Statistics 5350/7110
Time Series Analysis

Lecture 10
Decomposing Time Series

Professor Robert Stine

Admin Issues

- Questions
- Assignments
  - Q2 due today
- Quick review
  - Scatterplot smoothing
    Moving averages, kernel smoothing
    Polynomials, smoothing splines, lowess and loess
  - Cross-validation and the choice of bandwidth
  - Global constant controls amount of smoothing

Today's Topics

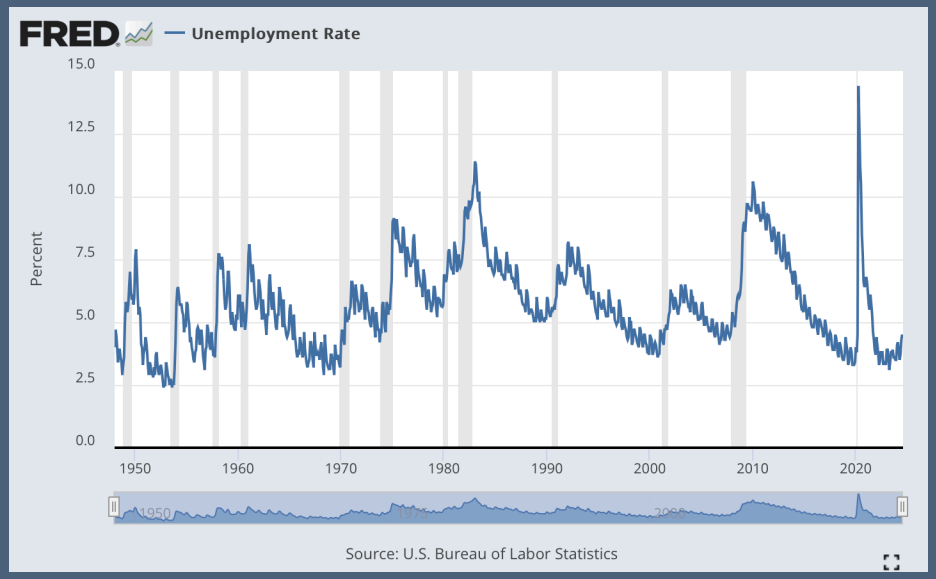
Text, §3.3
FPP, Chap 6

- Decomposing a time series
  - Official statistics often remove seasonal patterns from data
  - Rational: Don't want users confusing a seasonal pattern with "substantive" changes
    e.g. Unemployment in US goes up in the summer when schools let out
  - Typically descriptive or treated as "preprocessing"
- Techniques
  - Median polish (robust alternative to anova)
  - Generalized additive model (GAM)
  - Black boxes with many tunable attributes, relying on judgement
    - Loess-based STL (a descendant of X-11 developed at US Census)
- Examples with R
  - Data preparation from on-line source (FRED)
  - Unemployment data
  - Climate (southern oscillation index) shown in Rmd file.

Seasonal Adjustment

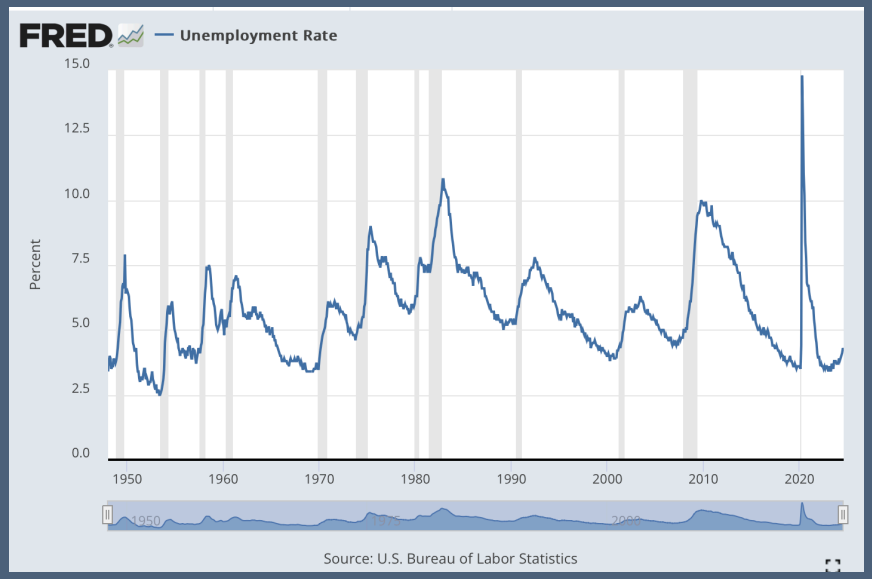
- Monthly US unemployment
  - Shaded intervals in FRED graph locate recessions in the US
  - Covid spike
  - How do they go from the raw values to the seasonally adjusted values?

