Statistics 5350/7110
Forecasting

Lecture 21
Building SARIMA Models

Professor Robert Stine

Preliminaries

- Questions?
- Assignments
  - Assignment 5 posted
  - Due next Thursday
- Quick review
  - Building SARIMA models
  - Adding explanatory factors
  - Calendar variables

Lecture_20.Rmd & Lecture_21.Rmd (today)

Today's Topics

Text, Chapter 5

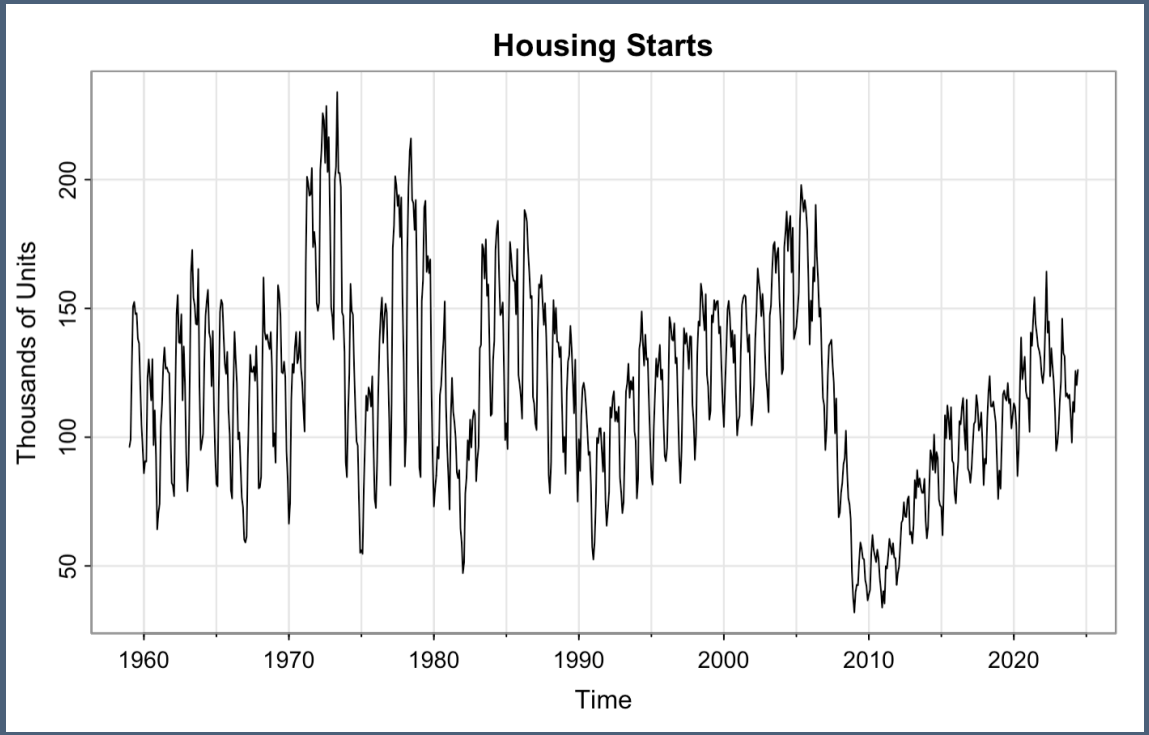
- Building SARIMA models
  - Identifying a model
  - Incorporating exogenous factors

- Examples
  - Housing construction
  - Sales in furniture stores (data from Assignment)

New Housing Construction

Housing Time Series

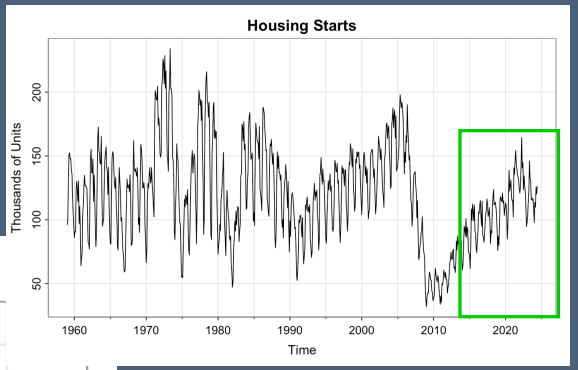
- New housing construction
  - Thousands of units, privately owned
  - Clearly seasonal with non-stationary trends

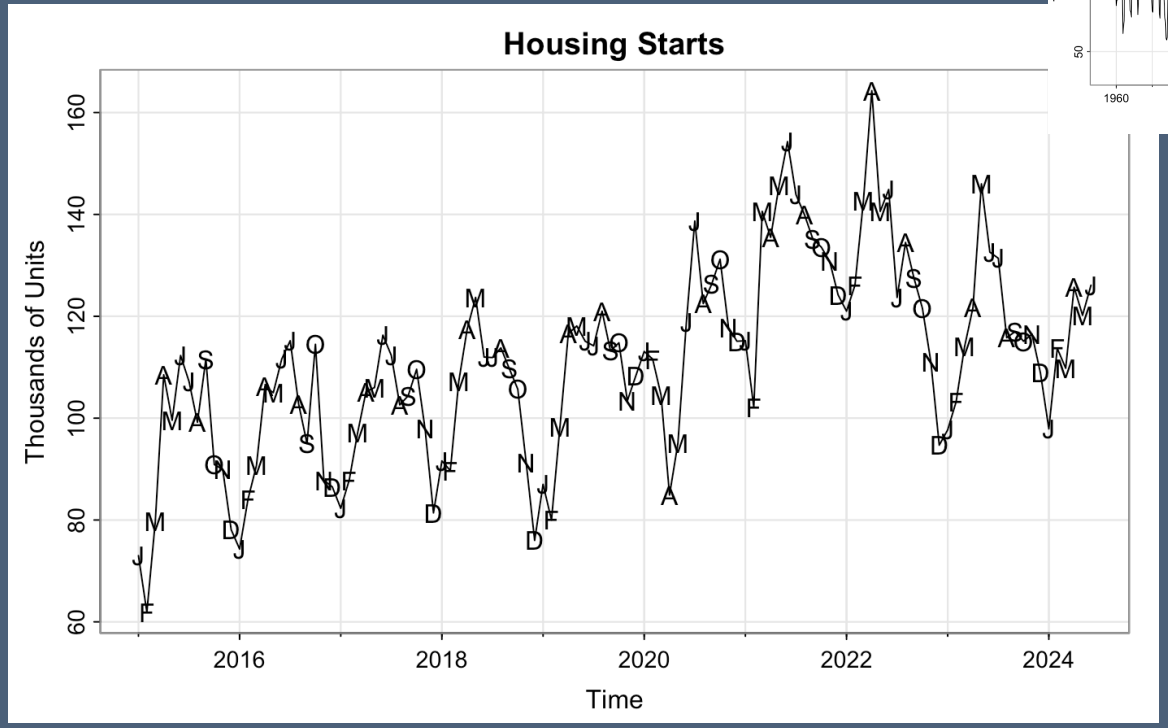


FRED HOUSTNSA

Housing Time Series

- New housing construction
  - Thousands of units, privately owned
  - Zoom in on recent trend to examine seasonality
  - Is seasonal pattern stable?

