Model Fitting and Forecasting U.S. Rent Data

Max Sullivan

Department of Mathematics, Boise State University, Boise, Idaho 83725

May 4, 2022

Abstract

I examine the Consumer Price Index of Housing in four U.S. regions, two size class divisions, and national average data. I use Ordinary Least-Squares regression on these time series, determine if the data is significantly trended, and compare the relative trend of each with the national data. I fit an ARMA model to each set of OLS residuals, run model diagnostics, and forecast the future values for each region and class size. Comparing the various sets of data in each category, I determine if there is significant difference between subcategories and national data, and across subcategories, and discuss the implications of these results for policymakers, political groups, and citizens.

1 Introduction

The rental prices in the U.S. have been steadily increasing for years, and the national median gross rent was \$1,097 in 2019 [2]. These prices have increased significantly especially in regions of smaller size, as many people emigrate from areas of large population, with high costs of living, such as California, where the cost of living has increased significantly very quickly. Various studies look at the growth of rent prices in larger cities, and in these high density population centers, but much less covered is the increase in rent prices in places like Boise. Boise rent prices climbed 12.4% between Jan. 2020 and Jan. 2021 [5]. The other top 4 highest growth areas were Bakersfield, CA, Fresno, CA, Gilbert, AZ, and Chesapeake, VA. Four of these top 5 are in the western region, and with the exception of Fresno, all 5 have populations under 500,000 (with Fresno population 525,000).

This study leads to questions about how the increase in rent compares between various regions and population sizes. Obviously, the information in Suppe's article is localized to Boise, since it is written for the audience of Boise residents. Analysis at the state level shows similar trends, see [4]. However, I believe a similar and more rigorous analysis of this subject over larger regions of the country would be interesting to the general public, and political figures and groups.

I will be analyzing the percent increase in housing costs, from publicly available data, over the past 12 years (starting at the peak of the great recession in 2010), to analyze which regions and population sizes have the most significant trends in increased housing costs. Additionally, using Ordinary Least-Squares regression and ARMA modelling of the OLS residuals, I will forecast the future increases to rent in various regions and population sizes in the US.

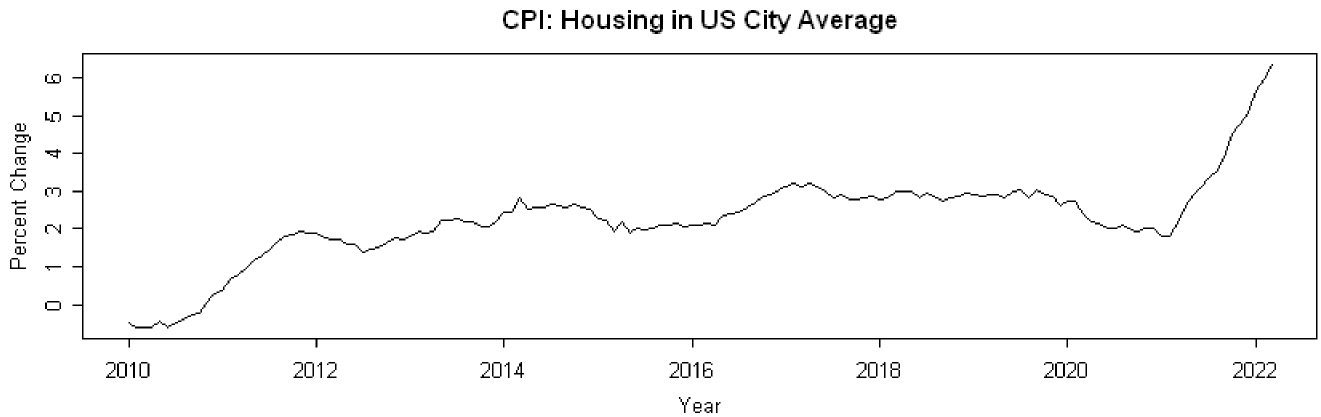
2 The Data

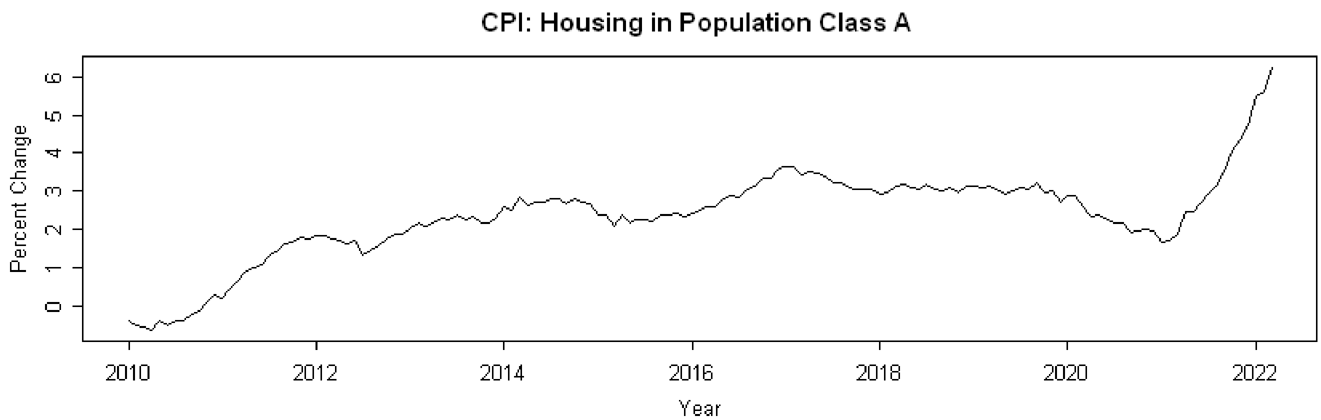
Unless otherwise stated, all data comes from the Bureau of Labor Statistics (BLS) or the U.S. Census Bureau. The data was accesses through the BLS website, and the Federal Reserve of St. Louis Economic Data (FRED) website, which collects it's data from the BLS and U.S. Census Bureau.

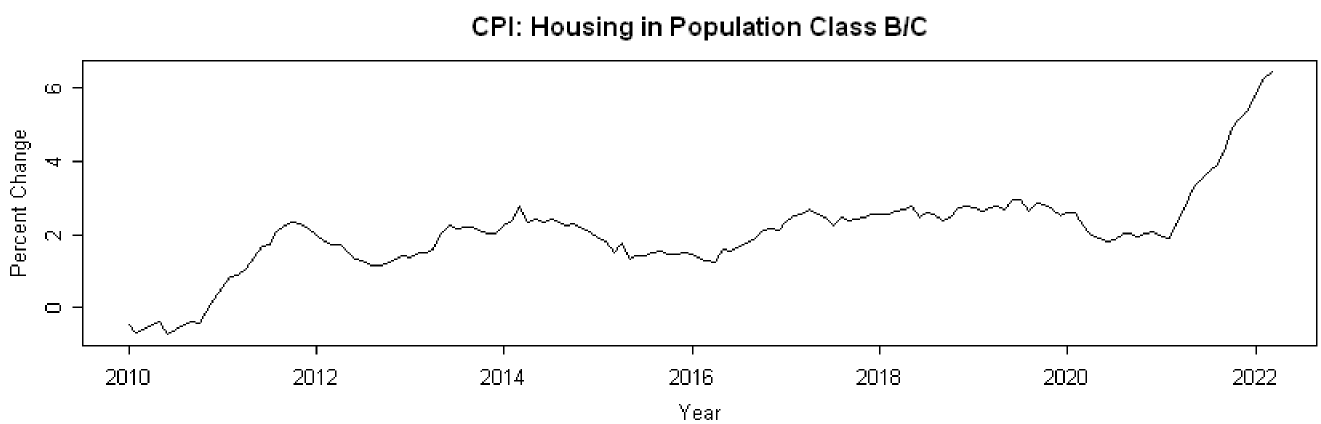
Consumer Price Index (CPI) is defined by the BLS as "a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services." [3]

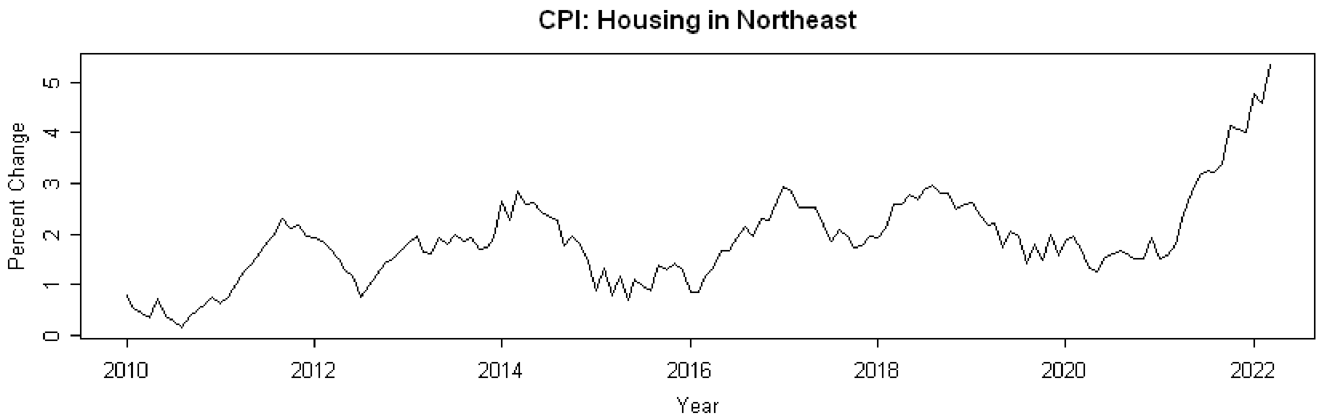
Each data set I have analyzed is a CPI with certain material conditions. Each series is the "Consumer Price Index for All Urban Consumers: Housing" with varied regions, and population sizes. There are four geographic regions treated by the BLS data: Northeast, South, Midwest, and West. There are also two currently used divisions of size treated by this data: Class Size A (populations above 1,500,000) and Class Size B/C (populations between 50,000 and 1,500,000).

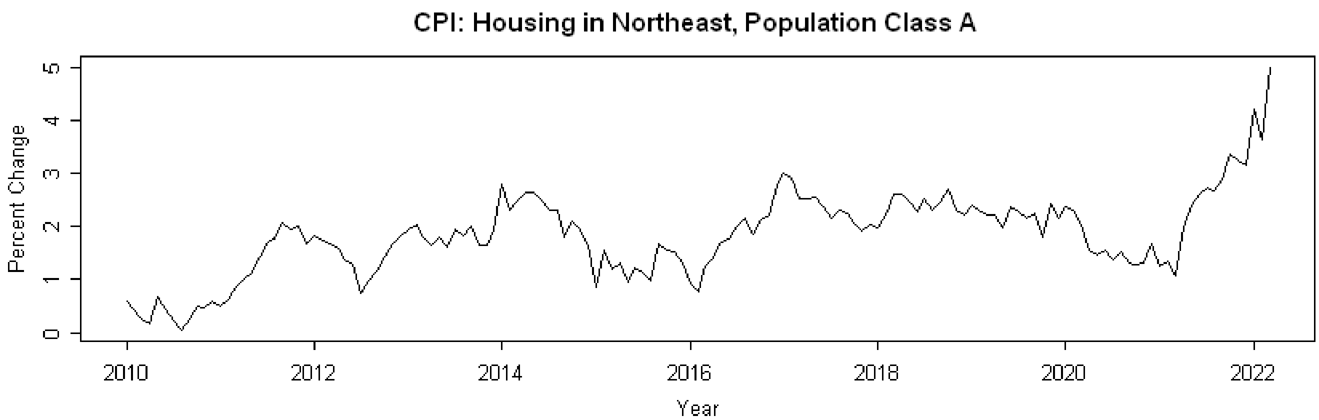
The start time for each series varies, therefore I have decided to restrict each series of data to a start of Jan. 2010, to make sure every series has the same number of observations over the same interval of time. The data is recorded monthly as percent changes from the previous year. This was done to standardize the data, since different series start at different years. Below are visual summaries of the national data, and all 4 regional data series'.

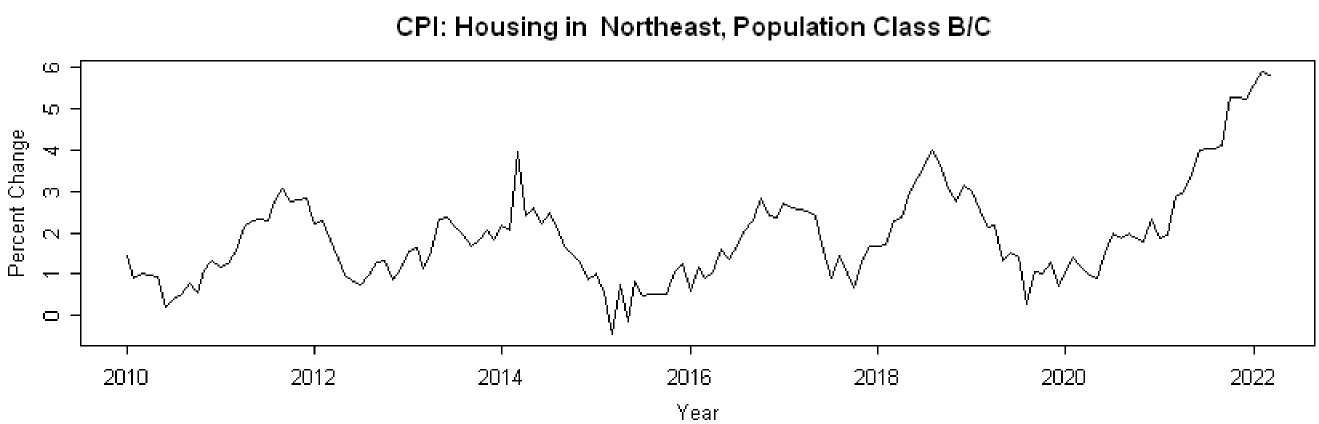

CPI: Housing in US City Average
Percent Change
Year
0
1
2
3
4
5
6
2010
2012
2014
2016
2018
2020
2022


CPI: Housing in Population Class A
Percent Change
Year
0
1
2
3
4
5
6
2010
2012
2014
2016
2018
2020
2022

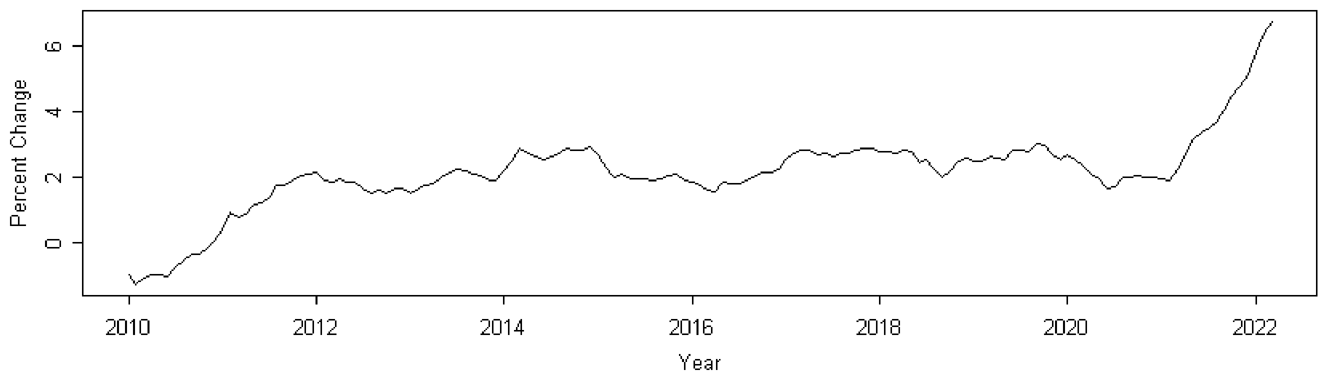

CPI: Housing in Population Class B/C
Percent Change
Year
0
2
4
6
2010
2012
2014
2016
2018
2020
2022


