California Wildfires & Insurance: Calculating the amount of money that the government of California will need to pay between 2022 and 2025 inclusive in order to provide minimal property insurance coverage for wildfires

Pages: 25[1]

[1] 2 pages of reference included in page number; Appendix with data is after reference page

Table of Contents:

Introduction and Rationale

Finance and accounting have always been topics that have fascinated me. In addition, I have spent a lot of time learning about science subjects such as Biology and Earth Science. I had always wanted to find a way to combine these two passions of mine, and about 2 years ago, while I was preparing for a debate topic, I found the perfect way to do so. That was when I discovered the field of fiscal green policy.

Of course, there is usually a large discrepancy between the optimal policies targeted towards sustainability and those that are actually implemented. A key reason for this is the amount of money necessary to implement such policies. Thus, I decided that I would like to calculate the exact amount necessary to implement one such policy.

After coming to this decision, I needed to choose an environmental issue and its corresponding governmental policy. While there are a variety of environmental issues, I decided that I wanted to focus on the California wildfires. There were a few main reasons for this decision. First, I knew that the increase in both the severity and frequency of natural disasters is a huge environmental issue today. A recent heat wave in the area that I lived had made this issue very apparent. Second, I knew that there was a lot of data available surrounding the California wildfires as opposed to other natural disasters. This would make my research a lot easier. Lastly, after some research, I discovered California had already enacted an insurance plan, FAIR, that worked to help those affected by wildfires. This would be the perfect policy to do my calculations on as even though I knew that this insurance plan would be unable to help everyone who needed insurance, I still wanted to know the expenses of this insurance policy in an ideal world.

Overall, my exploration would use historical data to model both the cost of the California wildfires in property damages and the current amount of damages from said wildfires already covered by insurance to calculate the amount of extra funds necessary to achieve minimum insurance coverage for all property damages from wildfires.

Aim and Hypothesis

As established previously, the aim of this exploration will be to calculate the amount of money that the government of California will need to pay between 2022 and 2025 in order to provide minimal property insurance coverage for wildfires. I chose the timeframe between 2022 and 2025 because I knew that extrapolation, or predicting values outside of a data set, would become more and more inaccurate the further my prediction values were from my collected data. Thus, to minimize error, I choose a relatively narrow timeframe that started as close as possible to the years that I had data for.

I decided to achieve this aim through two models. First I would create a model for the property damage costs of California wildfires from past data. Then, I would create a model for the amount of insurance currently paid out for the wildfires, again using past data. Both of these models would be in relation to time, specifically the number of years after a certain date.

The data I used would need to be data collected by others, as I myself had no way to obtain the information I needed. Preferably, the data would be from the state government of California in order to maximize the accuracy and minimize the chance of error.

From my knowledge of natural disasters based on prior research, I expected that the property damage costs of the California wildfires would follow some sort of exponential curve. Since e is the most common base within the real world, the relationship would most likely be defined by the function $D(t_1) = A_1 e^{B_1 t_1} + C_1$ where $D(t_1)$ would be the damage cost for that year,

t_1 would be the number of years elapsed since a certain date, and A_1, B_1, and C_1 would be constants.

For the model for the amount of insurance currently paid, I expected the curve to follow the one for property damage costs as it is logical that the higher the damage costs, the greater the amount of insurance paid. Thus, the relationship would also be defined by an exponential function with base e. Therefore, we can write this function as $I(t_2) = A_2 e^{B_2 t_2} + C_2$ where $I(t_2)$ would be the insurance paid for that year, t would be the number of years elapsed since a certain date, and A_2, B_2, and C_2 would be constants.

In order to successfully compare the two models, the timeframe for each would need to be the same. Thus, t_1 would actually equal t_2 and the two functions would actually be $D(t) = A_1 e^{B_1 t} + C_1$ and $I(t) = A_1 e^{B_1 t} + C_1$.

To test my hypothesis, I collected data from 2010 to 2017 as this was the range of years for which data for the cost of property damages of wildfires and data for amount paid by insurance could both be found.

<u>Data Collection</u>

The data I used came from databases provided by the California government. A key issue to note is that when compiling data, I excluded the data from the Butte Fire in 2015 as these fires were significantly more impactful than the average California wildfire. Had I not subtracted them, they would have skewed the data greatly as fires on such a large scale are extremely uncommon, even after accounting for climate change.

Wildfire Data

The wildfire data I used came from the California Redbooks, which are compilations of statistics related to the California wildfires. Since the Redbooks are organized by year, to compile my data, I flipped through the Redbooks for the time period from 2010 to 2017 and copied down the total property damages into a spreadsheet. I then divided each damage cost by 1,000,000 to obtain the damage costs in millions of dollars so that the numbers would be easier to work with. The final data is shown below[2].

