Zillow Home Value Index Time Series Analysis

STAT 626
Summer 2023
Team 4

Agenda

- Group Introduction

- Review Data and Transformations

- Initial Model Fit and Diagnostics

- Additional Model Fit and Diagnostics

- Conclusion

Team Members

Elizabeth Storm

MS Statistics
Graduating May 2024
St Paul, MN
Actuary in P&C
Insurance

Theory

Allison Moore

MS Statistics
Graduating May 2024
Washington, D.C.

Programing

Jialin Lai

2nd year doctoral
Student, Curriculum
and Instruction

College Station

Writing

Rashmi Sai Burugupalli

MS Data Science

Graduating August 2023
College Station

Data Visualization

Nick Klein

MS Statistics
Graduating May 2024
Iowa

Computation

Recap of Data and Project Goals

Complete a time series analysis to use historical home prices to forecast future home values.
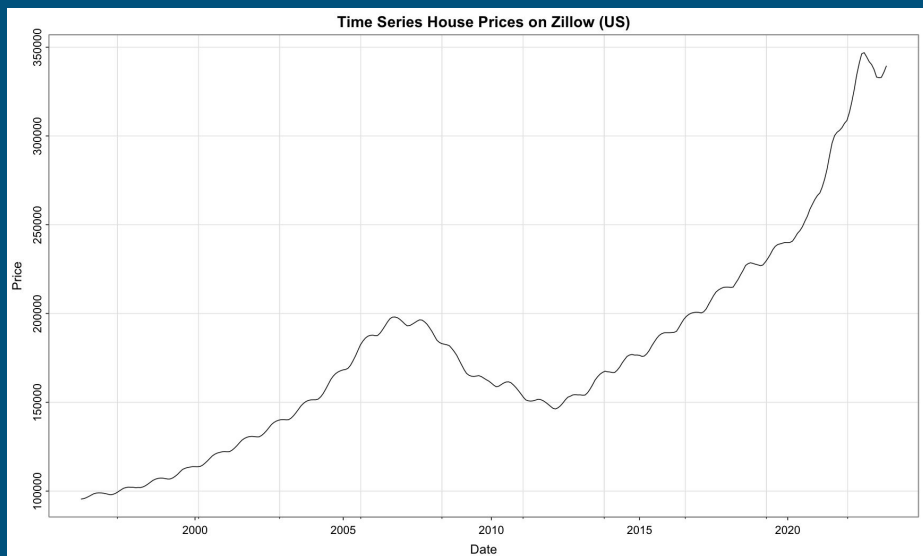
Application Areas:

- Buyers / Sellers
- Economists
- Financial Institutions

Zillow Home Value Index (ZHVI):
- A measure of the typical home value, reflecting the typical value for homes in the 35th to 65th percentile range
- Summarized across all home types (Single-family residence, condo & co-op)

Exploratory Data Analysis

Time Series House Prices on Zillow (US)

- Long term increasing trend
- Yearly cycles (January through August)
- Prior to 2020, minimal short-term volatility

Time Series House Prices on Zillow (US)
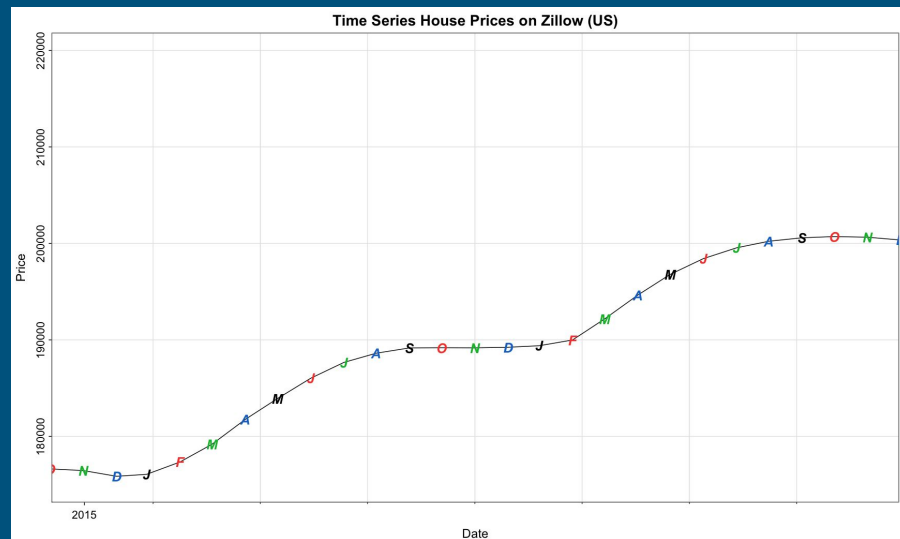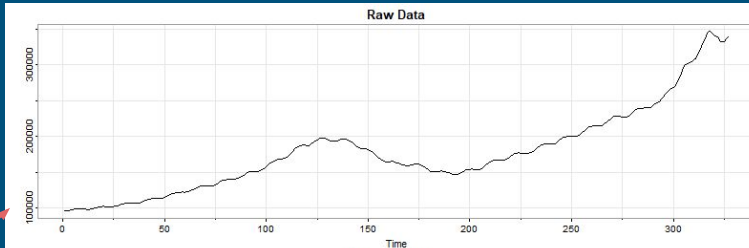
Data Transformations

1. **Raw**
   a. ADF: 0.35752
   b. P: 0.99

2. **Deseason**
   a. ADF: -0.022547
   b. p: 0.99
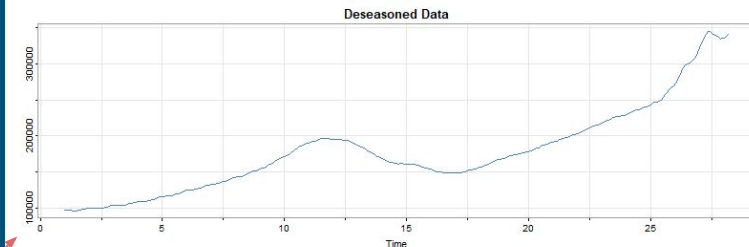
3. **Detrend**
   a. ADF: -4.885
   b. P: 0.01