CPI: Housing in Northeast
Percent Change
0
1
2
3
4
5
2010
2012
2014
2016
2018
2020
2022
Year

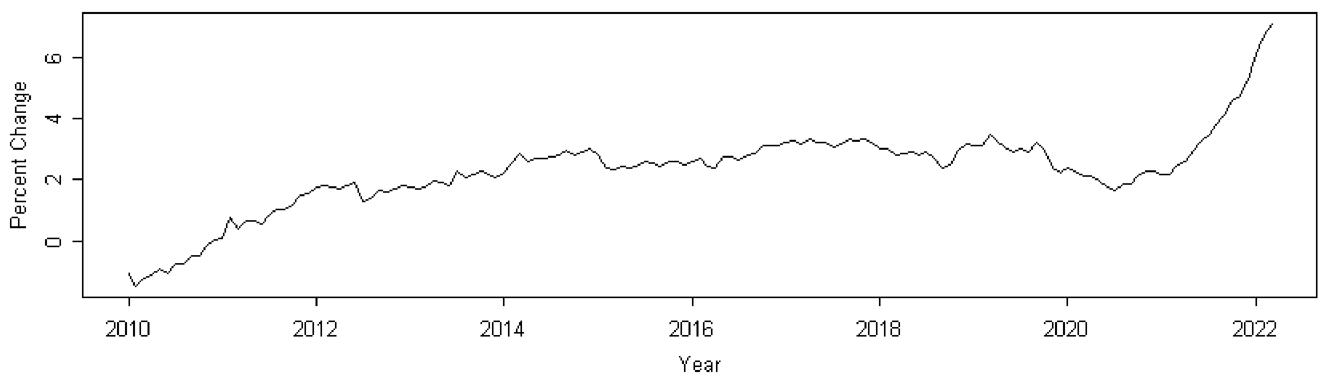

CPI: Housing in Northeast, Population Class A
Percent Change
0
1
2
3
4
5
2010
2012
2014
2016
2018
2020
2022
Year


CPI: Housing in Northeast, Population Class B/C
Percent Change
0
1
2
3
4
5
6
2010
2012
2014
2016
2018
2020
2022
Year

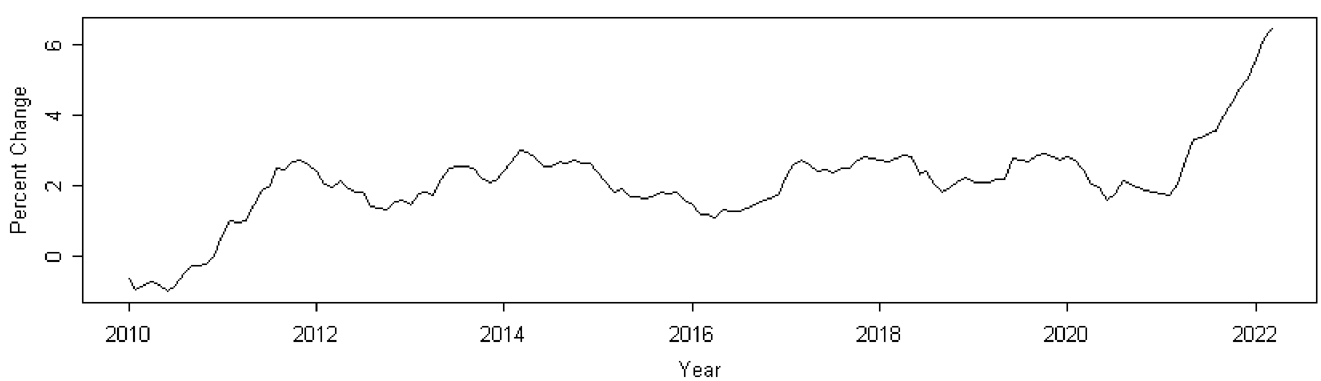
CPI: Housing in South

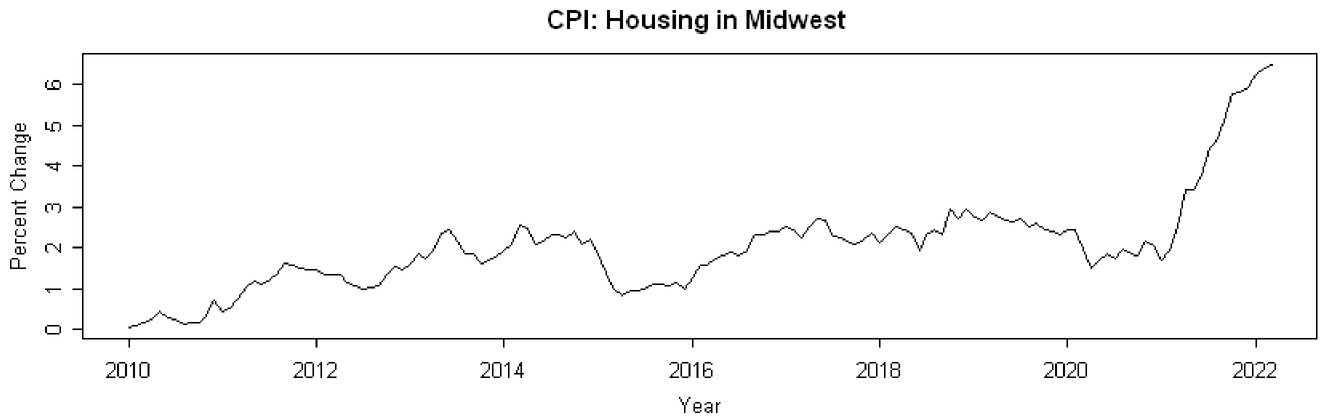


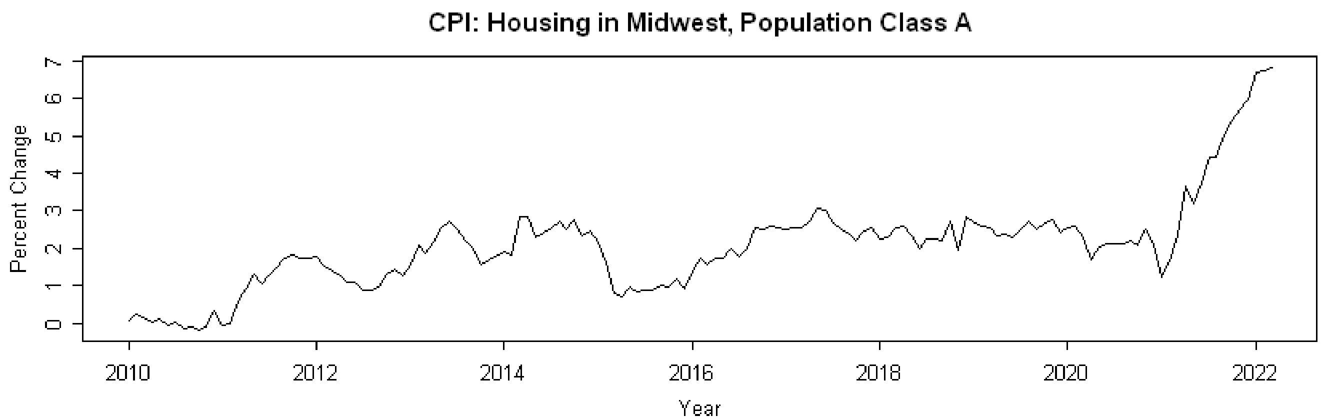
CPI: Housing in South, Population Class A

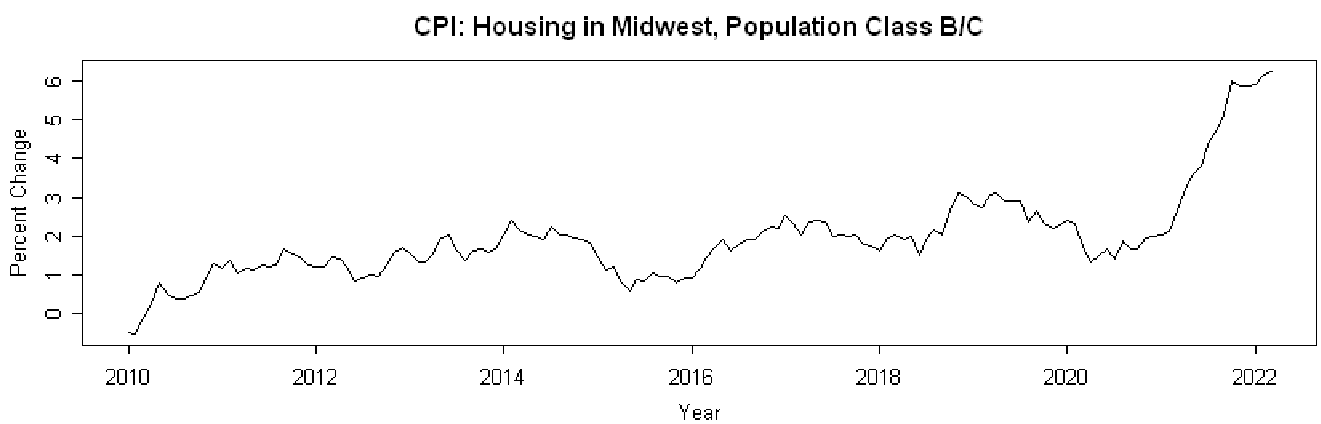


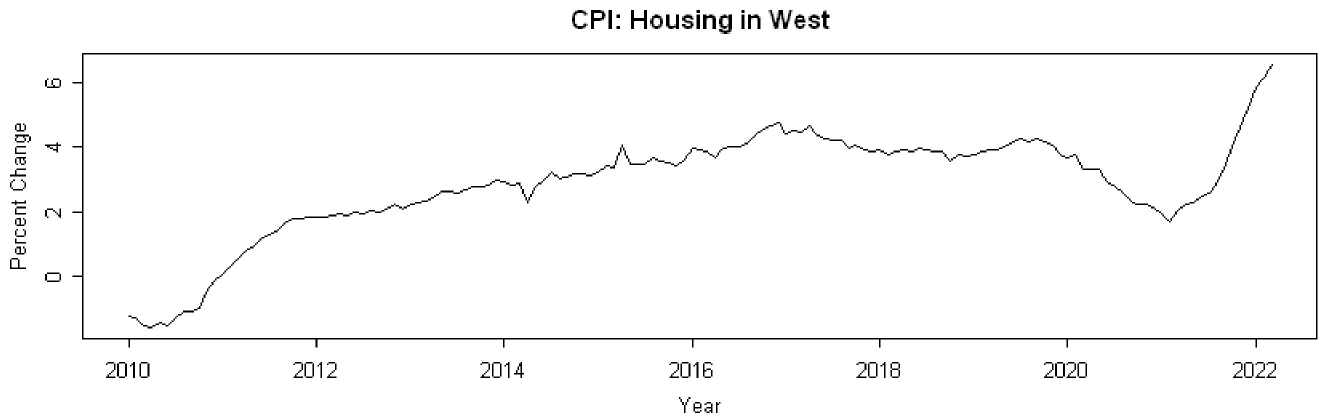
CPI: Housing in South, Population Class B/C

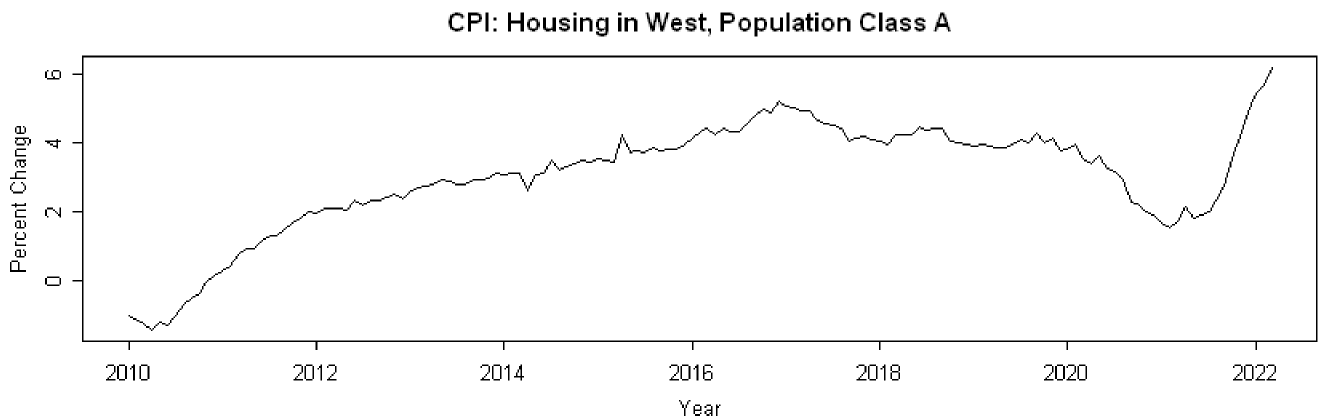


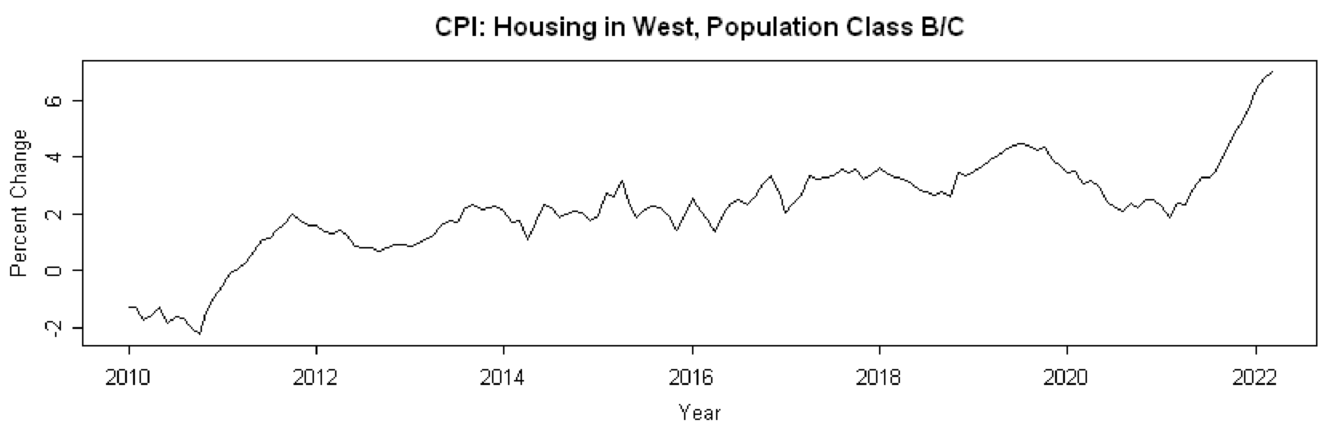

CPI: Housing in Midwest
Percent Change
0
1
2
3
4
5
6
2010
2012
2014
2016
2018
2020
2022
Year


CPI: Housing in Midwest, Population Class A
Percent Change
0
1
2
3
4
5
6
7
2010
2012
2014
2016
2018
2020
2022
Year


CPI: Housing in Midwest, Population Class B/C
Percent Change
0
1
2
3
4
5
6
2010
2012
2014
2016
2018
2020
2022
Year


CPI: Housing in West
Percent Change
Year
0
2
4
6
2010
2012
2014
2016
2018
2020
2022


CPI: Housing in West, Population Class A
Percent Change
Year
0
2
4
6
2010
2012
2014
2016
2018
2020
2022


CPI: Housing in West, Population Class B/C
Percent Change
Year
-2
0
2
4
6
2010
2012
2014
2016
2018
2020
2022

3 Research Questions

The common questions we want to be able to answer after analyzing this data will involve how the rents have changed over time, how they compare across population sizes, regions, and to national data. Below are the specific questions an average consumer might be interested in, that I have decided to analyze. for each [region] the 4 regions can be inserted, as shorthand for the multiple comparisons to be made.