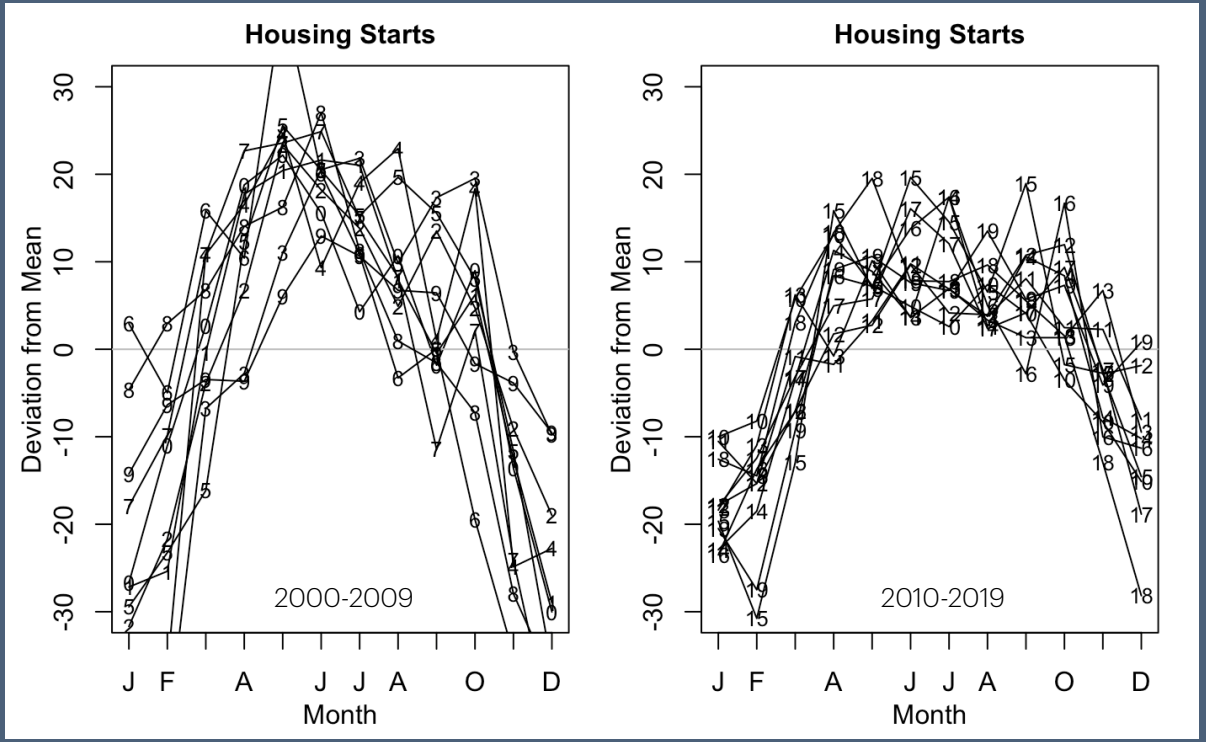




FRED HOUSTNSA

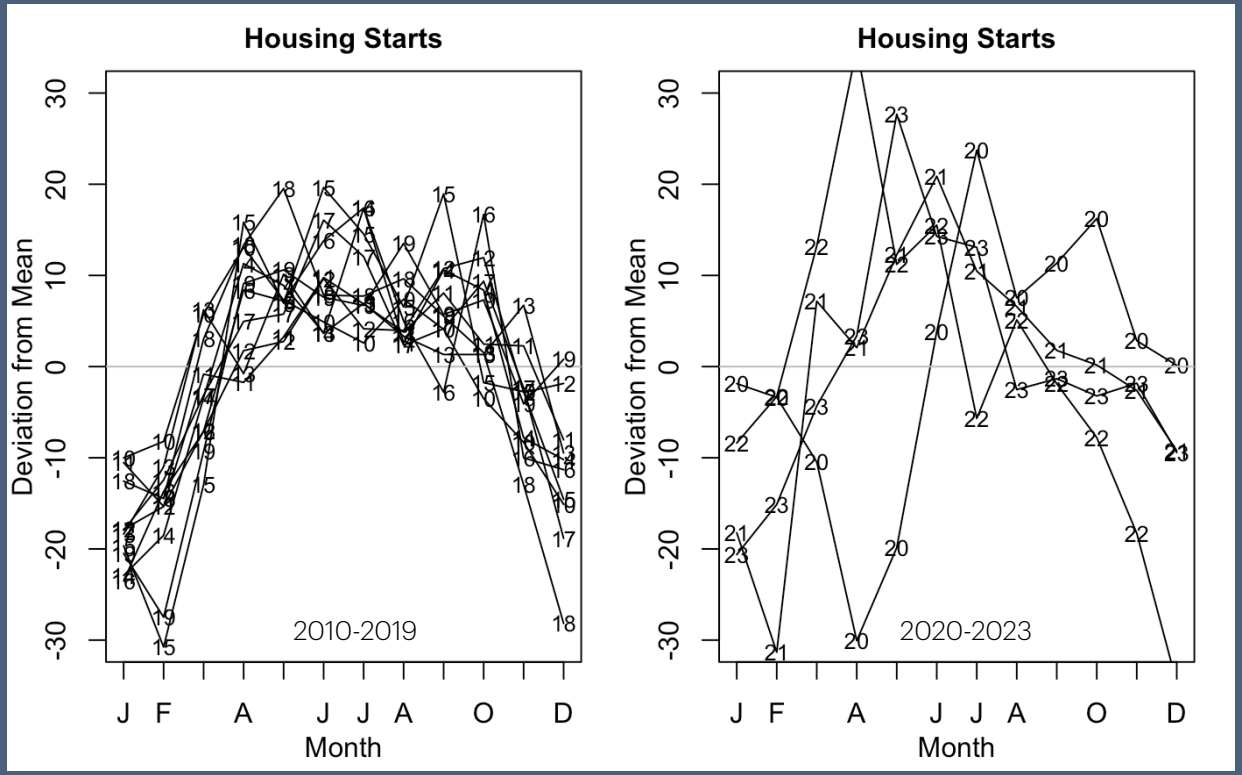
Housing Starts

- Seasonal stability
  - Overlay deviations from annual mean



Housing Starts

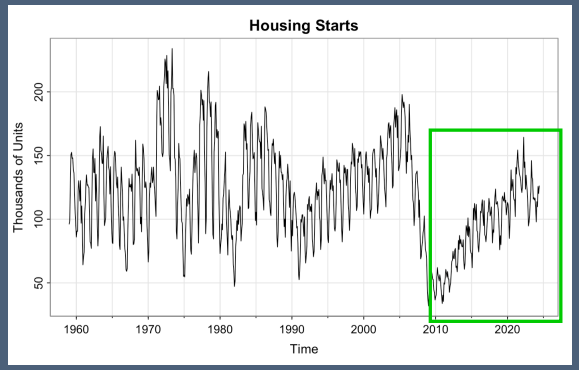
- Seasonal stability
  - Overlay deviations from annual mean

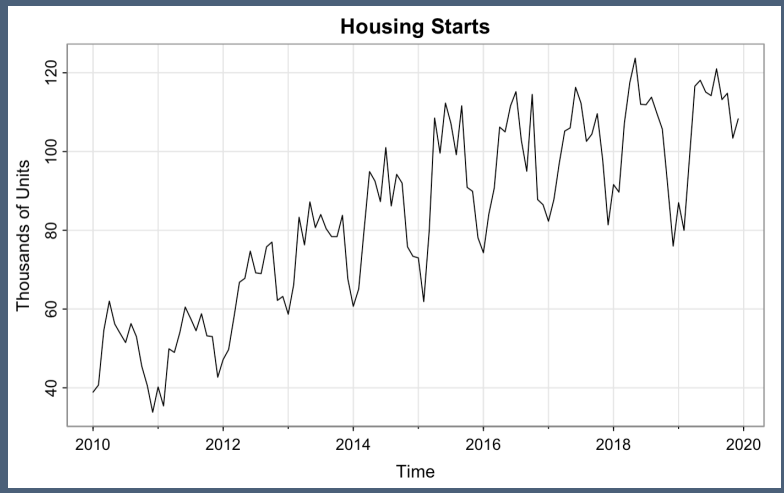


Housing Starts

- Recognize nature of changes
  - Different seasonality
  - Collapse of construction during 2008-2009
- Focus on period since housing bubble burst…
  - Use data of 10 years, 2010-2019
  - Objective: Predict 2 years out
- Revised focus reveals…
  - Softening pattern prior to Covid
  - Magnitude of seasonal variation prompts question:

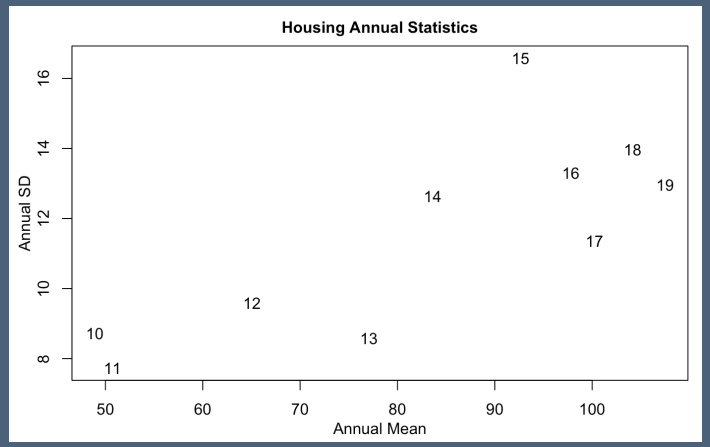
Should we be modeling on log scale?