Raw



?

Seasonally Adjusted



Time Series Decomposition

- Represent a time series X_t as sum of three components (or product)

$$X_t = T_t + S_t + N_t$$

  - Trend… slowly varying, smooth
  - Seasonal… periodic
  - Noise, remainder, irregular… everything that's not trend or seasonal. Often a stationary process.
- Additive or multiplicative
  - Distinguish from relationship between mean and variation
  - Logs convert multiplicative to additive
  - Substantive insight: How you think seasonal factors work
    Example: Does travel in the summer increase by additive amount or percentage gain?
- Questions
  - Estimate the trend first or the seasonal first? Simultaneously?
  - How to estimate either?

Decompose with Median Polish

- Tabular arrangement
  - Monthly: Organize time series into table with 12 columns, one for each month. Rows represent years.
  - Estimate row (annual) and column (month) effects using medians
  - Like a two-way analysis of variance, with medians replacing means

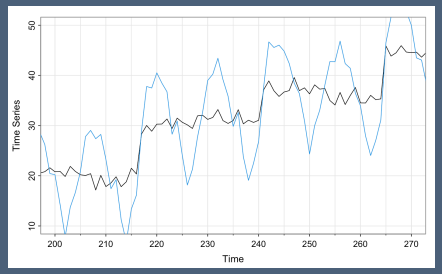
$$X_{rc} = \mu + \alpha_r + \beta_c + e_{ij} \quad \Rightarrow \quad X_{year,month} = \mu + \alpha_{year} + \beta_{month} + e_{y,m}$$

  - Ambiguous (need to impose constraints)
- Neat aspects
  - Simple, intuitive, and very fast
  - Robust to outliers by using median rather than mean
  - Anticipates more sophisticated methods with better trend estimates
  - Handy diagnostic plot to check for additivity
- Weakness
  - Not such a great model for overall trend (piecewise constant), particularly if trend is strong

Tabular arrangement anticipates STL algorithm

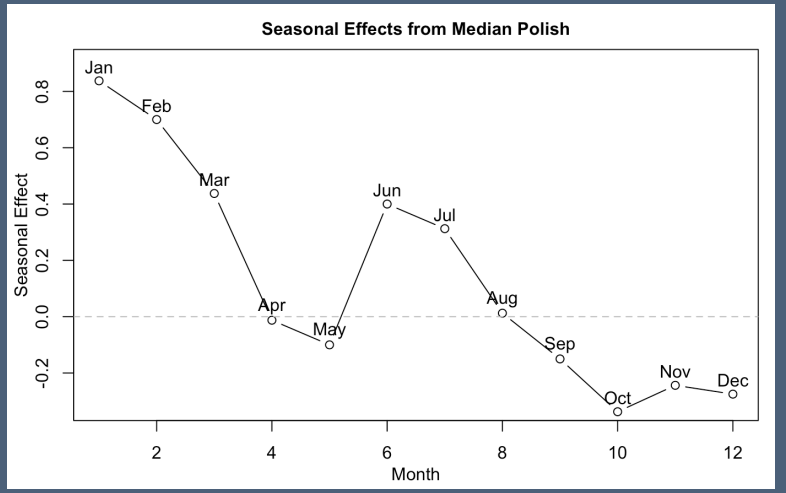
Iterative calculation, unlike sweeping out means

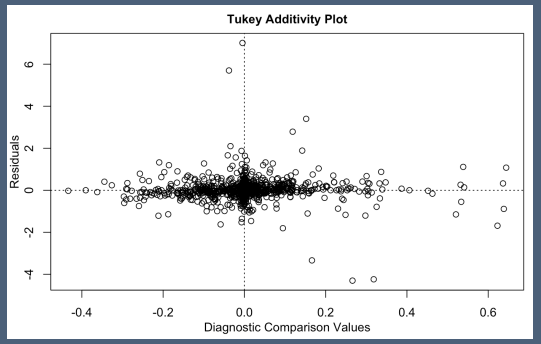
Tukey was famous for clever diagnostic plots (e.g. one degree of freedom test for non-additivity)

