MARIN
ECONOMIC FORUM

Land Prices, Policy and Connections

October 2017

Provided by Marin Economic Forum
555 Northgate Dr, Suite 235
info@marineconomicforum.org

MARIN
ECONOMIC FORUM

Contents

MARIN
ECONOMIC FORUM

Executive Summary

Land prices are regional and are affected by multiple factors. This study looks specifically at how county-level, home-price data may be correlated and cause changes to other counties' land prices. Home sales have more observations than bare land or commercial property sales, and have comparable data used across counties regularly to compare real estate pricing. The core Bay Area counties of San Francisco, San Mateo, Santa Clara, Alameda, and Contra Costa are compared to data for the North Bay counties of Sonoma and Napa as to their effects on Marin County and Marin County affects each county's median home prices.

Marin County is the focus of this study. Data from 1996 to 2017 are used to test this basic hypothesis at the countywide level using three measures that culminate with a test for causality. Commuting patterns, the fact that these counties share both residents as workers and also transportation systems, provide further logic that real estate prices would be linked from one county to another. The real question is how counties have relative influence one over the other.

Correlation is the statistical relationship between two series of data like the slope of a line; Marin County's median home prices are highly correlated to all counties compared. For Marin County, correlations range between +0.85 and +0.97, suggesting a very strong relationship between Marin County home/land prices and each county in our sample.

The next results are for the coefficient of determination (R^2), or how one county's variability in home prices helps determine the variability in another county's home prices. Because this statistic is simply the correlation relationship squared, high correlations imply high "goodness of fit" in estimation models. For Marin County, the Bay Area counties show more explanation of how Marin County prices vary than the North Bay counties, but Marin County also explains variability in other counties' home prices.

This two-way explanation does not tell us, for example, if San Francisco County influences Marin County or if Marin County influences San Francisco.

Neither of these statistics are themselves indicators of "causality", or how older values of one county's home prices may influence another's values in the current market. Causality is a hotly-debated topics in economics. The data and graphics in this study show simple causality tests using lagged values of each county with Marin County for three major results:

1. Marin County home prices are influenced by the core Bay Area counties more than the North Bay Counties;
2. There is bi-directional causality in almost every case, which is a classic result when data are this highly correlated; and
3. Marin County cities show more influence from southern to northern cities than the reverse.

Does Marin County follow San Francisco or Sonoma and Napa counties in the evolution of its land prices using home prices as a proxy? That is the central question of this study; we see later the data suggest the hypothesis is correct. The basic evidence shown in this study suggests that Marin County home/land prices are driven more by core Bay Area prices than by North Bay prices otherwise. Logic, anecdotes from the real estate industry, and the home price data all support this conclusion.

MARIN
ECONOMIC FORUM

Land Pricing and Marin County

Introduction

Land prices are regional and are affected by multiple factors. This study looks specifically at how county-level, home-price data may be correlated and cause change in other counties' land prices. The goal of this project is to provide land price comparisons for Marin County to other parts of the Bay Area using data on developed land in San Francisco, San Mateo, Santa Clara, Contra Costa, Alameda, Sonoma, and Napa counties. However, there are contrasts within Marin County: land price dynamics in Novato are likely different than in San Rafael. Data for this study comes from Zillow Research (a publicly-available source).
Land comes in three basic forms: bare land, land with commercial improvements, and land with residential improvements.

Marin County is the focus of this study for three reasons. First, it is a relatively expensive place to live and has one of the highest median home price levels in the United States. Second, Marin County is adjacent to San Francisco County; along with San Mateo County, Marin County is seen as a place where people who work in San Francisco live beyond the city. Third, there are statements that Marin County is more rural than suburban or urban in its land price fundamentals. Does Marin County act more like San Francisco or Sonoma and Napa counties in the evolution of its land prices using home prices as a proxy? That is the central question of this study.

There are three measures used here to show comparisons of home prices from 1996 to 2017. The first measure is simple correlations, or how price trends for each county connected over time. Correlation is the statistical relationship between two series of data like the slope of a line; Marin County's median home prices are highly correlated to all counties compared. This result corroborates the basic idea that real estate prices are regional.

The second result comes from the coefficient of determination, or how one county's variability in home prices helps determine the variability in another county's home prices. This "goodness-of-fit" test is very statistically significant for county home prices as compared to one another; because this statistic is simply the correlation squared, high correlations imply high goodness of fit. However, neither of these statistics are themselves indicators of the third result called "causality", or how older values of one county's home prices may influence another currently.

There may currently be a housing "crisis" as a result of the housing supply and demand factors in Marin County. Pro-housing legislators, the Governor's Office, fair-housing advocates, and others can use this data to speak in realities about land prices.

In summary, this report has the following elements:

- Data description and methodology, including data on commuting that links labor and housing markets across the Bay Area and North Bay;
- Show the correlation, coefficients of determination and causality tests;
- Conclusions and recommendations; and
- Appendix: Methodology explained.

Linking Land Prices and Commuting

We assume throughout this brief analysis that residential property values are indicative of what bare land and commercial properties would be worth otherwise for three reasons:

1. Residential property prices increase land values as competition for other uses and basically have one use: provide a home;
2. Commercial property needs residential property as a complement and may have many uses; and
3. As residential property values rise, incentives to build use up available land otherwise, increasing bare land's value to slow down incentives/profits to build more homes.

The economics of land values follow simple supply and demand principles and are regional. Marin County also has metered its use of available land by designating properties as protected from development. As a result, there has been a reduction of available land. San Francisco City and County, with limited land as defined by its municipal boundaries, faces a similar fate; building has been vertical, which Marin County has not done en masse.

