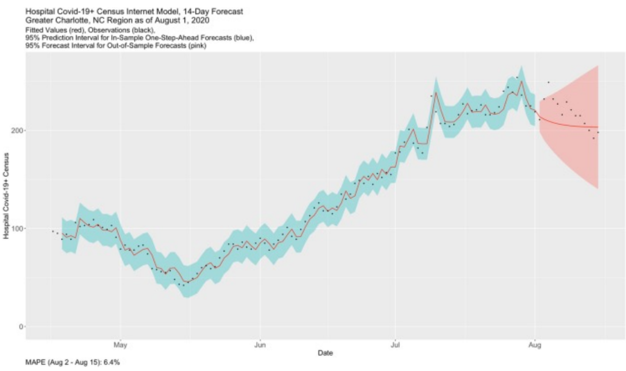
FORECASTING IN ECONOMICS & BUSINESS ECOM90024

LECTURE 3: DECOMPOSING TIME SERIES – SMOOTHING & SEASONALITY

1

EVALUATING FORECASTS

- In the previous lecture, we discussed how we can evaluate models based on their goodness of fit using statistics such as AIC, BIC and R^2
- It is important to recognize that these statistics tell us about the ***in-sample*** performance on the model.
- But if our goal is to generate the best forecasts possible, we will need to evaluate the ***out-of-sample*** performance!
- This motivates us to consider forecast accuracy metrics.



2

MEASURES OF FORECAST ACCURACY

- There are several commonly reported measures of forecast accuracy. All involve some sort of summation of forecast errors.
- Let $\hat{y}_i$ represent the predicted value of our variable of interest.
- Let y_i represent the actual value of our variable of interest (that this observation belongs in the test set).
- The prediction/forecast error is given by

$$e_i = y_i - \hat{y}_i$$

3

MEASURES OF FORECAST ACCURACY

1. Root mean square error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} e_i^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

Note here that N refers to the number of observations in the test set.

- Scaled in the same units as the data
- Akin to the out-of-sample standard error of the regression
- Most commonly used and reported metric.
- Highly influenced by extreme values.

4

MEASURES OF FORECAST ACCURACY

2. Mean absolute error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|e_i| = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

Note here that N refers to the number of observations in the test set.

- Scaled in the same units as the data
- Overcomes the issue of extreme value domination

5

MEASURES OF FORECAST ACCURACY

3. Mean absolute percentage error (MAPE)

$$MAPE = 100\times\frac{1}{N}\sum_{i=1}^{N}\left|\frac{e_i}{y_i}\right| = 100\times\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Note here that N refers to the number of observations in the test set.

- Scaled in percentages so unit insensitive
- Measures how large the predictive error is relative to the actual value
- Undefined if actual observation y_i is equal to zero and can get explosive when y_i is close to zero.

6

MEASURES OF PREDICTIVE ACCURACY

- Mean absolute scaled error (MASE) for time series data

$$MASE = \frac{\frac{1}{N}\sum_{i=1}^{N}|e_i|}{\frac{1}{T-1}\sum_{t=2}^{T}|y_t - y_{t-1}|}$$

Note here that N refers to the number of observations in the test set and T refers to the number of observations the training set.

- The naïve prediction in a time series context is simply the value of the variable of interest in the previous time period.

7

MEASURES OF FORECAST ACCURACY

- Note that these measures are statistical metrics. They are simply measures of distances (or scaled/relative distances). In optimization and decision theory, these are known as ***loss functions***.
- In many cases, these metrics will favor different predictive models.
- The impact of prediction errors on the business may not be reflected in certain metrics.
- There may be some business contexts which may call for a ***user-specific loss function***.

8

MEASURES OF FORECAST ACCURACY

- For instance, you may be faced with a business context in which an overprediction might have a greater impact on the business compared to an underprediction. So, you may need to incorporate that asymmetry into your measure.
- **Example:** Predicting demand has asymmetric implications on inventory management.
- **Example:** Predicting an asset price for the purpose of assessing loan collateral.

9

INTERVAL FORECASTS

- To produce a more informative forecast, we could calculate an ***interval forecast*** which specifies a range of values within which we could expect the future to be realized according to some (pre-specified) probability.
- In order to produce a confidence interval around our point forecast, we need to make an additional assumption about the distribution of the regression errors. Specifically, let's consider the following simple regression model:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

$$\varepsilon_t \sim_{iid} N(0, \sigma^2)$$

- Here we are assuming that the errors ε_t are independently and identically distributed as Normal with a mean of 0 and variance σ^2.

10

INTERVAL FORECASTS

- We know from the previous lecture that given some observations $\{y_t, x_t\}_{t=0}^{T}$, we can compute estimates $\hat{\beta}_0, \hat{\beta}_1$ of the parameters β_0, β_1
- Then, using these estimates, we can compute fitted/predicted values of Y_t when the predictor/explanatory variable takes value x_h

$$E\left[Y \middle| X = x_h\right] = \hat{y}_h = \hat{\beta}_0 + \hat{\beta}_1 x_h$$

- Having computed these fitted and predicted values, we can proceed to construct an intervals around them that have the statistical properties that we desire.

11

CONFIDENCE VS. PREDICTION INTERVALS

- When constructing interval forecasts, we have two options available to us. We can either construct ***confidence intervals*** around our point forecasts or ***prediction intervals***.
- Which one we use will depend on the object that we are interested in forecasting.
- If we are interested in forecasting the value of the conditional mean of our variable of interest (i.e., the dependent variable) we would construct a confidence interval. This interval represents a range of values within which we expect the true conditional mean $E\left[Y \middle| X = x_h\right]$ to lie with a specific probability.
- If we are interested in forecasting a specific value/realization of our variable of interest, we would construct a prediction interval. This interval represents a range of values within which we expect our dependent variable Y_h to lie with a specific probability.