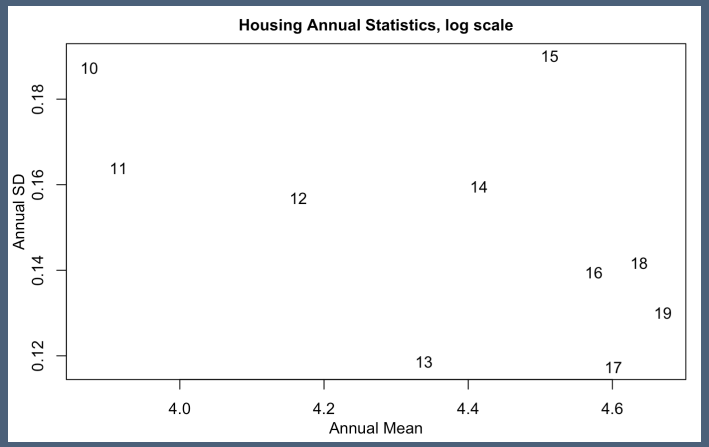




Housing: To Log or Not To Log?

- Substantive thinking
  - Will eventually difference data for stationarity
  - Differences of logs nicely interpretable as percentage change
- Data analysis
  - Examine relationship of mean to variance
  - Would like these to be "separable"

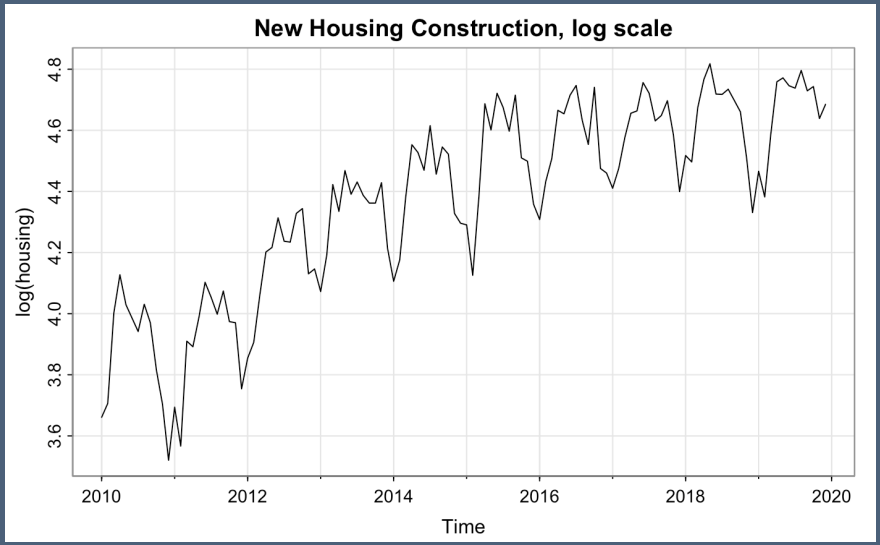


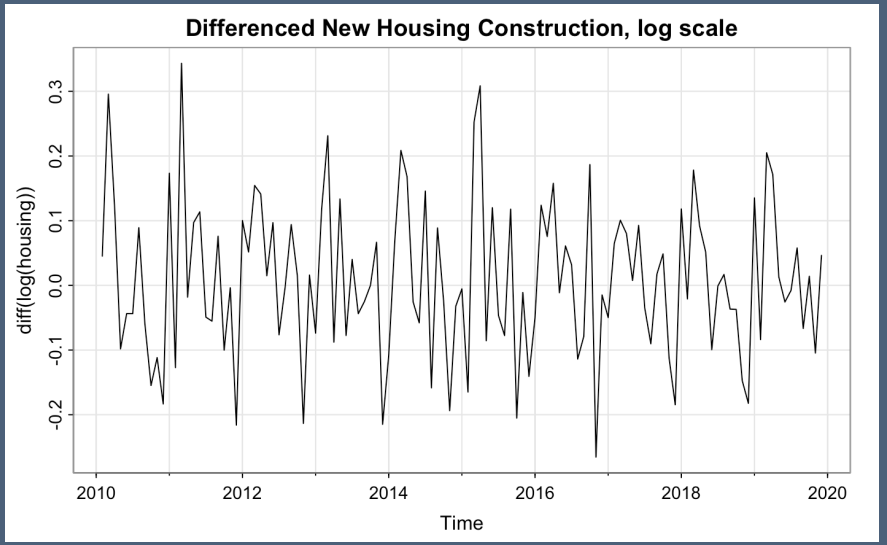


Conclude?

Housing: Differences

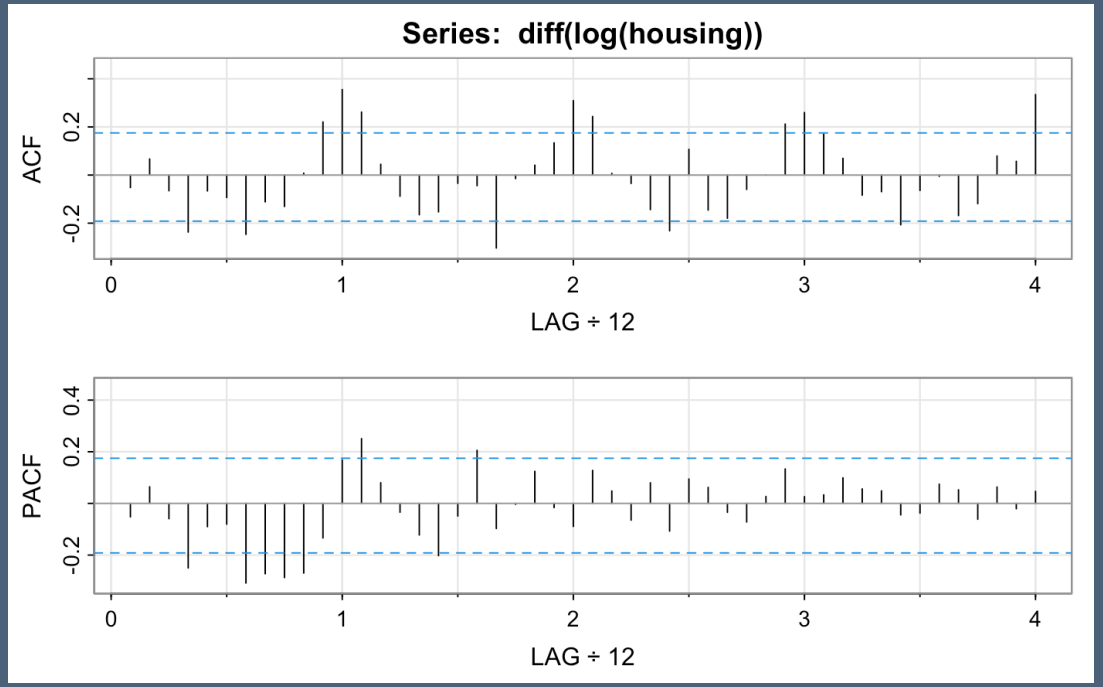
- Sequence plots
  - Log counts
  - Differenced log counts
- Patterns worth noticing in the differences?
  - Particularly with regard to seasonality





Housing: Model Identification

- ACF and PACF
  - Differenced log housing starts
- Weaker patterns
  - Magnitudes of the autocorrelations are smaller than in other examples
- Model
  - Initial choices?
- Anything else
  - What else should we be thinking about before ARIMA modeling?