NOT STATIONARY

NOT STATIONARY
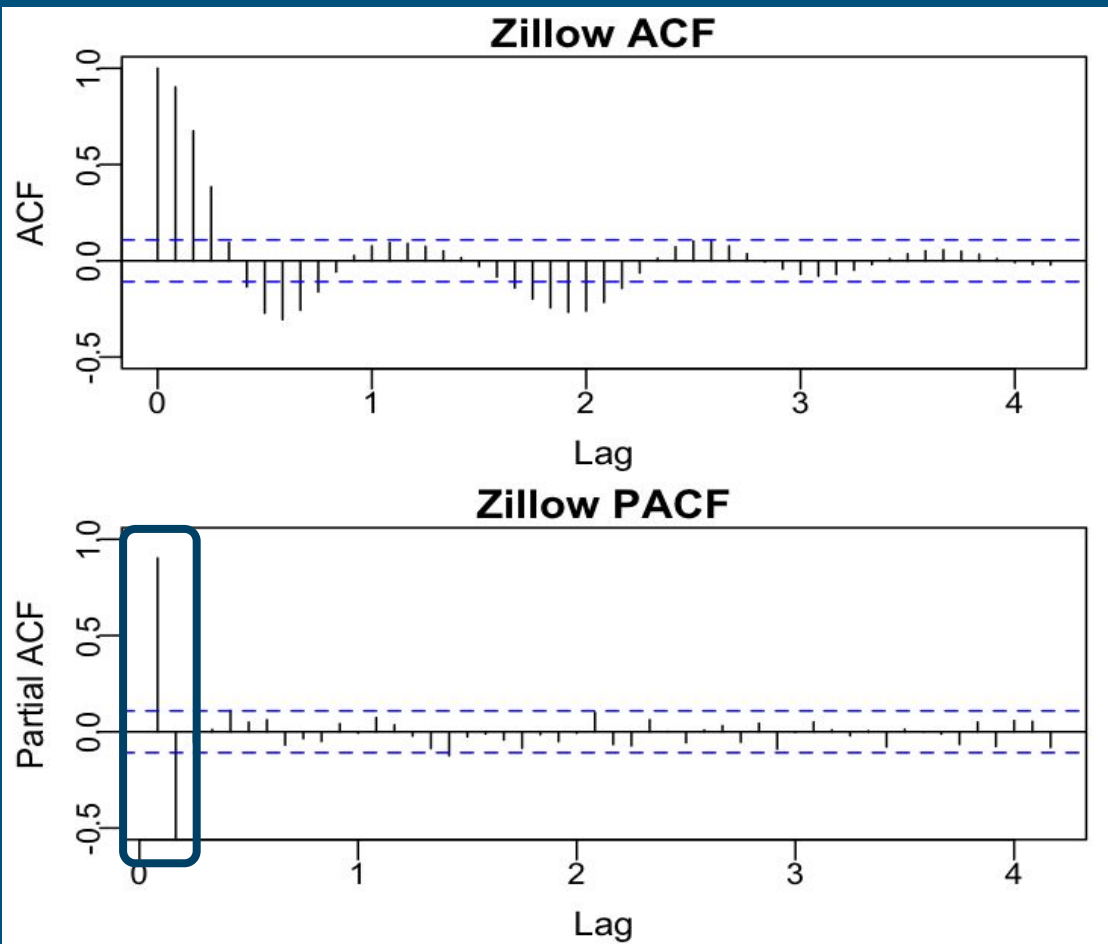
STATIONARY
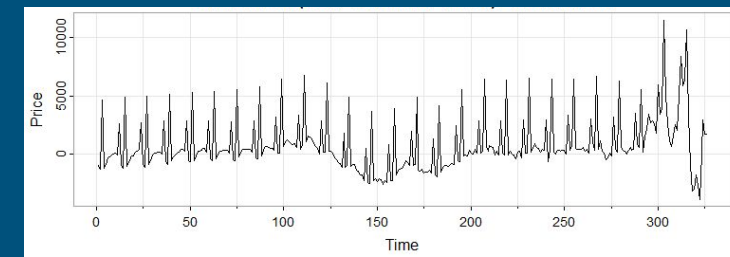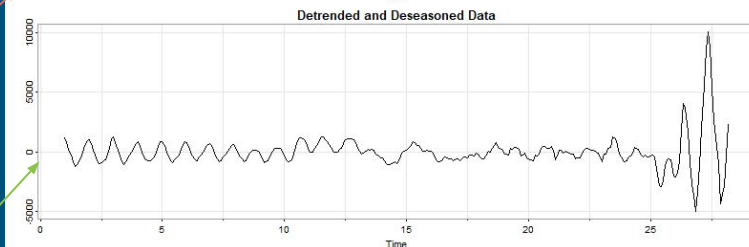
Data Transformations
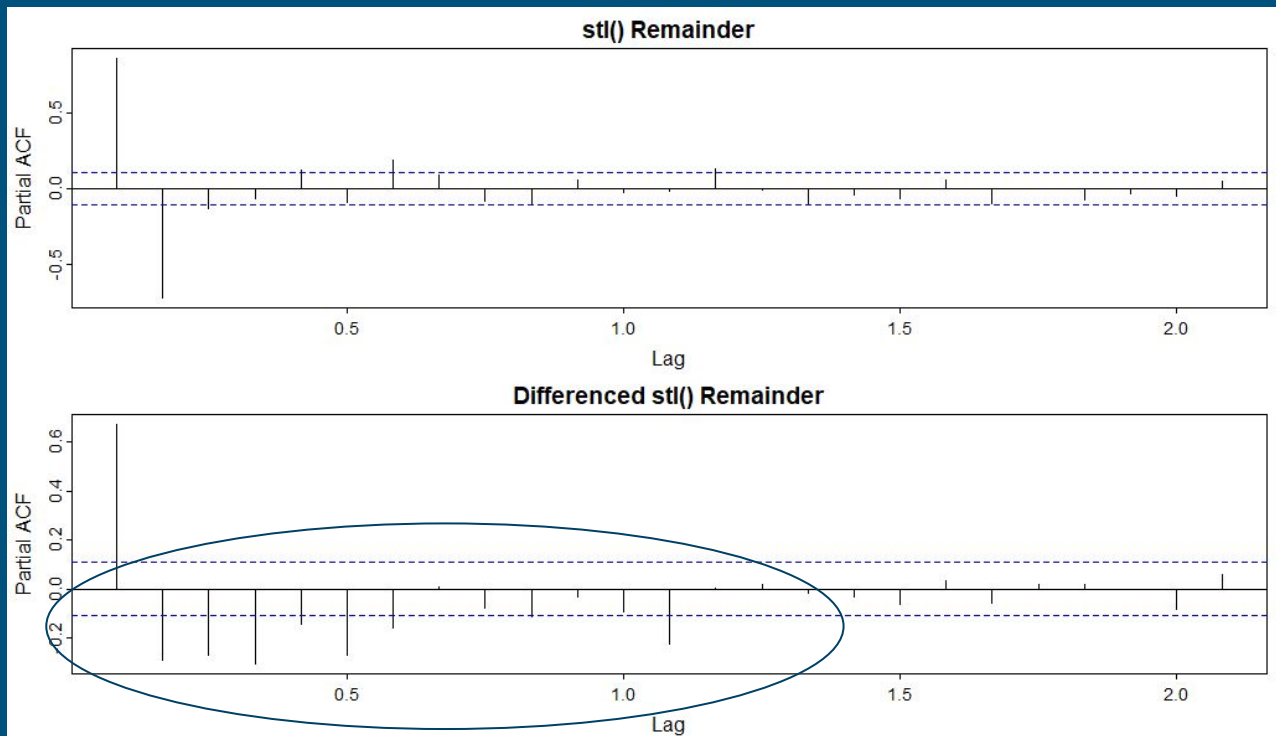
Correlograms