- What is the average percent increase in rent across this time period for [region]?
- How do the percent increases for [region] compare to the national percent increase?
- How do percent increases for [region 1] compare to [region 2]?
- What will the future percent changes in rent be for [region]?

For each region, these can be further analyzed by population class size.

These questions are the ones I decided to focus on, because the general public everywhere in the US is concerned about the rising cost of rent. It is also of particular interest to the general public where the cost of rent is increasing the fastest. In every region and population class, percent change of rent from the previous year increased drastically during the last 6 months, so many political entities and researchers have analyzed the causes. These research questions and their results could be important support for arguments made by policy makers about building affordable housing units, or implementing policies to help struggling renters.

4 Analysis Methods

To answer these questions, I will be doing the following analysis steps for each set of data.

- Run Ordinary Least Squares regression on each data set, and determine the linear trend (and test for significance).
- Graph the OLS Residuals for each series.
- Fit the OLS residuals with an ARMA model, choosing the best fitting model based on AICC values.
- Forecast future values using the ARMA model for each series.

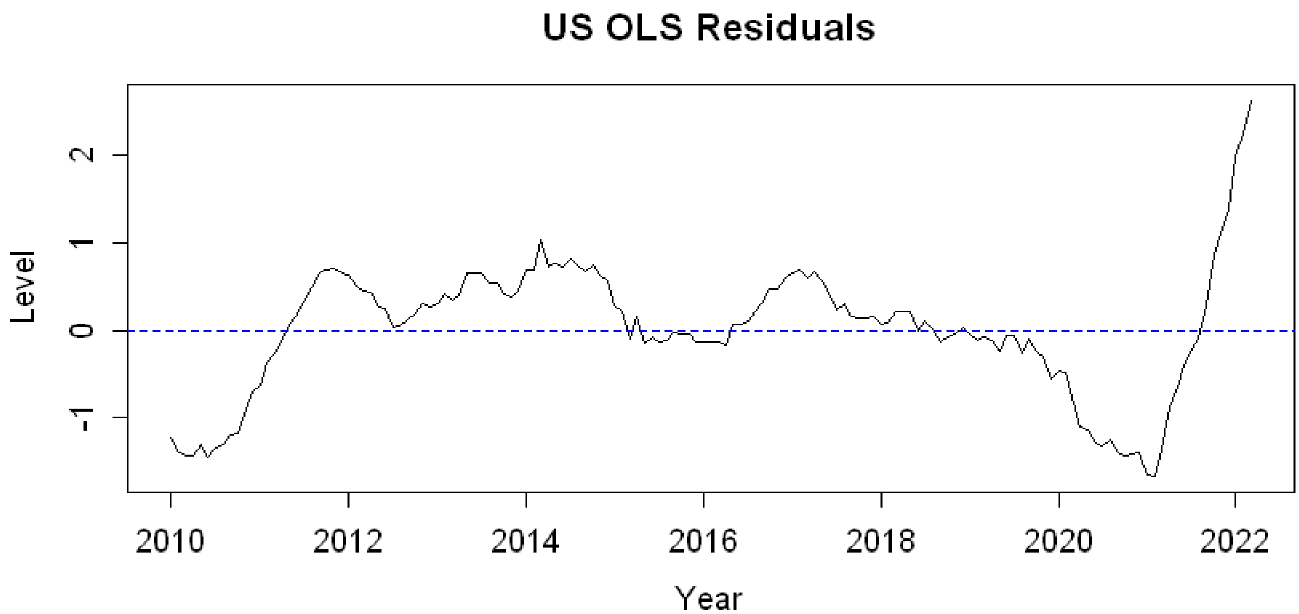
These analysis methods are fairly standard, the extension I am exploring is using these methods on all 15 series and comparing them. The results of this comparison, as mentioned above, could be of interest to policymakers, if one region outperforms another.

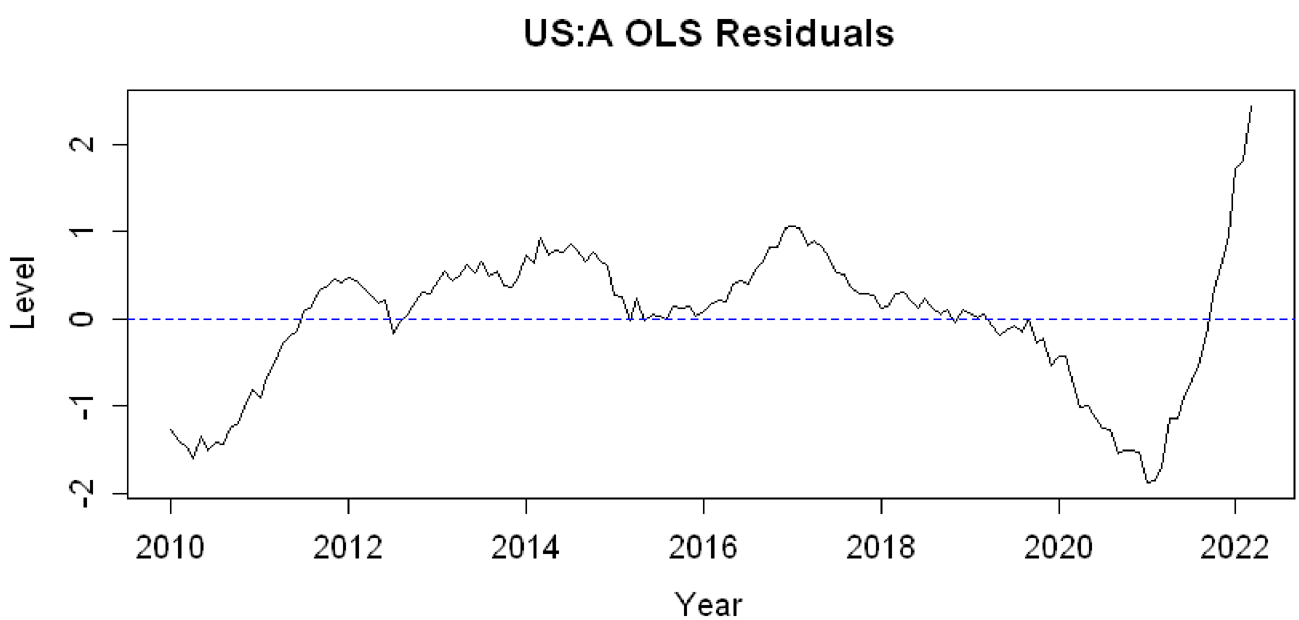
I will summarize the results of each analysis step in tables, for ease of review, and comment on the significance of each in the next section.

4.1 OLS Regression

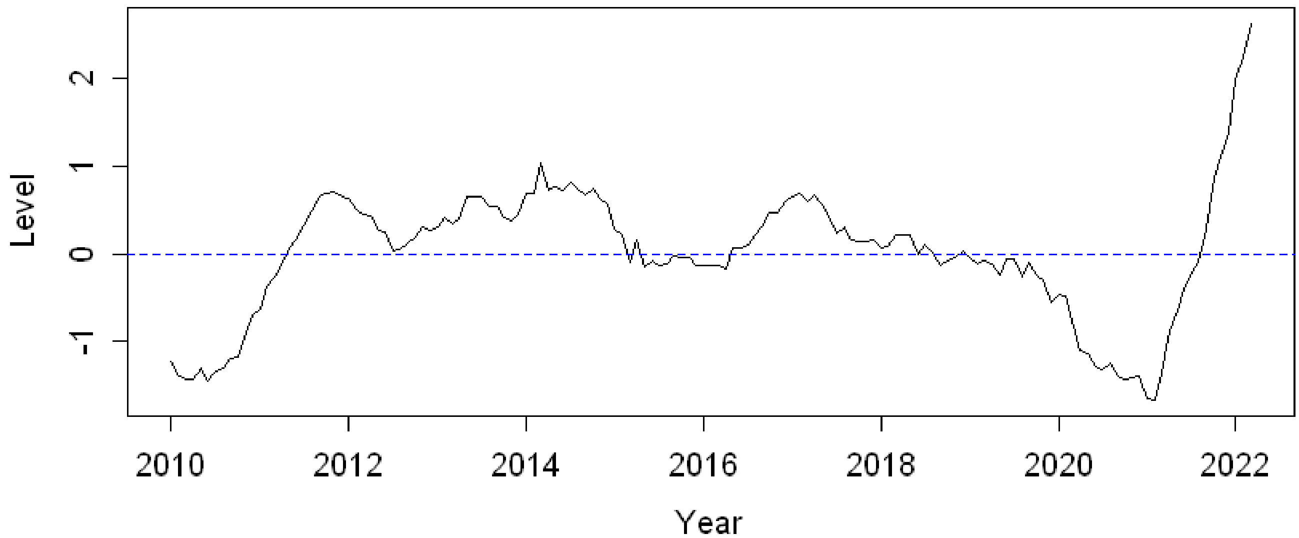
A summary of the OLS regression results, and the OLS residual plots.

OLS Regression Results		
Region:Pop.	β_1	p-value
US	0.020364	2.2e-16
US:A	0.020047	2.2e-16
US:B/C	0.020465	2.2e-16
NE	0.01231	1.251e-15
NE:A	0.011237	2.26e-16
NE:B/C	0.012299	2.062e-08
S	0.020241	2.2e-16
S:A	0.023774	2.2e-16
S:B/C	0.017649	2.2e-16
MW	0.020855	2.2e-16
MW:A	0.021662	2.2e-16
MW:B/C	0.020040	2.2e-16
W	0.026423	2.2e-16
W:A	0.023669	2.2e-16
W:B/C	0.03224	2.2e-16