12

CONFIDENCE INTERVAL – SIMPLE LINEAR REGRESSION

- It can be shown that the $(1-\alpha)\%$ ***confidence interval*** for the conditional mean of the dependent variable in a simple linear regression when the predictor/explanatory variable is $X = x_h$ is given by

$$\hat{y}_h \pm t_{(1-\alpha/2,T-2)}\sqrt{\frac{\sum_{t=1}^{T}(y_t-\hat{y}_t)^2}{T-2}\times\left(\frac{1}{n}+\frac{(x_h-\bar{x})^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2}\right)}$$

- Where $t_{(1-\alpha/2,T-2)}$ is the value of a Student t random variable with $T-2$ degrees of freedom that corresponds to a tail probability of $\alpha/2$

13

PREDICTION INTERVAL – SIMPLE LINEAR REGRESSION

- It can be shown that the $(1-\alpha)\%$ ***prediction interval*** for the value of the dependent variable in a simple linear regression when the predictor/explanatory variable is $X = x_h$ is given by

$$\hat{y}_h \pm t_{(1-\alpha/2,T-2)}\sqrt{\frac{\sum_{t=1}^{T}(y_t-\hat{y}_t)^2}{T-2}\times\left(1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{\sum_{t=1}^{T}(x_t-\bar{x})^2}\right)}$$

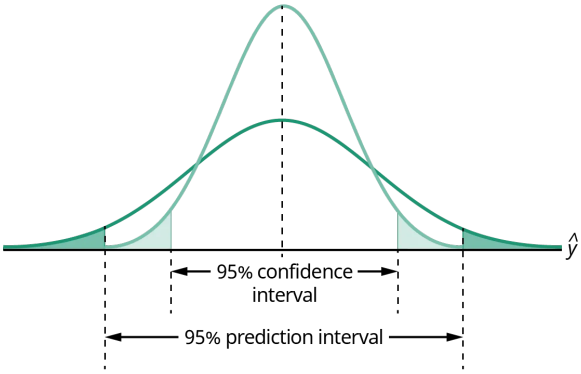
- Where $t_{(1-\alpha/2,T-2)}$ is the value of a Student t random variable with $T-2$ degrees of freedom that corresponds to a tail probability of $\alpha/2$

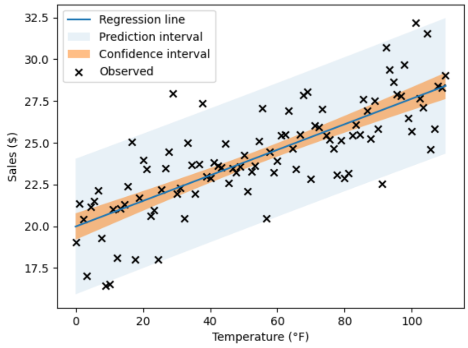
- Note that the prediction interval will always be wider than the confidence interval. This reflects the fact that there is much less uncertainty involved when trying to predict a conditional mean versus a single observation of a variable of interest. (i.e., $E[Y]$ vs. Y!)

14

PREDICTIVE ASSESSMENTS

- In general, prediction/forecast intervals are wider than confidence intervals as there is much greater uncertainty associated with estimating the value of a random variable compared to estimating the value of a parameter.





15

MATRIX REPRESENTATION OF MULTIPLE LINEAR REGRESSION

- Recall that a multiple linear regression model with k explanatory variables is written as,

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \varepsilon_t$$

- Given a set of observations on the above variables, we may organize them using data matrices, so that

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{k,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,T} & \cdots & x_{k,T} \end{bmatrix}$$

16

MATRIX REPRESENTATION OF MULTIPLE LINEAR REGRESSION

- Then, the OLS estimates of the intercept and slope coefficients can be easily computed using matrix multiplication

$$\boldsymbol{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$$

- These computations are super easy to do in R. Computers are very fast at manipulating matrices!

- Once we have our OLS estimates, then we can simply compute the fitted values using the following formula:

$$\hat{\boldsymbol{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_T \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{k,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,T} & \cdots & x_{k,T} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} = \boldsymbol{x}\boldsymbol{b}$$

17

MATRIX REPRESENTATION OF MULTIPLE LINEAR REGRESSION

- Similarly, to compute a predicted value $\hat{y}_h$ for a given set of values for the explanatory variables $\boldsymbol{x}_h = [1 \quad x_{1,h} \quad \cdots \quad x_{k,h}]'$, we write

$$\hat{y}_h = [1 \quad x_{1,h} \quad \cdots \quad x_{k,h}] \begin{bmatrix} b_0 \\ \vdots \\ b_k \end{bmatrix} = \boldsymbol{x}_h'\boldsymbol{b}$$

- As you can see, the matrix representation allows us to express everything in a more compact form.