- ACF trailing off
- PACF clear cut after h=2
- Indicates AR(2) as a good starting point

Coefficients from AR(2)

- Fit using the SARIMA function

- Model produced:

$$x_{t=1.5801_{(0.0363)}}x_{t-1} - 0.752_{(0.0364)}x_{t-2} + w_t$$

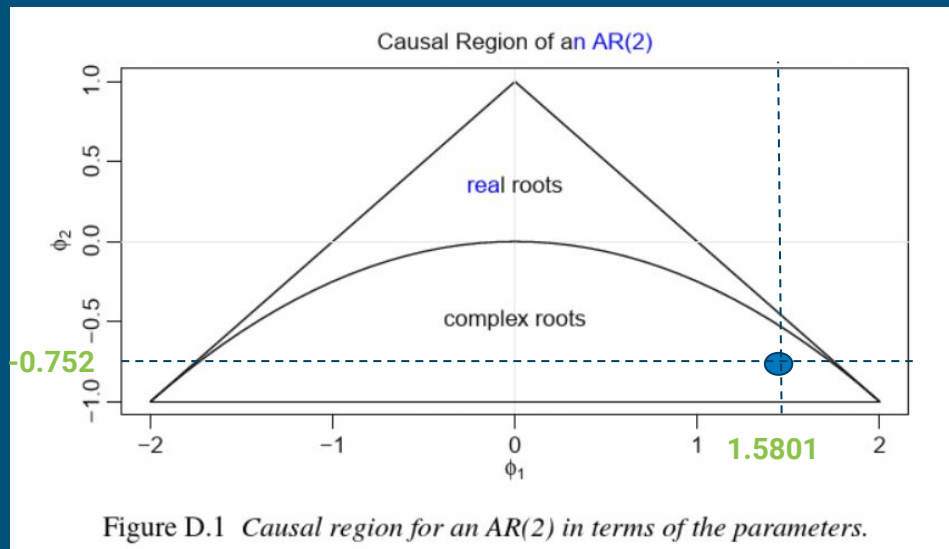- SARIMA model produces two AIC values: 4828.52 and 14.766

Unit Root Computation

Finding the Roots

$$x_{t=1.5801_{(0.0363)}} x_{t-1} - 0.752_{(0.0364)} x_{t-2} + w_t$$
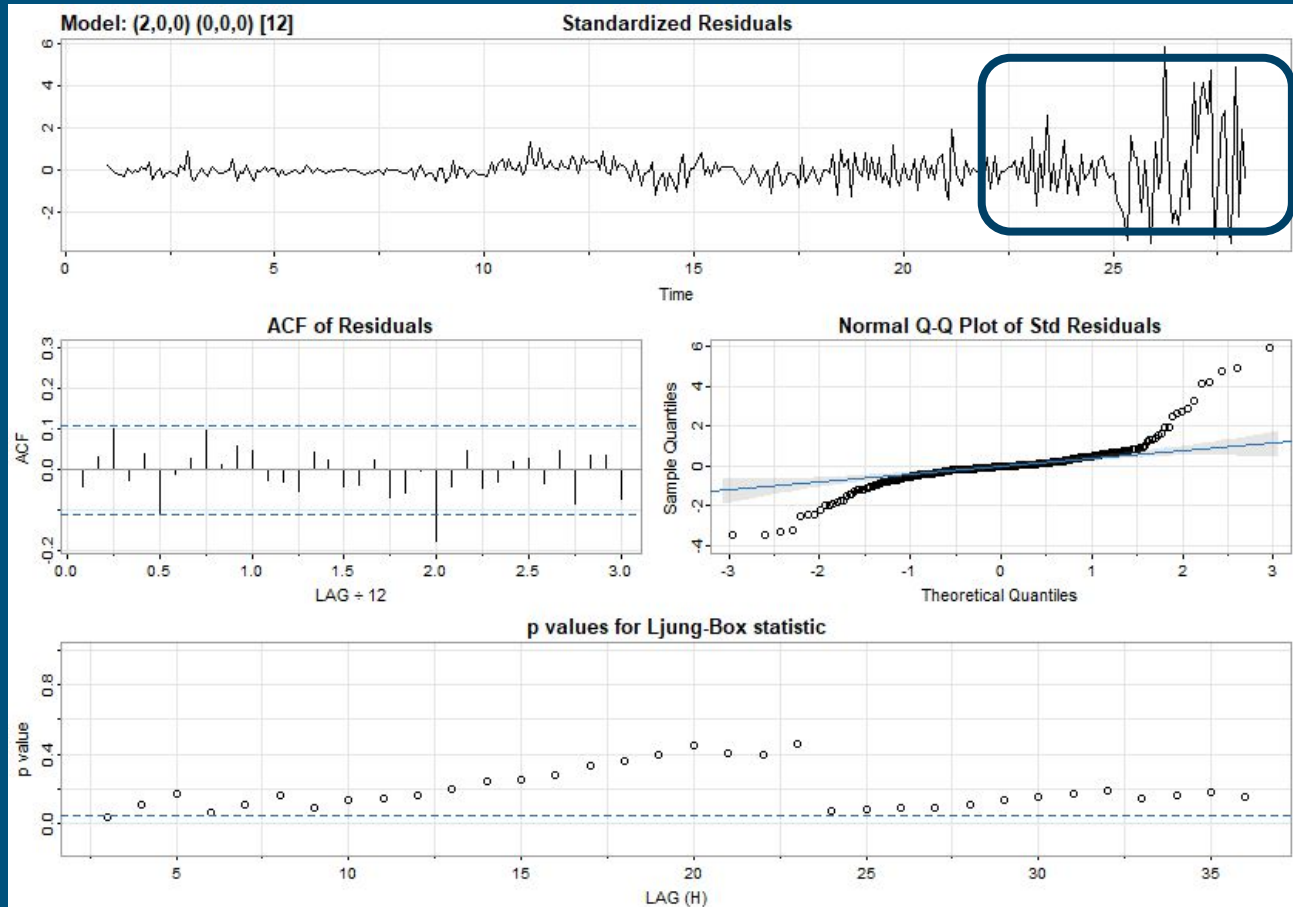
$$(1 - 1.5801B + 0.752B^2)\hat{x}_t = w_t$$