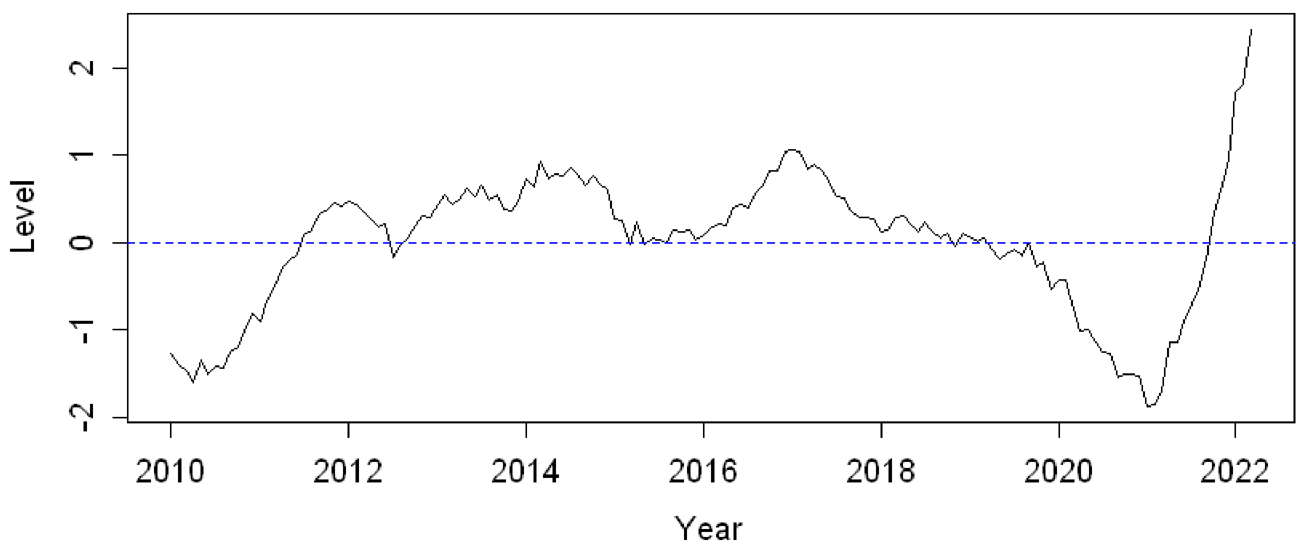

US OLS Residuals
Level
Year
2010
2012
2014
2016
2018
2020
2022
-1
0
1
2

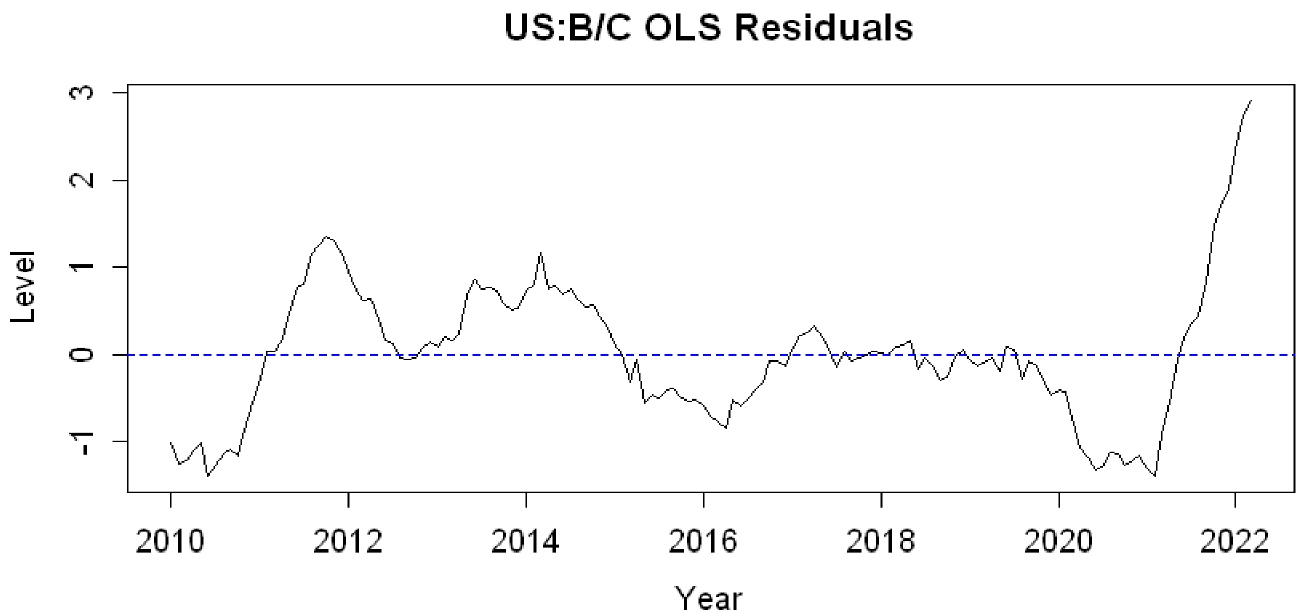

US:A OLS Residuals
Level
Year
2010
2012
2014
2016
2018
2020
2022
-2
-1
0
1
2

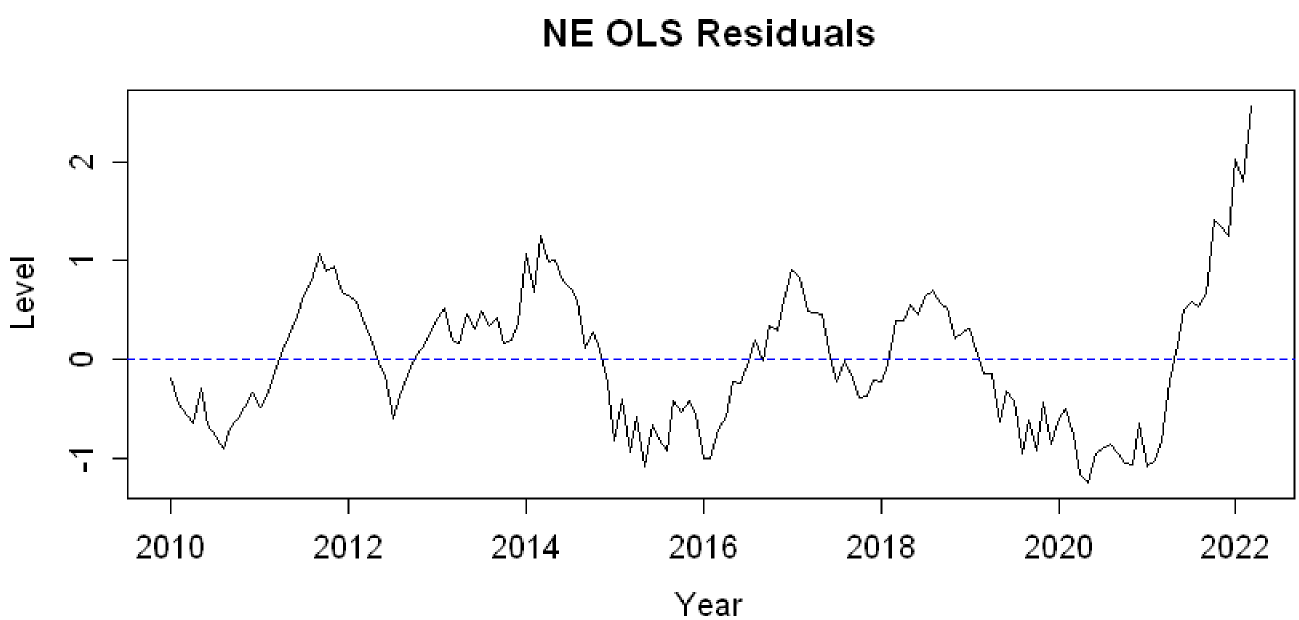
US OLS Residuals

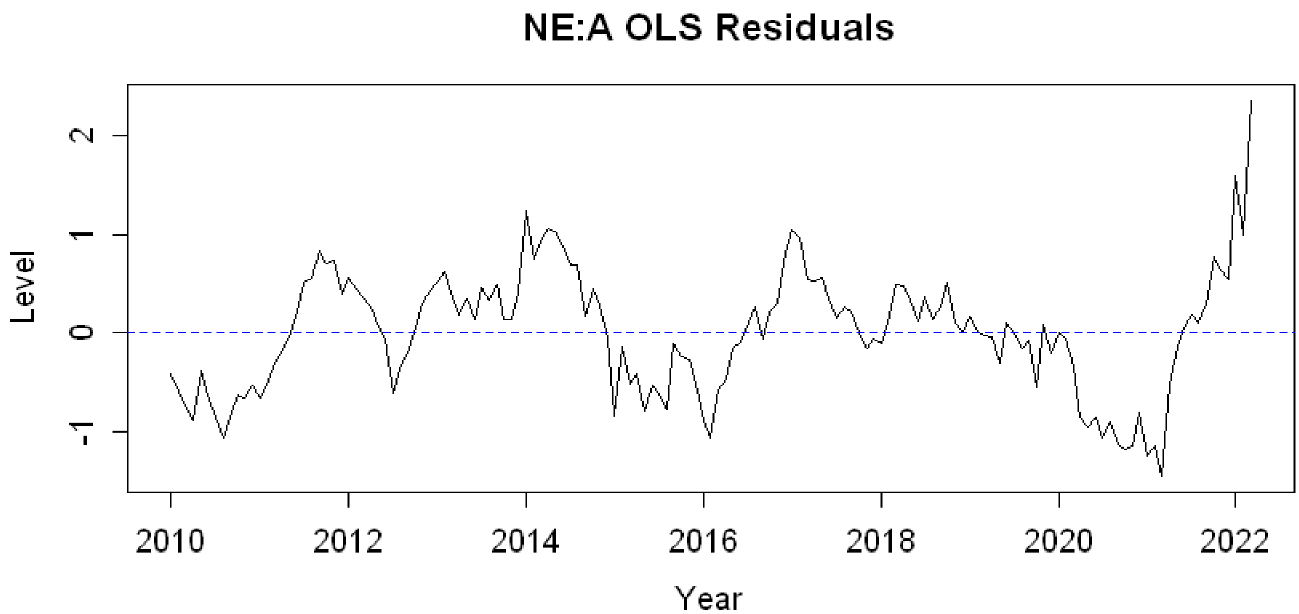


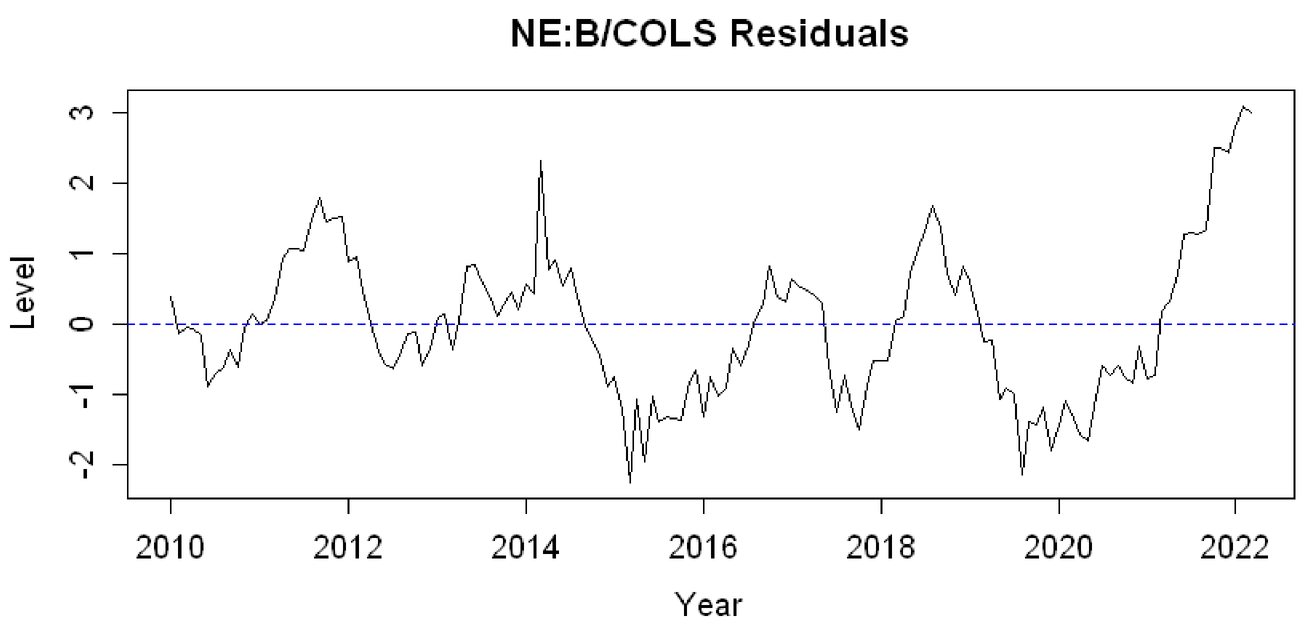
US:A OLS Residuals

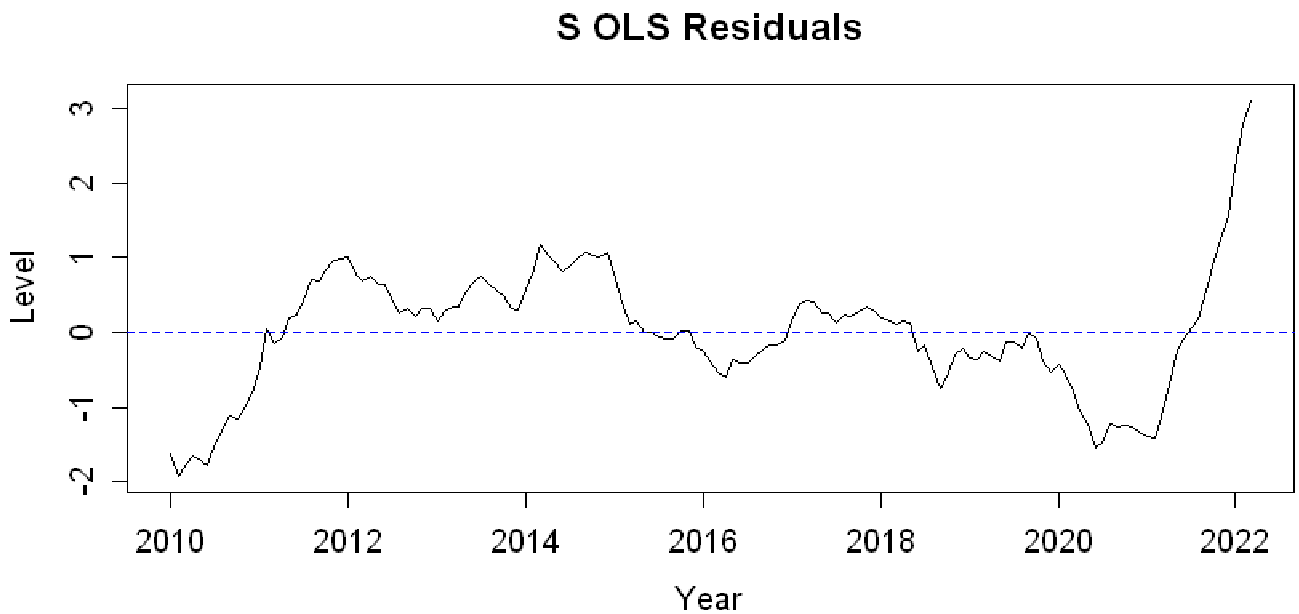


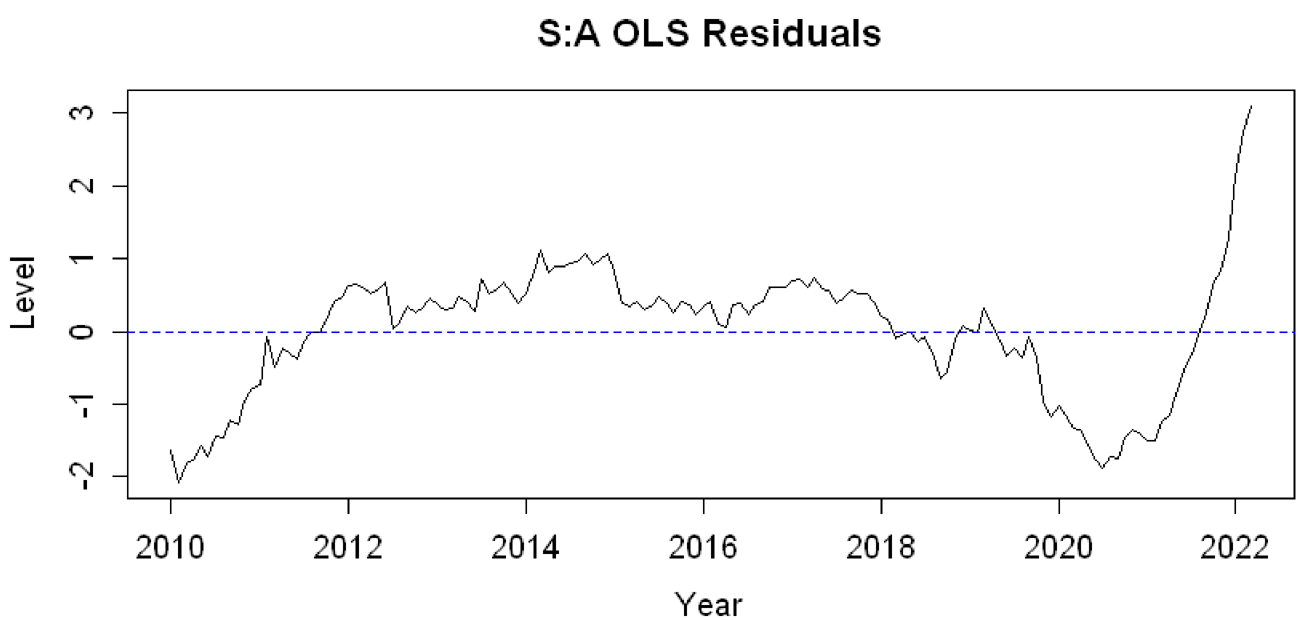

US:B/C OLS Residuals
Level
3
2
1
0
-1
2010
2012
2014
2016
2018
2020
2022
Year

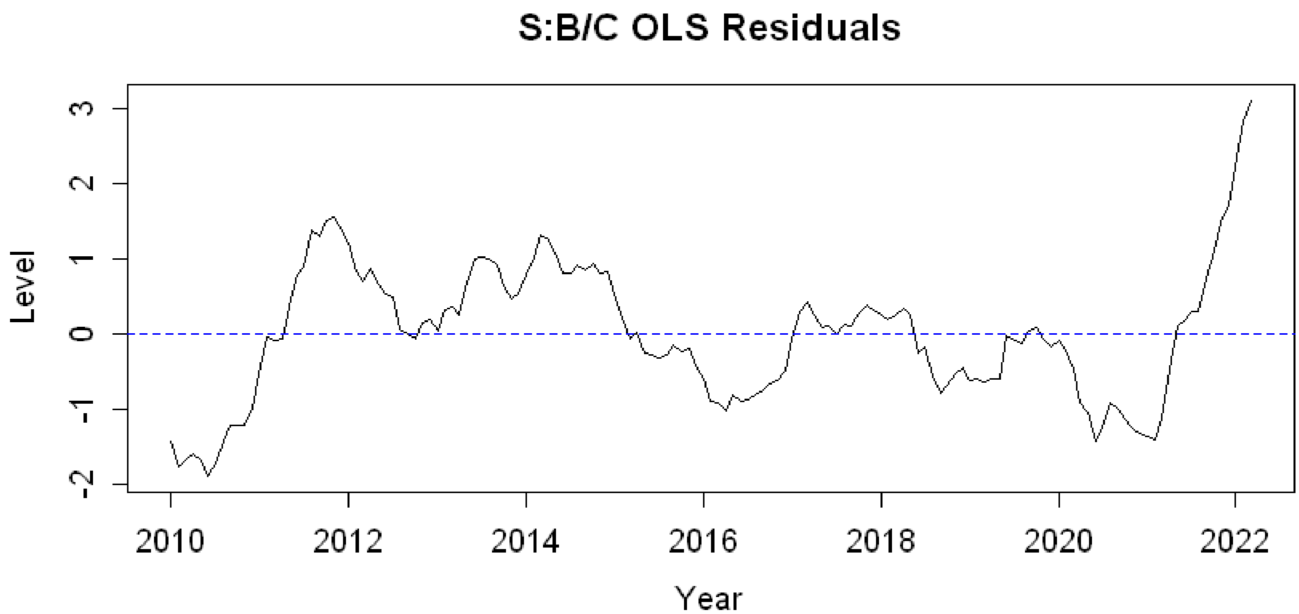

NE OLS Residuals
Level
2
1
0
-1
2010
2012
2014
2016
2018
2020
2022
Year

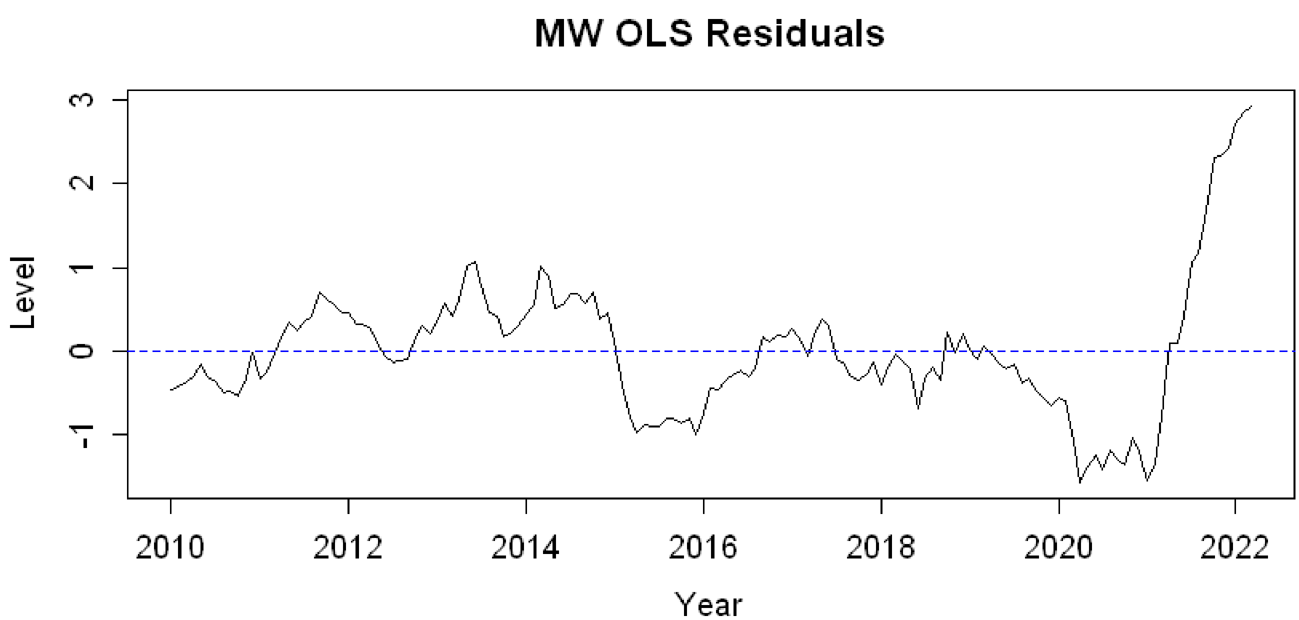

NE:A OLS Residuals
Level
2
1
0
-1
2010
2012
2014
2016
2018
2020
2022
Year

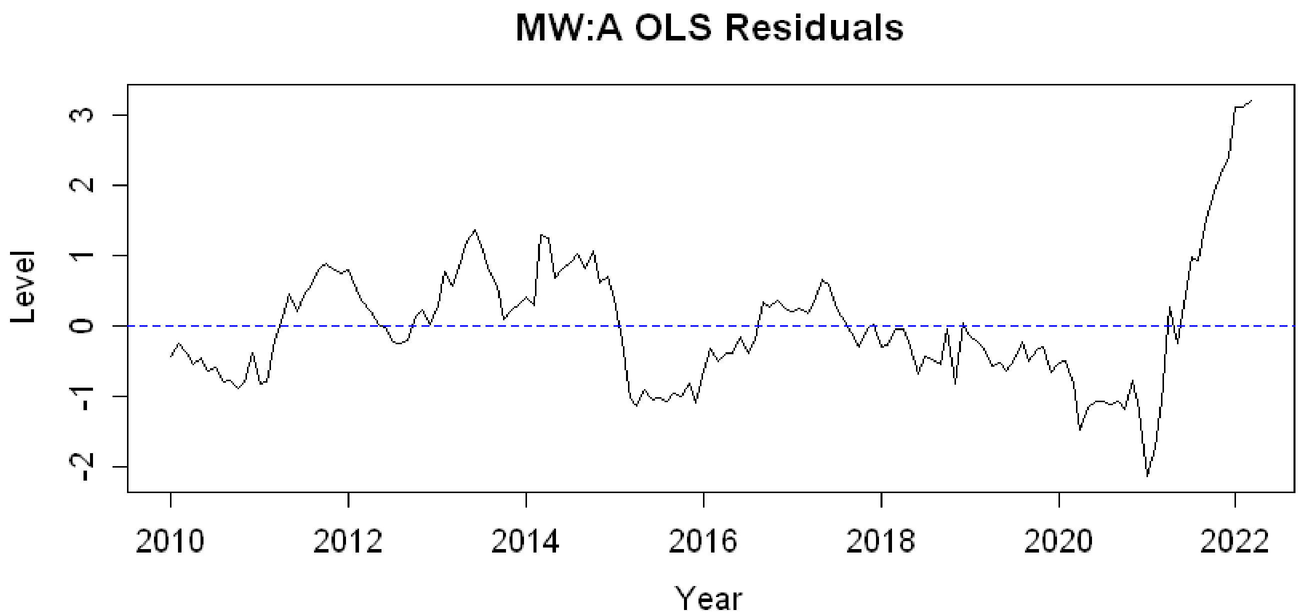

NE:B/COLS Residuals
Level
3
2
1
0
-1
-2
2010
2012
2014
2016
2018
2020
2022
Year

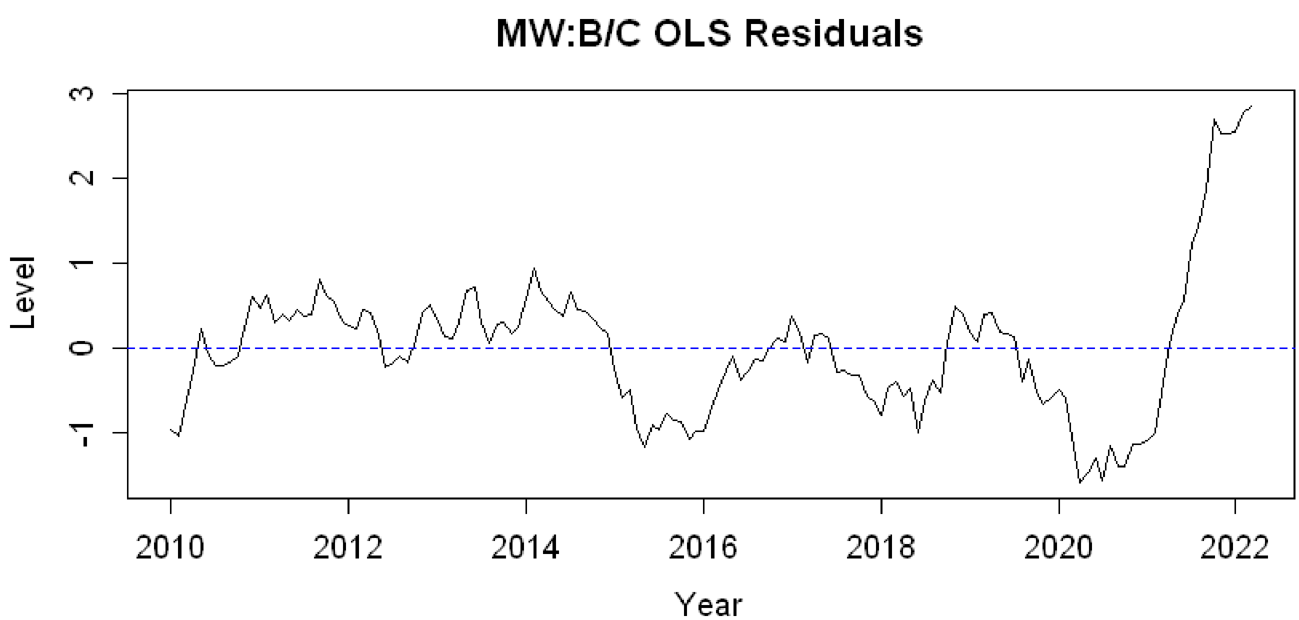

S OLS Residuals
Level
Year
-2
-1
0
1
2
3