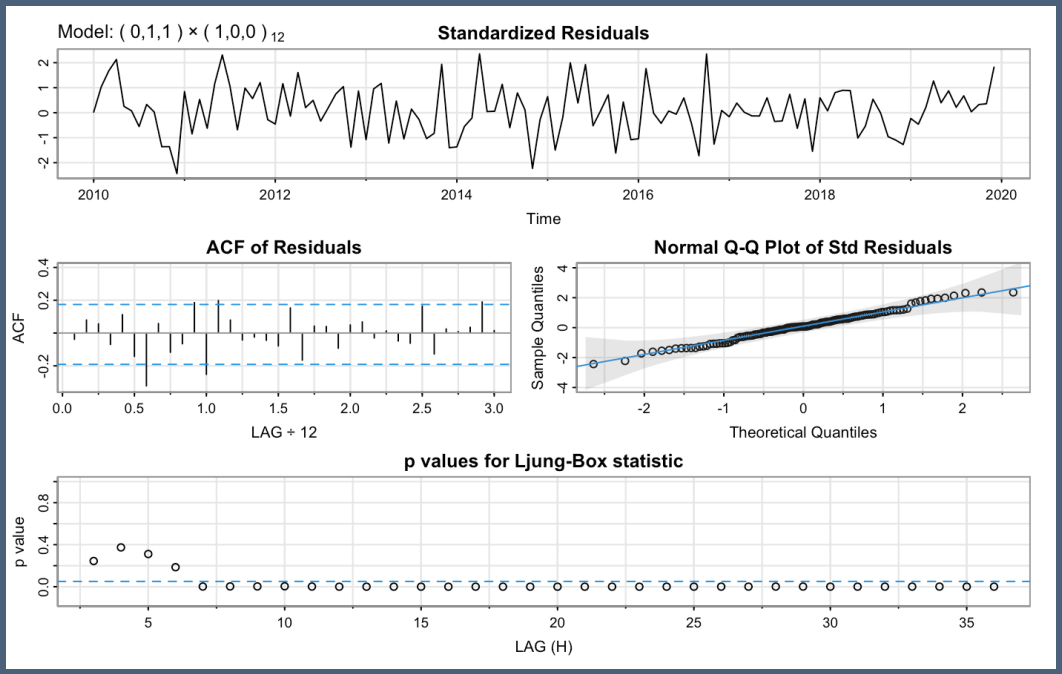


Build non-stochastic features such as length of month, dates available for construction.

Housing: SARIMA Modeling

- Initial model
  - p=0, d=1, q=1, P=1, D=0, Q=0
  - Includes monthly calendar features (days in month, weekdays in month)

```
Coefficients:
            Estimate     SE t.value p.value
ma1          -0.5801 0.0986 -5.8829  0.0000
sar1          0.7063 0.0733  9.6384  0.0000
n_weekdays    0.0043 0.0098  0.4393  0.6613
n_days        0.0382 0.0242  1.5798  0.1169

sigma^2 estimated as 0.009408587 on 115 degrees

AIC = -1.671128  AICc = -1.668179  BIC = -1.554
```

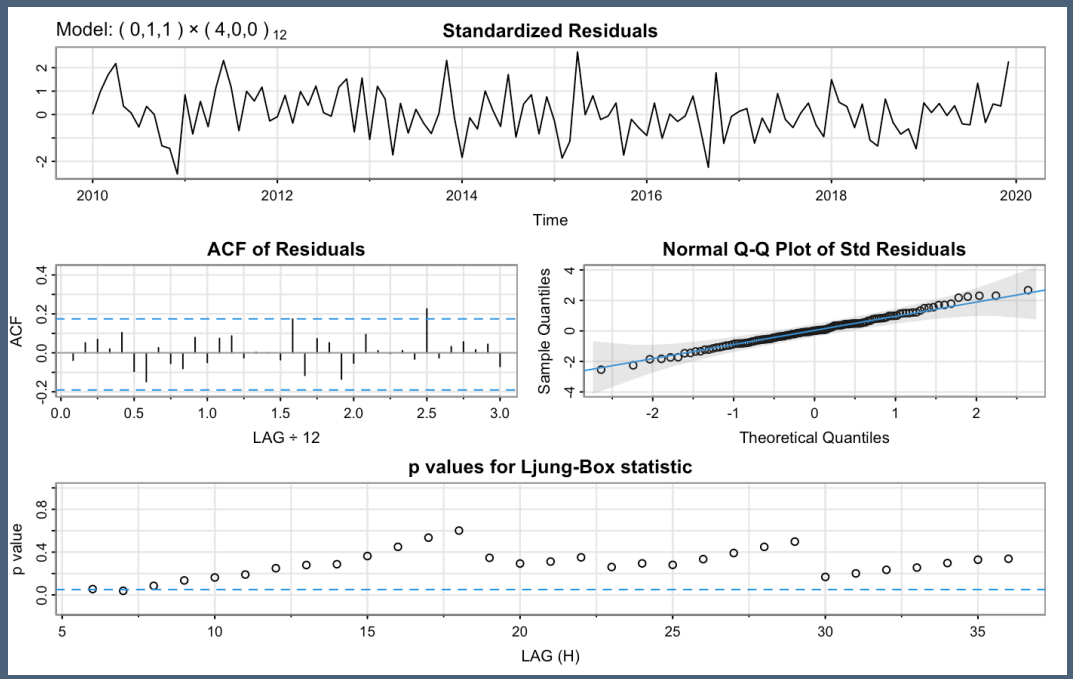


Housing: SARIMA Modeling

- Revised model
  - p=0,d=1,q=1, P=**4**, D=0, Q=0, S=12
  - Continue to include monthly calendar features

Increase P gradually and check fit each time

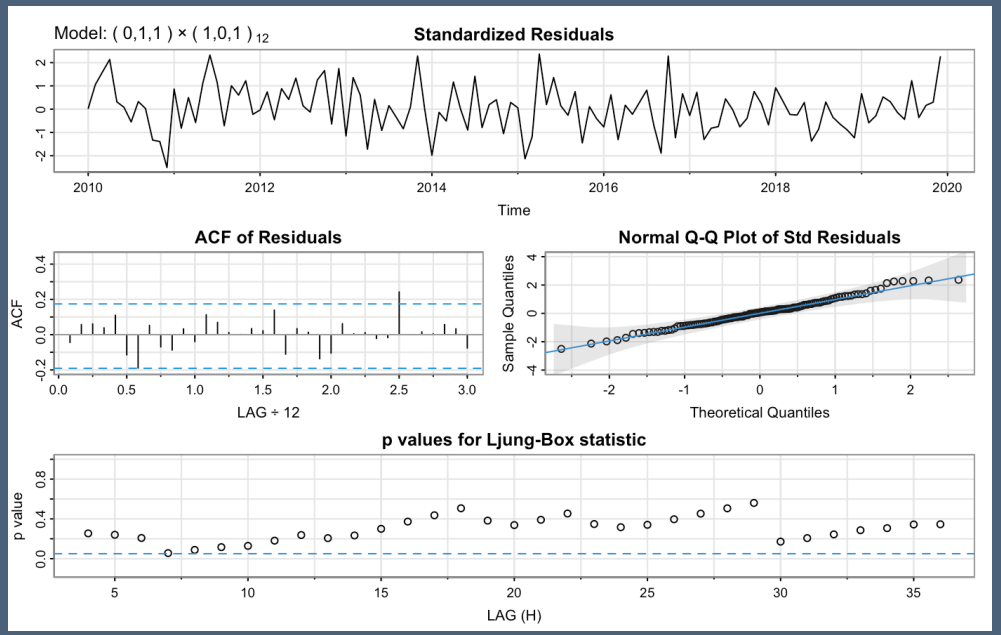
```
Coefficients:
          Estimate     SE t.value p.value
ma1         -0.5944 0.0782 -7.6025  0.0000
sar1         0.2062 0.0971  2.1241  0.0359
sar2         0.1275 0.0862  1.4795  0.1418
sar3         0.1747 0.0985  1.7742  0.0788
sar4         0.4095 0.0992  4.1269  0.0001
n_weekdays   0.0087 0.0068  1.2845  0.2016
n_days       0.0331 0.0298  1.1114  0.2688

sigma^2 estimated as 0.006228125 on 112 degrees

AIC = -1.906268  AICc = -1.897789  BIC = -1.719
```