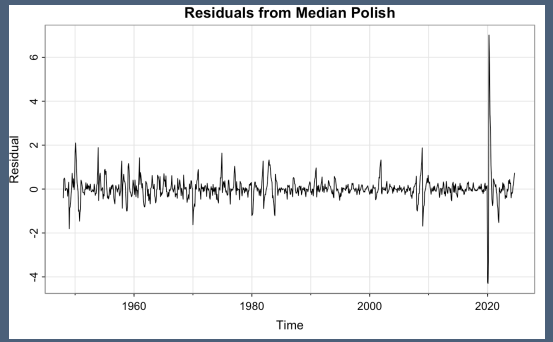


Decomposition by Median Polish

- Apply to Unemployment
  - Arrange data into a table with 77 rows and 12 columns
  - Deal with missing values (incomplete in 2024)
- Seasonality
  - Column medians estimate seasonality
  - Peaks in January and June (post holiday, summer vacation)
- Diagnostic analysis
  - Diagnostic shows no evidence of multiplicative effect
  - Outliers associated with Covid easy to spot



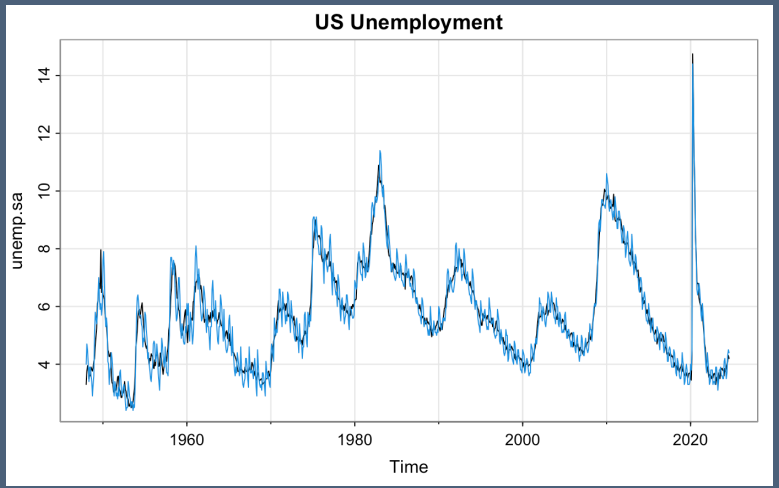




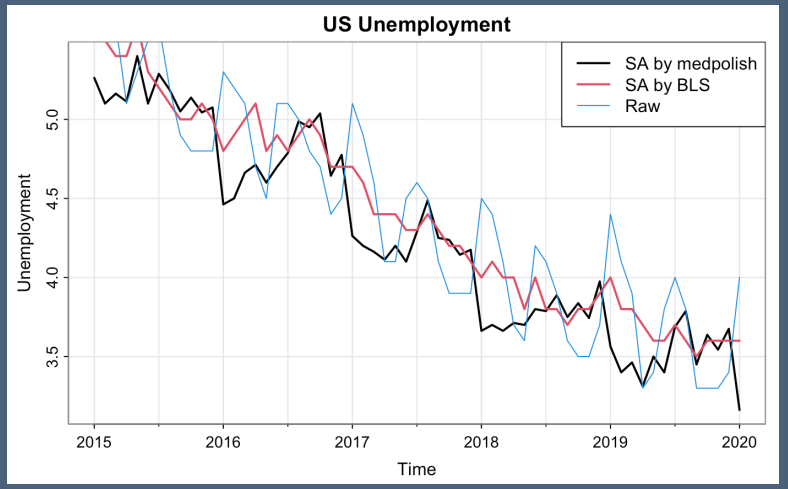
This whiplash of residuals in 2020 is a by-product of the limited trend fit by a median polish, namely one value for each year (piecewise constant).

Seasonally Adjusted Unemployment

- Seasonally adjusted time series
  - Use fit from median polish
  - Reconstruct data without seasonal terms
  - Hard to see effects over full time span



- Zoom in
  - Time series for 2015 - 2020
  - Include official seasonally adjusted unemployment rate
- Comparison
  - Suppress seasonal jumps
  - Medpolish trend inadequate (e.g. 2018 flat)