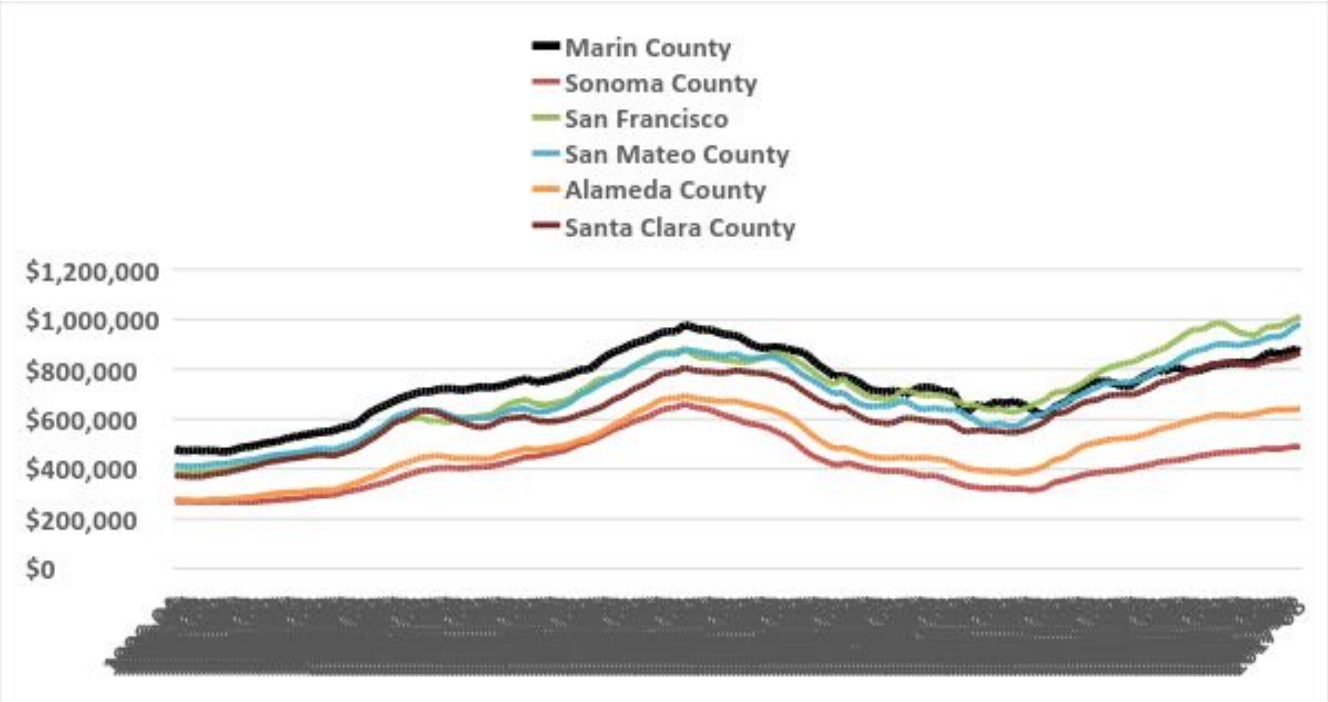
It is difficult to disentangle commute/traffic/transportation issues from housing choice. The economics of each are connected. If the price of commuting increased, say if gas prices rose sharply, some workers may not want to commute as much without a pay increase. If wages were to rise as a result of commute costs rising, this could fuel higher housing prices because wages would not just rise for those that commute, they would rise for other workers as well.

Commuting and Labor Market Links

Marin County, which is connected physically to three counties (Contra Costa, San Francisco and Sonoma) has common links with Sonoma County in terms of population, agricultural heritage and roadway connections, but has labor markets that are connected to the core Bay Area counties otherwise. Commuting data show this. Marin County has been an alternative place to live outside of San Francisco, San Mateo, Contra Costa, and Alameda counties, versus Sonoma, Napa and Solano counties to access Bay Area employers.

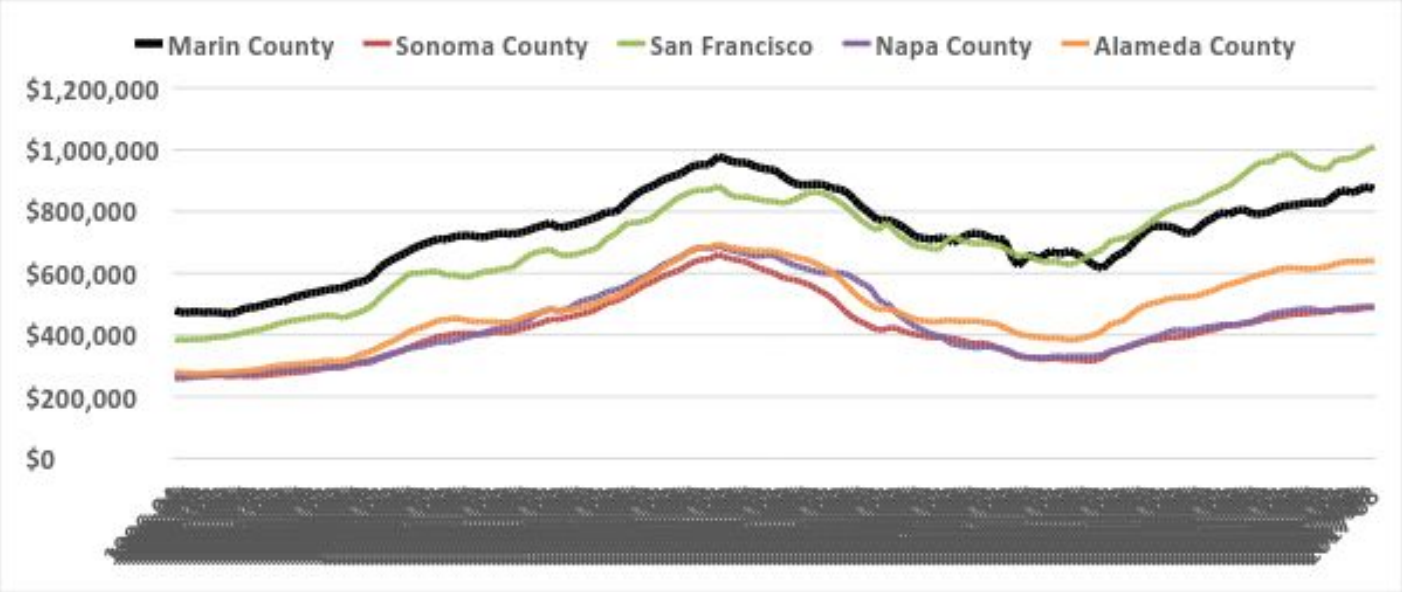
Housing costs, wages, and the costs of commuting are all "determinants" of land price and the quantity of land available. Figures 1a and 1b show the connections to San Francisco and the core Bay Area versus Sonoma and Napa counties in inflation-adjusted terms. The key to these figures is the premium paid to live in Marin County and counties such as San Francisco and San Mateo is consistent over time versus the other Bay Area and North Bay Counties. This simple, observational test suggests the central question of this study has the answer yes.

Figure 1a: Home Prices 1996 - 2017, 2009$, Marin, Sonoma, Alameda, Contra Costa, Santa Clara, San Mateo and San Francisco Counties, All Home Types



Source: Zillow Research, California Department of Finance

Figure 1b: Home Prices 1996 - 2017, 2009$ Marin, Sonoma Napa, Alameda, and San Francisco Counties, All Home Types



Source: Zillow Research, California Department of Finance

Logic suggests that larger housing and labor markets should direct smaller, regional labor and housing market changes; small counties and municipalities are subject to larger markets. Commute pattern data are from 2002 to 2015 as of October 2017, and are a mix of Census, Bureau of Labor Statistics and Internal Revenue Service data called the Longitudinal Employment and Housing Dynamics (LEHD) database (the latest data was published in Sept 2017 for data from 2002 to 2015, see http://onthemap.ces.census.gov for more)

The hypothesis above is that Marin County is influenced more by core Bay Area county land/home prices than by other North Bay counties. Commuting plays a role in people's choices of where to live regionally and also influence the cost of housing. Figures 2 and 3 provides these estimates focused on Marin County for those that live in Marin County (Figure 2) and those that live outside of Marin County, but work in Marin County (Figure 3).