Year	Damage Costs ($)	Damage Costs (Millions of $)
2010	3,397,442	3.397442
2011	7,222,651	7.222651
2012	28,213,200	28.213200
2013	29,799,753	29.799753
2014	20,034,168	20.034168
2015	1,561,836,666	1561.836666
2016	148,266,893	148.266893
2017	2,135,125,902	2135.125902

Insurance Data

The insurance data that I used came directly from the California Department of Insurance. The department had split insurance data between insurance for rental property and for owner occupied property. Thus, to get the total amount of insurance premiums written, I summed up the two amounts. I then divided the numbers by 1,000,000,000 so that the premiums were expressed in billions of dollars in order to make the numbers smaller and easier to work with.

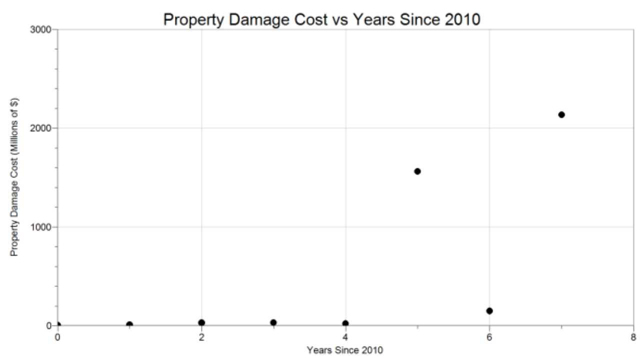
The chart below shows the final edited data[3].

[2] For full data, view appendix
[3] For full data, view appendix

Year	Written Premium ($)	Written Premium (Billions of $)
2010	1,183,826,611	1.183826611
2011	1,205,065,924	1.205065924
2012	1,199,836,085	1.199836085
2013	1,164,746,872	1.164746872
2014	1,175,207,546	1.175207546
2015	1,156,563,715	1.156563715
2016	1,158,130,056	1.158130056
2017	1,044,438,632	1.044438632

<u>Modeling: Property Damage Costs</u>

Plotting the points from the data for property damage costs, we get the following graph.



Immediately, we notice that the data point for 2016 is an outlier. Thus, when determining a function to model the graph, particular care needs to be taken to ensure that the outlier does not overly skew the model.

Exponential Model

From our hypothesis, we predicted that the function for the property damage costs would follow an exponential relationship with a function of $D(t) = A_1 e^{B_1 t} + C_1$.

Looking at our data and plotted points, a few key things can be determined:

1. **$B_1 t$ is positive.**

   If $B_1 t$ was negative it would reflect the graph of e^x over the y-axis resulting in $y \to 0$ as $x \to \infty$. Clearly, this is not the case as in this graph, the y-value increases as the x-value increases.

2. **A_1 is positive.**

   If A_1 was negative, it would reflect the graph of e^x over the x-axis resulting in $y \to -\infty$ as $x \to \infty$. Again, this not true since as previously stated, the y-value increases as the x-value increases in this graph.

3. **The y-intercept is (0, 3.397422).**

   Looking at our table, we see that the total property damages in millions of dollars in 2010 was 3.397422. Since 2010 is 0 years after 2010, this point falls on the y-axis of our graph.

To find our constants, we can substitute points into the graph. As previously noted, the year 2016 was an outlier, and so we want to avoid using that specific point. Additionally, we want to choose points that are far away from each other so that function represents the entire graph and not just a small subsection. Since we will need a total of 3 points to solve for 3 unknowns, it would be best to choose the data points where $x = 0, 3,$ and 7.

We choose to first use our y-intercept because the x-value of this point is 0, which makes solving much easier. Thus, we get

$$3.397422 = A_1 e^{B_1(0)} + C_1$$

Which means that

$$A_1 = C_1 - 3.397422$$

Plugging this back into our original equation, we get that

$$D(t) = (C_1 - 3.397422)e^{B_1 t} + C_1$$

Next, we plug in the data point (3, 29.799753) into the above equation to get

$$29.799753 = (C_1 - 3.397422)e^{B_1(3)} + C_1$$

Thus,

$$C_1 = \frac{-29.799753 - 3.397422e^{3B_1}}{-1 - e^{3B_1}}$$

And the original equation becomes

$$D(t) = (\frac{-29.799753 - 3.397422e^{3B_1}}{-1-e^{3B_1}} - 3.397422)e^{B_1 t} + \frac{-29.799753 - 3.397422e^{3B_1}}{-1-e^{3B_1}})$$

Lastly, we plug in the data point (7, 2135.125902) into our equation to get

$$2135.125902 = (\frac{-29.799753 - 3.397422e^{3B_1}}{-1-e^{3B_1}} - 3.397422)e^{B_1(7)} + \frac{-29.799753 - 3.397422e^{3B_1}}{-1-e^{3B_1}})$$