2010
2012
2014
2016
2018
2020
2022

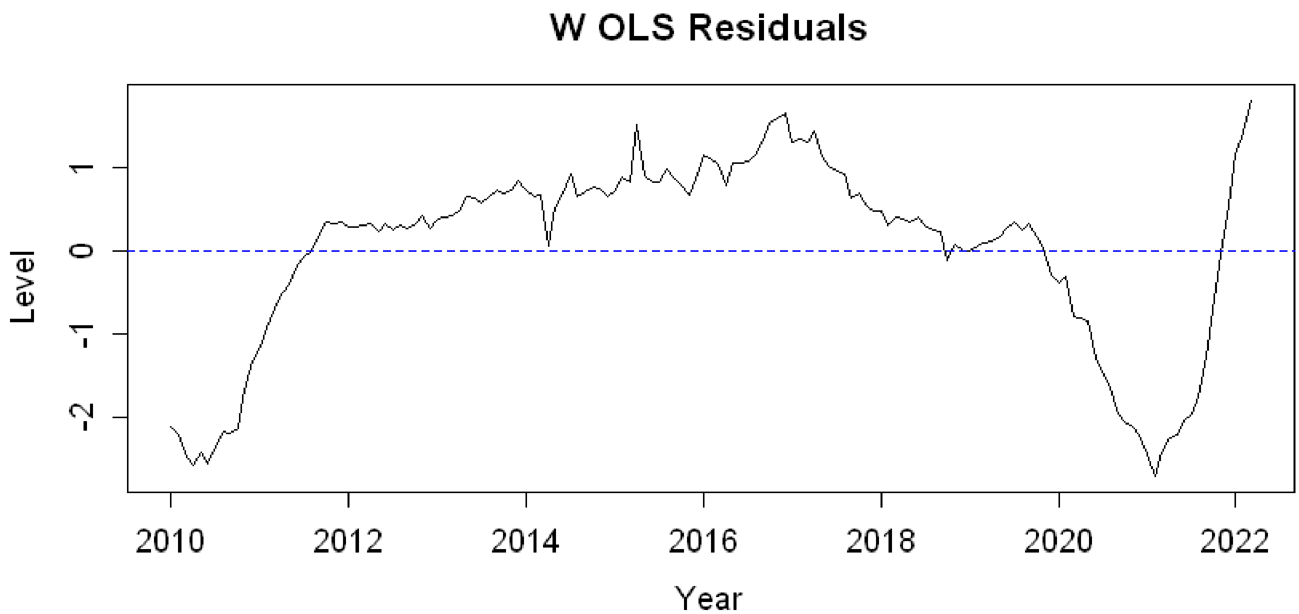

S:A OLS Residuals
Level
Year
-2
-1
0
1
2
3
2010
2012
2014
2016
2018
2020
2022

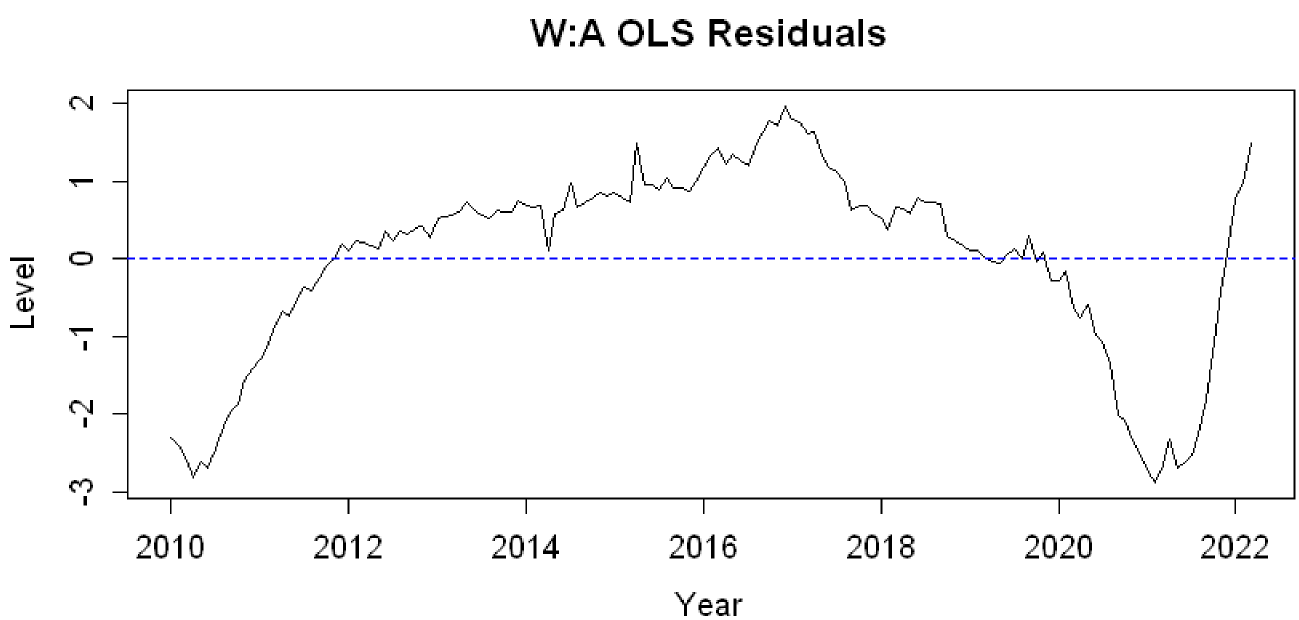

S:B/C OLS Residuals
Level
Year
3
2
1
0
-1
-2
2010
2012
2014
2016
2018
2020
2022

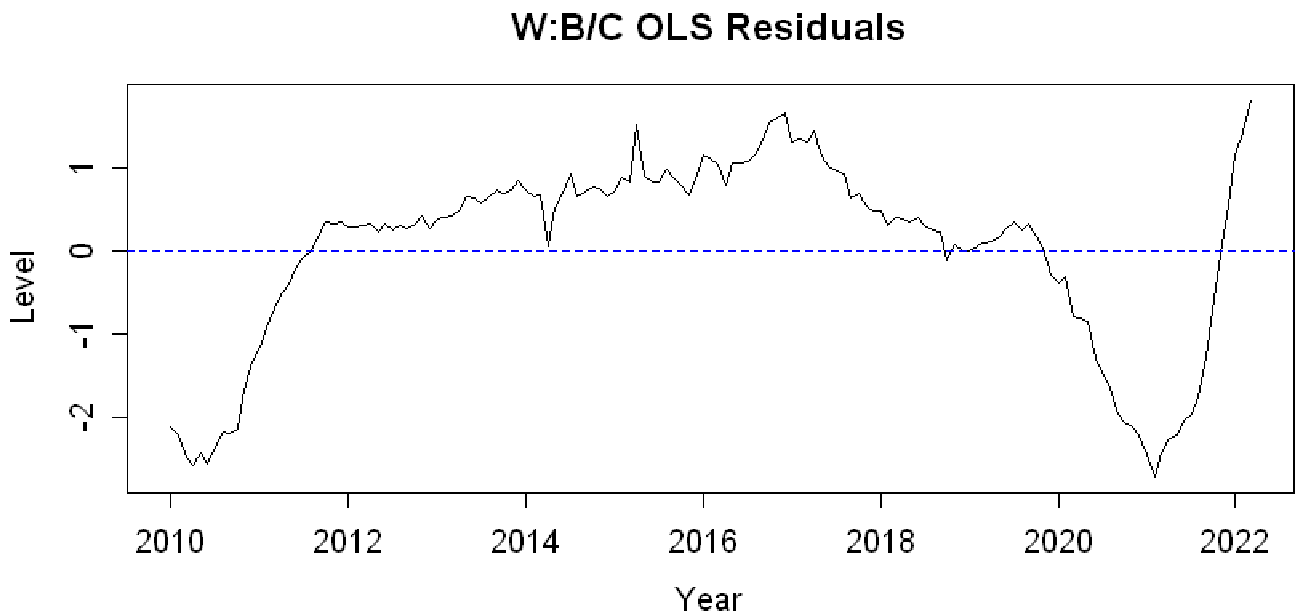

MW OLS Residuals
Level
Year
3
2
1
0
-1
2010
2012
2014
2016
2018
2020
2022


MW:A OLS Residuals
Level
Year
3
2
1
0
-1
-2
2010
2012
2014
2016
2018
2020
2022


MW:B/C OLS Residuals
Level
Year
3
2
1
0
-1
2010
2012
2014
2016
2018
2020
2022


W OLS Residuals
Level
1
0
-1
-2
2010
2012
2014
2016
2018
2020
2022
Year


W:A OLS Residuals
Level
2
1
0
-1
-2
-3
2010
2012
2014
2016
2018
2020
2022
Year



4.2 OLS Residuals Model Fitting

The OLS residuals are clearly not entirely white-noise, there are still some major trends happening, so the model fitting will be crucial to the forecasting of future values. Below are the summaries for AICC values of ARMA(p,q) model for each series.

US AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-104.3
ARMA(2,0)	-124.4
ARMA(3,0)	-148.2
ARMA(0,1)	173.9
ARMA(0,2)	71.9
ARMA(0,3)	4.5
ARMA(1,1)	-115.2
ARMA(2,1)	-159.5
ARMA(1,2)	-129.2

US:B/C AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-57.6
ARMA(2,0)	-68.9
ARMA(3,0)	-74.8
ARMA(0,1)	191.9
ARMA(0,2)	98.7
ARMA(0,3)	41.5
ARMA(1,1)	-63.9
ARMA(2,1)	-91.8
ARMA(1,2)	-67.2

NE:A AICC	
ARMA(p,q)	AICC
ARMA(1,0)	80.6
ARMA(2,0)	74.8
ARMA(3,0)	71.6
ARMA(0,1)	191.8
ARMA(0,2)	129.4
ARMA(0,3)	110.2
ARMA(1,1)	76.7
ARMA(2,1)	73.1
ARMA(1,2)	73.2

S AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-56.7
ARMA(2,0)	-97.2
ARMA(3,0)	-95.6
ARMA(0,1)	195.3
ARMA(0,2)	82.5
ARMA(0,3)	19.3
ARMA(1,1)	-90.4
ARMA(2,1)	-101.3
ARMA(1,2)	-90.8

US:A AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-80.2
ARMA(2,0)	-85.8
ARMA(3,0)	-120.4
ARMA(0,1)	194.2
ARMA(0,2)	84.9
ARMA(0,3)	37.3
ARMA(1,1)	-82.9
ARMA(2,1)	-78.5
ARMA(1,2)	-103.7

NE AICC	
ARMA(p,q)	AICC
ARMA(1,0)	50.9
ARMA(2,0)	48.1
ARMA(3,0)	39.8
ARMA(0,1)	207.6
ARMA(0,2)	127.1
ARMA(0,3)	107.7
ARMA(1,1)	49.3
ARMA(2,1)	46.3
ARMA(1,2)	41.4

NE:B/C AICC	
ARMA(p,q)	AICC
ARMA(1,0)	193.3
ARMA(2,0)	189.7
ARMA(3,0)	190.9
ARMA(0,1)	328.8
ARMA(0,2)	268.7
ARMA(0,3)	250.7
ARMA(1,1)	190.5
ARMA(2,1)	191.4
ARMA(1,2)	190.1

S:A AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-15.1
ARMA(2,0)	-18.9
ARMA(3,0)	-19.2
ARMA(0,1)	228.5
ARMA(0,2)	142.6
ARMA(0,3)	92.0
ARMA(1,1)	-17.9
ARMA(2,1)	-35.6
ARMA(1,2)	-16.7

S:B/C AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-8.3
ARMA(2,0)	-40.0
ARMA(3,0)	-38.9
ARMA(0,1)	222.9
ARMA(0,2)	117.9
ARMA(0,3)	55.4
ARMA(1,1)	-31.6
ARMA(2,1)	-40.3
ARMA(1,2)	-36.0

MW:A AICC	
ARMA(p,q)	AICC
ARMA(1,0)	96.0
ARMA(2,0)	97.9
ARMA(3,0)	99.8
ARMA(0,1)	246.4
ARMA(0,2)	188.8
ARMA(0,3)	156.2
ARMA(1,1)	97.9
ARMA(2,1)	94.4
ARMA(1,2)	100.0

W AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-26.2
ARMA(2,0)	-35.4
ARMA(3,0)	-51.1
ARMA(0,1)	289.3
ARMA(0,2)	183.9
ARMA(0,3)	121.9
ARMA(1,1)	-30.9
ARMA(2,1)	-62.7
ARMA(1,2)	-39.5

W:B/C AICC	
ARMA(p,q)	AICC
ARMA(1,0)	97.8
ARMA(2,0)	93.9
ARMA(3,0)	95.8
ARMA(0,1)	258.4
ARMA(0,2)	183.9
ARMA(0,3)	146.5
ARMA(1,1)	94.5
ARMA(2,1)	84.9
ARMA(1,2)	95.9

MW AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-9.8
ARMA(2,0)	-11.5
ARMA(3,0)	-9.7
ARMA(0,1)	189.2
ARMA(0,2)	131.9
ARMA(0,3)	68.5
ARMA(1,1)	-11.3
ARMA(2,1)	-15.6
ARMA(1,2)	-9.2

MW:B/C AICC	
ARMA(p,q)	AICC
ARMA(1,0)	23.2
ARMA(2,0)	20.7
ARMA(3,0)	22.7
ARMA(0,1)	202.3
ARMA(0,2)	150.8
ARMA(0,3)	89.4
ARMA(1,1)	20.9
ARMA(2,1)	15.6
ARMA(1,2)	23.0

W:A AICC	
ARMA(p,q)	AICC
ARMA(1,0)	-5.8
ARMA(2,0)	-9.2
ARMA(3,0)	-29.4
ARMA(0,1)	315.9
ARMA(0,2)	213.8
ARMA(0,3)	144.2
ARMA(1,1)	-6.9
ARMA(2,1)	-5.5
ARMA(1,2)	-19.7

The model ARMA model that best fits each OLS residual based on the AICC values is given below.

Best Fit Model	
Region:Pop	ARMA(p,q)
US	ARMA(2,1)
US:A	ARMA(3,0)
US:B/C	ARMA(2,1)
NE	ARMA(3,0)
NE:A	ARMA(3,0)
NE:B/C	ARMA(2,0)
S	ARMA(2,1)

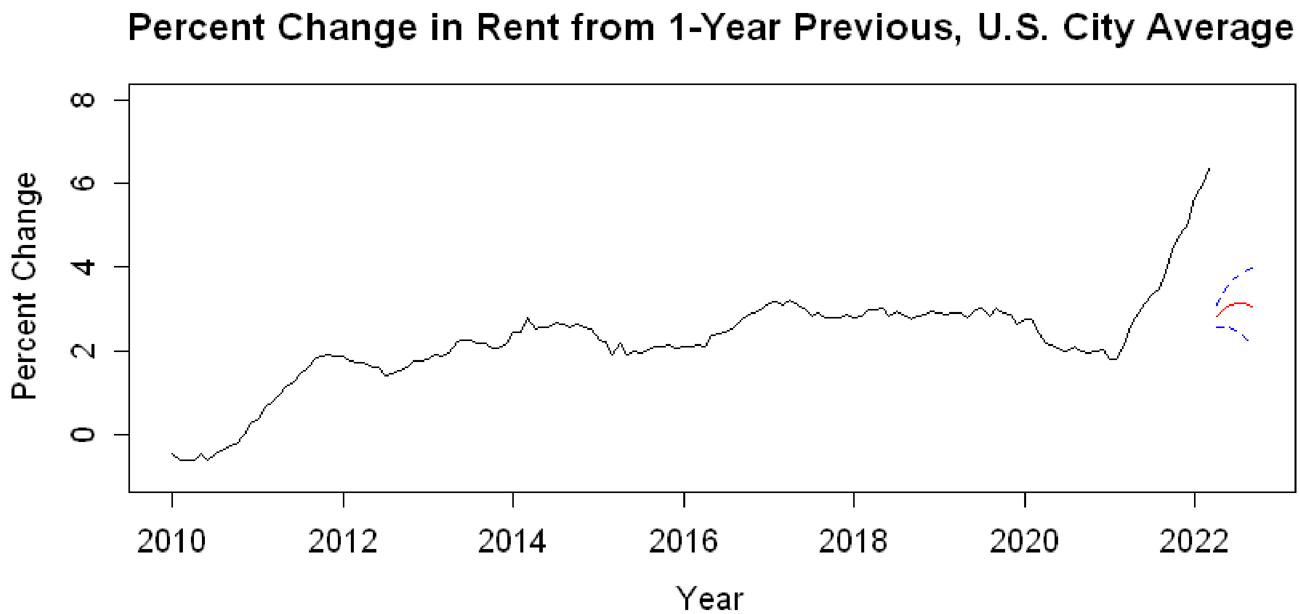
Best Fit Model	
Region:Pop	ARMA(p,q)
S:A	ARMA(2,1)
S:B/C	ARMA(2,0)*
MW	ARMA(2,1)
MW:A	ARMA(2,1)
MW:B/C	ARMA(2,1)
W	ARMA(2,1)
W:A	ARMA(3,0)
W:B/C	ARMA(2,1)

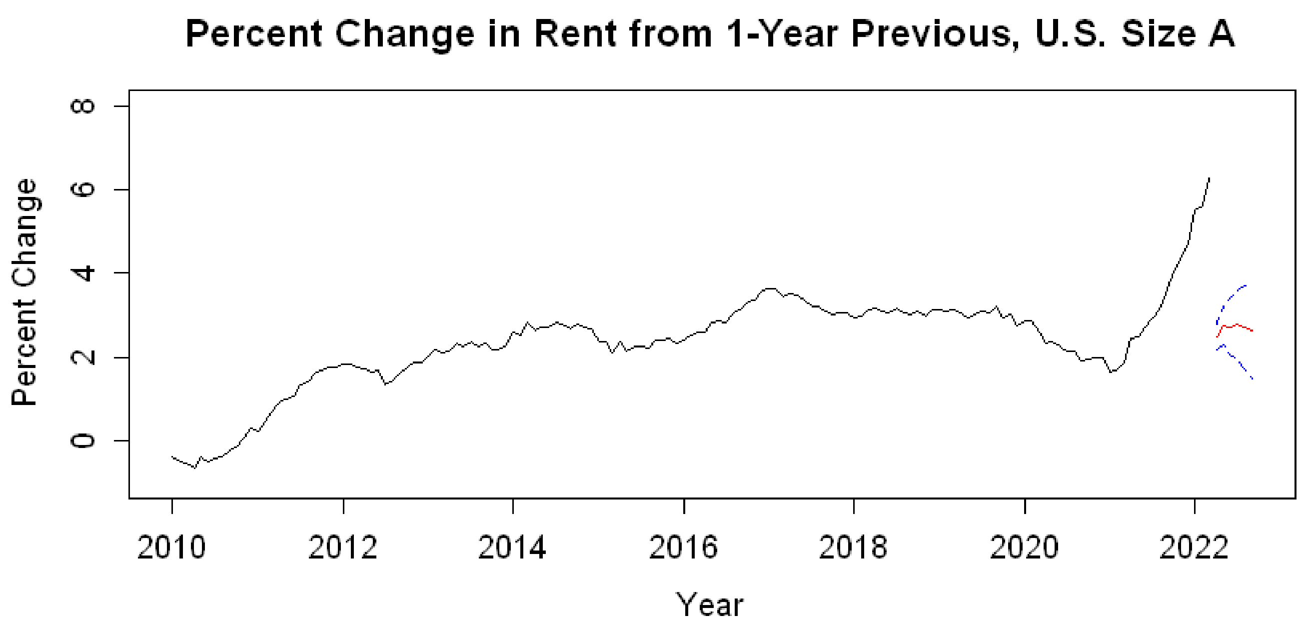
*Note ARMA(2,1) has an AICC value of -40.3 which is smaller than the AICC value for ARMA(2,0) (-40.0) however since these values are so close, I chose the simpler model. Other such choices could have been made, but in the interest of maintaining a cohesive methodology, this is the only instance where the model without the smallest AICC value was chosen.

The parameter values for $\phi_1, \phi_2, \theta_1, \theta_2$ and σ^2 for each model can be found in the R code notebook. Additionally, each set of residuals was checked using the Ljung-Box portmanteau test, which is detailed in the R code notebook, showing that each model is a good fit for the given series'. Since model diagnostics have passed, we can begin forecasting future values.

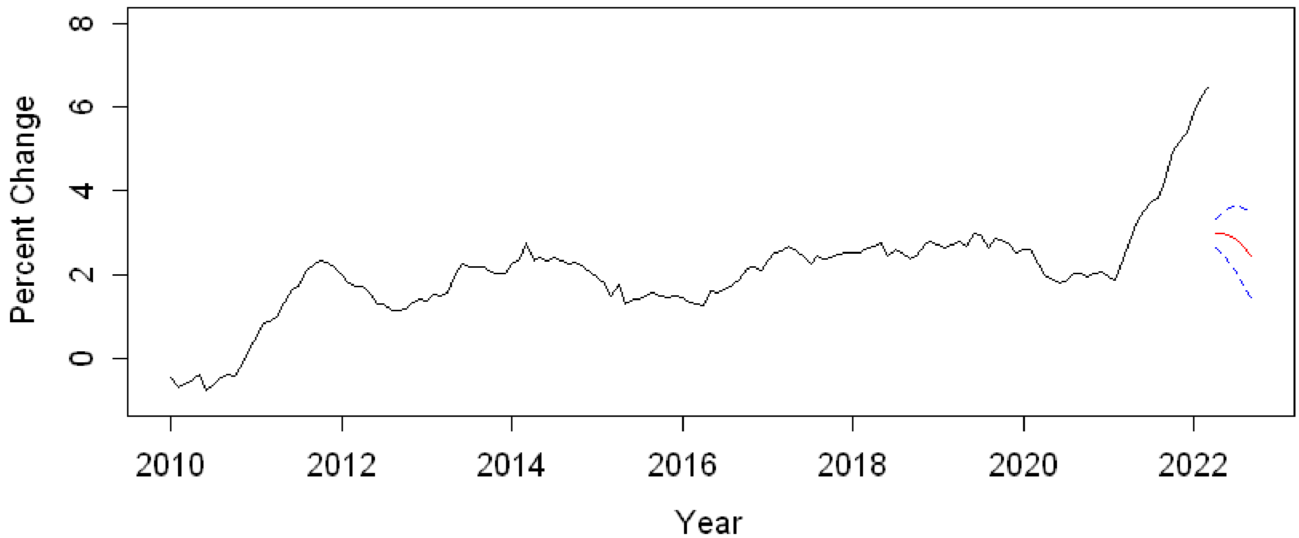
4.3 Forecasting

The model diagnostics for each set of residuals with ARMA models fitted to them have passed, which allows us to forecast future values. I decided to forecast the next 6 months, to September 2022. Below are visuals of the forecasted values, with a 95% accuracy margin.