18

CONFIDENCE INTERVAL – MULTIPLE LINEAR REGRESSION

- Then, it can be shown that the $(1-\alpha)\%$ ***confidence interval*** for the conditional mean of the dependent variable in a multiple linear regression when the predictor/explanatory variable is $\boldsymbol{x}_h = [1 \quad x_{1,h} \quad \cdots \quad x_{k,h}]'$ is given by

$$\hat{y}_h \pm t_{(1-\alpha/2, T-k-1)} \sqrt{\left(\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T-k-1}\right) \boldsymbol{x}_h{}'(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}_h{}'}$$

- Where $t_{(1-\alpha/2, T-k-1)}$ is the value of a Student t random variable with $T-k-1$ degrees of freedom that corresponds to a tail probability of $\alpha/2$

19

PREDICTION INTERVAL - MULTIPLE LINEAR REGRESSION

- It can also be shown that the $(1-\alpha)\%$ ***prediction interval*** for the value of the dependent variable in a multiple linear regression when the predictor/explanatory variable is $\boldsymbol{x}_h = [1 \quad x_{1,h} \quad \cdots \quad x_{k,h}]'$ is given by

$$\hat{y}_h \pm t_{(1-\alpha/2, T-k-1)} \sqrt{\left(\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T-k-1}\right) (1 + \boldsymbol{x}_h{}'(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}_h{}')}$$

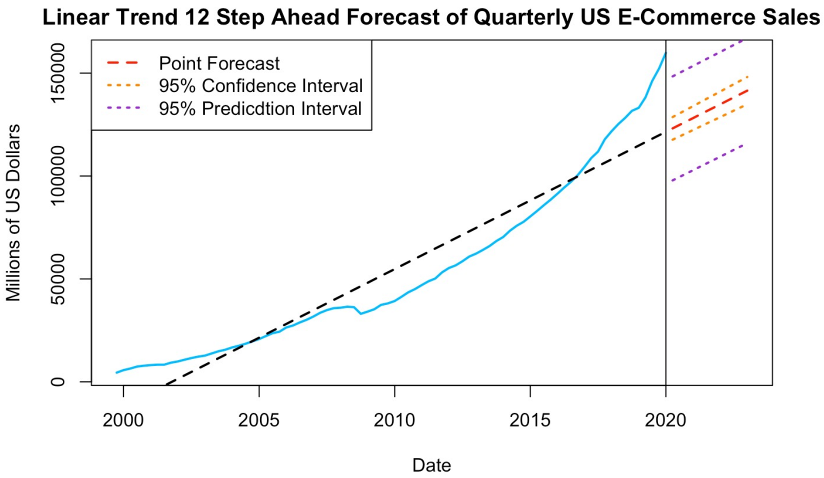
- Where $t_{(1-\alpha/2, T-k-1)}$ is the value of a Student t random variable with $T-k-1$ degrees of freedom that corresponds to a tail probability of $\alpha/2$

- Fortunately, we don't have to compute these objects manually! The ***predict()*** function in R will do this for us.

20

CONFIDENCE AND PREDICTION INTERVALS FOR LINEAR TREND MODEL

- Let's revisit the linear trend model that we estimated last week. The 95% confidence and 95% prediction intervals are depicted as follows:

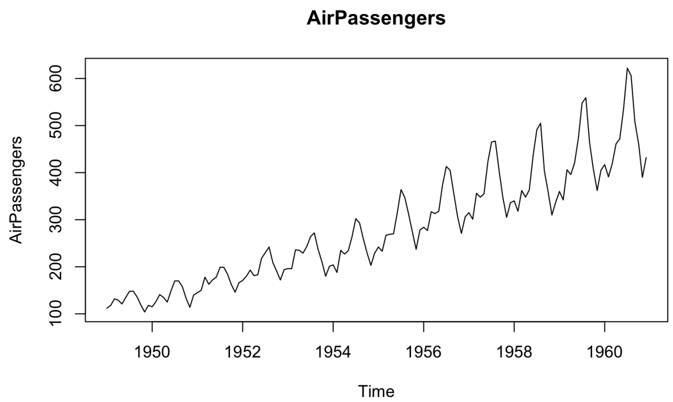


21

SEASONALITY

- In addition to trends, many economic and financial time series exhibit strong seasonality, that is, patterns that repeat every year.

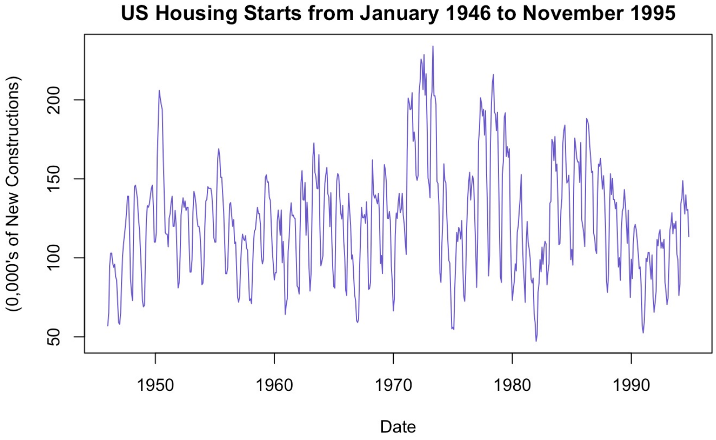
- Seasonality often manifests as result of technological, preferential and institutional fluctuations that are linked to the calendar, such as holidays that occur at the same time every year.