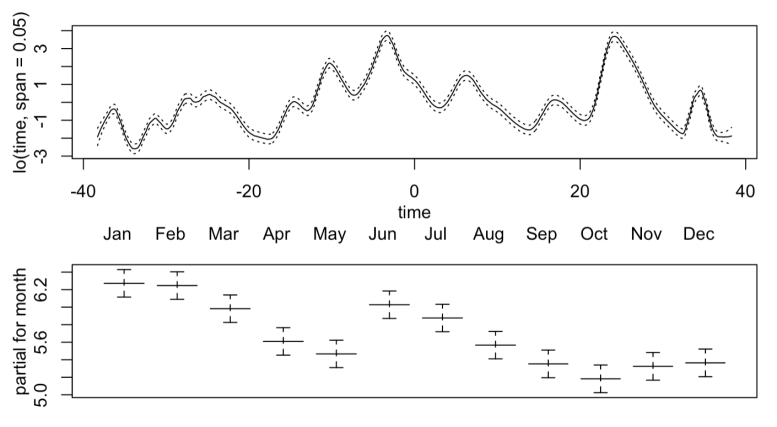


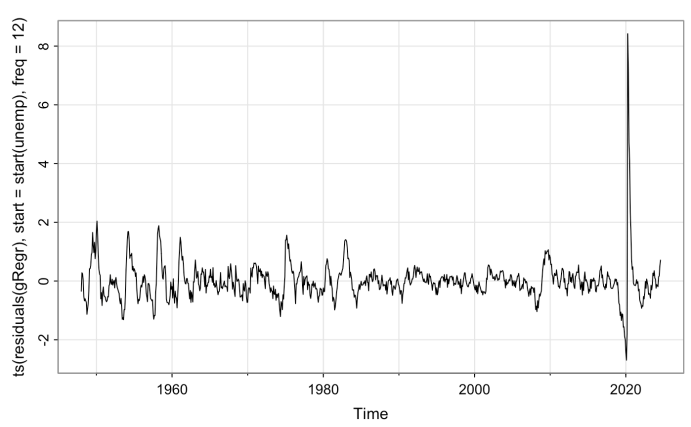
Decomposition by Regression

- Improve on the median polish
  - Allow a slowly varying smooth trend rather than piecewise constant
  - Possible smooths from splines, moving averages, or loess.
- Generalized additive model
  - Combine smooth features with standard regressors
    - Estimate the trend with loess
    - Estimate the seasonal component from monthly dummy variables
  - Provides standard errors for estimates (but these are not reliable due to autocorrelation)

GAM fit results:
Explore various spans

Seasonality resembles median polish result

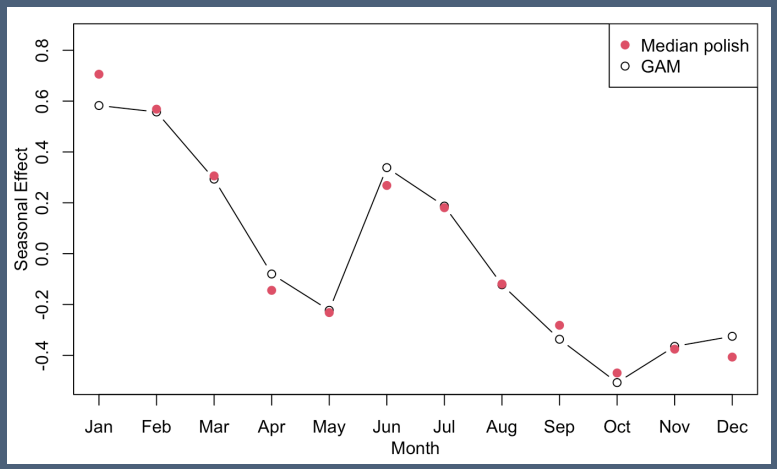




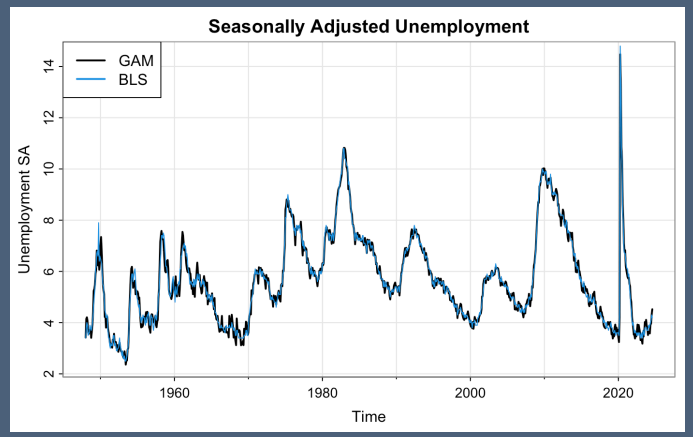
Whiplash persists since this model leaks the future into the past

