Trend-seasonal decomposition and exponential smoothing models

Data and Tasks

Data and Tasks: Plan

- Time series is a data type

Data and Tasks: Plan

- Time series is a data type
- Tasks for one row

Data and Tasks: Plan

- Time series is a data type
- Tasks for one row
- Tasks for multiple rows

What is a time series?

Time series

A sequence of observations ordered in time

$$0, 0, 5, 7, 102, 53, 23$$

What is a time series?

Time series

A sequence of observations ordered in time

$$0, 0, 5, 7, 102, 53, 23$$

Time series

A sequence of random variables ordered in time

$$y_1, y_2, y_3, y_4, \ldots, y_T$$

Tasks for one series

- Predict the following values

Tasks for one series

- Predict the following values
- Restore missing values in the middle of a series

Tasks for one series

- Predict the following values
- Restore missing values in the middle of a series
- Restore individual observations from aggregated ones

Tasks for one series

- Predict the following values
- Restore missing values in the middle of a series
- Restore individual observations from aggregated ones
- Detect point of discord (or structural break)

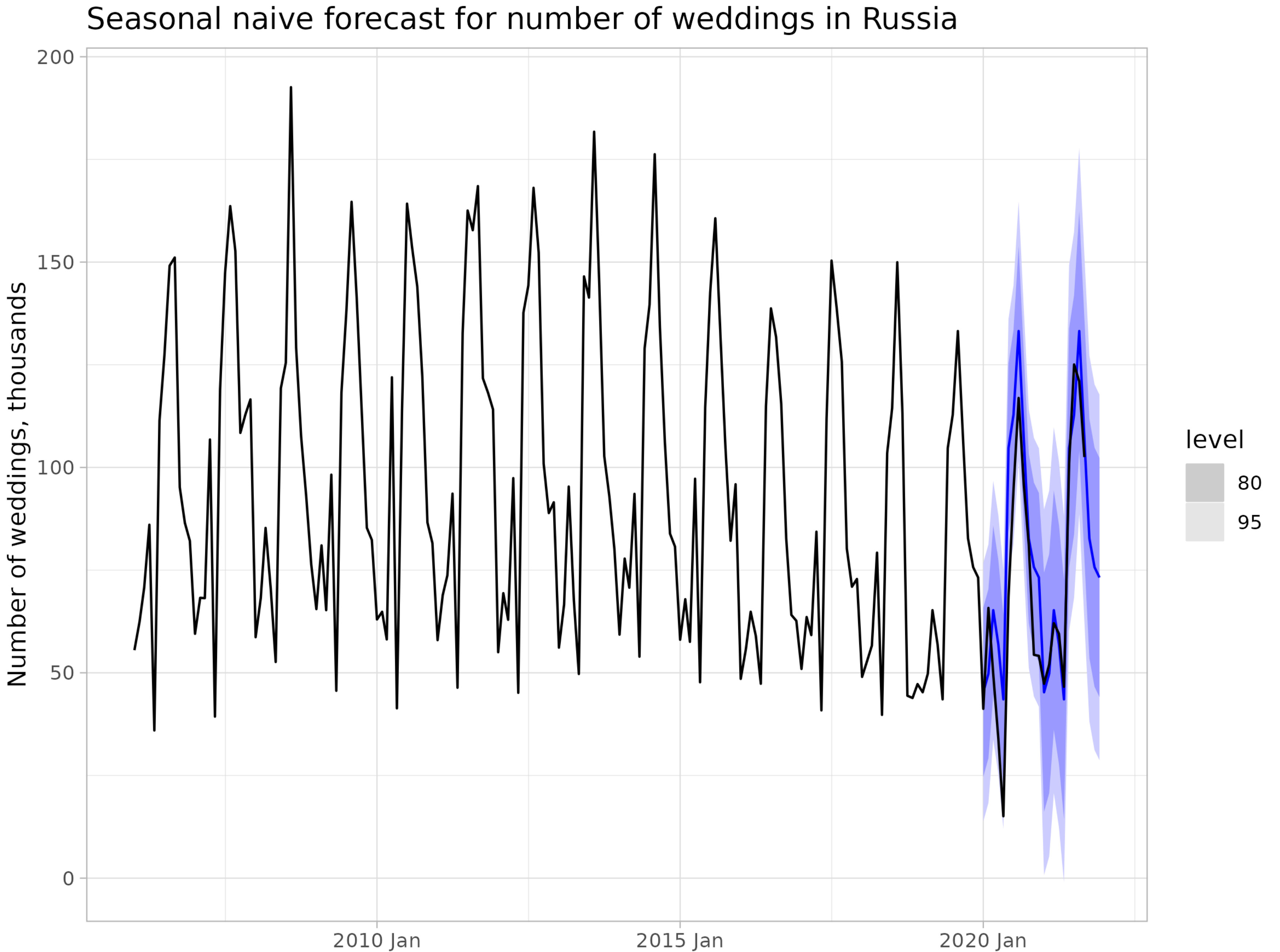
Tasks for one series

- Predict the following values
- Restore missing values in the middle of a series