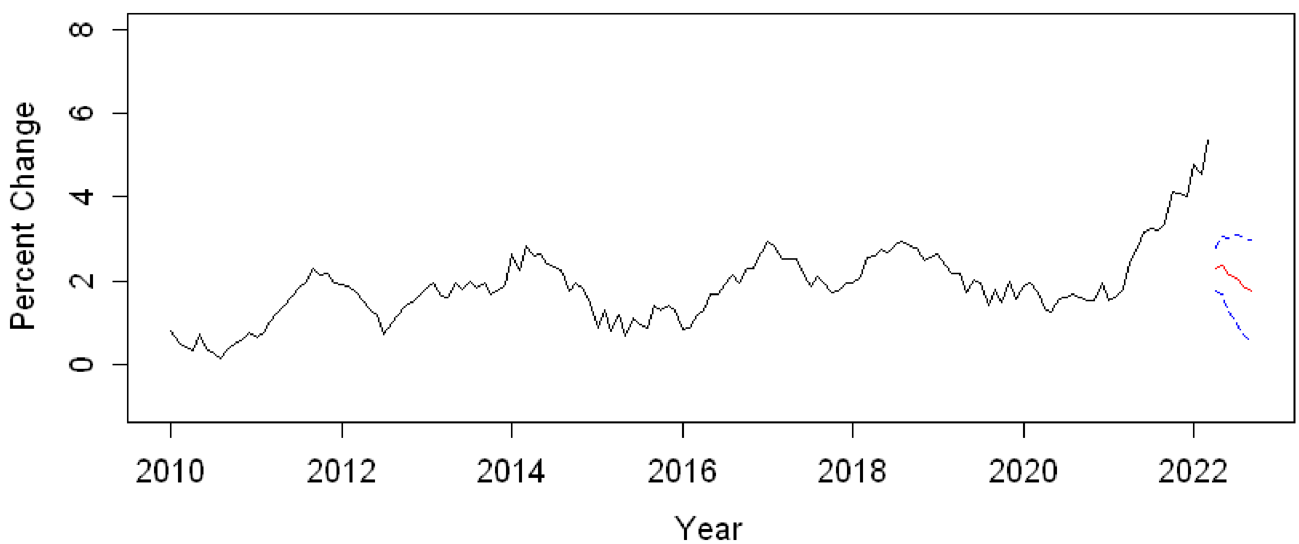


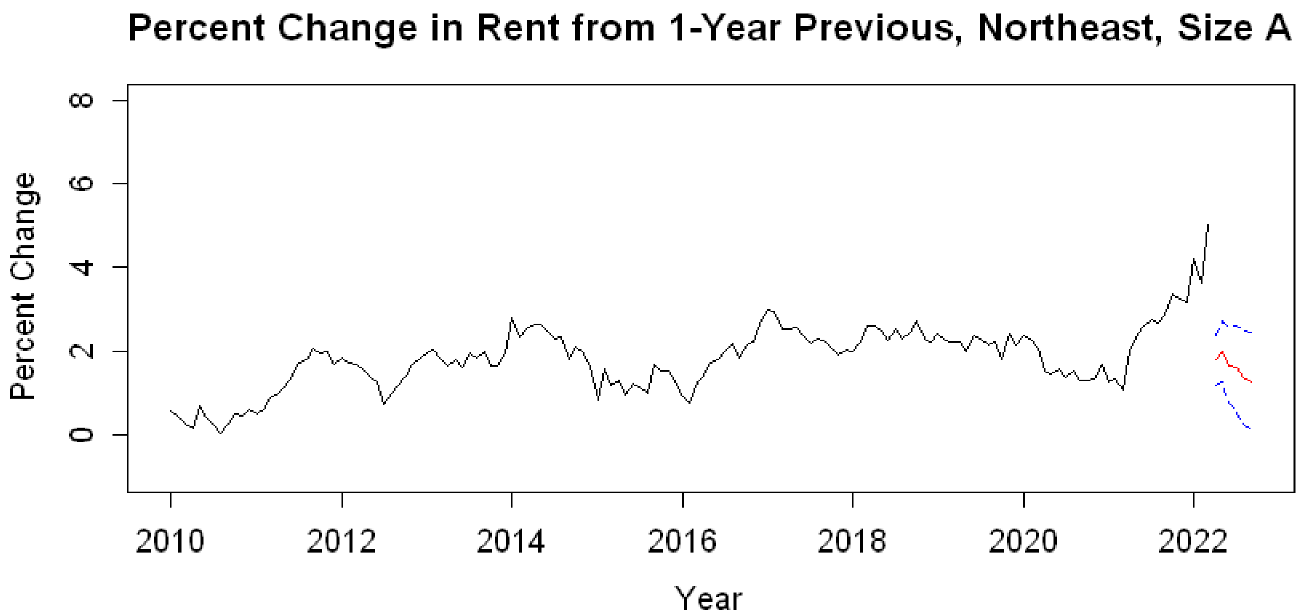


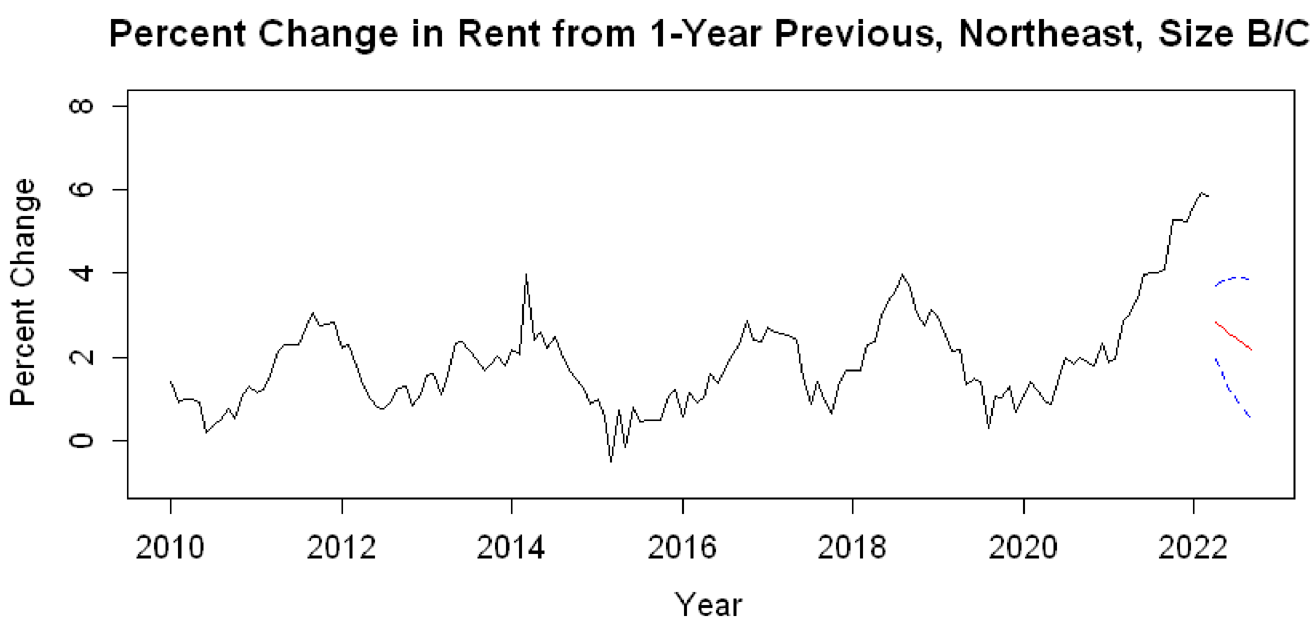
Percent Change in Rent from 1-Year Previous, U.S. Size B

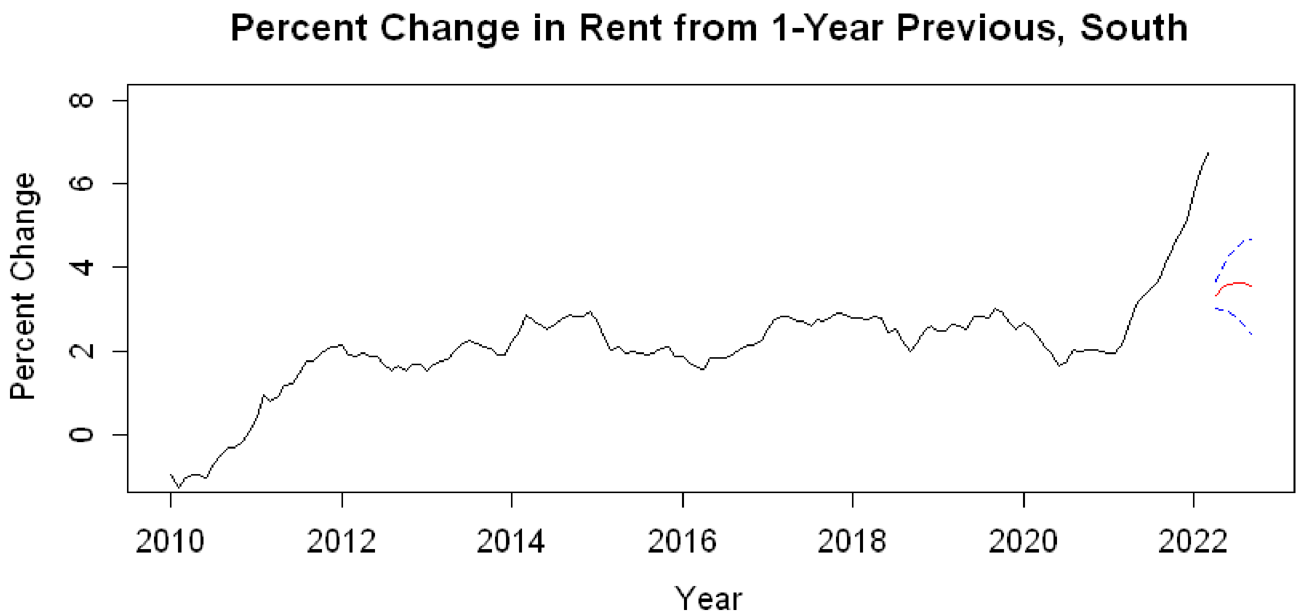


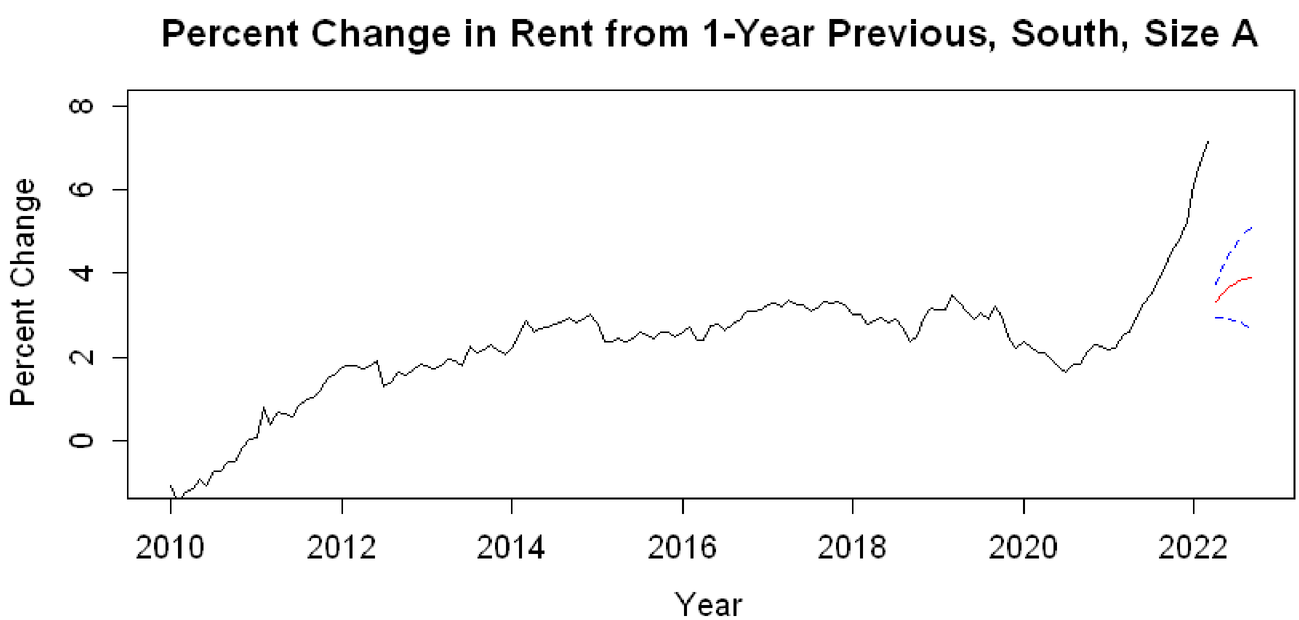
Percent Change in Rent from 1-Year Previous, Northeast



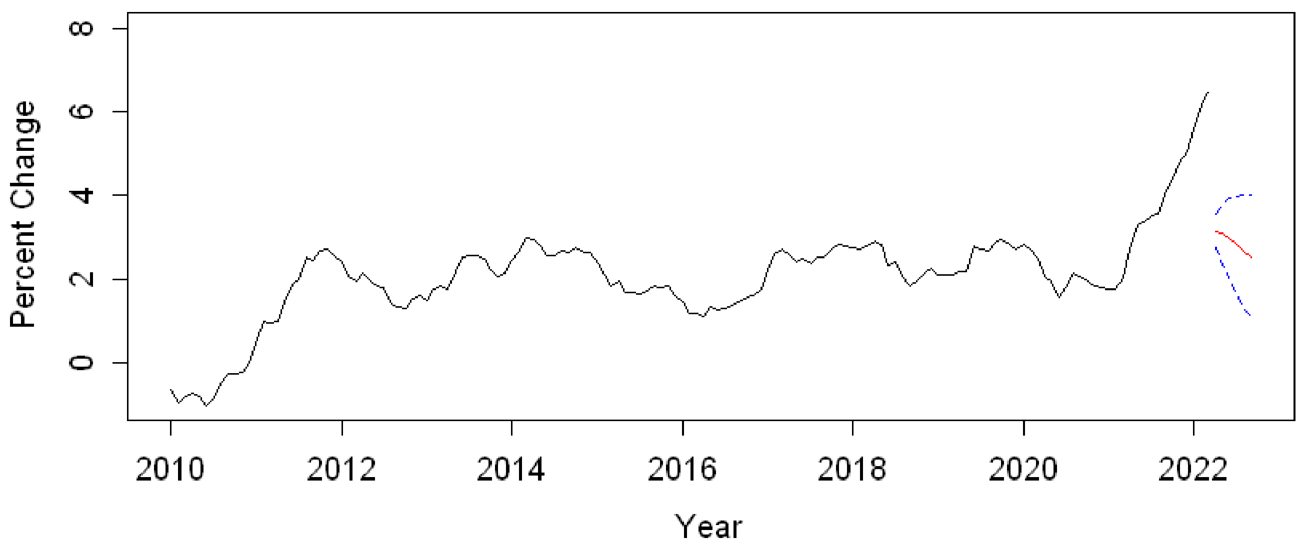

Percent Change in Rent from 1-Year Previous, Northeast, Size A
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year


Percent Change in Rent from 1-Year Previous, Northeast, Size B/C
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year

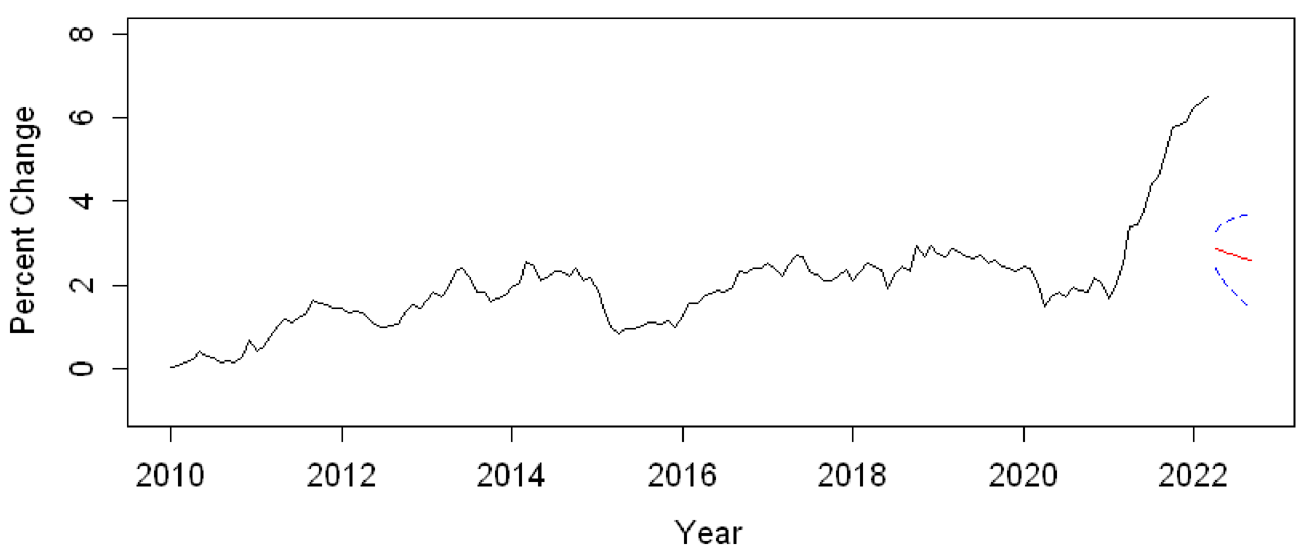

Percent Change in Rent from 1-Year Previous, South
Percent Change
8
6
4
2
0
2010
2012
2014
2016
2018
2020
2022
Year