Which evaluates to

$$B_1 \approx 1.10658$$

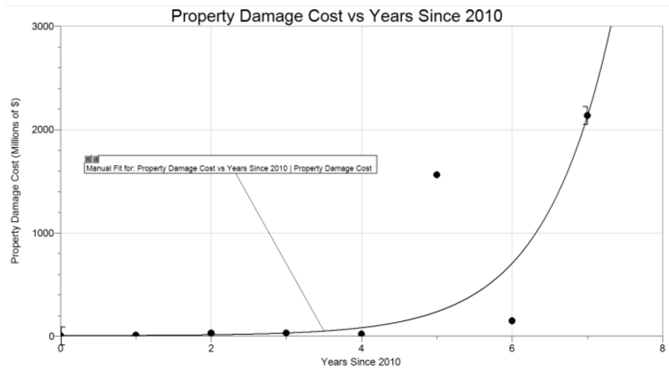
$$C_1 \approx 4.31887$$

$$A_1 \approx 0.921448$$

Substituting these values into the function gives us

$$D(t) \approx 0.921448e^{1.10658t} + 4.31887$$

Plotting this function provides the following graph



Solely from looking at these points, we notice that the function is not an entirely accurate representation of the data, but it does follow the trend of the data points as well as our predicted end behavior. Therefore, it is necessary to look at other possible functions before deciding which function is most representative of the data.

Linear Model

Next, we consider the linear model. Although the shape of the graph does not look linear, the sum of squared residuals for this model can be used as a base for error comparison. If the data were to be represented by a linear function, the function would be $I(t) = At^2 + B$. A would be positive as the y-values of the data points decrease as their x-values increase. The value of B would be 3.397442. Thus, the function would be

$$I(t) = At + 3.397442$$

Plugging in the point (7, 2135.125902), which will provide the most general trend, we get

$$2135.125902 = A(7) + \ 3.397442$$

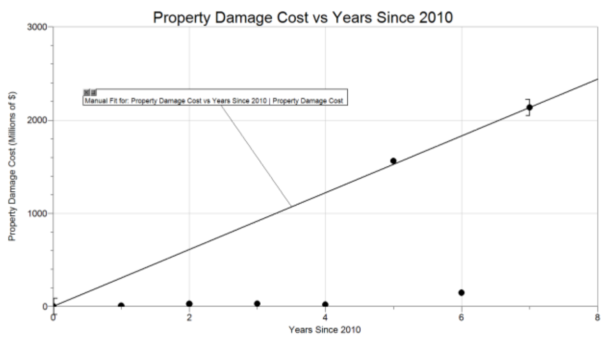
And so

$$A \approx 304.5326371$$

And our function is

$$I(t) = 304.5326371t + 3.397442$$

Graphing the function, we get



Clearly, a linear function is not the best way to model the property damage.

Quadratic Model

If the function for the property damage costs follows a quadratic relationship, the relationship would follow a function of $D(t) = At^2 + Bt + C$.

Again, we notice a few things by looking at our data and plotted points:

1. **A is positive.**

   If A was negative the graph would be concave rather than convex and thus as $x \to \infty$, $y \to -\infty$. That would not fit the end behavior seen by our data points.

2. **The y-intercept at (0, 3.397422) is the minimum of the function.**

   Since our graph is convex rather than concave, we know that the vertex must be the minimum of the function. We assume that (0, 3.397422) is the vertex as out of the all the points we have, it is the minimum value.

To find the value of C within our function, we simply need to plug in our y-intercept. Hence,

$$3.397422 = A(0)^2 + B(0) + C$$

And

$$C = 3.397422$$

Next, to find the value of B, we consider the $D'(t)$. Taking the derivative of $D(t)$, we find that $D'(t) = 2At + B$. Since (0, 3.397422) is the minimum, the derivative at that point will be equal to 0. Thus, we get

$$0 = 2A(0) + B$$

And

$$B = 0$$

Therefore, our function is now

$$D(t) = At^2 + 3.397422$$

Lastly, to find A, we can plug in another point from a dataset. For the same reasons discussed when finding the exponential model, we want to choose a point as far away from the y-intercept as possible. Hence, we plug in the point (7, 2135.125902) to get

