Mini-Project 2 – Interpolation and Curve Fitting

Zachary Moser

October 6, 2024

Abstract

In this project we will look at the Interpolation, linear curve fit, and non-linear curve fit of a set of data. We will find the interpolant which will connect a set of data so predictions can be made in between points. The linear curve fit and non-linear curve fit will not directly resemble the graph of data points making it less useful for interior points; however, useful for predicting future points. We can also find the R^2 value of the non-linear curve fit which tells us how accurately the curve fits the data.

1 Introduction

For this paper we will be looking at Housing prices in Philadelphia, PA, Dubois, PA, and Boston, MA from January 2000 through January 2024. We will interpolate, and find curve fits for each city. Interpolating involves finding a vandermonde matrix, solving for coefficients using python, and then plotting the graphs on a specific portion of the domain. For this project we will look at the change in housing prices in these cities from January of 2007 to June of 2009. This was a volatile time for the housing market because of the stock market crash. Next, we will find a linear curve fit for the data. To do this we will use the same vandermonde matrix as before; however, this time we will only go to the 3rd power which will give us a matrix with 4 columns. Finally, for the non-linear curve fit, we will define a cubic function to fit to our data. We will use python to optimize our fit and plot the data alongside the curve fit.

In the following sections we will look at the each step in further detail and interpret the graphs and what they tell us about the housing market.

2 Problem Statement

The goals for this project are to formulate interpolation and curve fitting problems into linear systems, solve these systems in python, use python to fit linear and nonlinear functions to data, and investigate the graphs.

Our first problem will be finding an interpolant for each set of data from January 2007 to June 2009 which we chose to look at because it will capture the effects that the stock market crash had on the housing market. As we can see in Figures : 1, 2, and 3 these functions are mostly decreasing along this time frame. For our linear curve fitting problem we will be looking at all data from January 2000 to 2024. We will analyze slope of the graphs for each city to see which was growing the fastest at different times. We can see this in figures 4, 5, and 6. We will then look at nonlinear curve fitting. For nonlinear curve fitting, we will extrapolate to see the possible future housing markets in each of these cities. The R^2 value shows us the accuracy of the curve so it will also tell us how accurate the future predictions might be. In figures 7, 8, and 9 we can see the curve fits and the extrapolation to around January 2030.

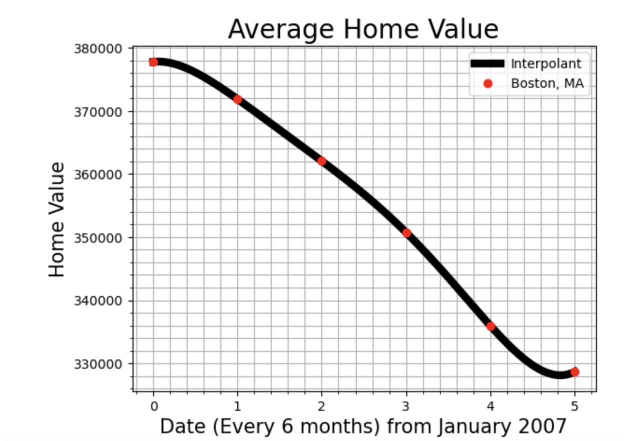

Figure 1: Interpolant of Boston data.

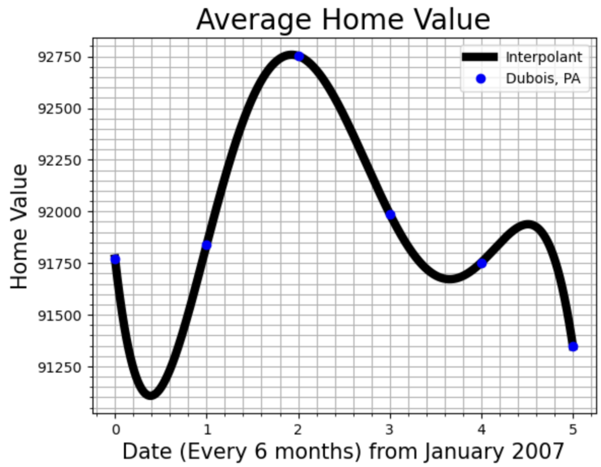

Figure 2: Interpolant of Dubois data.