Z = 1.0506 ± 0.4754i

Complex number but conclude both are great than 1, outside the unit root, and therefore AR(2) model is causal

Figure D.1 *Causal region for an AR(2) in terms of the parameters.*

Diagnostics

- Plot of residuals
- ACF
- QQ plot
- Ljung-Box/Q test

Forecasts

Figure A

Figure B

Zillow Vaue Index- House Price Forecasting

Figure C

Zillow Vaue Index- House Price Forecasting
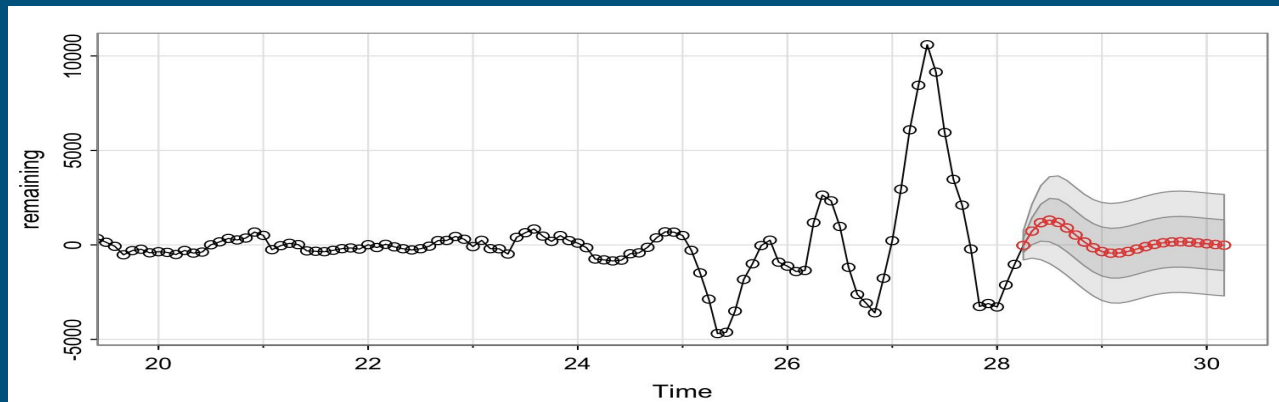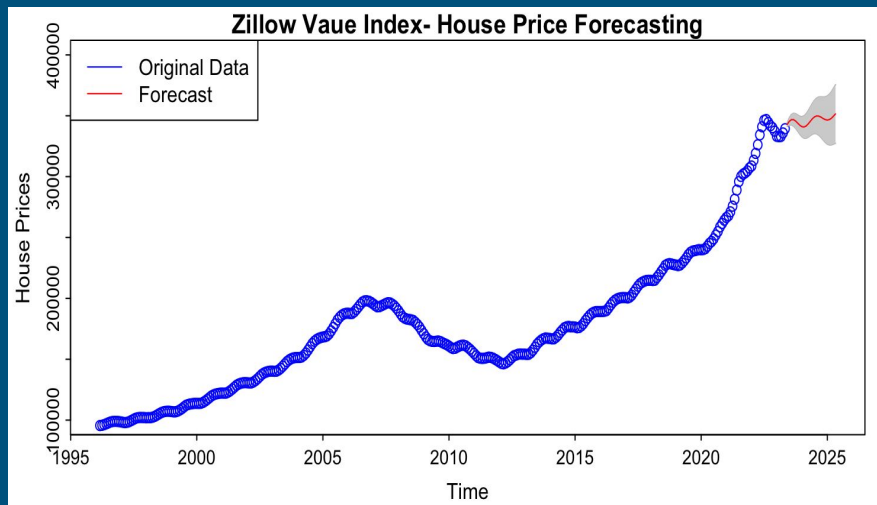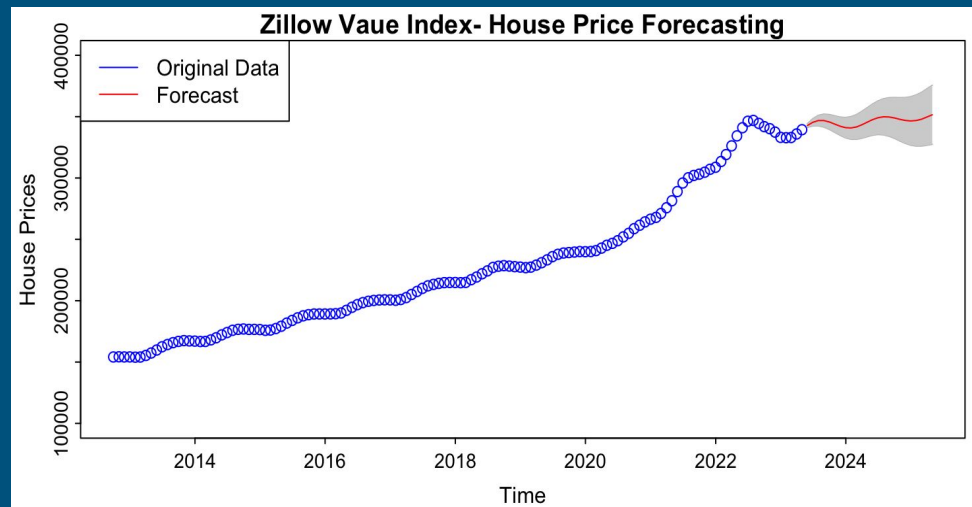
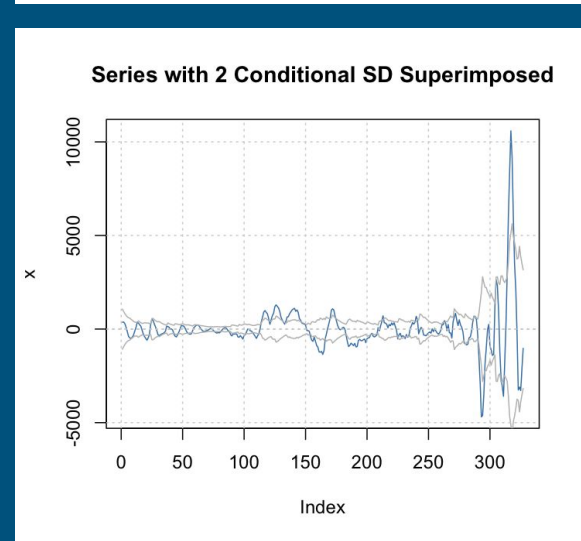# Additional Models
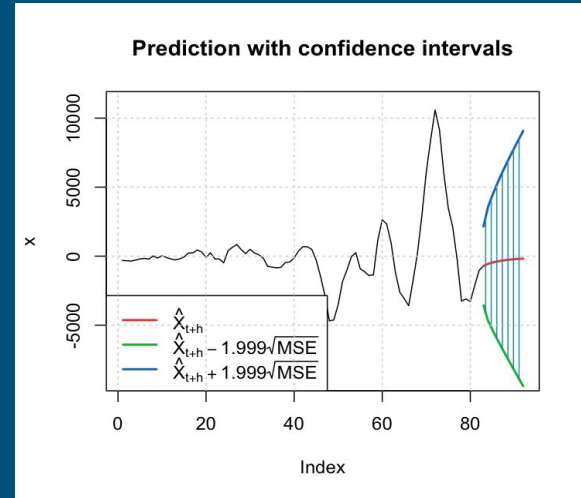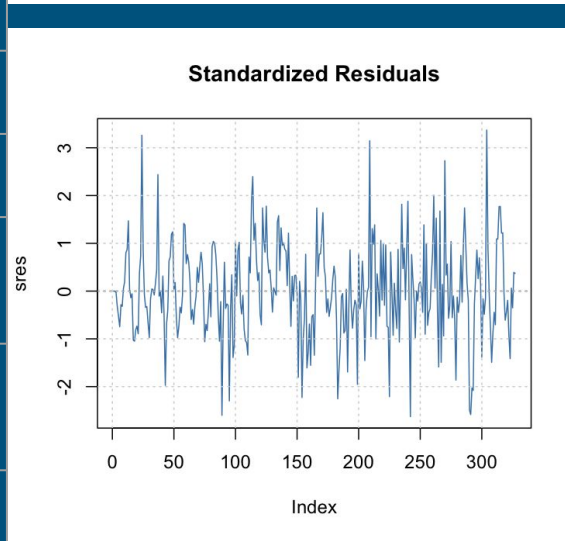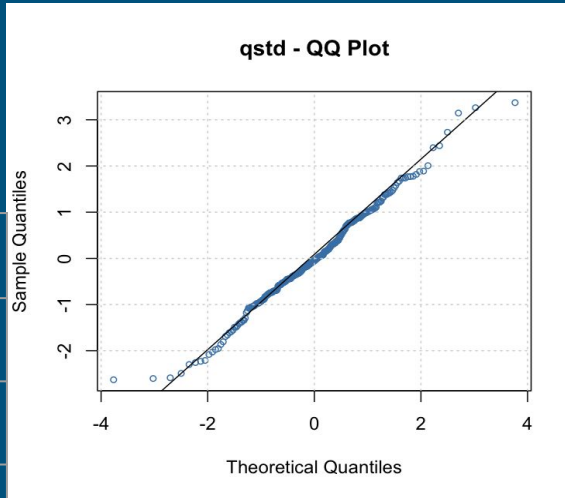
GARCH