22

SEASONALITY

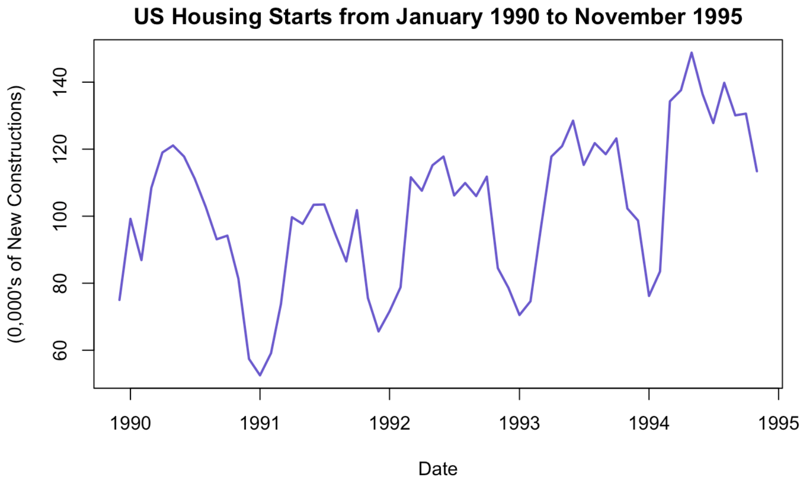
- To illustrate, let's build a seasonal forecasting model for housing starts (the number of home constructions) in the United States.
- Our dataset comprises of monthly observations from January 1946 to November 1995
- There is no apparent long-term trend.



23

SEASONALITY

- Zooming in on the period from 1990 to 1994 we observe that there is a clear seasonal pattern. What do you think is driving this?
- How should we go about modeling these types of fluctuations?



24

SEASONALITY

- One way to look at these seasonal fluctuations is to think of them as season specific averages. If the fluctuations are very similar at common points in time (i.e., same month, quarter, day of week), then we can model these season specific averages as seasonal factors.

- That is, we can think of each observation in our housing start data as the sum of the seasonal factor that corresponds to that month plus some idiosyncratic variation.

- For instance, the observation for January 1990 can be thought of as comprising of the seasonal factor corresponding to the month of January plus a component that is specific to that particular observation:

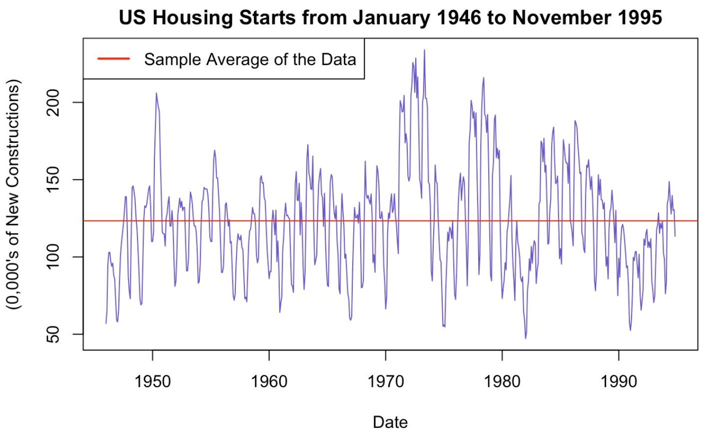
$$Y_{Jan1990} = Jan + \varepsilon_{Jan1990}$$

25

SEASONALITY

- A simple way for us to estimate a seasonal factor is to compute them as the average of the observations that correspond to that season.

- Thus, the January factor could be estimated by the average of all the January observations.



26

SEASONALITY

- The standard technique for estimating seasonal factors in this way through the use of ***seasonal dummy variables*** in a regression. These are indicator variables that take value 1 if the observation belongs to the season represented by the dummy and 0 otherwise.

- The form of a seasonal dummy will depend on the frequency in which the data is observed and the structure of the seasonal fluctuation. To illustrate, let s denote the number of seasons in a year and for simplicity let's assume that we are working with a time series that is observed quarterly (such as GDP) and that the first observation occurs in the first quarter. If $s = 4$, then we create four dummy variables:

$$\begin{aligned}
D_1 &= \{1,0,0,0,1,0,0,0,1,0,0,0, \dots\} \\
D_2 &= \{0,1,0,0,0,1,0,0,0,1,0,0, \dots\} \\
D_3 &= \{0,0,1,0,0,0,1,0,0,0,1,0, \dots\} \\
D_4 &= \{0,0,0,1,0,0,0,1,0,0,0,1, \dots\}
\end{aligned}$$

- In the case of our housing starts, the data is observed monthly so if we were modelling our seasonality as occurring at this frequency, $s = 12$

27

SEASONALITY

- The deterministic seasonal component of a time series can then be specified and estimated in the following way:

$$Y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- The inclusion of seasonal dummies in a regression model creates an intercept that varies in a deterministic manner throughout the seasons within each year.

- The parameters γ_i are called ***seasonal factors*** and summarize the seasonal pattern over the year.

- It is important to note that if you wish to include the exhaustive set of seasonal dummies in your regression, you will have to set your intercept term equal to zero. Failing to do this will lead to perfect multicollinearity as one regressor will be a perfect linear combination of the others.