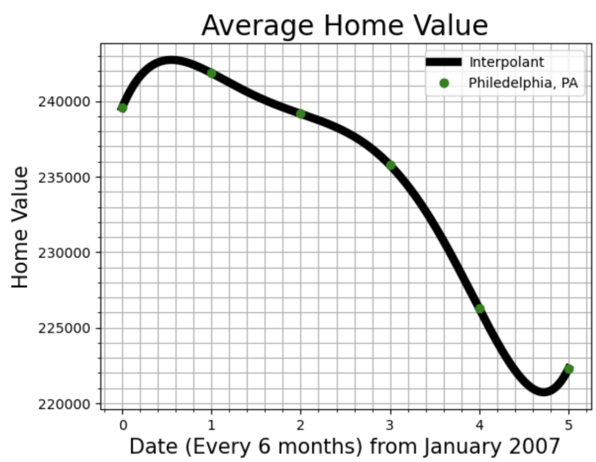

Figure 3: Interpolant of Philadelphia data.

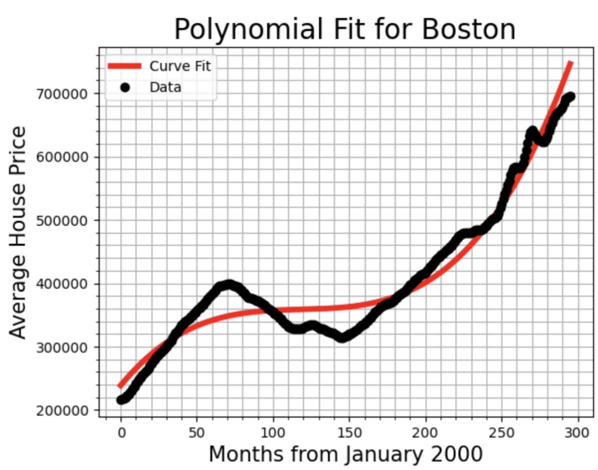

Figure 4: Linear fit of Boston Data

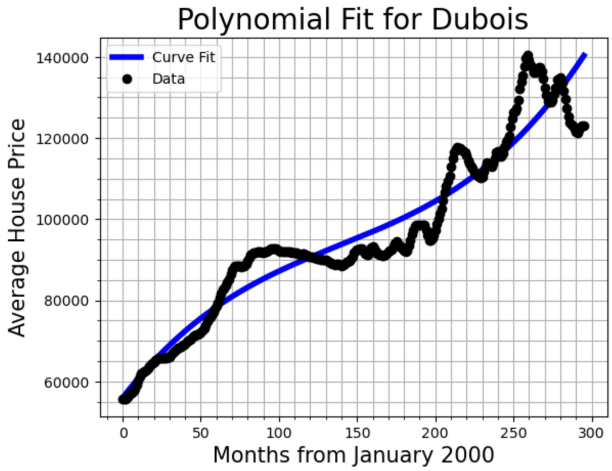

Figure 5: Linear fit of Dubois data.

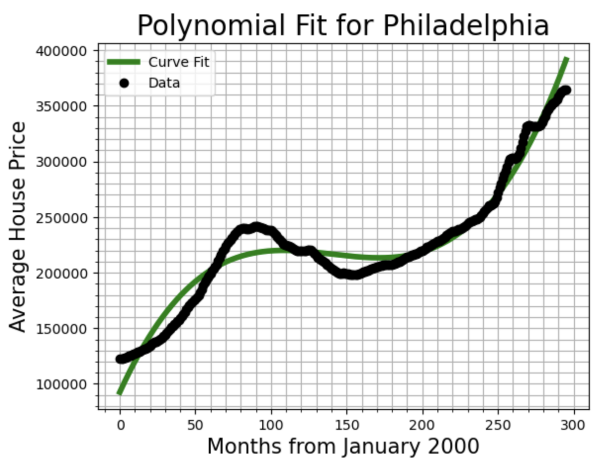

Figure 6: Linear fit of Philadelphia data.

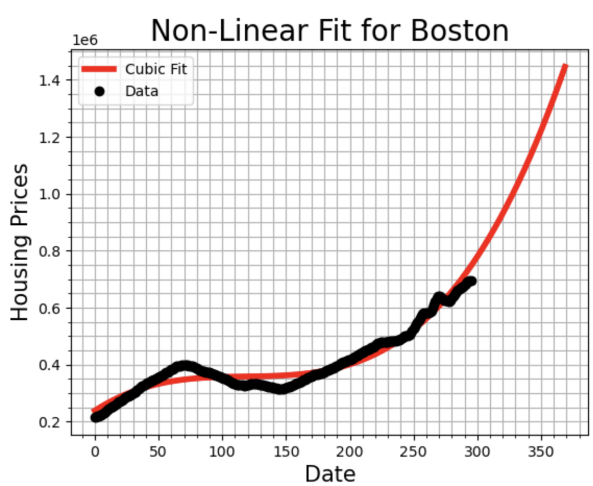

Figure 7: Nonlinear fit of Boston data.