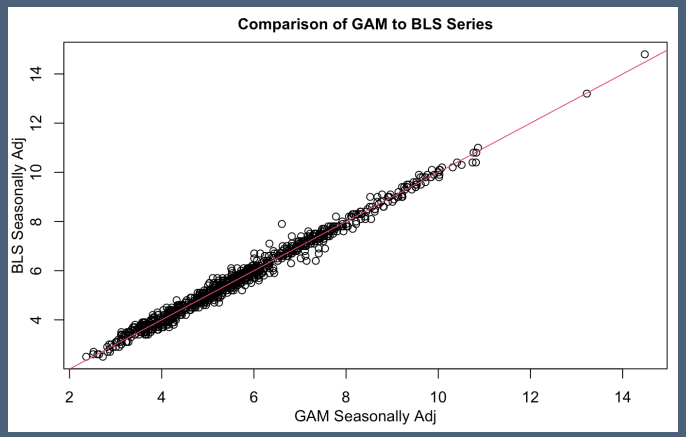
Comparisons

- Seasonal terms similar to those from median polish
  - Both scaled so average seasonal effect is zero.
  - GAM also provides std. error (albeit not reliable in this application)



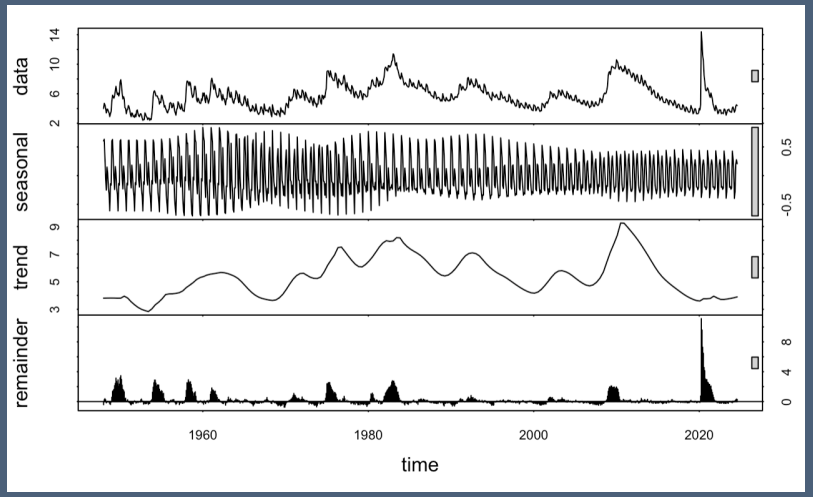
- Similar to BLS seasonal adjustment



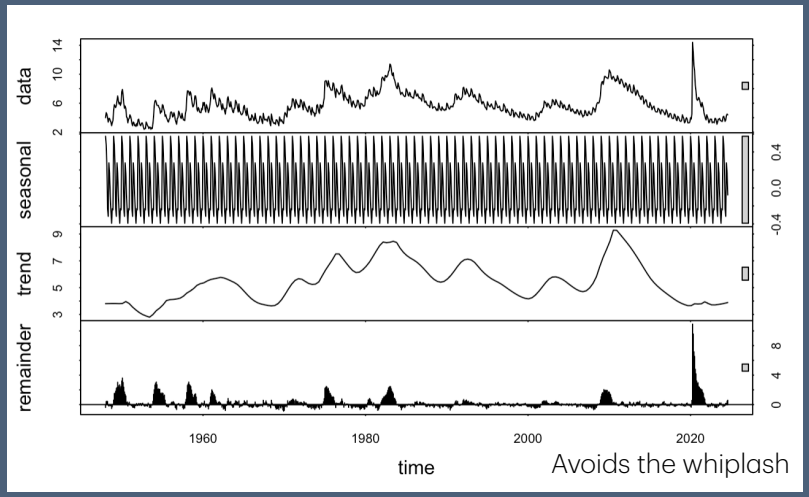


Decomposition using STL

- STL automatic decomposition
  - Estimates trend and seasonal using loess
  - Seasonal component can be allowed to vary over time
- Seasonal patterns
  - Fixed periodic effect, as in dummy variable model
  - Alternatively using loess fit to sub-series (e.g. sequence of January values)