$$2135.125902 = A(7)^2 + 3.397422$$

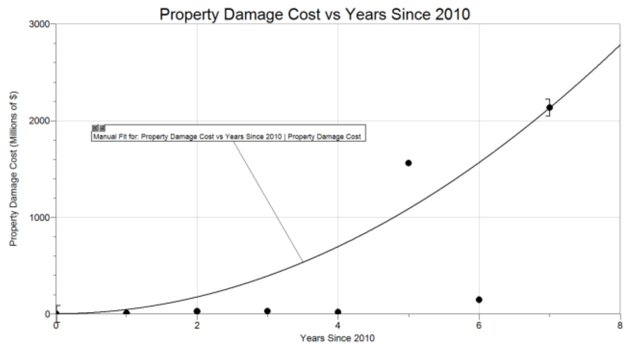
And

$$A \approx 43.5047$$

This means that our overall equation is

$$D(t) \approx 43.5047t^2 + 3.397422$$

Plotting this function gives us the following graph



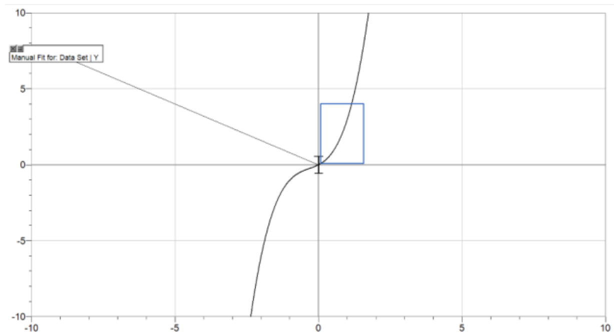
Compared to the exponential function, the quadratic function is less representative of earlier data points. On the other hand, the quadratic function is more representative of later data points as it is highly unlikely that the property damage cost will increase at the rate projected by the exponential function. Next, we test a cubic function. Doing so allows us to compare the sum of squared residuals of polynomial functions with increasing degrees and thus, determine if increasing the degree of the model will lead to a significant change in model accuracy.

Cubic Model

If the function for the property damage costs follows a cubic relationship, the relationship would follow a function of $D(t) = At^3 + Bt^2 + Ct + D$.

By looking at our data and plotted points:

1. **The points are best represented by the section of a general cubic function boxed below.**



2. **A is positive.**

   If A was negative the boxed section of the graph would be concave rather than convex and thus as $x \to \infty,\ y \to -\infty$. That would not fit the end behavior seen by our data points.

3. **The y-intercept at (0, 3.397422) is the turning point of the function.**

   If the turning point were at a point such that the x-value was less than 0, the graph would have a very steep slope in the first quadrant, which is not representative of our data. On the other hand, if the turning point were at a point such that the x-value was greater than 0, there would be a negative amount of property damage in 2010, which is not possible. Overall, this means that $D''(0) = 0$ at this point. Taking the derivative of $D(t)$ twice, we get that

$$D'(t) = 3At^2 + 2Bt + C$$

   And

$$D''(t) = 6At + 2B.$$

   Thus,

$$D''(0) = 6A(0) + 2B = 0$$

   Which means that

$$B = 0$$

   And

$$D(t) = At^3 + Ct + D$$

In addition, we know that $D(0) = 3.397422$ which means that $D = 3.397422$.

4. **The function $D(t)$ is best represented as $D(t) = At^3 + 3.397422$.**

   We notice that within our data points, the minimum occurs when $x = 0$. However, it is highly likely that this is not actually a local minimum as the property damage cost of years before 2010 should show a decreasing tend. Thus, there is no local minimum overall. Similarly, we notice that the maximum within our data points occurs when $x = 7$. Yet, this is likely not a maximum as property damage costs of years after 2017 should show an increasing trend. Thus, there is no local maximum within our graph either. The only cubic function that satisfies these requirements is one that follows the equation $f(x) = Ax^3 + B$.

To find the value of A within our function, we simply need to plug in a value. We choose a value as far away from our y-intercept as possible in order to represent the entire trend. Thus, we plug in (7, 2135.125902) to get

$$2135.125902 = A(7)^3 + 3.397422$$

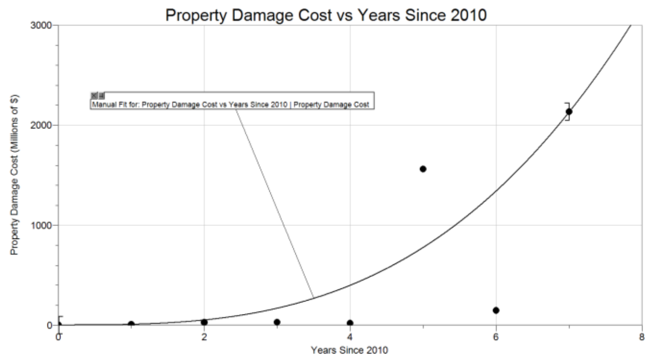
Which means that

$$A \approx 6.214951778$$

This means that our overall equation is

$$D(t) \approx 6.214951778t^3 + 3.397422$$

Plotting this function gives us the following graph



Compared to the quadratic function, the cubic function seems to be more representative of the smaller x-values. However, to compare the accuracy, non-linear regression will need to be calculated.