Housing: SARIMA Modeling

- Switch to seasonal moving average
  - Lots of AR suggests maybe better with MA
  - p=0,d=1,q=1, P=**1**, D=0**,** Q=**1**, S=12
  - Includes monthly calendar features… still not significant
  - Nice fit, but coefficients are "special"

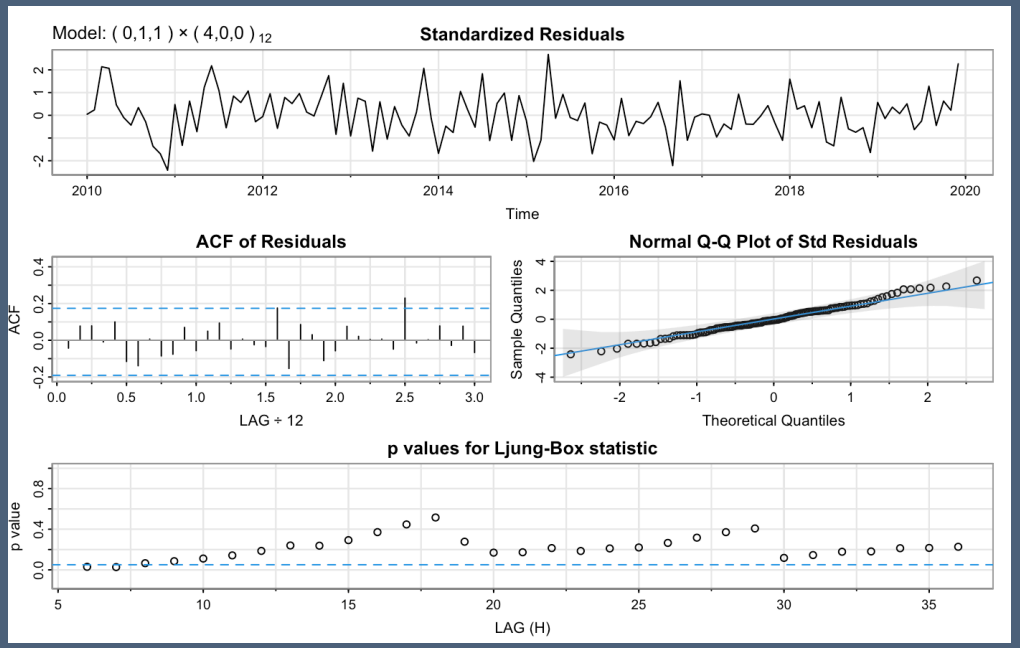
```
Coefficients:
           Estimate     SE   t.value p.value
ma1         -0.5929 0.0773   -7.6739  0.0000
sar1         0.9998 0.0009 1086.7900  0.0000
sma1        -0.9706 0.0718  -13.5270  0.0000
n_weekdays   0.0088 0.0073    1.2157  0.2266
n_days       0.0380 0.0298    1.2774  0.2040

sigma^2 estimated as 0.005666564 on 114 degrees

AIC = -1.935206  AICc = -1.930744  BIC = -1.795
```



Housing: SARIMA Modeling

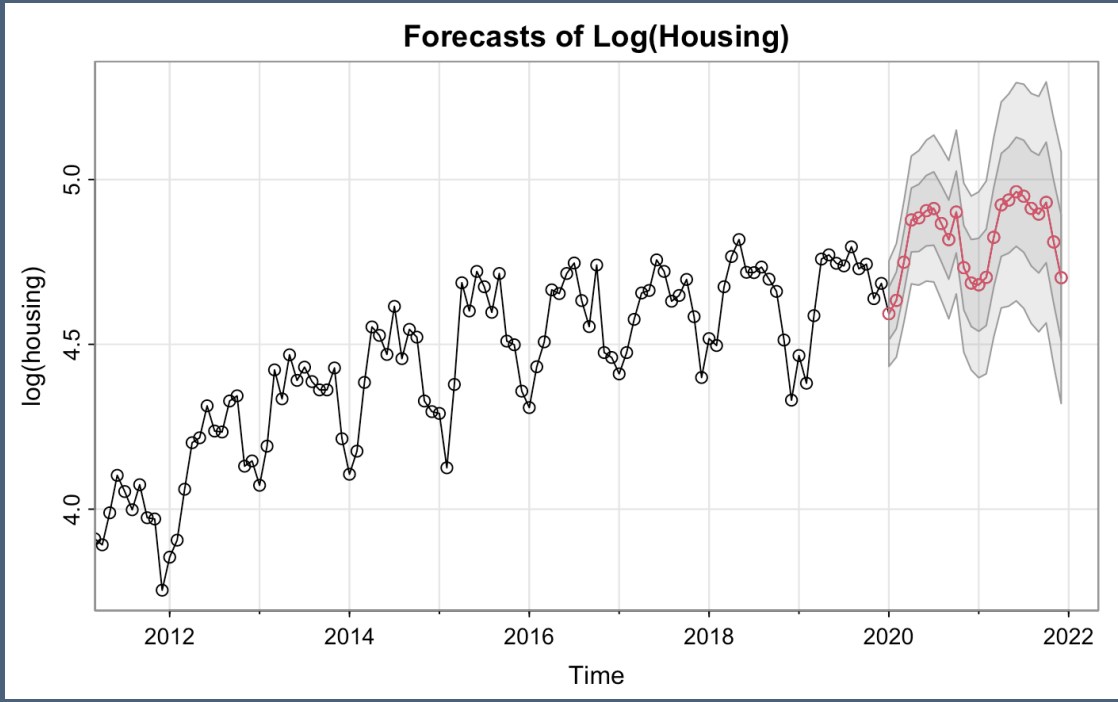
- Final model
  - p=0, d=1, q=1, P=4, D=0, Q=0, S=12
  - Removes monthly calendar features (one at a time, never significant)
  - Similar error variance to other models, simpler form
  - Residuals ≈ Normal, residual ACF mostly small

```
Coefficients:
         Estimate     SE t.value p.value
ma1       -0.6041 0.0775 -7.7967  0.0000
sar1       0.2237 0.0955  2.3415  0.0210
sar2       0.1265 0.0867  1.4590  0.1473
sar3       0.1543 0.0989  1.5594  0.1217
sar4       0.4136 0.0969  4.2670  0.0000
constant   0.0071 0.0131  0.5438  0.5876

sigma^2 estimated as 0.006397382 on 113 degrees

AIC = -1.896211  AICc = -1.889909  BIC = -1.732
```