28

SEASONALITY

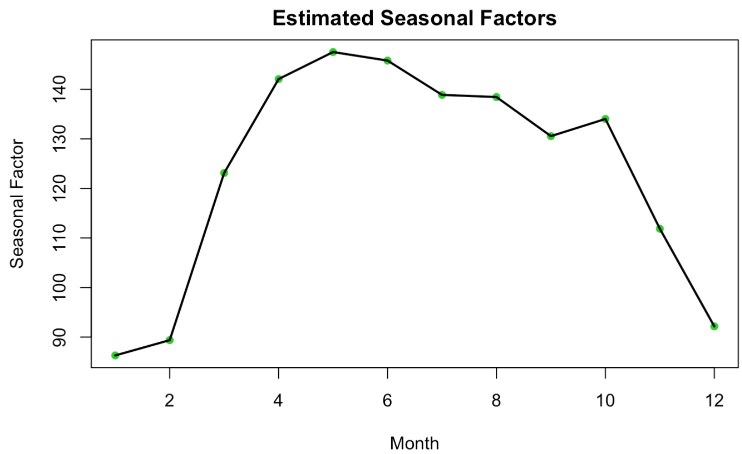
- Here are the estimation results from our seasonal dummy model applied to the housing starts data.
- The estimated coefficients here are the season specific averages where each season corresponds to a month of the year.
- You can verify this in Excel!

```
## Call:
## lm(formula = y$hstarts ~ 0 + M1 + M2 + M3 + M4 + M5 + M6 + M7 +
##     M8 + M9 + M10 + M11 + M12)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.976 -17.504  -2.158  13.447  90.155
##
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## M1    86.294      3.952   21.84   <2e-16 ***
## M2    89.382      3.952   22.62   <2e-16 ***
## M3   123.116      3.952   31.16   <2e-16 ***
## M4   142.076      3.952   35.95   <2e-16 ***
## M5   147.527      3.952   37.33   <2e-16 ***
## M6   145.802      3.952   36.90   <2e-16 ***
## M7   138.882      3.952   35.14   <2e-16 ***
## M8   138.445      3.952   35.03   <2e-16 ***
## M9   130.553      3.952   33.04   <2e-16 ***
## M10  134.020      3.952   33.91   <2e-16 ***
## M11  111.865      3.952   28.31   <2e-16 ***
## M12   92.158      3.993   23.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.66 on 575 degrees of freedom
## Multiple R-squared:  0.9544, Adjusted R-squared:  0.9535
## F-statistic:  1004 on 12 and 575 DF,  p-value: < 2.2e-16
```

29

SEASONALITY

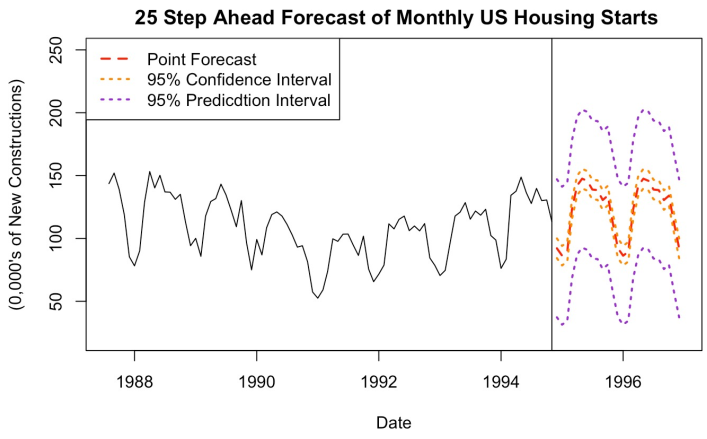
- Plotting the estimated coefficients, we can see that the month specific averages are very low in January and February, and then rise quickly and peak in May, after which they decline, at first slowly and then abruptly in November and December.



30

FORECASTING USING A SEASONAL DUMMY MODEL

- As was the case with our linear trend mode, we can use our model estimates to generate point and interval forecasts:



31

SMOOTHING METHODS

- Deterministic trends can often be difficult to implement when dealing with real world data.
- Apart from the visual impression that we have of the data, there is often very little that guides us in which functional form to use! (Linear, quadratic, exponential, cubic, quartic, etc.)
- Most real-world time trends do not behave according to precise mathematical functions that we have arbitrarily chosen.
- Also, deterministic functions do a very poor job of capturing cyclical fluctuations.
- An alternative approach is to compute trends as smoothed observations. This is a more data-oriented approach.

32

SMOOTHING METHODS

- Recall that in last week's lecture we discussed the following decomposition:

$$Y_t = T_t + S_t + C_t + \varepsilon_t$$

- Where,

$$T_t = trend\ component$$

$$C_t = cyclical\ component$$

$$S_t = seasonal\ component$$

- Smoothing methods are one way of extracting the trend and cycle component from the observed time series.

33

SMOOTHING VIA MOVING AVERAGE

- A straightforward way to smooth time series data is by computing what is known as a moving average (sometimes referred to as local averaging).

- By averaging across a range (or window) of observations, we can smooth out the short fun fluctuations (such as seasonality and randomness) to extract the longer run fluctuations.

- Given a time series $\{y_t\}_{t=1}^{T}$, a moving average of order m is defined as:

$$MA(m)_t = \frac{1}{m}\sum_{j=-k}^{k} y_{t+j}$$

- This operation produces a smoothed observation of the time series at time t by taking the average of the k observations surrounding period t.

34

SMOOTHING VIA MOVING AVERAGE

- The choice of m will depend on the kinds of short run and seasonal fluctuations that are present in the time series. For instance, if there exist seasonal patterns that occur within a 12 period intervals, then we will need an averaging window that is at least as wide as 12 periods.

- If m is too small, there will be insufficient smoothing and our smoothed series will still contain random and seasonal fluctuations.