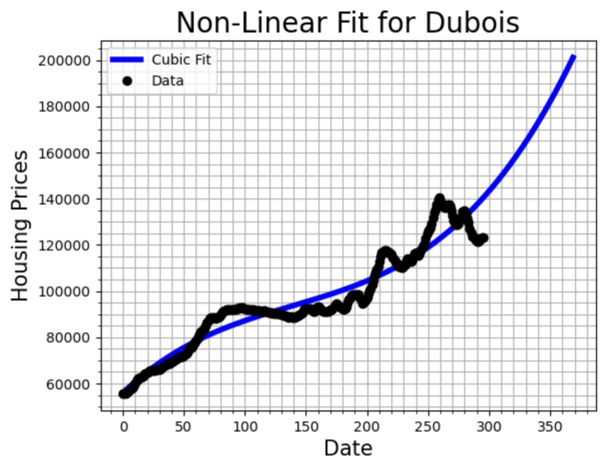

Figure 8: Nonlinear fit of Dubois data.

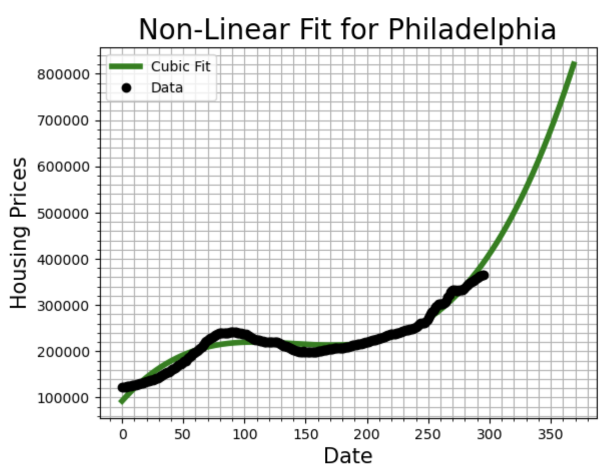

Figure 9: Nonlinear fit of Philadelphia data

3 Methodology

In order to start solving for the interpolant we must use the $A * c = y$ formula where A is a vandermonde matrix, c is matrix of coefficients, and y is a matrix of y values or in our case housing prices.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 & x_3^5 \\ 1 & x_4 & x_4^2 & x_4^3 & x_4^4 & x_4^5 \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 & x_5^5 \\ 1 & x_5 & x_5^2 & x_5^3 & x_5^4 & x_5^5 \end{bmatrix} * \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{bmatrix} = \begin{bmatrix} y_{84} \\ y_{90} \\ y_{96} \\ y_{102} \\ y_{108} \\ y_{114} \end{bmatrix}$$

This matrix equation was then solved for c in python using $linalg.inv(A)@y$. Then c was flipped and plotted to form the graphs 1, 2, and 3.

Next, we started solving for the linear fit of our data. This was done by using the formula $A * c = y$ once again. This time; however, we will use a different vandermonde matrix to solve for c. This time we are looking at all y values so we will not truncate the y coordinates. This will lead to many more rows for our matrices. We will also be less worried about a perfect fit so we will reduce our vandermonde matrix to x^3 being it's largest value. We will write our matrix equation as:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ - & - & - & - \\ 1 & x_{294} & x_{294}^2 & x_{294}^3 \end{bmatrix} * \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ - \\ c_{294} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ - \\ y_{294} \end{bmatrix}$$

The next step was to find the normal equations so we could solve the equation for c. We found the normal equation of A with: $A_norm = A.T@A$ where $A.T$ is the transpose of A. Then we found the normal equation for the others as: $y_norm = A.T@y$ for each city(y-value). These normal eqautions were plugged into LA.solve and then the graphs were plotted as seen in 4, 5, and 6.

Finally, we solved for the nonlinear curve fit. To do this we simply defined a cubic function:

$$f(x, a, b, c, d) = a * x^3 + b * x^2 + c * x + d$$

We then used python to find the optimal curve fit of this function and our data.