Figure 2: Live in Marin, Employed Elsewhere or in Marin County, 2007 – 2015

County	2007	2008	2009	2010	2011	2012	2013	2014	2015
Marin	42,222	42,618	41,255	40,481	40,738	40,559	41,230	41,674	42,161
San Francisco	21,922	23,168	22,397	23,519	25,043	25,797	26,199	26,692	26,918
Alameda	6,897	7,030	6,482	6,865	7,127	7,197	7,276	7,370	7,377
Sonoma	4,776	4,468	4,439	4,369	4,931	4,709	5,020	5,046	5,147
Contra Costa	4,337	4,314	4,160	4,222	4,297	4,222	4,204	4,247	4,492
San Mateo	3,677	3,736	3,586	3,523	3,711	3,849	3,921	4,029	3,962
Santa Clara	3,024	3,277	2,897	3,041	3,045	3,348	3,438	3,657	3,603
Sacramento	1,829	1,965	1,788	2,064	1,944	1,934	1,905	1,899	1,942
Los Angeles	777	1,210	1,444	1,505	1,383	1,287	1,546	1,679	1,524
Solano	1,127	1,134	1,206	1,215	1,233	1,257	1,340	1,293	1,288
All Other Locations	5,769	6,736	6,665	6,603	6,976	6,848	7,453	8,039	7,365
Total All Jobs	96,357	99,656	96,319	97,407	100,428	101,007	103,532	105,625	105,779

Source: onthemap.ces.census.gov

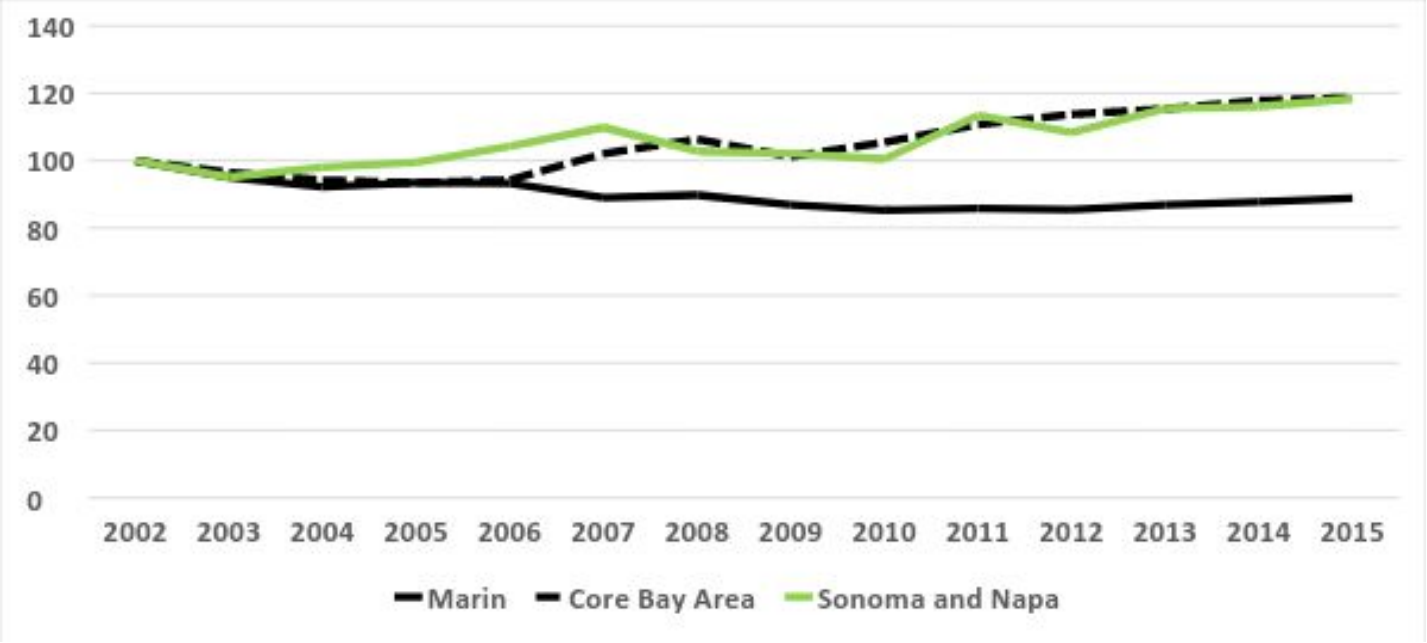
Figure 3: Live in Marin or Elsewhere, Work in Marin County, 2007 – 2015

County	2007	2008	2009	2010	2011	2012	2013	2014	2015
Marin	42,222	42,618	41,255	40,481	40,738	40,559	41,230	41,674	42,161
Sonoma	14,306	15,212	14,409	14,264	14,541	15,038	15,327	15,249	15,439
Contra Costa	8,894	8,949	8,611	8,749	9,145	9,155	9,984	10,123	10,856
San Francisco	10,393	9,735	9,437	8,924	9,497	9,590	9,474	9,250	10,168
Alameda	5,786	6,470	6,343	6,274	6,389	6,518	6,760	6,985	7,553
Solano	4,403	4,382	4,295	4,327	4,435	4,694	4,772	4,791	5,001
San Mateo	3,424	3,331	3,187	3,259	3,352	3,242	3,297	3,454	3,484
Santa Clara	2,961	2,652	2,828	2,546	2,630	2,615	2,877	2,907	2,794
Sacramento	2,203	2,158	2,061	2,153	2,272	2,345	2,339	2,351	2,380
Napa	1,932	1,922	1,804	2,014	2,025	2,103	2,101	2,313	2,316
All Other Locations	6,857	7,779	8,588	8,484	8,766	9,105	10,011	10,542	10,319
Total All Jobs	103,381	105,208	102,818	101,475	103,790	104,964	108,172	109,639	112,471

Source: onthemap.ces.census.gov

Figure 4 shows how those that live in Marin County and work there also has faded slowly since 2002. The core Bay Area counties are the dominant places where Marin County residents work. Sonoma County is the main exception, and remains the main county where people that work in Marin County beyond local residents live (as shown in Figure 3).

Figure 4: Live in Marin, Employed Elsewhere or in Marin County, 2002 – 2015, Index 2002 = 100



Source: onthemap.ces.census.gov

Data and Statistical Tests

Given commutes, the land price data should show a larger, statistical relationship between Marin County housing prices in the Bay Area counties versus North Bay counties. To test such a hypothesis, statistical theory has many tests and some concerns. This study uses median home/land sales prices. We need to use test statistics that focus on data over time or "time series" data.

Causality is the real relationship sought by this study, but it is difficult to show in practice. In simple terms, causality is when a statistical relationship is significant only one way, or when one variable acts as an independent influence on another, but the reverse is not a significant relationship.[1] Causal relationships are difficult to find in economic variables that typically feed off each other (what we call later "bi-directional" causality). However, we can say a few things about land prices based on how markets work:

- Large markets are more likely to influence smaller markets than the opposite;

[1] Please see the appendix for more details on methodology.