Prophet

Piecewise

GARCH

Why?

- Evidence of increasing variance in AR(2) diagnostics

- Evidence of squared returns being autocorrelated
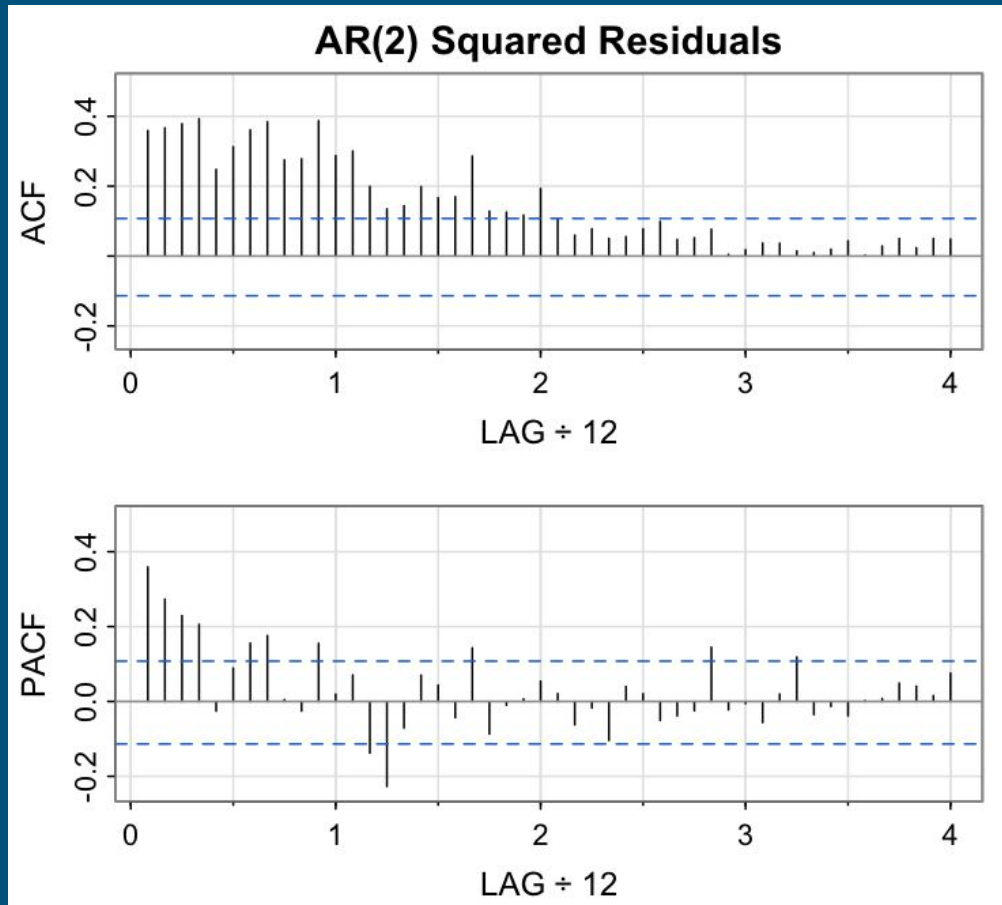
GARCH Fit