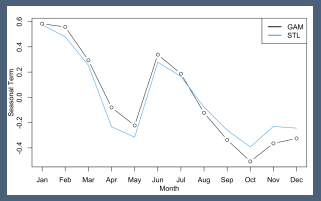


Smooth



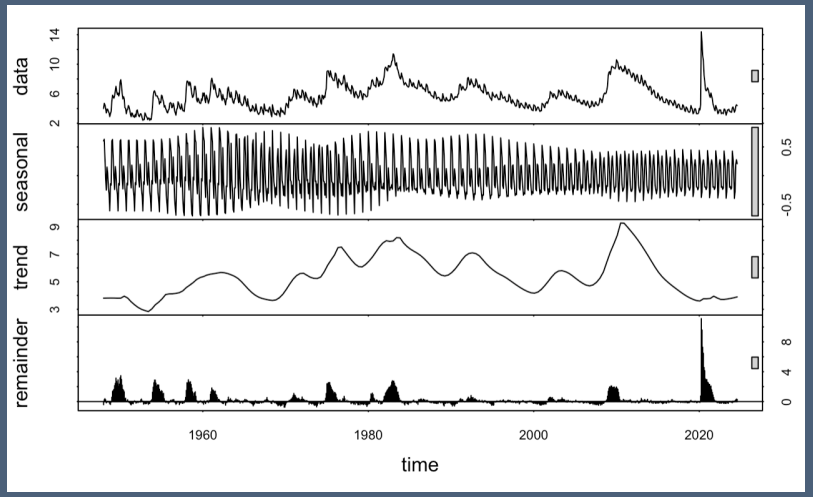
Periodic

Compare to GAM estimates

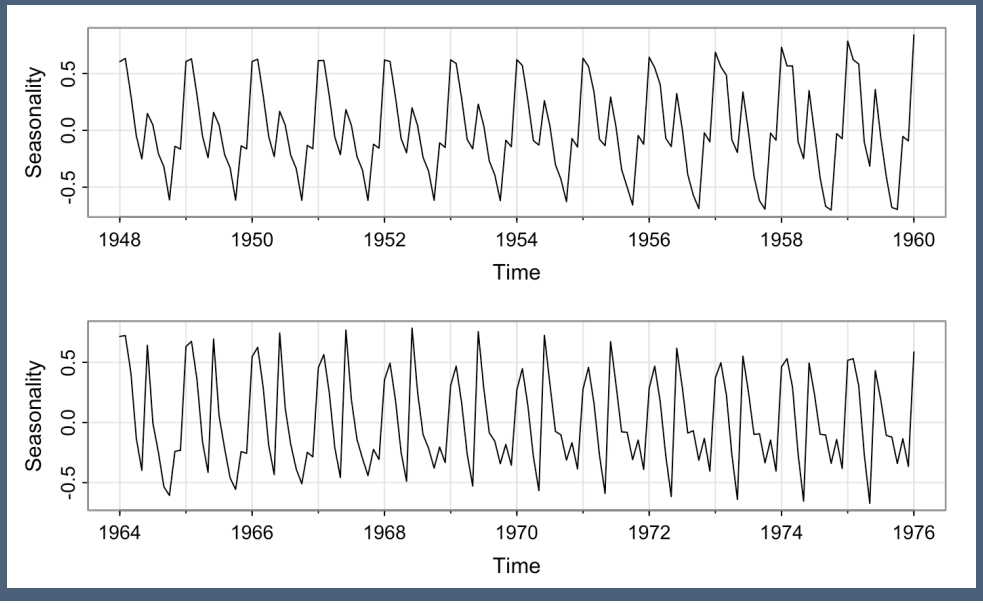


Decomposition using STL

- Flexible seasonal patterns
  - Loess fit to sub-series (e.g. sequence of January values)
  - Allow subtle changes to seasonality
- Example
  - Pre-1960: Peak seasonal effect in January
  - Post-1964: Peak seasonal effect shifts to June



Smooth



Discussion

- Ambiguity between “trend” and “seasonal”
  - What’s a trend? What’s seasonal?
  - Is that a trend or a seasonal variation?
- Seasonal rigidity
  - Periodic seasonal modeling is rigid, assuming constant effect over the years
  - Effect might be proportional, suggesting multiplicative model
  - Employment example shows could be important to allow drift
- Inference
  - Are these changes in seasonality large, relative to random variation?
  - Easy to lose sight of confidence intervals
  - Provided by a GAM model, though dubious here due to autocorrelation
- Revisions
  - Decomposition methods use data from the future to seasonally adjust data in the past!
  - With more recent data, often change impression of what happened previously

What next?

- ARMA models for time series
  - autoregressive, moving average (maybe better autoregressive, moving sum)
  - a.k.a., Box-Jenkins models
  - You've already seen them in their simplest guise...
    Combine autoregression with a moving average of errors