Non-Linear Regression

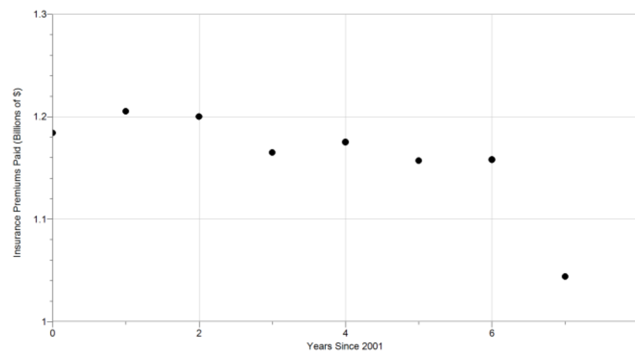
To compare the accuracies of each model, we turn to regression. We use squared residuals because residuals themselves have different signs and adding them would thus mitigate the amount of error. The chart below shows the values.

		Exponential Model			Linear Model			Quadratic Model			Cubic Model		
x-value	Observed Value	Predicted Value	r	r^2	Expected Value	r	r^2	Expected Value	r	r^2	Expected Value	r	r^2
0	3.397442	5.240318	-1.842876	3.396192	3.397442	0	0	3.397422	0.00002	0	3.397422	0.00002	0
1	7.222651	7.0153275	0.207324	0.042983	307.93008	-300.707	90424.96	46.902122	-39.6795	1574.46	9.6123738	-2.38972	5.710775
2	28.2132	12.745114	15.46809	239.2617	612.46272	-584.25	341347.5	177.41622	-149.203	22261.54	53.117036	-24.9038	620.201
3	29.799753	29.799819	-0.000066	0	916.99535	-887.196	787116	394.93972	-365.14	133327.2	171.20112	-141.401	19994.35
4	20.034168	81.373224	-61.33906	3762.48	1221.528	-1201.49	1443587	699.47262	-679.438	461636.6	401.15434	-381.12	145252.6
5	1561.836666	237.33113	1324.506	1754315	1526.0606	35.77607	1279.927	1091.0149	470.8218	221673.1	780.26639	781.5703	610852.1
6	148.266893	708.94762	-560.6807	314362.9	1830.5933	-1682.33	2830222	1569.5666	-1421.3	2020093	1345.827	-1197.56	1434150
7	2135.125902	2135.1152	0.010702	0.000115	2135.1259	0.000002	0	2135.1277	-0.0018	0.000003	2135.1259	0	0
				2072683			5493978			2860566			2210875

We notice that the exponential model is most representative of our data as it has the smallest r^2 value. In addition, we notice that as the degree of the polynomial function increases, the model becomes more and more accurate. However, the increase in accuracy gets smaller and smaller with each degree increase, meaning that a cubic model is already a very accurate representation of our data points. Overall, we prefer the cubic model representation as our function as even though the exponential model has a smaller r^2, we notice that the model's rate of change in the later values is extremely high. It is unlikely that such a rate of change actually occurs and may lead to an overly high prediction for the property damage costs.

<u>Modeling: Insurance Premiums</u>

Plotting the points from the data for property damage costs, we get the following graph.



The first thing we notice is that our hypothesis that the graph for insurance premiums would follow a curve similar to that of property damage costs is incorrect. Although an exponential function might still be the best model for these data points, the function itself will certainly be very different from the one we created for property damage costs.

Exponential Model

As established earlier, if the model for currently paid insurance premiums follows an exponential relationship, the function would be represented by the equation $I(t) = A_2 e^{B_2 t} + C_2$.

We follow a process similar to the one we took to model the cost of property damages. Looking at our data and plotted points, we determine that:

1. **$B_2 t$ is positive.**

If $B_1 t$ was negative the graph would still be convex, which is not representative of the datapoints that we have.

2. **A_2 is negative.**

   If A_1 was positive, $y \to \infty$ as $x \to \infty$. Our data points show that $y \to -\infty$ as $x \to \infty$. Thus, we need to reflect the graph of e^x over the x-axis by multiplying A_2 by -1.

3. **The y-intercept is (0, 1.183826611).**

   Looking at our table, we see that the amount paid by insurance in billions of dollars in 2010 (our $x = 0$) was 1.183826611.

To find our constants, we repeat the same process as we did earlier with the damage costs. This time, we choose the data points where $x = 0, 4,$ and 7 since the data point for $x = 3$ represents a small outlier/dip in the graph.