Test	Statistic
AIC	13.65
BIC	13.73
Ljung-Box (R,Q(10))	52.58*
Ljung-Box (R,Q(15))	62.45*
Ljung-Box (R,Q(20))	64.71*
Ljung-Box (R^2,Q(10))	7.73
Ljung-Box (R^2,Q(15))	11.08
Ljung-Box (R^2,Q(20))	11.82

qstd - QQ Plot

Prediction with confidence intervals

$\hat{X}_{t+h}$
$\hat{X}_{t+h} - 1.999\sqrt{MSE}$
$\hat{X}_{t+h} + 1.999\sqrt{MSE}$
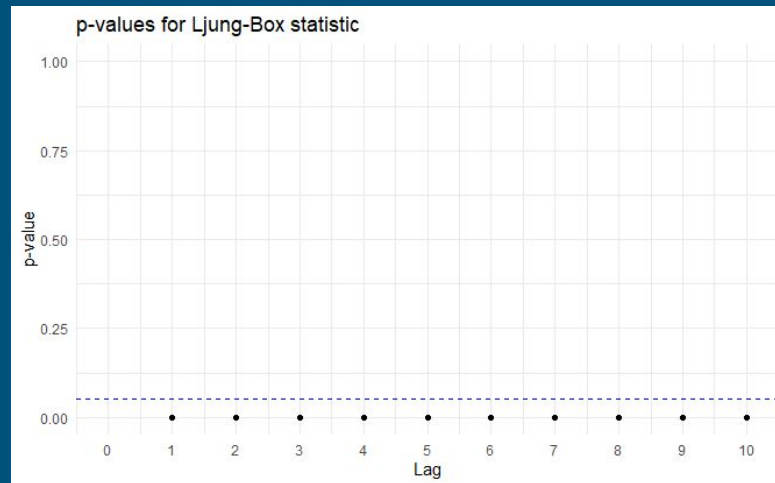

Standardized Residuals
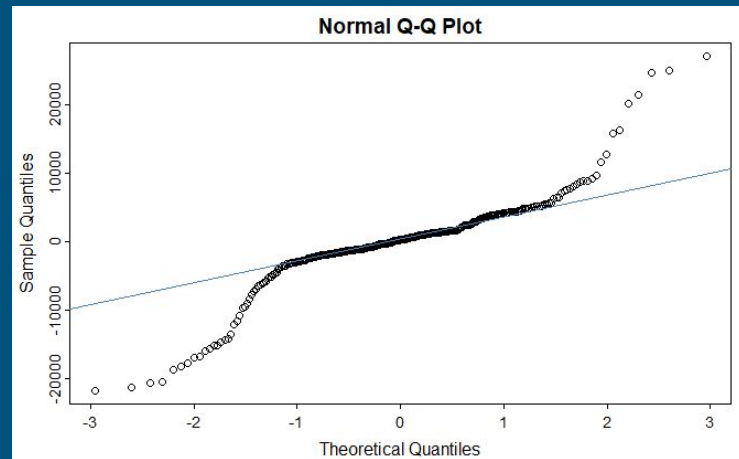

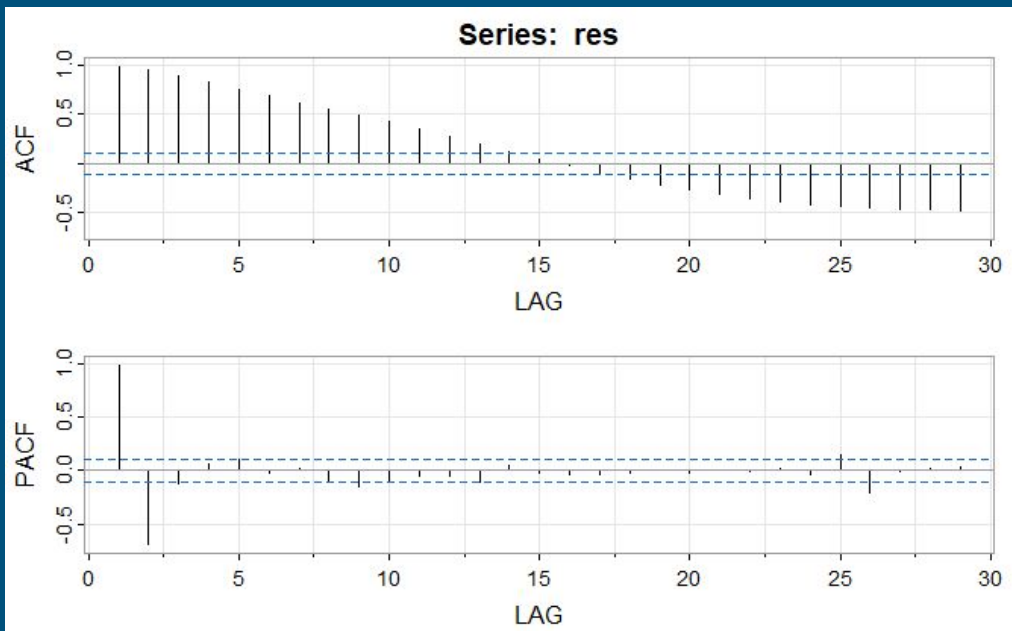Series with 2 Conditional SD Superimposed