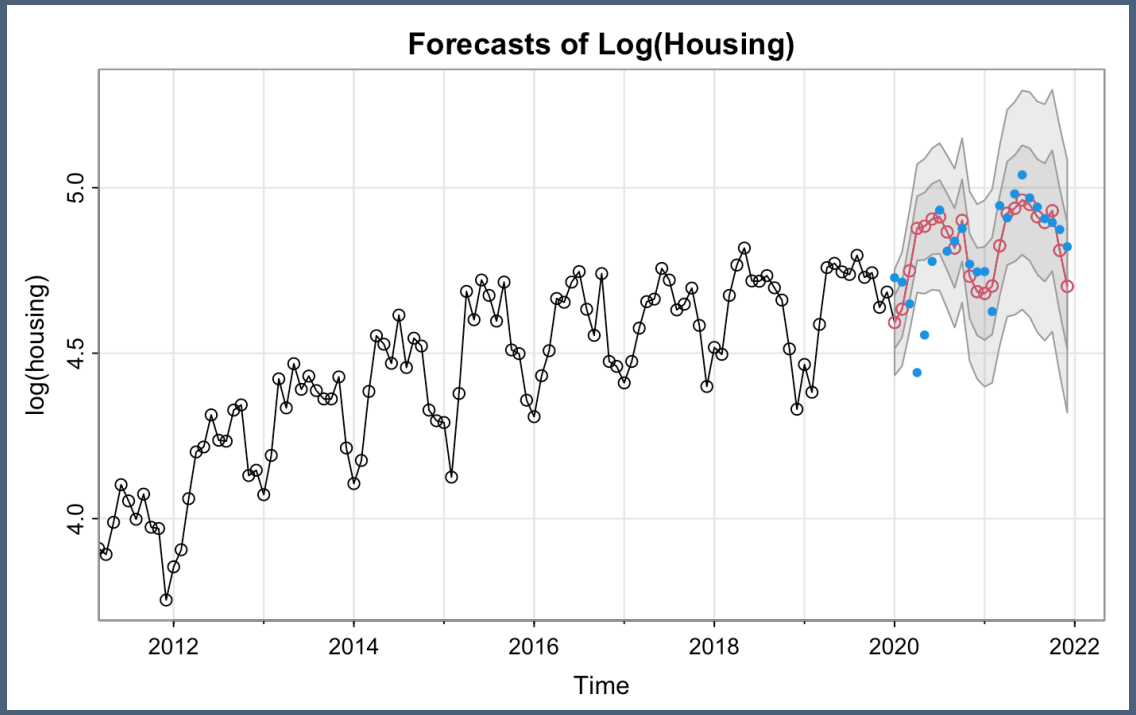
Housing: SARIMA Forecasts

- Forecast two years out
  - Predicts growth in subsequent years
  - Obviously miss Covid drop (which wasn't so big for housing)



Housing: SARIMA Forecasts

- Forecast two years out
  - Add two years of actual data
  - Anticipate recovery in 2021 after Covid… Luck?

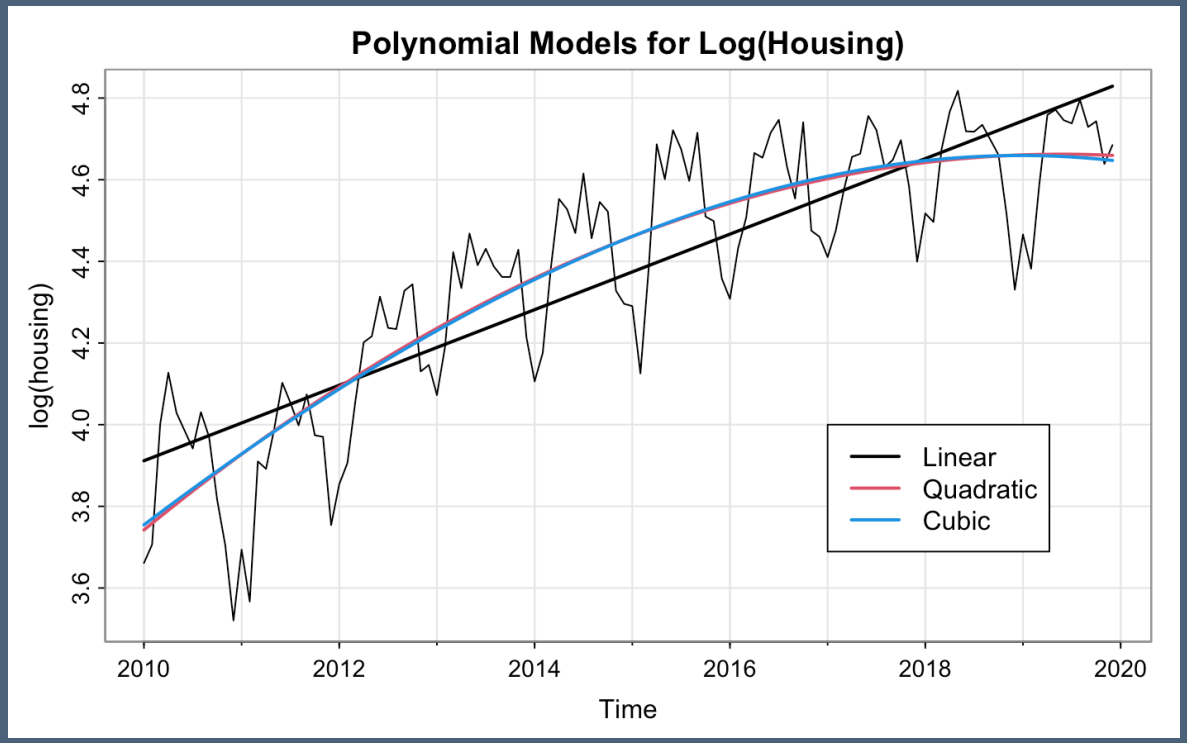


How do you think a regression model with trend would fare in this example?

Regression Model

Regression Trend Models

- What would you use?

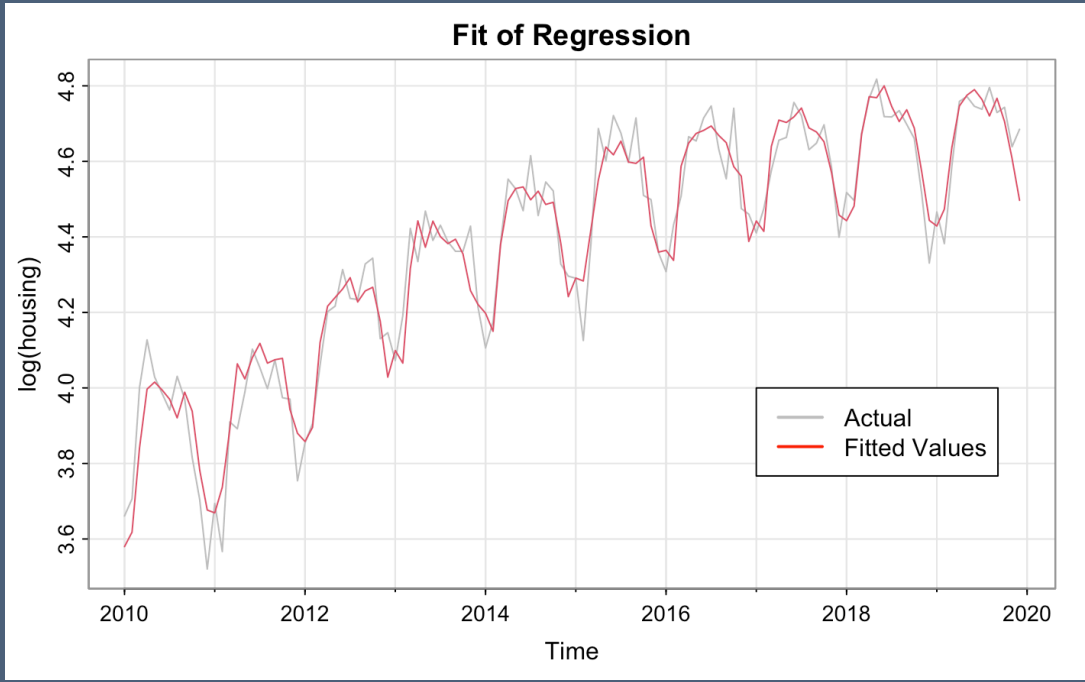


Regression Trend Model

- Quadratic regression model
  - AR(1) for residuals
  - Many seasonal coefficients are similar (April-July)
  - Claims better fit than SARIMA (0.0054 vs 0.0064)