First plugging in our y-intercept and solving for A_2 we get

$$A_2 = C_2 - 1.183826611$$

Hence our original equation is

$$I(t) = (C_2 - 1.183826611)e^{B_2 t} + C_2$$

Next, plugging the data point (4, 1.175207546) and solving for C_2 we get

$$C_2 = \frac{-1.175207546 - \ 1.183826611e^{3B_2}}{-1 - e^{3B_2}}$$

And the original equation becomes

$$I(t) = (\frac{-1.175207546 - 1.183826611e^{3B_2}}{-1-e^{3B_2}} - 1.183826611)e^{B_2 t} + \frac{-1.175207546 - 1.183826611e^{3B_2}}{-1-e^{3B_2}})$$

Finally, we plug in the data point (7, 1.044438632) into our equation and solving for B_2 we get

$$B_2 \approx .721402$$

Which means that

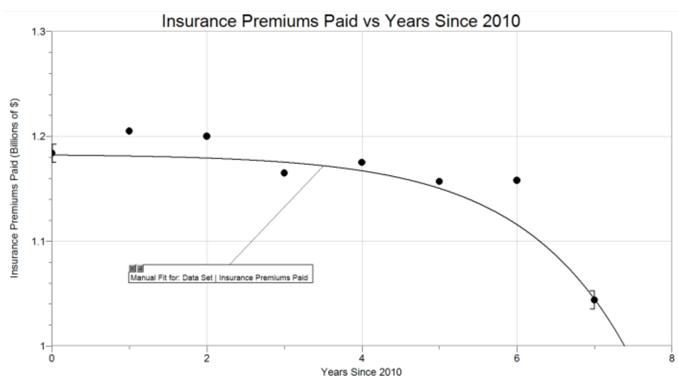
$$C_2 \approx 1.18294$$

$$A_2 \approx -.000887$$

Thus, our function is

$$I(t) \approx -.000887e^{.721402t} + 1.18294$$

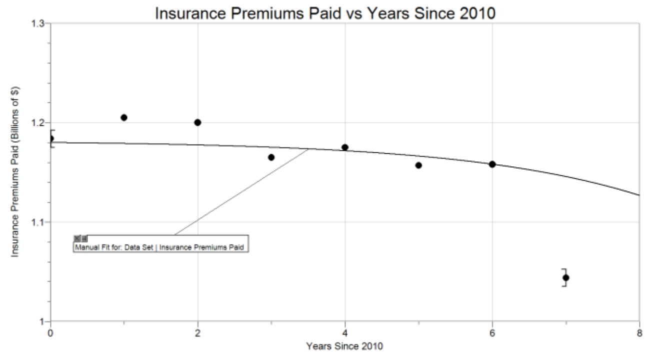
And we get the following graph



Although this seems like a pretty good model for our data, we that the x-intercept of our function is 2018, meaning that according to our model, no insurance was paid out in 2018, which is not true. Therefore, we conclude that either $t = 7$ is probably an outlier. Hence, we need to repeat the same process we performed above, but this time, we use $t = 6$ as our final point instead. We get the following function

$$I(t) \approx -.001897e^{.421462t} + 1.18193$$

And graph



This graph is a much more reasonable representation of our data set as even though with enough extrapolation the amount paid will still reach zero, the time when that happens will be outside our target time period of 2022-2025. However, we still need to consider other functions, mainly the polynomial functions. We can solve for the linear and quadratic functions and run a regression test as before to find the most accurate model.

Linear Model

Lastly, we consider the linear model. If the data were to be represented by a linear function, the function would be $I(t) = At^2 + B$. A would be negative as the y-values of the data points decrease as their x-values increase. The value of B would just be 1.183826611. Thus, the function would be

$$I(t) = At + 1.183826611$$

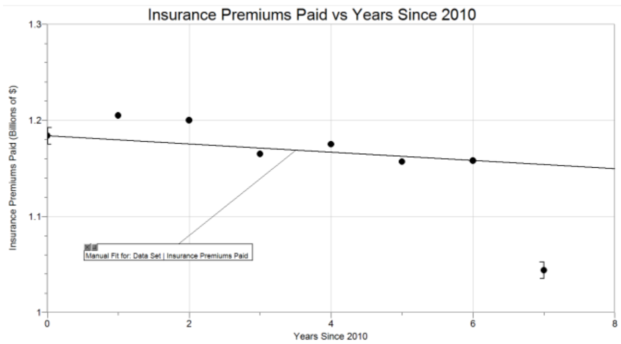
Plugging in the point (6, 1.158130056) and solving for A, we get

$$A \approx -0.00428276$$

And our function is

$$I(t) = -.00428276t + 1.183826611$$

Graphing the function, we get



Although a linear function appears to also be a very good way to model the amount of insurance paid, we need to compare it with other models using the sum of the squared residuals.

Quadratic Model

If the function for the insurance costs paid is quadratic relationship, the relationship would follow a function of $I(t) = At^2 + Bt + C$.

We make the following observations:

1. **A is negative.**

   If A was positive the graph would be convex rather than concave.

2. **The point at (1, 1.205065924) is the maximum of the function.**

   Since our graph is concave, the vertex must be the maximum of the function. We assume that (1, 1.205065924) is the vertex as it is the maximum value within our data points.