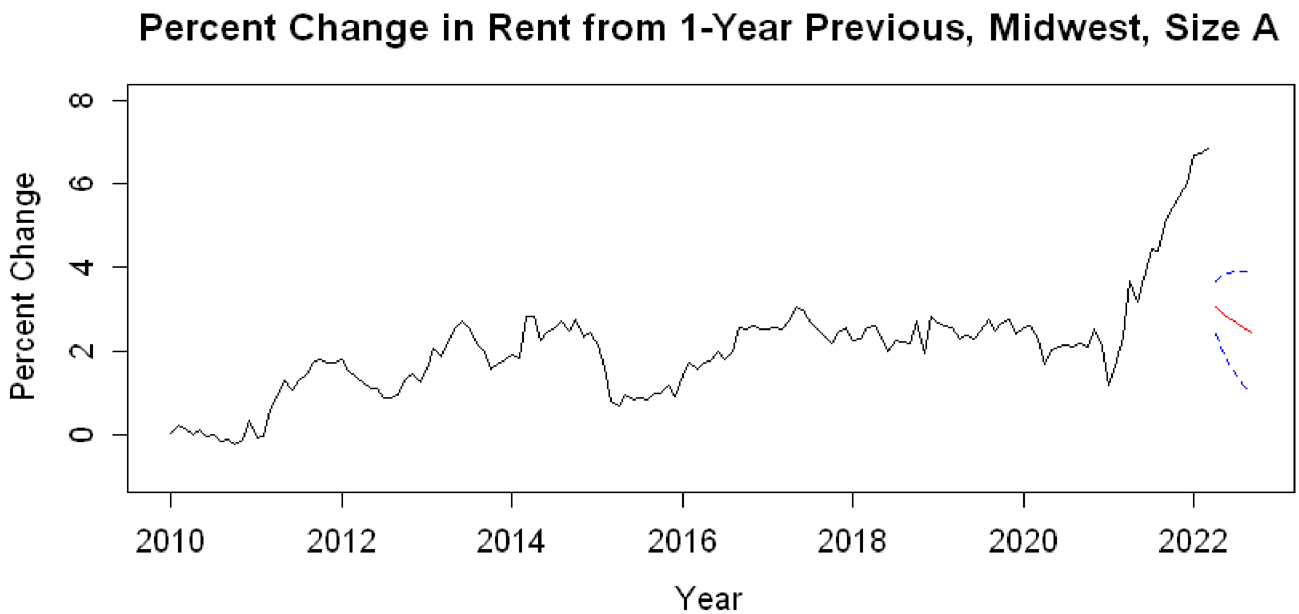

Percent Change in Rent from 1-Year Previous, South, Size A
Percent Change
8
6
4
2
0
2010
2012
2014
2016
2018
2020
2022
Year

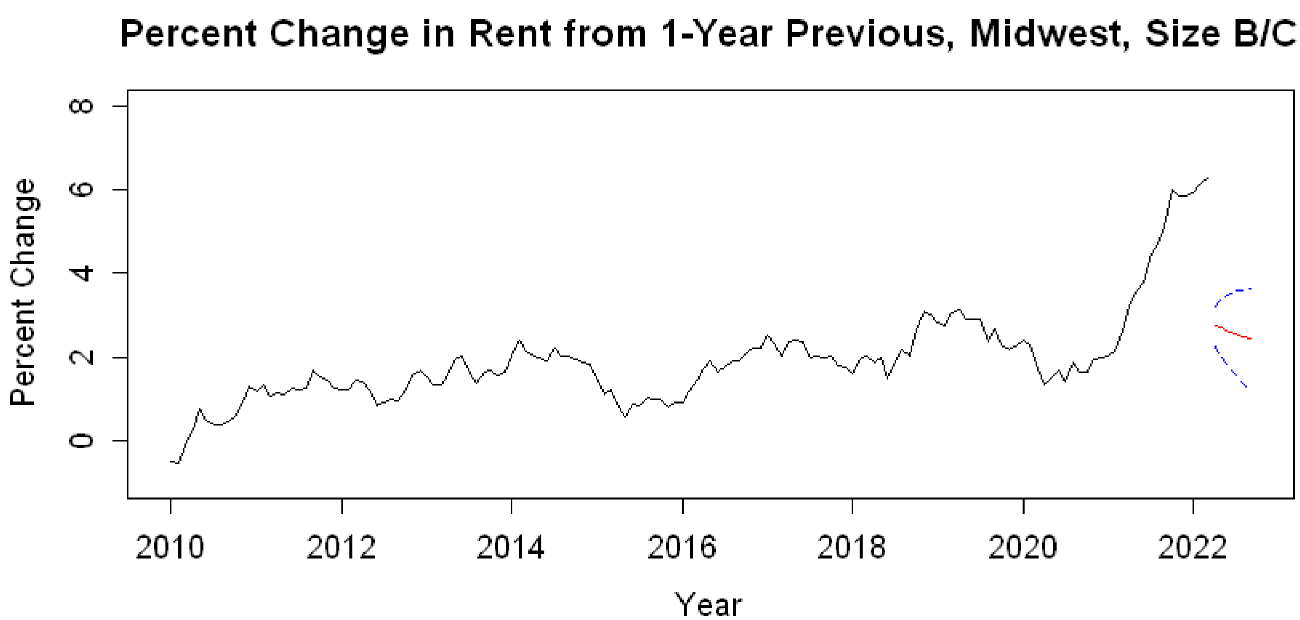
Percent Change in Rent from 1-Year Previous, South, Size B/C

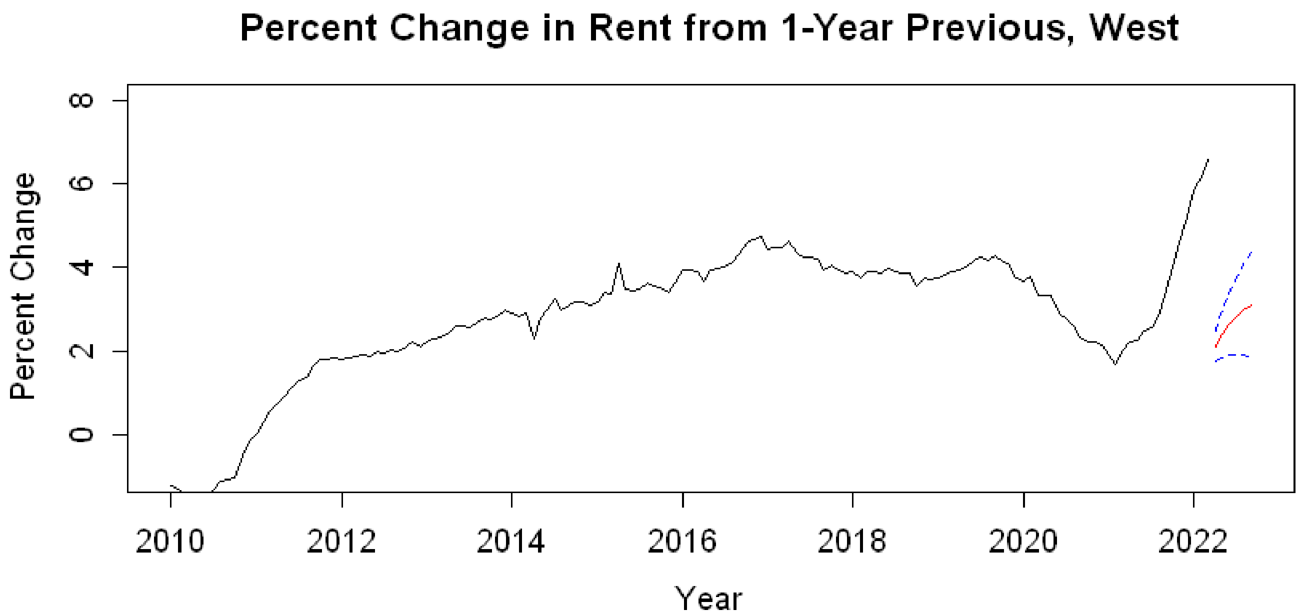


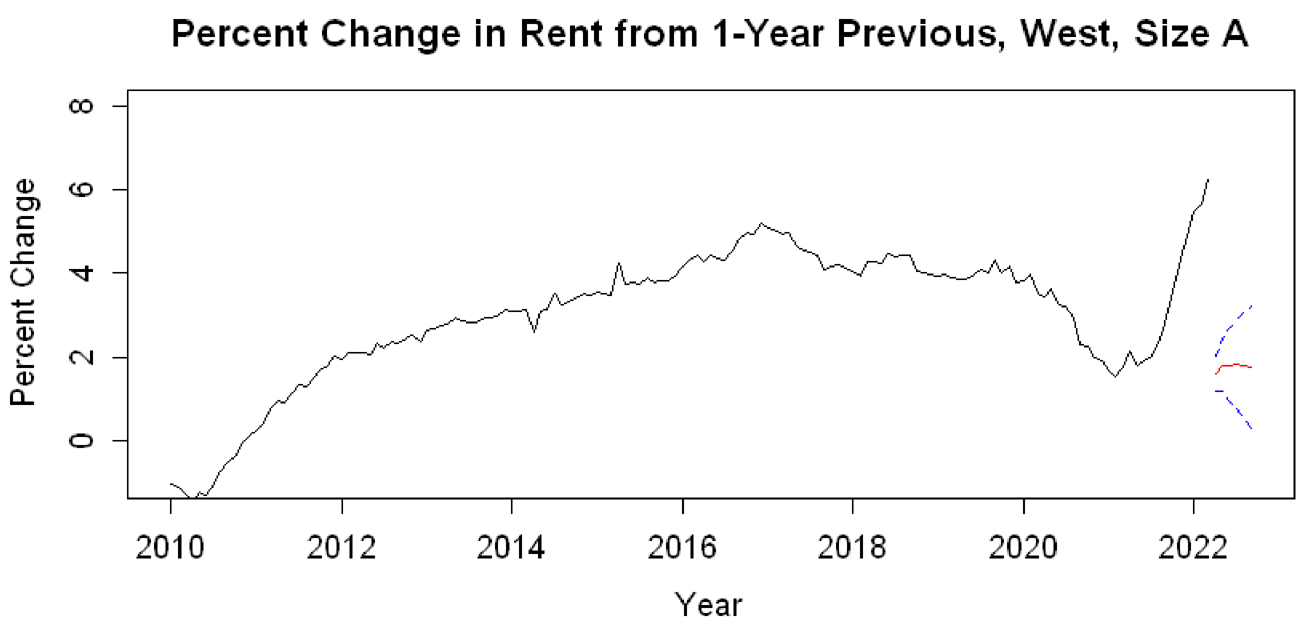
Percent Change in Rent from 1-Year Previous, Midwest

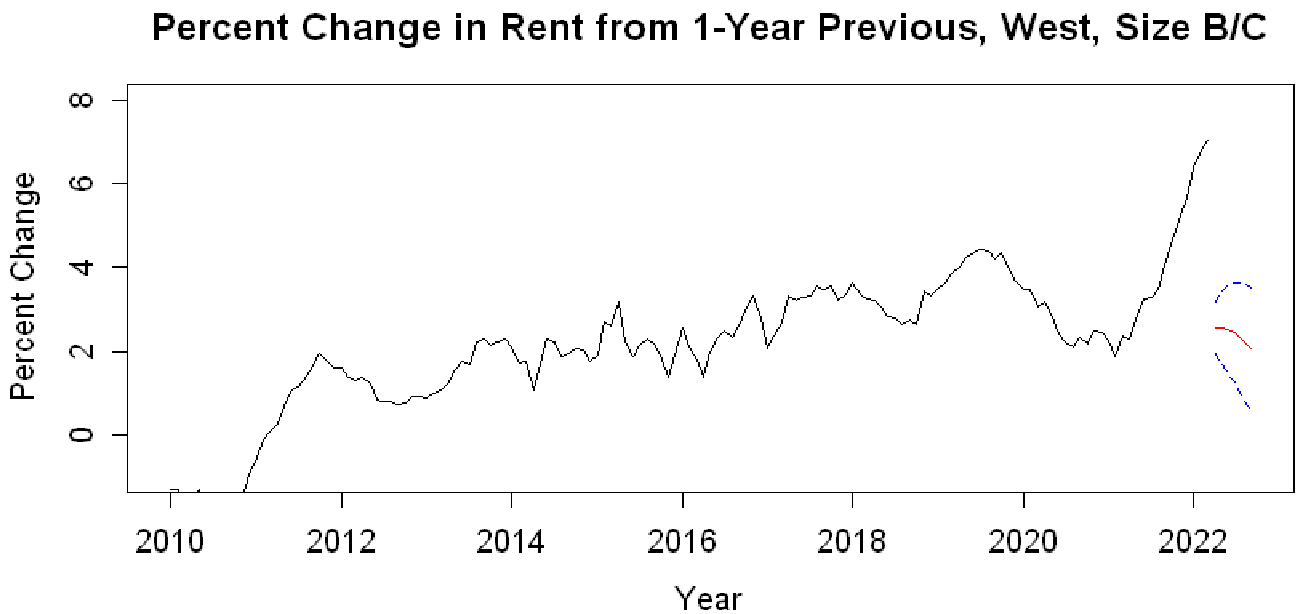



Percent Change in Rent from 1-Year Previous, Midwest, Size A
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year


Percent Change in Rent from 1-Year Previous, Midwest, Size B/C
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year


Percent Change in Rent from 1-Year Previous, West
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year


Percent Change in Rent from 1-Year Previous, West, Size A
Percent Change
0
2
4
6
8
2010
2012
2014
2016
2018
2020
2022
Year



A summary of the predicted values and errors is given in the table below.

5 Conclusions and Discussions

From these results we can make the following observations, to answer our research questions.

- The national data suggests population size B/C regions had higher overall percentage growths than size A.
- The regions with overall lower percentage growth than the national growth were Northeast, significantly, and South, slightly. Midwest and West both had higher than the national average trends to their percent increase over the entire period.
- The smallest growth, was Northeast population size A, and the largest growth was West size B/C.
- The projected future values all suggest a decrease must occur soon, since the last several months of extremely rapid increase are unsustainable.

I believe this analysis is useful, and provides an interesting perspective on which regions and population sizes have better control over rent inflation, and which are experiencing more dramatic change. This could be used for further analysis at county and city level within each region, and an examination of policy within regions, to determine causes.

This project started out with a smaller scope, of analyzing the trends of Boise rent price increase, and comparing them to the national prices. The available data was limited for that specific information since Census data for median rent prices was yearly, and only went back to 2010. More analysis could be done on that subject, with more time to access private data at the county level. Comparing that data to the general economic data for Boise [1] could be an interesting direction of further research of this type.

References

[1] Boise City Area Economic Summary. Bureau of Labor Statistics. Accessed 04-18-22.

[2] Explore Census Data. U.S. Census Bureau Official Website. Accessed 04-11-22.

[3] Federal Reserve Economic Data. St. Louis FRED official website. Accessed 04-18-22.

[4] Idaho residential rent and rental statistics. Department of Numbers. Accessed 04-14-22.

[5] Ryan Suppe. Boise rents went up faster than anywhere in the us in the past year. BoiseDev, Feb 2021. Accessed 04-03-22.