- Restore individual observations from aggregated ones
- Detect point of discord (or structural break)
- Decompose series to a trend and seasonal parts

Tasks for one series

- Predict the following values
- Restore missing values in the middle of a series
- Restore individual observations from aggregated ones
- Detect point of discord (or structural break)
- Decompose series to a trend and seasonal parts
- …

Forecasting



Tasks for multiple series

- Use additional series when studying the target series

Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related

Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related
- Measure cause and effect relationships

Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related
- Measure cause and effect relationships
- Classify the new series into one of the existing classes

Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related
- Measure cause and effect relationships
- Classify the new series into one of the existing classes
- Understand which series are close to each other

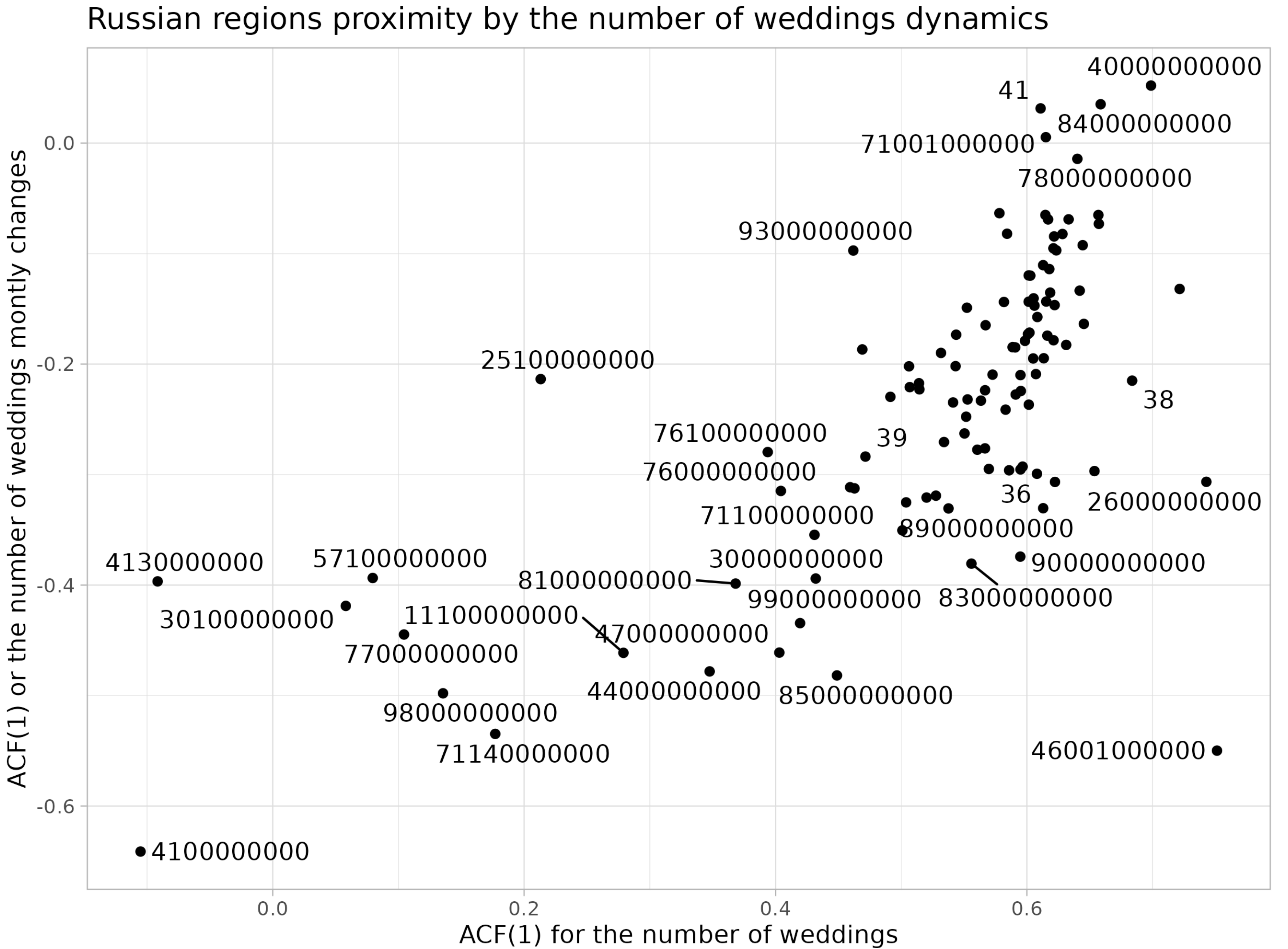
Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related
- Measure cause and effect relationships
- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