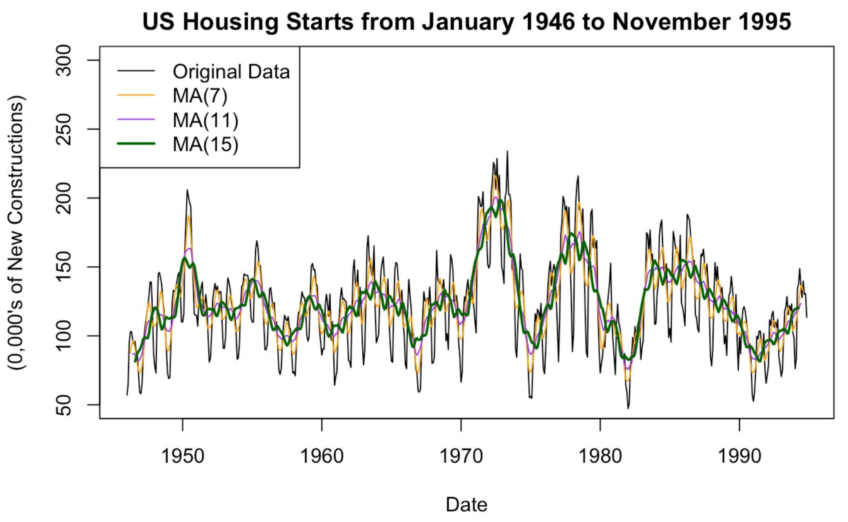
- If m is too large, there will be too much smoothing which will obscure the very long run trend and cyclical fluctuations we are trying to isolate!

35

SMOOTHING VIA MOVING AVERAGE

- Computing moving averages of orders $m = 7{,}11{,}15$, we can see that the higher the m, the greater the degree of smoothing.

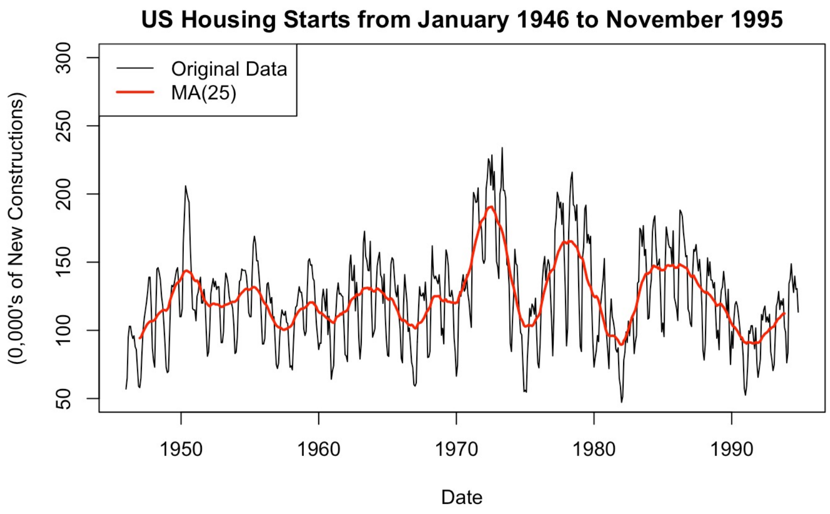
- At these orders of averaging, the seasonal fluctuations are still clearly present.



36

SMOOTHING VIA MOVING AVERAGE

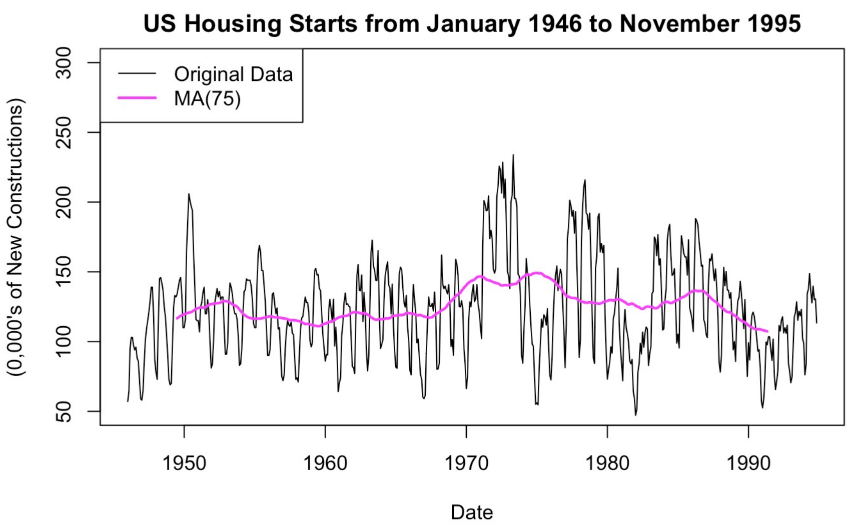
- Let's try widening our smoothing window so that $m = 25$
- Here we have a reasonably smoothed series that preserves the long run trend and cyclical fluctuations



37

SMOOTHING VIA MOVING AVERAGE

- When we make our smoothing window too big, we end up with an oversmoothed series that is uninformative.



38

SMOOTHING VIA MOVING AVERAGE

- So far, we have just chosen m rather arbitrarily. We can be a bit more thoughtful!
- In the presence of seasonal fluctuations, we should choose m to be equal to a multiple of the periodicity of the seasonal fluctuation. That is, if there is a seasonal pattern that takes place over 5 time periods, then our smoothing window should be a multiple of 5.
- We typically choose m to be an odd number to ensure symmetry (i.e., there are an equal number of observations above and below the point in time that we are computing the smoothed observation)
- However, often we are have seasonal fluctuations that occur across even numbered time periods. To average out these fluctuations appropriately, we would need to compute a moving average of moving averages.