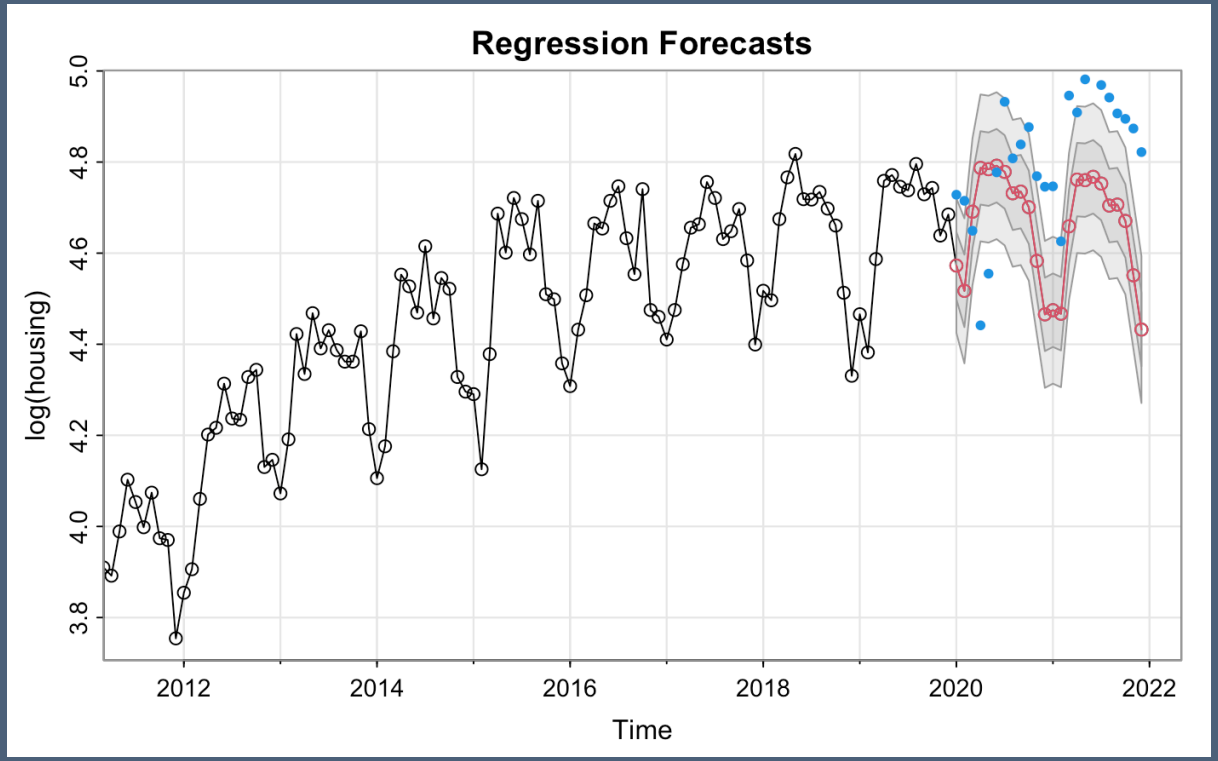
	Estimate	SE	t.value	p.value
ar1	0.4016	0.0873	4.6025	0.0000
intercept	4.3540	0.0284	153.3129	0.0000
trend	0.0731	0.0049	14.9807	0.0000
trend2	-0.0098	0.0015	-6.5622	0.0000
monthFeb	-0.0056	0.0278	-0.2016	0.8406
monthMar	0.1887	0.0329	5.7274	0.0000
monthApr	0.2936	0.0348	8.4408	0.0000
monthMay	0.2950	0.0355	8.3088	0.0000
monthJun	0.3050	0.0358	8.5257	0.0000
monthJul	0.2931	0.0359	8.1734	0.0000
monthAug	0.2477	0.0358	6.9167	0.0000
monthSep	0.2530	0.0356	7.1099	0.0000
monthOct	0.2205	0.0349	6.3081	0.0000
monthNov	0.1044	0.0333	3.1372	0.0022
monthDec	-0.0112	0.0286	-0.3921	0.6958

```
sigma^2 estimated as 0.005452987 on 105 degree

AIC = -2.105582  AICc = -2.06712  BIC = -1.73
```



Housing: Regression Forecasts

- Forecast two years out
  - Compare to actual data
  - Forecasts are too low… Quadratic trend doesn't continue

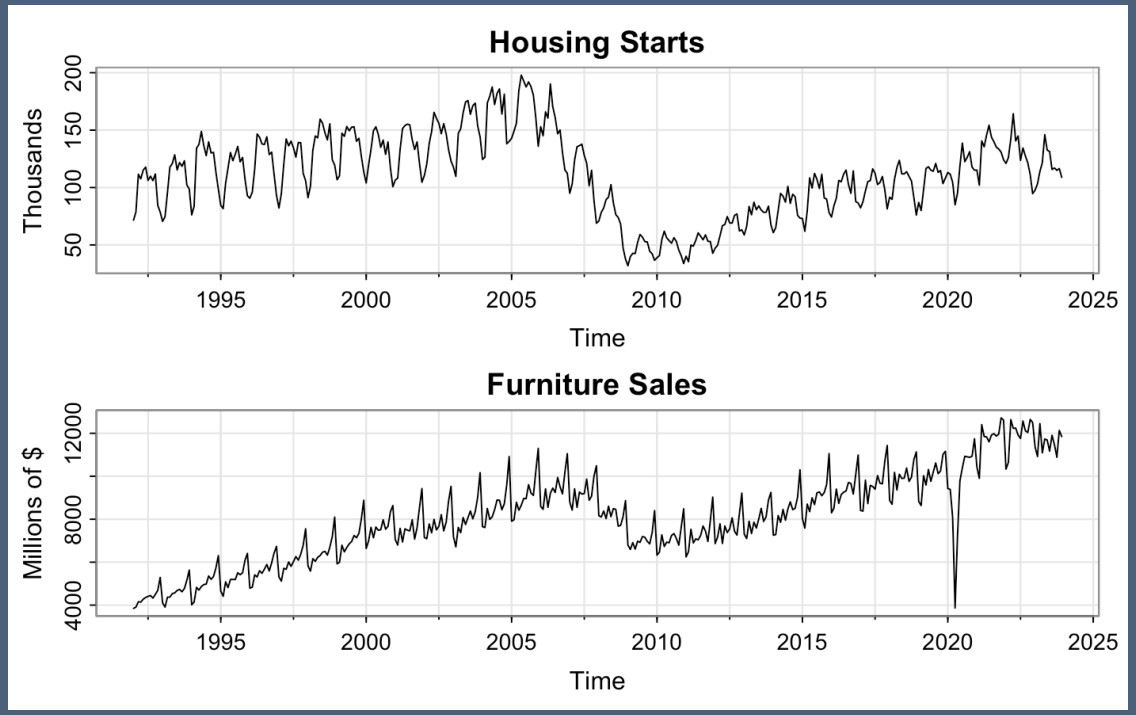


Furniture Sales

Are housing starts a leading indicator of subsequent sales?