Prophet

Transformed Data
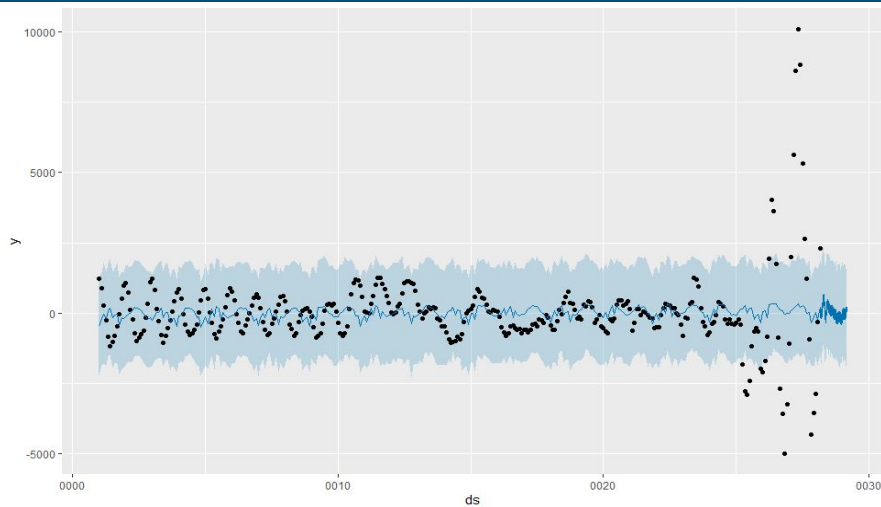
Raw Data
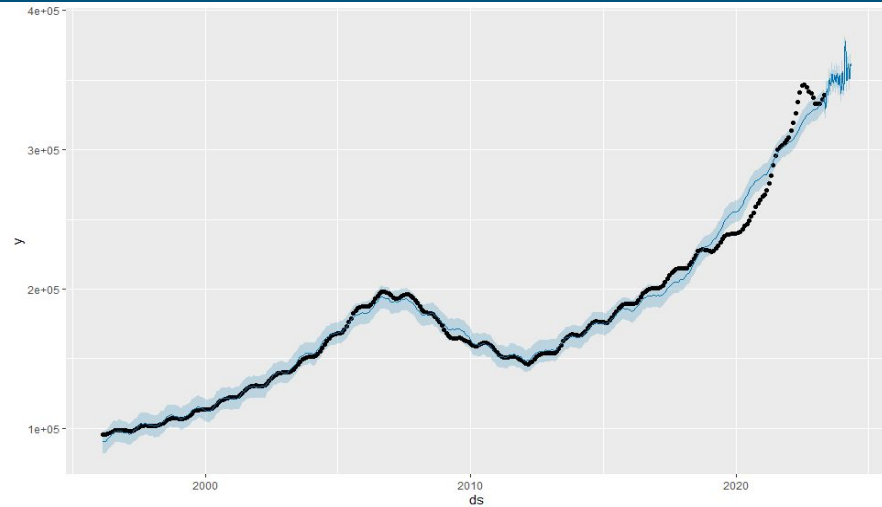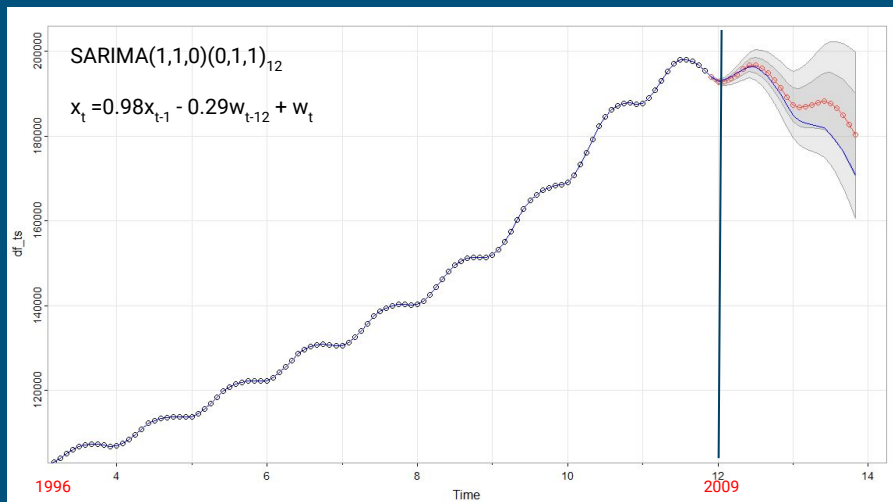
Prophet- Raw Residuals

Piecewise Model

Time Series House Prices on Zillow (US)
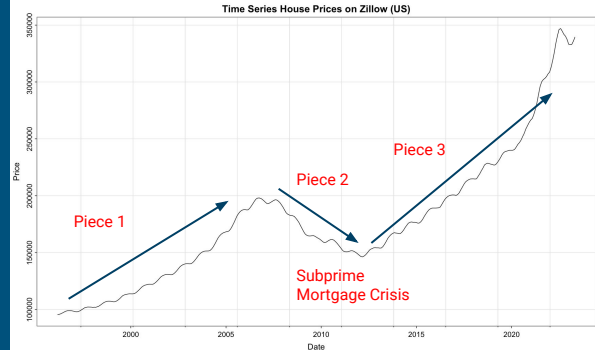
- Considered approach due to the subprime mortgage crisis
- Model fit across different periods are similar