- The core Bay Area markets have more jobs and act as a place people come to work in net versus leave to go to work, using Alameda, Contra Costa, Santa Clara, San Mateo, and San Francisco and the core Bay area counties; and
- Sonoma County has almost 80 percent of its working residents employed inside the county, hence home prices in the North Bay are likely influenced by internal and external factors.

Let's look at the data and test statistics.

Comparison of Data

Figures 1a and b showed a simple way to consider how the inflation-adjusted, median home prices have moved in Marin County versus North Bay and core Bay Area counties.[2] Figure 5 has simple correlations for these counties' median home prices from 1996 to 2017.

Figure 5: Correlation among Eight Bay Area Counties, 1996 – 2017, 257 monthly observations

	Alameda	Contra Costa	Marin	Napa	San Francisco	San Mateo	Santa Clara	Sonoma
Alameda	100.0%	94.1%	98.7%	92.9%	96.7%	98.4%	98.1%	96.1%
Contra Costa		100.0%	92.7%	98.7%	82.9%	87.0%	86.0%	99.3%
Marin			100.0%	93.0%	96.2%	97.3%	97.2%	95.4%
Napa				100.0%	82.8%	86.3%	85.2%	98.3%
San Francisco					100.0%	99.4%	99.3%	86.8%
San Mateo						100.0%	99.7%	90.3%
Santa Clara							100.0%	89.4%
Sonoma								100.0%

Sources: Author's Calculations and Zillow Research

Notice that the correlations are very high in Figure 5, but not all are as close to 100 percent (which is perfect positive correlation) as others. San Mateo and Santa Clara counties are over 99 percent correlated with San Francisco's median home price, as is Contra Costa or Sonoma counties.

Marin County, Alameda, San Mateo, Santa Clara, and San Francisco counties are highly correlated. Figure 6 shows data for the "coefficient of determination" or what is known as R^2. In statistical analysis, R^2 is related to the analysis of variance, or the variability of data around its middle measures, the average and median. When correlation is close to 100 percent, R^2 (which is simply the correlation coefficient in Figure 5 squared) is also close to 100 percent.

Figure 6: Coefficient of Determination, Current Values

	Alameda	Contra Costa	Marin	Napa	San Francisco	San Mateo	Santa Clara	Sonoma

[2] A quick comparison helps eliminate any concerns about multiple datasets. Zillow Research published data on all homes (condos and single-family homes or SFHs) since 1996. We use the San Francisco Consumer Price Index (CPI) as published by the California Department of Finance monthly.

Alameda	100.0%	88.6%	97.4%	86.4%	93.5%	96.9%	96.3%	92.4%
Contra Costa		100.0%	86.0%	97.5%	68.7%	75.7%	74.0%	98.7%
Marin			100.0 %	86.5%	92.5%	94.8%	94.4%	91.1%
Napa				100.0 %	68.5%	74.5%	72.6%	96.6%
San Francisco					100.0%	98.7%	98.6%	75.3%
San Mateo						100.0%	99.5%	81.6%
Santa Clara							100.0%	79.8%
Sonoma								100.0%

Sources: Author's Calculations and Zillow Research

Two concerns of using data over time include data are "spuriously" correlated because they are growing over time due to phenomena like inflation, and not necessarily due to the fundamental aspects of each data set. Because of these other variables, the correlation between Marin County and San Francisco County home prices may be suspect. Such a spurious relationship makes the coefficient of determination and the F-test suspect.

These data sets do have the characteristics of what statisticians call a "unit root" or a relationship from time period to time period that may be driving data outcomes all by itself. However, because median home price data in each county here have the same characteristics, we can call them "cointegrated" (where being integrated means that lagged time periods have a major effect on the current time period) such that all county home prices have the same data characteristics. Because of this, the bad things are nullified in the Granger analysis below, and have little effect on the results.

Why Figure 6 is Important

Notice that the percentages in Figure 6 have become lower. Correlation measures how data move with each other; R^2 tells a story of how the variability of data help to explain the variability of another data set. These relationships are a pre-cursor to testing for a causal relationship. The more one set of data can explain the variability of another data set, the more we can say they are related to each other. However, a causal test is in the lagged values, or how land prices in San Francisco nine months ago may help explain land prices in Marin County today.

The Importance of Lagged Values

This section investigates possible casual relationships by comparing lagged, median home price data to each other. The basic hypothesis of this study can be tested in a Granger Causality test. Such a test compares current data (say Marin County home prices in September 2017) to older or "lagged" values of home prices in other places (say San Francisco home prices in January 2017). The Granger Causality test takes one step beyond that. Such a test also considers the other direction of

change as a "control": San Francisco's current home prices are tested against lagged values of Marin County home prices to see if there are true, effects in one direction or if the two markets feed off each other such that neither causes the other to change **independently**.

This concern is at the heart of statistical tests for causal relationships, but these relationships are difficult to pin down in practice. By testing the lagged values from each direction of potential cause, we control for which side may have more influence on that variability.

The Granger Causality tests help explain what data are "independent" and which is "dependent" on the other in a comparison of two data sets, with some amount of error. We use the five (5) percent margin of error here. If San Francisco home prices are independent of Marin County but help determine home prices in Marin County (older values of Marin County home prices do not help explain the current values for San Francisco, but older values of San Francisco do help explain current home prices in Marin County in a statistically-significant way), we can make an inference that San Francisco home prices "cause" Marin County prices to change with some amount of error.

From Zillow Research, we start with 257 monthly observations of estimated median home prices as of July 2017. Because these data include a full mix of home types, it acts as a strong sample for land prices, as land prices are part of the "home" price (as the home is simply an addition or "improvement" to bare land). When we use older values in a Granger test, we lose some of the observations (what a statistician would call loss of "degrees of freedom"), reducing the test's viability. For example, if we thought nine months ago prices changing in San Mateo County would finally ripple over the Golden Gate Bridge to Marin County, the number of observations would fall to 246. In all cases, these data match each other in terms of degrees of freedom.[3]

This large amount of degrees of freedom, a reflection of the sample size, means a low value of the test statistic is needed to generate a statistically-significant result. The test statistic used here is the F-statistic, or a ratio of variances in each time series. More details are provided in the Appendix, but basically the F-statistic is based on the R^2 measure and helps to test for which variable explains movements in the other more completely. Figure 7 shows how the effects fade over time of all these counties on Marin County; Figures 8 through 13 look at the individual counties and their relationships to Marin County, where Figure 8 combines Sonoma and Napa counties in the same graph.

A test statistic value of greater than 1.26 (the vertical or "y" axis in each of Figures 7 through 13) is a statistically-significant relationship with five (5) percent error. As the test statistic rises, the statistical significance of the result rises. Notice that most of the results in Figures 7 through 13 show bi-directional causality, meaning both counties being compared have an influence on each other. However, Marin County has less influence on specific Bay Area counties, counties which have strong influences of Marin County through their lagged home prices.