39

SMOOTHING VIA MOVING AVERAGE

- To illustrate, let's consider a moving average of order $m = 4$, there are two possible ways to construct this:

1. $MA(4)_t = \frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1})$
2. $MA(4)_t = \frac{1}{4}(y_{t-1} + y_t + y_{t+1} + y_{t+2})$

- Neither option is symmetric, the first weights the past more while the second weights the future more. We can solve this issue by taking the average of the two to produce a moving average of moving averages:

$$\overline{MA(4)}_t = \frac{1}{2}\left[\frac{1}{4}(y_{t-2} + y_{t-1} + y_t + y_{t+1})\right] + \frac{1}{2}\left[\frac{1}{4}(y_{t-1} + y_t + y_{t+1} + y_{t+2})\right]$$

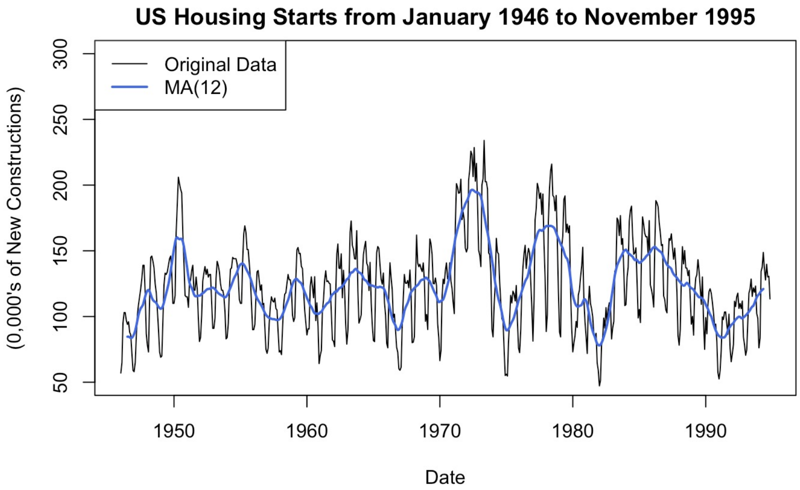
- Note that this is equivalent to an MA(5) with unequal weights

$$\overline{MA(4)}_t = \frac{1}{8}y_{t-2} + \frac{1}{4}(y_{t-1} + y_t + y_{t+1}) + \frac{1}{8}y_{t+2}$$

40

SMOOTHING VIA MOVING AVERAGE

- In the case of our housing starts data, we saw that the seasonal pattern in the data occurs over a 12-month period. Therefore, we should smooth our data using $m = 12$, where we center our moving average by taking the average in way described in the previous slide



41

CLASSICAL DECOMPOSITION

- Now that we understand how to compute seasonal and smoothed trend-cycle components directly, we can proceed to use the decompose() function to compute these components in a single line of code.

- We will first compute the classical additive decomposition

$$Y_t = T_t + S_t + C_t + \varepsilon_t$$

42

CLASSICAL DECOMPOSITION

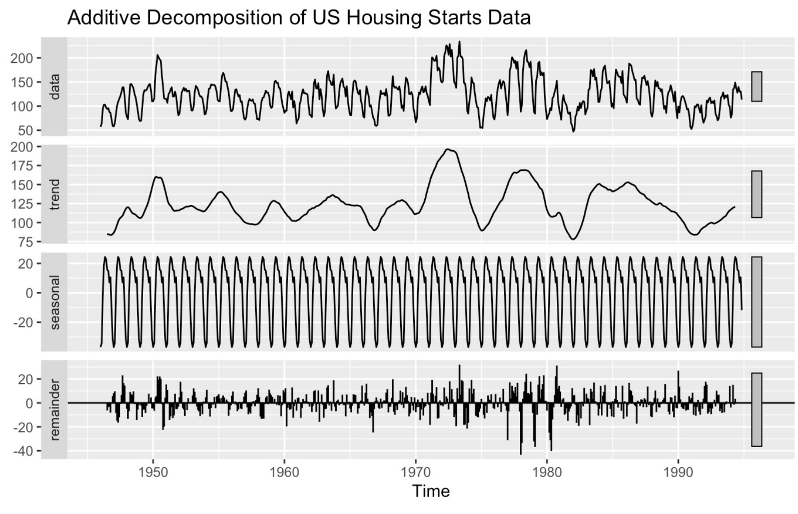
- The decompose function computes the additive decomposition via the following steps:

1. Setting the seasonal period m equal to the frequency in which the data is observed (e.g., $m = 4$ for quarterly data, $m = 12$ for monthly data, etc.)
2. Compute the trend-cycle component $\hat{T}_t + \hat{C}_t$ using an $MA(m)$ if m is odd and $\overline{MA(m)}$ if m is even.
3. Calculate the detrended series $y_t - \hat{T}_t - \hat{C}_t$
4. Calculate the seasonal component for each season as the average of the detrended values for that season (e.g., the average of all the January observations etc.) These seasonal component values are then adjusted to ensure that they add to zero. The seasonal component is obtained by stringing together these monthly values, and then replicating the sequence for each year of data. This gives $\hat{S}_t$
5. The remainder component is calculated by subtracting the estimated seasonal and trend-cycle components: $\hat{R}_t = y_t - \hat{T}_t - \hat{C}_t - \hat{S}_t$

43

CLASSICAL DECOMPOSITION

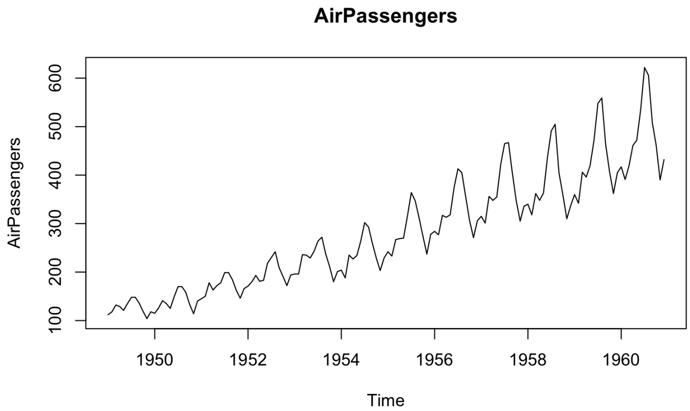
- Thus, the classical additive decomposition for our US housing starts data is given by:
- If the decomposition has done a good job of accounting for the time series structure of the data, the remainder should look like white noise (i.e., random!)