3. **The y-intercept is (0, 1.183826611).**

We repeat the same process we used to find the quadratic function for the damage costs. First, we plug in our y-intercept and solve for C to get

$$C = 1.183826611$$

Next, we solve for B in terms of A by using the derivative of $I(t)$ at the point (1, 1.205065924), $I'(t) = 0$ to get

$$B = -2A$$

Thus, our function is

$$I(t) = At^2 - 2At + 1.183826611$$

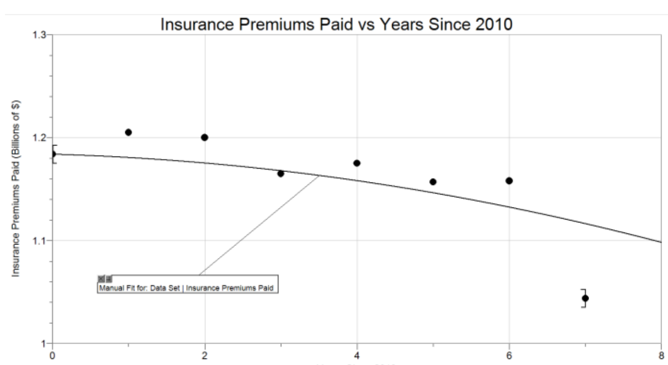
Since we cannot plug in the point where $t = 7$ as it is an outlier, we plug in the point (6, 1.158130056) and solve for A to get

$$A \approx -0.00107069$$

This means that our overall equation is

$$I(t) \approx -0.00107069t^2 - 0.00214138t + 1.183826611$$

With the following graph



Compared to the exponential model, the quadratic model fits more with the data including the outlier. On the other hand, the exponential model fits better with the data prior to the outlier. To make a better comparison, we turn to regression.

Non-Linear Regression

Solving for the sum of squared residuals, we get the following chart:

		Exponential Model			Linear Model			Quadratic Model		
x-value	Observed Value	Predicted Value	r	r^2	Expected Value	r	r^2	Expected Value	r	r^2
0	1.183827	1.180033	0.003794	0.000014	1.183827	0	0	1.183827	0.00002	0
1	1.205066	1.179039	0.026027	0.000677	1.179544	0.025522	0.000651	1.180615	0.024451	0.000598
2	1.199836	1.177523	0.022313	0.000498	1.175261	0.024575	0.000604	1.175261	0.024575	0.000604
3	1.164747	1.175213	-0.000066	0	1.170978	-0.00623	0.000039	1.167766	-0.00302	0.000009
4	1.175208	1.171692	0.003516	0.000012	1.166696	0.008512	0.000072	1.15813	0.017078	0.000292
5	1.156564	1.166325	-0.009761	0.000095	1.162413	-0.00585	0.000034	1.146352	0.010212	0.000104
6	1.15813	1.158145	-0.000015	0	1.15813	0	0	1.132433	0.025697	0.00066
				0.001296			0.0014			0.002267

Since our exponential model has the least r^2 value and shows a reasonable end behavior, we conclude that it is the best function to model our data points.

<u>Conclusion and Analysis</u>

To find the minimum amount that the California government needs to spend in order to ensure that all individuals are provided with some form of insurance given the current trends of both property damage costs and insurance costs paid, we can set the property coverage percentage to 50%. The minimum amount of insurance coverage is around 80% for most home insurance companies, but other the coverage for other property can range anywhere from 25% to 50%. Thus, 50% is a fair estimate for average total insurance coverage for wildfires, especially as California's FAIR insurance plan is mostly viewed as a last resort for those that cannot get insurance and does not fully cover all costs. The equation for the cost of insurance in millions of US dollars for expected property damages for all people will then be

$$D_{expected}(t) \approx (\frac{1}{2})(6.214951778t^3 + 3.397422)$$

$$D_{expected}(t) \approx 3.107475889t^3 + 1.698711$$

To find the increased amount of money that the California government needs to pay, we subtract the equation we got for insurance coverage already paid from the above equation. One thing we need to note is that we calculated insurance coverage already paid in billions of dollars while we calculated the damage costs in millions of dollars. Thus, we will need to multiply our function for insurance coverage by 1,000 when we subtract. Hence, we find that

$$C_{increased}(t) \approx\ 3.107475889t^3 + 1.698711 - 1000(-.001897e^{.421462t} + 1.18193)$$

The total amount in millions of US dollars that the State of California will need to pay between 2022 and 2025 inclusive can be found by

$$C_{increased}(12) +\ C_{increased}(13) + C_{increased}(14) + C_{increased}(15) \approx 28989.95861$$

Therefore, we can conclude that the amount of money that the government of California will need to pay between 2022 and 2025 inclusive in order to provide minimal property insurance coverage for wildfires is approximately $28989.95861 million or around $29 billion.