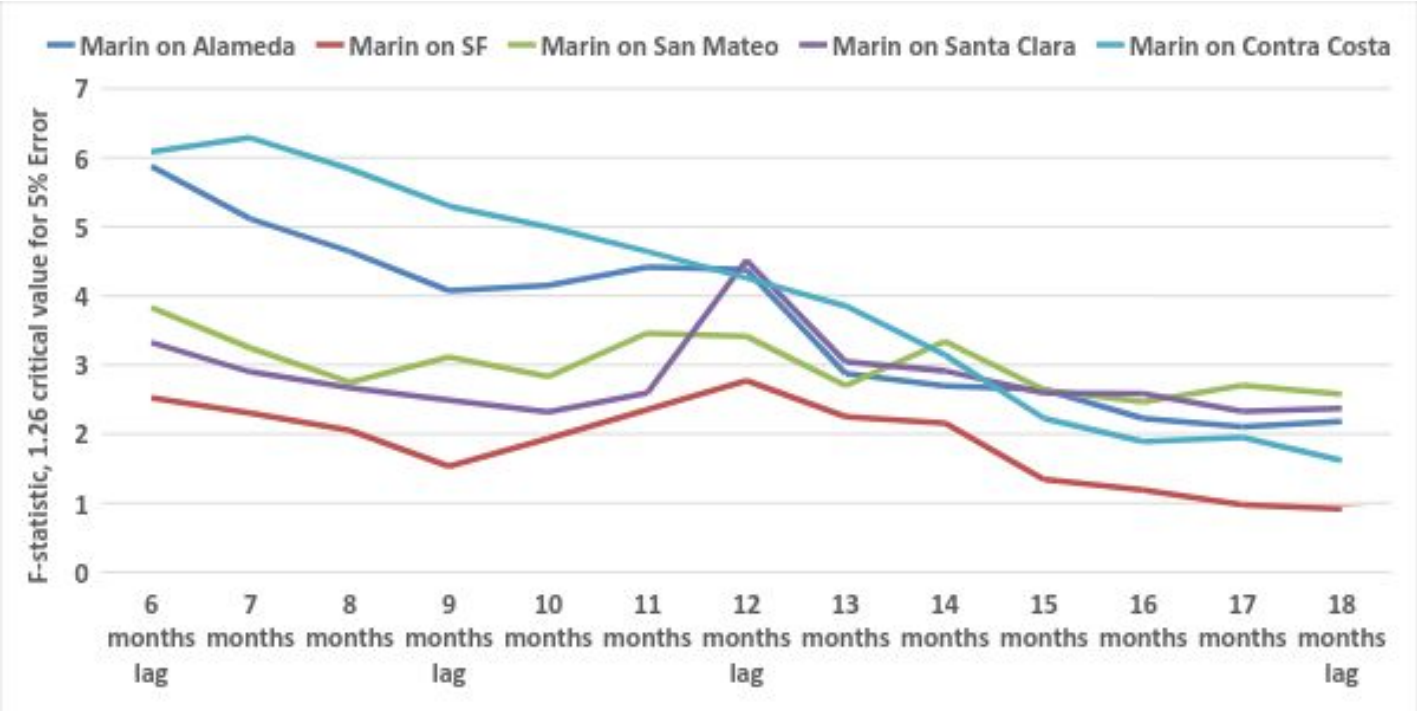
[3] Looking at monthly transactions individually and their distributions provide more volatility and not necessarily a measure of market.

Comparison of county-level data

The datasets compared in this study are best seen at the county level. The markets for land have prices determined by comparable properties and sales prices as new deals are made. County-level data provide more transactions (data points) and more variation potentially. The data for the Bay Area's nine counties are shown below in comparison form. We use monthly, median prices as the baseline frequency and measure. Figures 9 to 13 show tests for causal relationships between median home prices for Marin County and each of San Francisco, Alameda, Contra Costa, San Mateo, Sonoma, and Napa counties. Using data from Figures 7 and 8, Figures 9 through 13 tell their own stories, but a simple way to consider what each figure's results imply:

- Almost all the results are statistically significant, and the hypothesis that one county causes the other happens with less than 5 percent error, suggesting bi-directional causality;
- A gap between each line suggests one county has more causality than the other, even though there is bi-directional causality;
- Regional housing markets are going to eliminate a lot of the possibility for one place specifically to influence land prices regionally without some feedback; and
- Marin County, while a smaller county versus the larger, urban areas, still has some power in driving regional land prices.

Figure 7: Causality Test results for Marin County
How Marin County Drives Core Bay Area Counties' Home Prices



Sources: Zillow Research and Author's Calculations

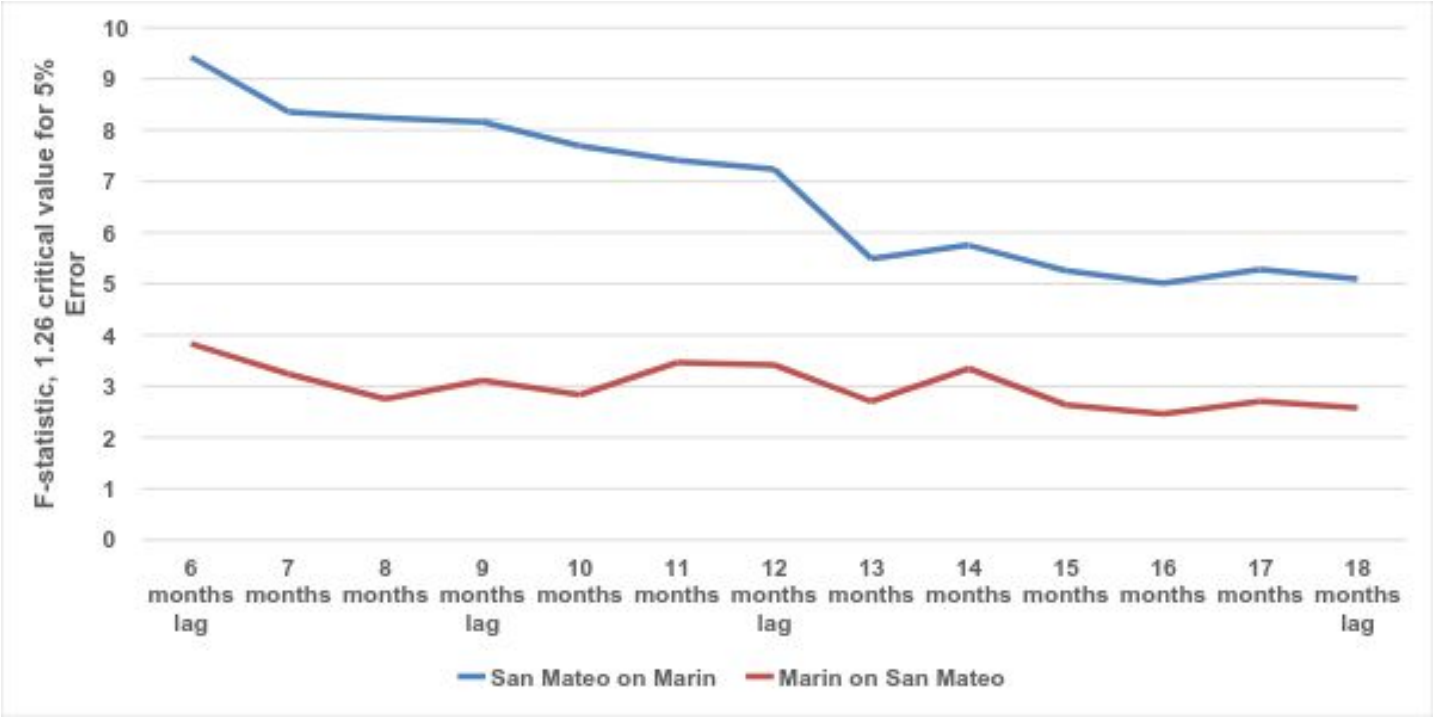
Figure 8: Causality Test results for Marin County
How Marin County Home Prices Affect and are Affected by North Bay Counties