44

CLASSICAL DECOMPOSITION

- The additive decomposition is the most appropriate if the magnitude of the seasonal fluctuations, or the variation around the trend-cycle, does not vary with the level of the time series.
- When the variation in the seasonal pattern, or the variation around the trend-cycle, appears to be proportional to the level of the time series, then a multiplicative decomposition is more appropriate.



45

CLASSICAL DECOMPOSITION

- A multiplicative decomposition is specified as:

$$Y_t = T_t \times S_t \times C_t \times \varepsilon_t$$

- The decompose function computes the multiplicative decomposition via the following steps:

1. Setting the seasonal period m equal to the frequency in which the data is observed (e.g., $m = 4$ for quarterly data, $m = 12$ for monthly data, etc.)
2. Compute the trend-cycle component $\hat{T}_t \times \hat{C}_t$ using an $MA(m)$ if m is odd and $\overline{MA(m)}$ if m is even.
3. Calculate the detrended series $y_t/(\hat{T}_t\hat{C}_t)$

46

CLASSICAL DECOMPOSITION

- Calculate the seasonal component for each season as the average of the detrended values for that season (e.g., the average of all the January observations etc.) These seasonal component values are then adjusted to ensure that they add to m. The seasonal component is obtained by stringing together these monthly values, and then replicating the sequence for each year of data. This gives $\hat{S}_t$

- The remainder component is calculated by dividing out the estimated seasonal and trend-cycle components

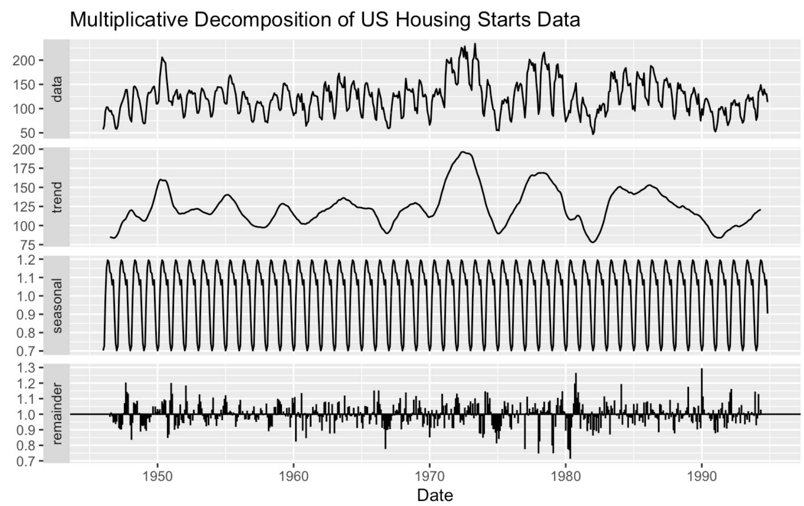
$$\hat{R}_t = \frac{y_t}{\hat{T}_t \times \hat{S}_t \times \hat{C}_t}$$

47

CLASSICAL DECOMPOSITION

- Here is the classical multiplicative decomposition for our US housing starts data:

- Note the scale of the seasonal indices and remainder.



48

CLASSICAL DECOMPOSITION

- Smoothed time series obtained via moving averages are useful for identifying and depicting the long-run fluctuations that are present in the data.
- We can use seasonal indices to de-seasonalize (i.e., seasonally adjust) a time series of interest.
- However, because we lose observations from either end of the series when computing centered moving averages, classical decompositions aren't very useful for forecasting the trend-cycle component.
- Also, moving averages have a tendency to over-smooth rapid rises and falls in the data as well as unusual values (i.e., outliers).

49

SIMPLE EXPONENTIAL SMOOTHING

- For time series that do not have a clear trend or seasonality, smoothing can be achieved by a method known as simple exponential smoothing.
- This method computes a smoothed observation as a weighted average of all past observations.
- Let's suppose that we have a time series $\{y_t\}_{t=0}^{T}$, then using simple exponential smoothing, the smoothed observation at time t, (which we denote as l_t) is given by the following geometric series

$$l_t = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots$$

- Where $0 < \alpha < 1$

50

SIMPLE EXPONENTIAL SMOOTHING

- This is a geometric series in which the smoothing parameter α controls the speed of decay of the weights.
- Small α gives less weight to the recent observation and slower decay.
- Large α gives more weight to the recent observation and faster decay.
- We can write the geometric series in a more compact form as,

$$l_t = \alpha y_t + (1 - \alpha) l_{t-1}$$

51

SIMPLE EXPONENTIAL SMOOTHING

- We can compute a simple exponentially smoothed series in the following way:
- Start with the initial values of the time series y_0 and y_1 then, we set $l_0 = y_0$ and compute:

$$l_1 = \alpha y_1 + (1 - \alpha) l_0$$

- Then, proceed to compute:

$$l_2 = \alpha y_2 + (1 - \alpha) l_1$$

- And proceed iteratively until

$$l_T = \alpha y_T + (1 - \alpha) l_{T-1}$$

Through recursive substitution we can see that this becomes the original geometric series that we presented initially a couple of slides ago!

52

NEXT WEEK

- We will explore exponential smoothing in a bit more detail
- Covariance stationarity
- Autoregressive models