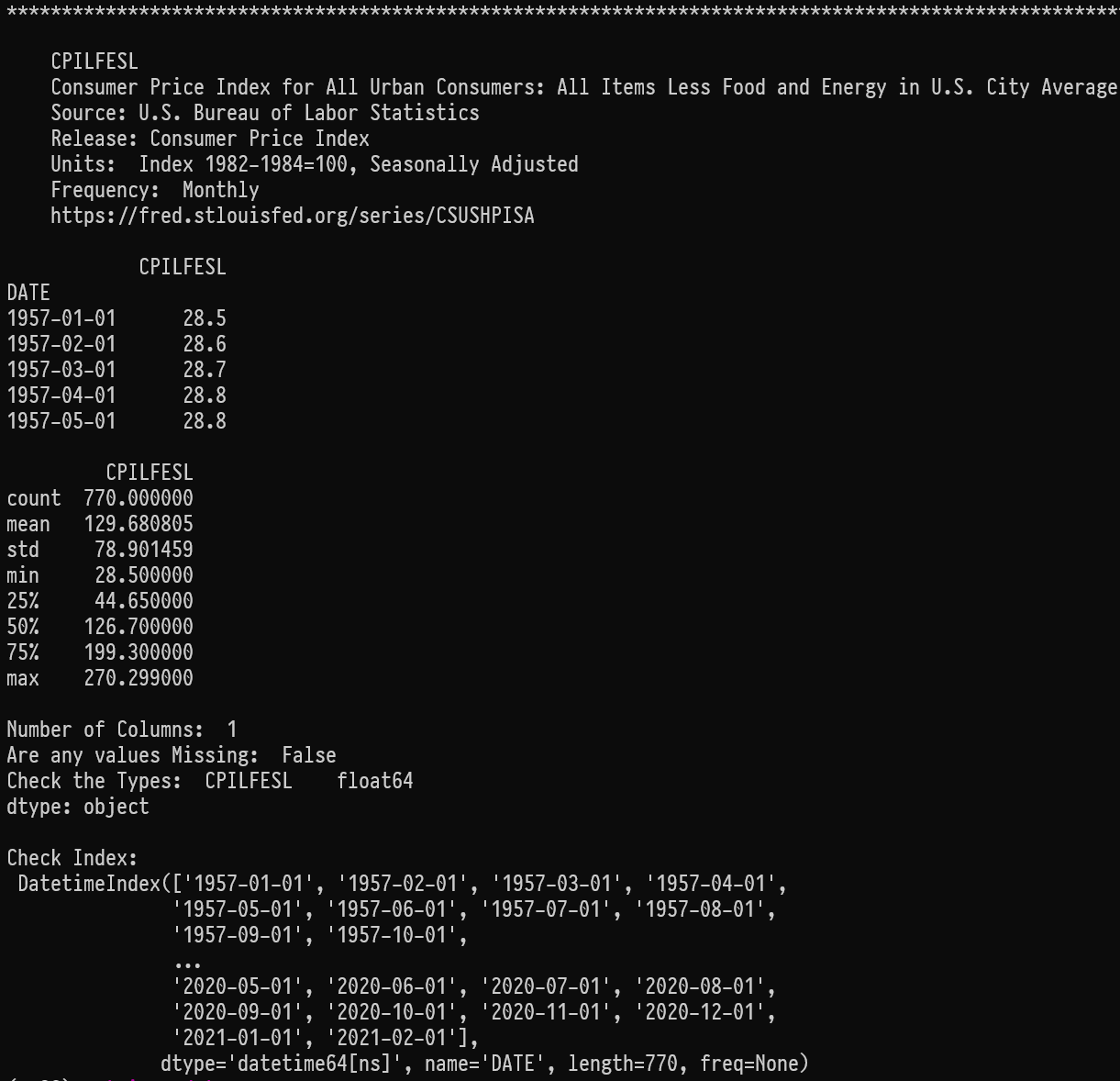


Figure Sample Output from FRED data Verification