Sources: Zillow Research and Author's Calculations

Figure 9: Marin and San Francisco Causal Relationships, 6 to 18 month lags

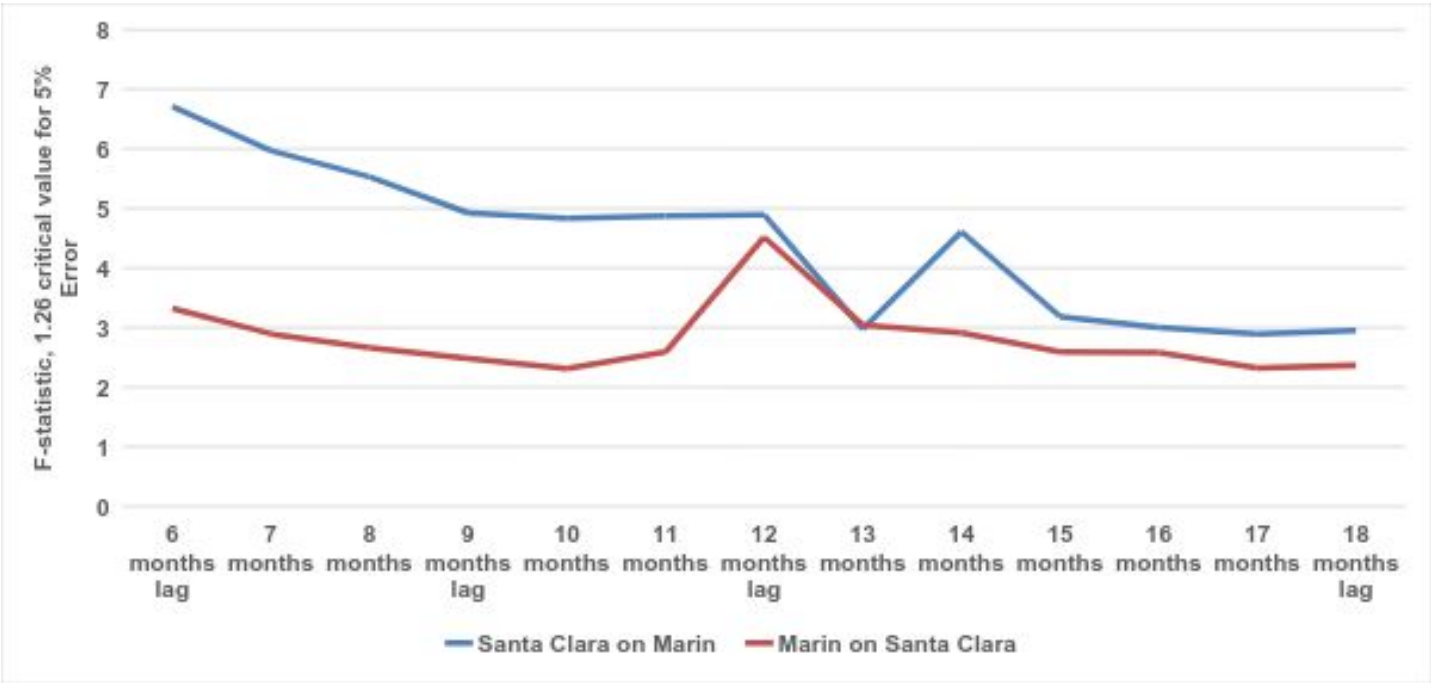
Sources: Zillow Research and Author's Calculations

Figure 10: Marin and San Mateo Causal Relationships, 6 to 18 month lags



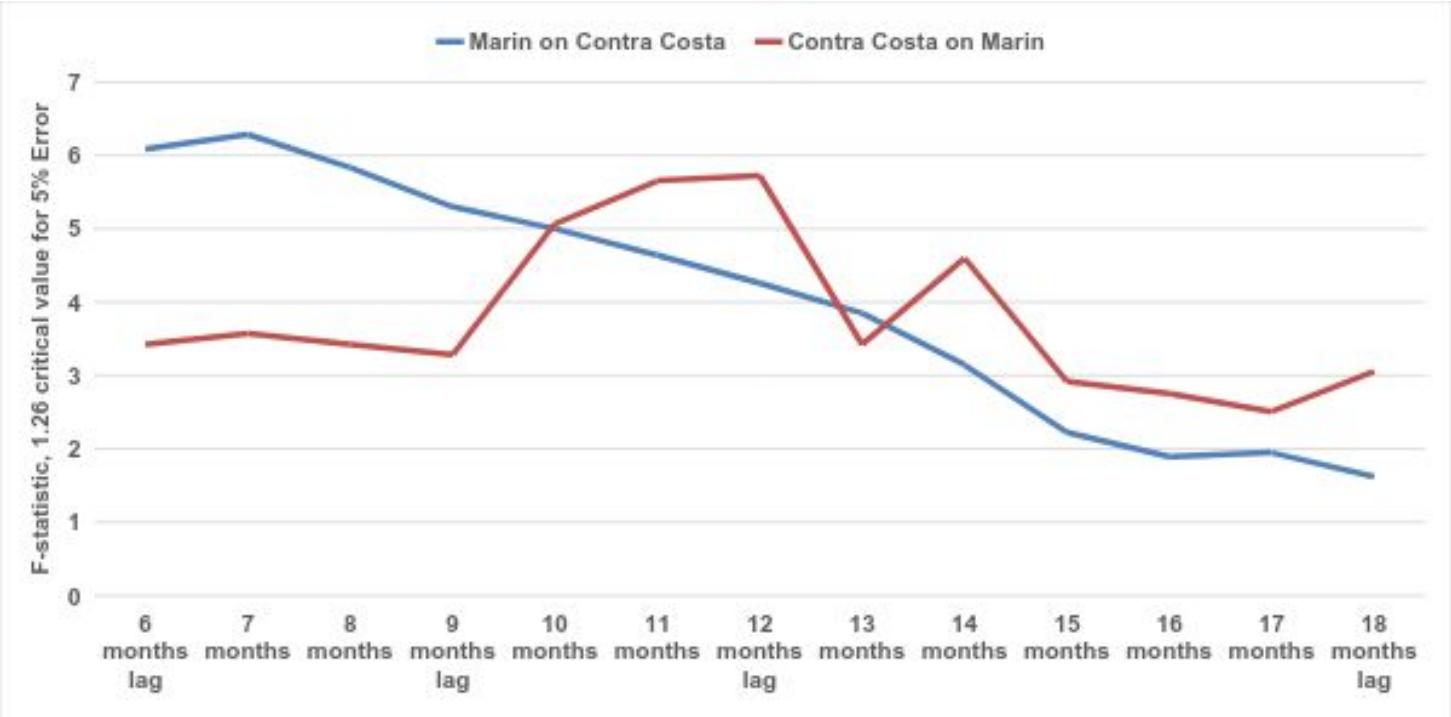
Sources: Zillow Research and Author's Calculations