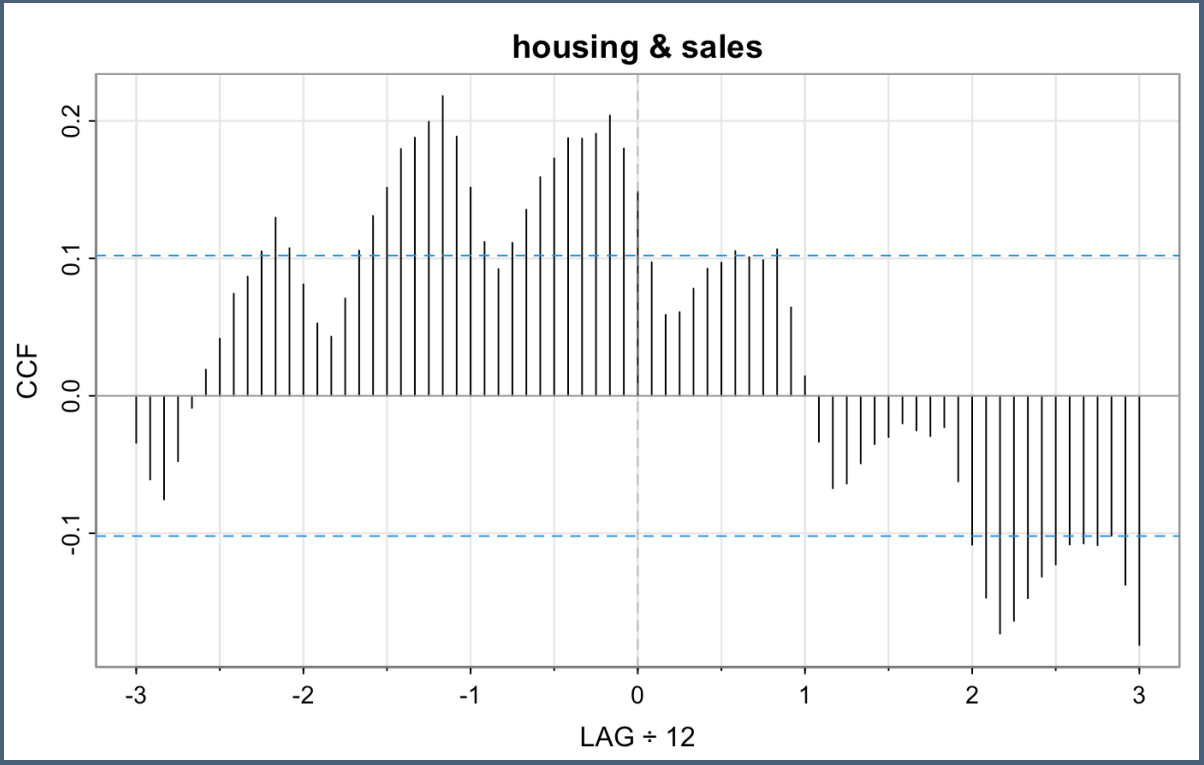
Furniture Sales and Housing Starts

- Leading indicator?
  - Does the volume of housing starts anticipate later sales at furniture stores?
  - How long does it take to build and then sell a home?



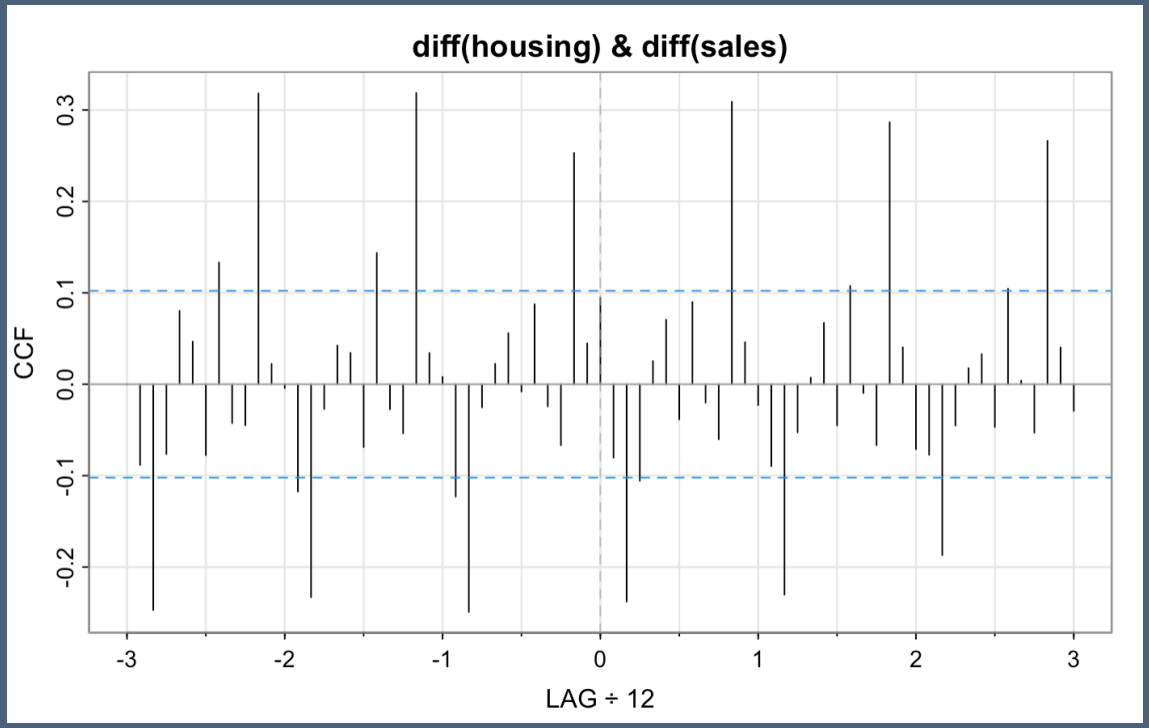
Cross-Correlations

- Cross-correlation, levels
  - Modest correlation ≈ 0.20 using starts and sales
  - Practical leading indicator leads by about 14 months (Is 2 months reasonable?)



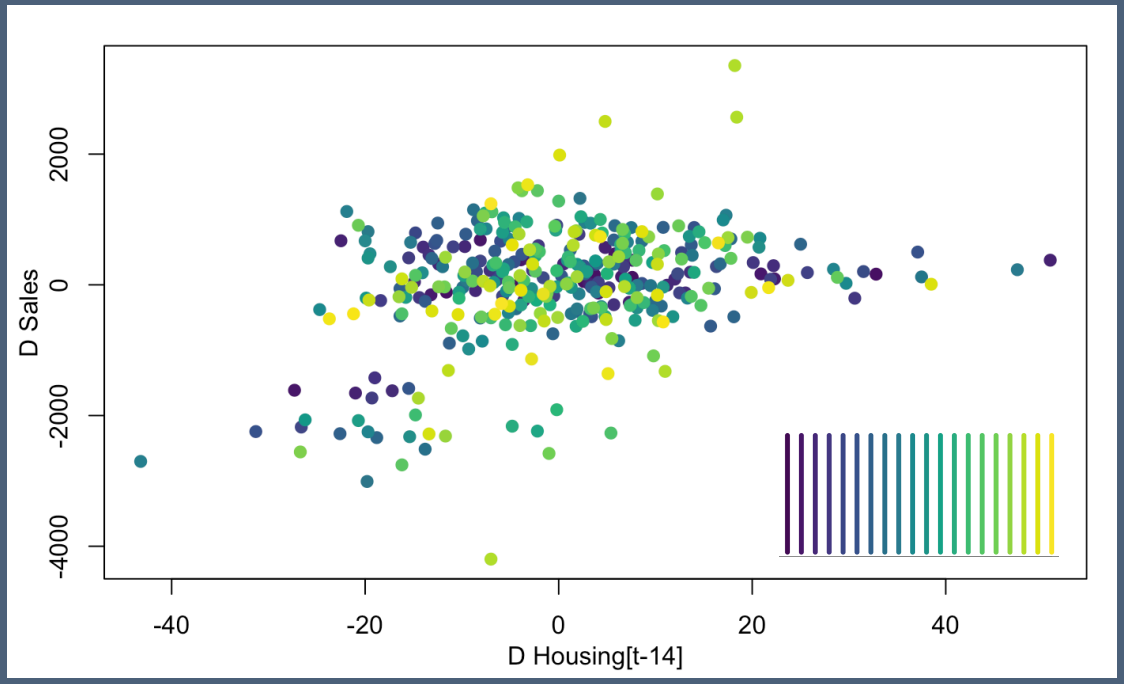
Cross-Correlations

- Cross-correlation, differences
  - Stronger (though still weak) correlation ≈ 0.30 using differences of starts and sales
  - Practical leading indicator leads by about 14 months (Is 2 months reasonable?)



Cross-Correlations

- Cross-correlation, differences
  - Stronger (though still weak) correlation ≈ 0.30 using differences of starts and sales
  - Practical leading indicator leads by about 14 months (Is 2 months reasonable?)
- Zoom in on correlation at 14 months
  - What's going on?



What's next?

- Different approach to understanding time series … Periodicity
  - Different way to approach stationary processes
  - Novel diagnostic method
  - More directed to science than for modern economics