4 Results

Our results were simply the graphs of our data and the interpolants or curve fit. These results can be found here: 1, 2, 3, 4, 5, 6, 7, 8, and 9. We will talk more about these results in section 5.

5 Discussion

Looking at our results for our interpolant graphs which was looking at housing prices from January 2007 to June 2009 will allow us to view trends and make assumptions of what happened in between our data points. By looking at the graph of Boston, MA (Figure 1), we can see a constant decline from January 2007 to June 2009 with the maximum decline coming in September 2008. Similarly, looking at Philadelphia,PA (Figure 3), we can see a slow decline in housing prices from April 2007 to May 2008 turns into a rapid decline from May 2008 until leveling out in May 2009. We can also see that the peak decline occurred in January 2009. Our final interpolant graph of Dubois, PA (Figure 2) shows us something very different from the other graphs. We can see a steady increase from March 2007 to January 2008 before we see a decrease from January 2008 to October 2008 and another from April 2009 to June 2009. We can also see that the peak decline in housing cost occurs in June 2009 which is different from the other 2 cities.

Looking at 1 shows exactly where each city saw their biggest decline. We can see that both cities in PA declined in 2009 while Boston declined slightly sooner in 2008. We also saw Boston start to decline sooner than the PA cities in the interpolant graphs 1, 2, and 3.

City	Greatest Decline
Boston	09/2008
Philadelphia	01/2009
Dubois	06/2009

Table 1: Table of Cities and their greatest decline during the stock market crash.

Next we will look at our graphs of linear curve fits. Looking at Boston (4), the curve fit doesn't look very good. The data decreases during the interval we spoke about in the interpolant graphs, but the linear curve fit doesn't show this decline that the data clearly shows.

Moving on to the linear fit of Dubois (5), we can see a lot of change in the data. That said, the curve fit summarizes the overall flow of the data well. The constantly changing data might have to do with Dubois having fewer homes than Boston and Philadelphia leading to sporadic data.

Finally, the Philadelphia curve fit looks good 6. The data is similar to Boston; however, this curve fit is much tighter to the data throughout the graph.

Next, we will look at the nonlinear curve fits. For these graphs we will extrapolate to look at future data and look at R^2 value that will tell us the accuracy of our curve fits. Looking at Boston (7), we can see average prices growing to 1,400,000 by 2029 which is about double what the average prices are now in 2024 according to our data. We also have an R^2 value of 0.95 which means this fit is accurate to the data.

Moving on to Dubois (8), we can see a slightly worse R^2 value (2). This is likely due to the sporadic data we referred to above. We can also see a slight rise in housing costs from 2024 to 2029 (3). This a slight rise compared to the other cities we are looking at. Considering the lower R^2 value this may not be as accurate as the other values.

Finally, we will look at the nonlinear curve fit for Philadelphia (9). We can see a huge increase here from 2024 to 2029 in table 3. We can also see a solid R^2 value of 0.95 in table 2. This R^2 value tells our curve fit is pretty accurate and, that we should also have a decent extrapolation.

Looking at table 3, we can see Philadelphia increased at the fastest rate with a 220 percent increase from 2024 to 2029. Boston followed close behind with a 200 percent increase, and Dubois had a 165 percent increase.

City	R^2 value
Boston	0.95
Philadelphia	0.95
Dubois	0.92

Table 2: Table of Cities and their R^2 values (The accuracy of their nonlinear curve fits).

City	Average Home Value in 20024	Average Home Value in 2029
Boston	700,000	1,400,000
Philadelphia	375,000	825,000
Dubois	124,000	204,000

Table 3: Table of Cities and their Average Home values in both 2024 and 2029.

6 Conclusion

Throughout this paper we have looked at the decline of Boston, Philadelphia, and Dubois during the stock market crash. We then looked at the data and tried to fit it linearly. Finally, we looked a non linear fit of the data for each of these cities. We saw more inaccuracies from Dubois due to the data being taken from less houses, but we also saw poor linear fits from Philadelphia and Boston.

This project could have been better if more cities were a part of the project. I enjoyed seeing the difference between Philadelphia and Dubois throughout the project, so doing one big city and one rural city from each state would've made for some interesting comparisons. I also would've liked to see if all smaller cities had inconsistencies in their data or if Dubois was an outlier.