Figure 11: Marin and Santa Clara Causal Relationships, 6 to 18 month lags



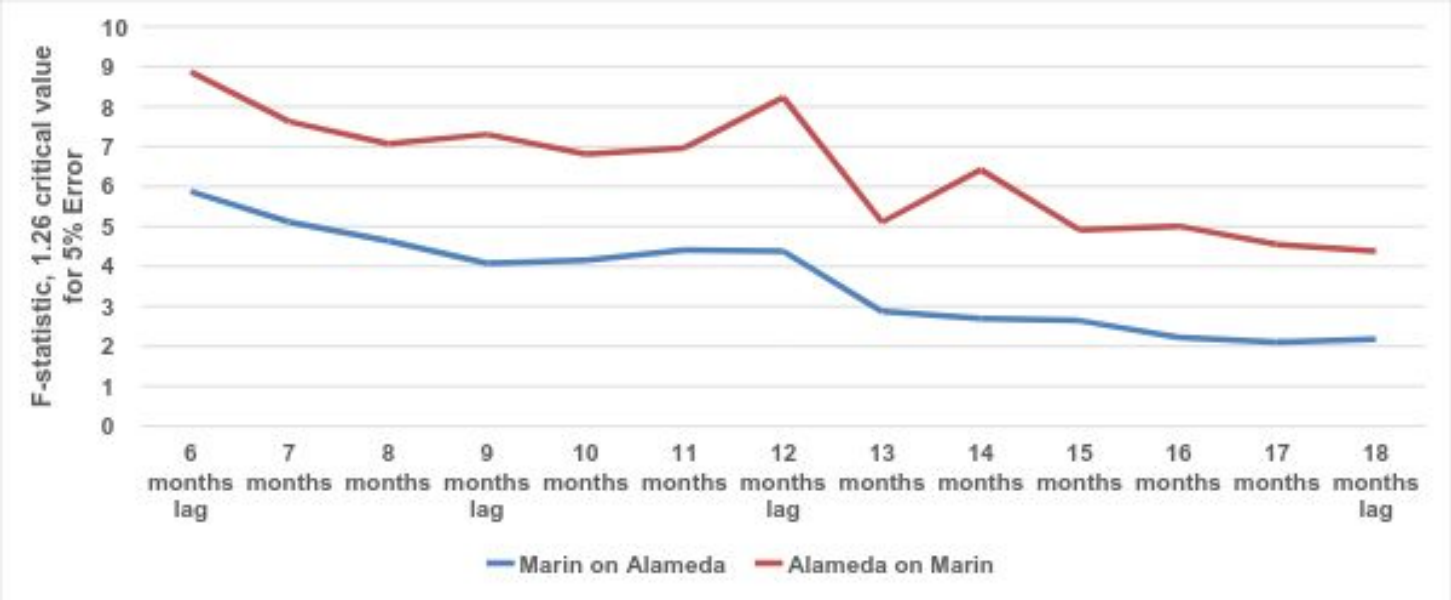
Sources: Zillow Research and Author's Calculations

Figure 12: Marin and Contra Costa Causal Relationships, 6 to 18 month lags



Sources: Zillow Research and Author's Calculations

Figure 13: Marin and Alameda Causal Relationships, 6 to 18 month lags



Sources: Zillow Research and Author's Calculations

Summary of Results

The f-statistic provides the statistical test for how these variables are related to each other over time. Like the R^2 statistic, the f-statistic is a goodness of fit test, or how well the choice of lagged data fit the model and ultimately help explain the variability of one county explaining the other county's home prices when using older data. In cases where the f-statistic is larger, the county that is seen as the "independent" variable using older values to explain current values of another county is said to Granger cause changes in home prices predictively.

For example, in Figure 13, lagged values of both Alameda and Marin counties explain current values of each other: there is bi-directional causality. Because Alameda's model has larger values of the f-stat, Alameda County has relatively more predictive causality on Marin County than the reverse. When Alameda County home prices change in July, we can infer that Marin County's home prices are likely to change the following July from that impulse more than when Marin County's prices change and affect Alameda County.

Looking at how Marin County is related to Sonoma and Napa counties versus the core Bay Area, the basic hypothesis of this study seems to be supported.

Similar Comparisons within Marin County: Larger Cities

Comparisons on median home price within Marin County can help understand price movements and connections rolling out from the Bay Area for four cities in Marin County: Sausalito, Mill Valley, San Rafael, and Novato. Figures 14 and 15 provide the same data as the countywide comparisons of the last section. Notice that the correlations are extremely high, with 100 percent being the maximum (basically suggesting the data are the same). Notice also a slight fade as we move from south to north, but a relatively higher (though slightly increased) level of correlation between Sausalito and Novato.

Figure 14: Correlations between Marin County Cities, 1996 to 2017

	Sausalito	Mill Valley	Novato	San Rafael
Sausalito	100.0%	98.4%	95.6%	97.7%
Mill Valley		100.0%	92.2%	96.5%
Novato			100.0%	97.6%
San Rafael				100.0%

Sources: Zillow Research and Author's Calculations

Figure 15: Coefficients of Determination between Marin County Cities, 1996 to 2017

	Sausalito	Mill Valley	Novato	San Rafael
Sausalito	100.0%	96.8%	91.4%	95.5%
Mill Valley		100.0%	85.0%	93.0%
Novato			100.0%	95.3%
San Rafael				100.0%

Sources: Zillow Research and Author's Calculations

The coefficients of determination in Figure 15 shows the variation in the Novato median home prices over time is 95.3 percent explained by variation in Sausalito home prices, but only 85.0 percent with Mill Valley. These comparisons, as discussed above, are only as good as the lagged variables in providing something closer to a causal relationship. What these figures tell is that the real estate markets in these cities are highly integrated as expected. The causal analysis above can be repeated for Marin County's cities. Figure 16 provides comparisons for the four cities described in the correlations and coefficient of determination figures in the last section.