Limitations and Extensions

- There was a very small set of data points that I could test. Therefore, the model is not extremely accurate. If given more time, I could possibly file a data request form with the Government of California to gain more data.
- Extrapolation itself always has certain issues, and since 2025 is relatively far away from the last datapoint I had, 2017, there could be a very large margin of error. Again, this margin of error would be minimized by an increase in the amount of data.
- I mainly ignored outliers within the data. A possible extension is to model functions with and without the outliers and then compare the different models.

- Since there were only a limited number of models that I could test within this exploration, there may be models that fit better that I did not consider.
- For the percentage of damages that the insurance covered, I used 50%. This was an estimate based on an outside source, and so it is probably not completely accurate.

References

"2 California Wildfires Caused Nearly $2 Billion in Damage." *Southern California Public Radio*, 28 Apr. 2016, https://archive.kpcc.org/news/2015/10/15/55058/valley-and-butte-fires-caused-nearly-2-billion-in/.

"2015 Incident Archive." *Cal Fire Department of Forestry and Fire Protection*, https://www.fire.ca.gov/incidents/2015/.

Bishop, Lindsay. "How Much Homeowners Insurance Do I Need?" *ValuePenguin*, ValuePenguin, 11 Nov. 2021, https://www.valuepenguin.com/how-much-homeowners-insurance-do-i-need.

Dixon, Lloyd, et al. "The Impact of Changing Wildfire Risks on California's Residential Insurance Market." *California's Fourth Climate Change Assessment*, 2018, https://www.rand.org/content/dam/rand/pubs/external_publications/EP60000/EP67670/RAND_EP67670.pdf.

Helsel, Phil. "California Wildfire That Killed 22 in Wine Country Was Caused by Homeowner Equipment, Not PG&E." *NBCNews.com*, NBCUniversal News Group, 25 Jan. 2019, https://www.nbcnews.com/news/us-news/california-wildfire-killed-22-wine-country-was-caused-homeowner-equipment-n962521.

Metz, Jason. "What to Know about Wildfire Insurance." *Forbes*, Forbes Magazine, 7 Sept. 2021, https://www.forbes.com/advisor/homeowners-insurance/wildfires/.

Payne, Paul, and Mary Callahan. "Valley Fire's Heartbreaking Toll." *Santa Rosa Press Democrat*, 12 Sept. 2016, https://www.pressdemocrat.com/article/news/a-year-later-valley-fires-massive-toll-in-lake-county-means-long-difficu/?gallery=0578211B-C284-40BA-8068-B1828EC01F4D.

Schofield, David B. "Going from a Pure Premium to a Rate." *CASACT*, 1998, https://www.casact.org/sites/default/files/database/studynotes_scholfield5.pdf.

Srinivasan, Anushka, et al. "California Wildfires." *SOA*, 14 Apr. 2021, https://www.soa.org/globalassets/assets/files/static-pages/research/arch/2021/arch-2021-2-california-wildfires.pdf.

United States, Congress, California Department of Forestry and Fire Protection, and Michael Mohler. *CAL Fire News Release*, 15 May 2019. https://www.fire.ca.gov/media/5121/campfire_cause.pdf.

United States, Congress, CALIFORNIA DEPARTMENT OF INSURANCE. *Dwelling Fire (Owner Occupied) Data Premiums, Exposures and Annual Average Written Premiums by Experience Year 2001-2017*. http://www.insurance.ca.gov/0400-news/0200-studies-reports/0250-homeowners-study/upload/A-DF-Owner-Occupied_v1.pdf.

United States, Congress, CALIFORNIA DEPARTMENT OF INSURANCE. *Dwelling Fire (Rental Dwelling) Data Premiums, Exposures and Annual Average Written Premiums by Experience Year 2001 - 2017*. http://www.insurance.ca.gov/0400-news/0200-studies-reports/0250-homeowners-study/upload/A-DF-Rental-2_v1.pdf.

United States, Congress, The Department of Forestry and Fire Protection. *Stats and Events*. https://www.fire.ca.gov/stats-events/.

Appendix

Wildfire Data:

Year	Raw Damage Costs ($)	Damage Costs After Subtracting Butte Fire ($)
2010	3,397,442	3,397,442
2011	7,222,651	7,222,651
2012	28,213,200	28,213,200
2013	29,799,753	29,799,753
2014	20,034,168	20,034,168
2015	3,061,836,666	1,561,836,666
2016	148,266,893	148,266,893
2017	2,135,125,902	2,135,125,902

Insurance Data:

Year	Raw Written Premiums ($)	Written Premiums After Subtracting Butte Fire ($)
2010	1,183,826,611	1,183,826,611
2011	1,205,065,924	1,205,065,924
2012	1,199,836,085	1,199,836,085
2013	1,164,746,872	1,164,746,872
2014	1,175,207,546	1,175,207,546
2015	1,456,563,715	1,156,563,715
2016	1,158,130,056	1,158,130,056
2017	1,044,438,632	1,044,438,632