SARIMA$(1,1,0)(0,1,1)_{12}$

$x_t = 0.98x_{t-1} - 0.29w_{t-12} + w_t$

AIC=12.7, ADF p-value = 0.01, Box-Ljung p-value = 0.03

SARIMA$(1,1,0)(0,1,1)_{12}$

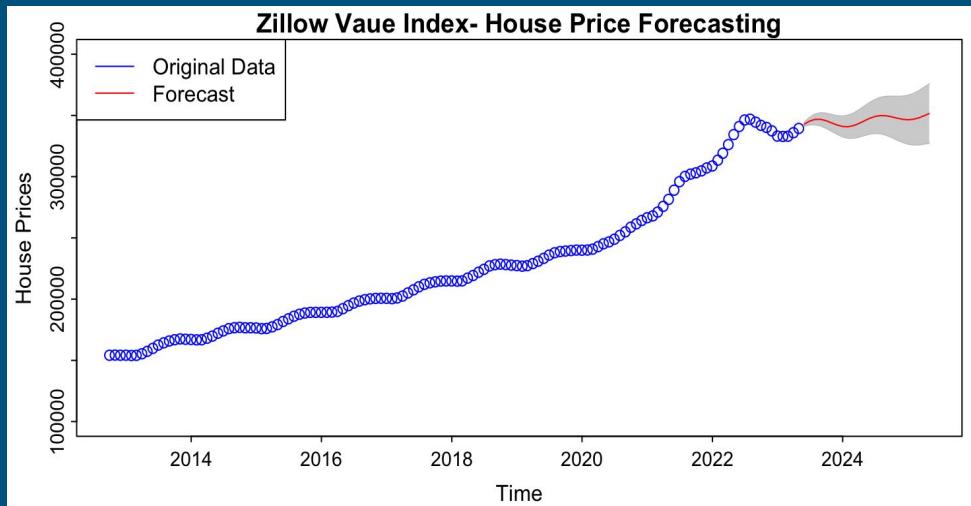$x_t = 0.89x_{t-1} - 0.93w_{t-12} + w_t$
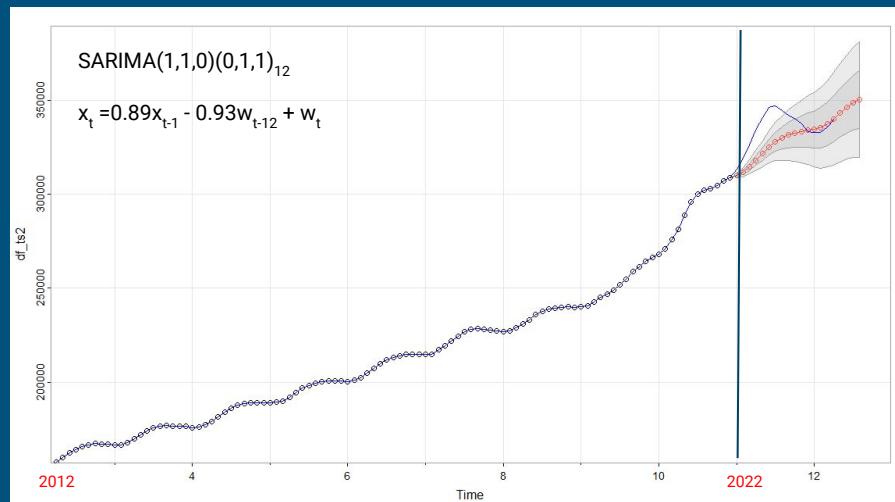
AIC=15.7, ADF p-value = 0.01, Box-Ljung p-value = 0.11

Conclusion

- Over-differencing induces dependencies

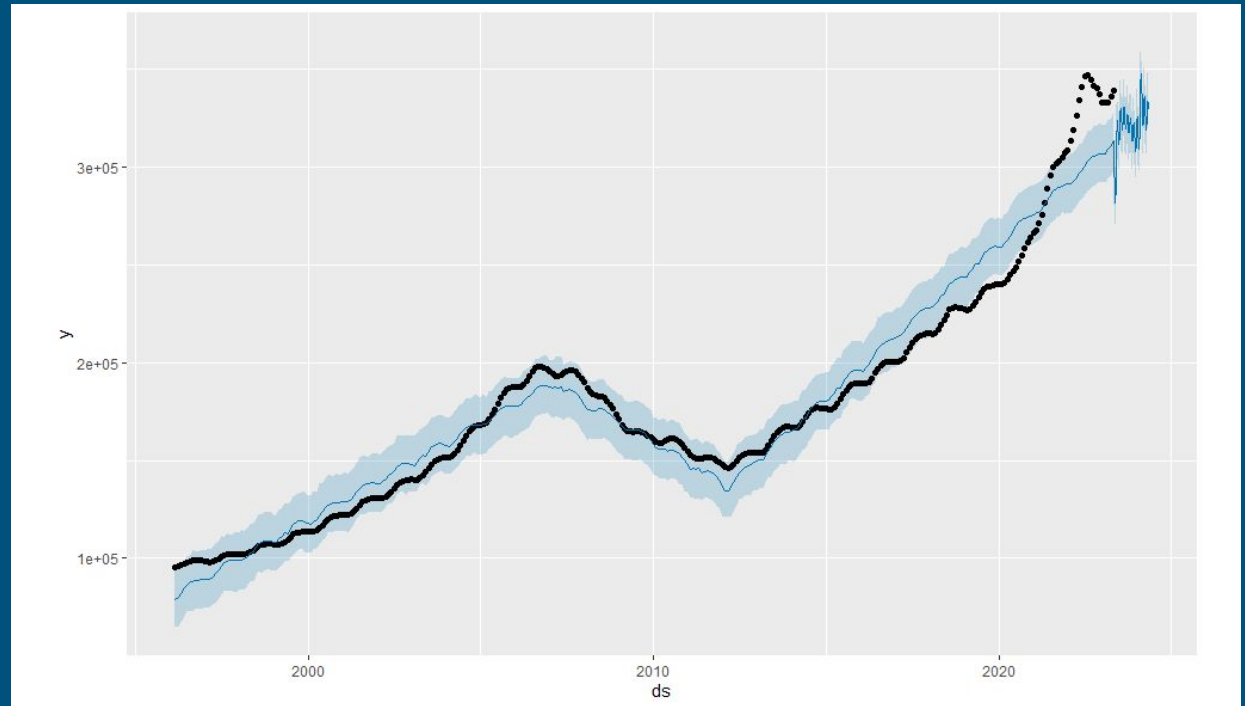- Many tools exist for Time Series analysis

- Keep it simple

Final Model: AR(2)

$$x_{t=1.5801_{(0.0363)}} x_{t-1} - 0.752_{(0.0364)} x_{t-2} + w_t$$

Zillow Vaue Index- House Price Forecasting

Appendix

A Piecewise Prophet (If need more time)

Timelines

M	T	W	Th	F	S	Su
24	**25**	**26**	**27**	**28**	**29**	**30**
		Team Mtg?		*HW6 due*	Team Mtg 3pm Potentially record	
		Addition Modeling & Analysis		Building Presentation		
31	**Aug 1**	**2**	**3**	**5**	**6**	**7**
Study for exam	Exam 7-8:30	Presentations 1-3; 7-8:30pm Back up Team Mtg 6pm Finalize paper	Presentations 4-6; 7-8:30pm	Presentations 7-10; 7-8:30pm **PAPER DUE**		

Timing for Presentation

4 min on Background & Data Transformation

5 min on Best fitting model including forecasts

   2 min on Unit Root

2-3 min Garch

2-3 min Prophet

2-3 min Piecewise

1 min Conclusion