Some patterns emerge that follow anecdotal evidence from the Marin County real estate industry. First, Sausalito home prices seem to have a larger effect on San Rafael and Novato home prices as time goes on. Notice as the number of lags (previous time periods) rise, Sausalito effect's on the other cities rise. Mill Valley and Novato have a similar relationship, but Mill Valley and San Rafael do not. Neither Novato nor San Rafael show much impact going forward, but there is bi-directional causality (where each feed off each other) as time goes on in almost every case. San Rafael and Mill Valley seem to have little effect on each other over time, which suggests that both are sub markets of other places that influence San Rafael and Mill Valley prices.

Figure 16: Comparison of Granger Results in Cities, Marin County

Sources: Zillow Research and Author's Calculations

Conclusions

This study shows comparisons of home prices as a proxy for land prices in the Bay Area counties. The focus is on how Marin County compares and follows core Bay Area (Contra Costa, San Francisco, San Mateo, Alameda, Santa Clara) versus the North Bay counties (Sonoma, Napa). The initial hypothesis is that Marin County is more like the core Bay Area counties in the behavior of land prices (where home prices act as a proxy) than Sonoma and Napa counties. The data and results help confirm this hypothesis.

As the data in this study show, Marin County's home prices are highly correlated with surrounding areas. For example, prices in San Francisco from 6 to 18 months ago may be statistically significant in their effects on Marin County prices today, but older values of Marin County may not be statistically significant in their effects on San Francisco prices today. Such a result is a simple test of causality and its direction for related data.

The results in this study show that Marin County prices are driven more by the core Bay area than North Bay counties. Such a result follows anecdotal evidence and basic logic. The summary results below show what counties have a causal effect on Marin County, counties that Marin County may influence alone, and where no or bi-directional causality may exist.

- **Marin County's median home price data moves more with core Bay Area home prices versus North Bay counties;**
  - **Smaller counties have smaller effects on their surrounding areas; and**
  - **These markets are regional.**
- **The correlation levels further show the regional aspects of these markets;**
  - **The causality analysis shows which counties have more effects on others; but**
  - **There is also a lot of bi-directional or feedback causality, which means the effects are limited.**

In short, this study's hypothesis seems to be statistically true. This is not proving anything. The data and statistical tests simply support the idea that Marin County land prices are driven by the Bay Area's urban areas more than there more suburban and rural areas (where Napa and Sonoma counties are examples of such areas).

Appendix: Methodology

Correlation Coefficient

The Pearson correlation coefficient (what is sometimes called "r") measures the linear correlation between two variables X and Y. It has a value between −1 and +1 and is a percentage of how much the mean or average of one data set moves with another, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. More formally, it is the ratio of the covariance (COV) of X and Y to each variable's standard deviation (Stdev).

$$r = COV(X, Y)/(Stdev(X) \times Stdev(Y))$$

Its importance for this study is to show the simple, linear relationships or movements of trends in each pairing of Marin County's real estate price trends with the other counties in this group. The second, important aspect is that the variability be explained in the data by a specific set of variables, if possible. The basis of regression analysis, or fitting a straight-line relationship to data that is otherwise non-linear, is the coefficient of determination.

Coefficient of Determination

While this statistic is simply the correlation coefficient multiplied by itself, literally R^2, it has a more significant idea behind it. The square of each standard deviation is the variables' variances. What R^2 measures is how much of one variables variability is measured by the other's variable. Another way of thinking about R^2 is measuring "goodness of fit" or how the choice of X "fits" as a way to explain Y independent of the movements of Y.

For example, two data sets X and Y have an average and a variance. The variability of X may help explain a large percentage in the variability of Y, some of it or very little. Like r, it is a percentage, but cannot be negative. R^2 lies between 0 and 1. When R^2 is equal to 1, how X varies around its average completely explains the variation in Y around its mean; if R^2 is close to zero, the variation of X does not explain any of the variation in Y. What that implies is that there are other variables that help explain its variation, including variation in Y explaining itself (a phenomenon called "serial correlation"). Causality tests can help tell us how variables that are independent of a variable Y help explain its variability. In the case of two variables, X and Y, the relationship is circular: high R^2 values need something more to determine if how X and Y explain each other is more about Y explaining X or the opposite if logic alone does not provide a simple answer.

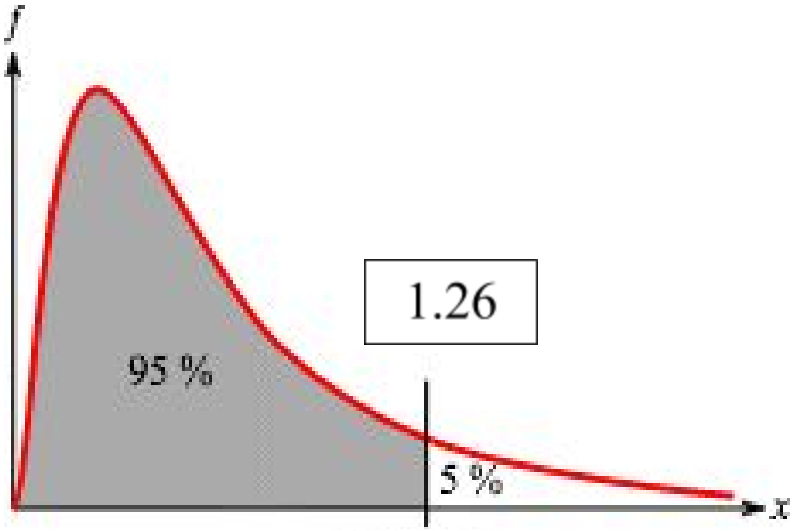
Granger Causality: Simple Form with Two Variables

The Granger causality test helps determine whether one data series is useful in forecasting another. Since the question of "true causality" is difficult to show in practice, statisticians assert that the Granger test finds only "predictive causality". A time series X is said to Granger-cause Y if it can be shown, using lagged

values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y.

The F-test is related to R^2, which connects the two in our study. Basically, when the F-stat (shown in the figures comparing "causal" relationships above) is larger, the more likely variable X in its older or "lagged" values drives current values of Y. Bi-directional causality happens when X and Y both have large F-stat values when compared to each other (results we see often in this study). Figure 17 shows the basic idea of the F-test, where the f-stat (the number in the figures above that represent the ratio of the variances of Y to the variance of X when there are lagged values), tests of the statistical validity of the basic hypothesis: there is no causal connection. When the f-stat is greater than 1.26, given the large amount of degrees of freedom, the data above easily exceed that level to accept the alternative hypothesis that there is predictive causality between each twosome of Marin County home prices and other county home prices or comparisons between Marin County's major cities with 5 percent error.

Figure 17: Basic Idea of the F-test using the f-statistic.



References

California Department of Finance (2017). "Inflation", Accessed at http://www.dof.ca.gov/Forecasting/Economics/Indicators/Inflation/

Census Bureau (2017). "Longitudinal Employment and Household Database", Accessed at: http://onthemap.ces.census.gov

Zillow Research (2017). "Zillow Data" Accessed at https://www.zillow.com/research/data/