Tasks for multiple series

- Use additional series when studying the target series
- Understand if series are related
- Measure cause and effect relationships
- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters
- …

Measuring series proximity



Models and algorithms

Models

- Explicit assumptions about the values $y_1, y_2, \dots, y_T$

ETS, ARIMA, ORBIT, PROPHET, …

Models and algorithms

Models

- Explicit assumptions about the values $y_1, y_2, \ldots, y_T$
- Estimation method: maximum likelihood, Bayesian approach

ETS, ARIMA, ORBIT, PROPHET, …

Models and algorithms

Models

- Explicit assumptions about the values $y_1, y_2, \ldots, y_T$
- Estimation method: maximum likelihood, Bayesian approach
- Point and interval forecasts, hypothesis testing

ETS, ARIMA, ORBIT, PROPHET, …

Models and algorithms

Algorithms

- Fuzzy assumptions about the values $y_1, y_2, \ldots, y_T$

STL, gradient boosting, random forest, …

Models and algorithms

Algorithms

- Fuzzy assumptions about the values $y_1, y_2, \ldots, y_T$
- A special instruction for actions

STL, gradient boosting, random forest, …

Models and algorithms

Algorithms

- Fuzzy assumptions about the values $y_1, y_2, \ldots, y_T$
- A special instruction for actions
- Point estimates without confidence intervals

STL, gradient boosting, random forest, …

Course Focus

Forecasting one-dimensional series using models

Series Components

Series Components: Plan

- Trend, cyclicity and seasonality

Series Components: Plan

- Trend, cyclicity and seasonality
- Additive and multiplicative decomposition

Series Components: Plan

- Trend, cyclicity and seasonality
- Additive and multiplicative decomposition
- A formal definition?

Looking for components

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Looking for components

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Trend — smoothly changing component of the series

Looking for components

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Trend — smoothly changing component of the series

Seasonal component — a component with a clear frequency and stable intensity

Looking for components

Additive series decomposition:

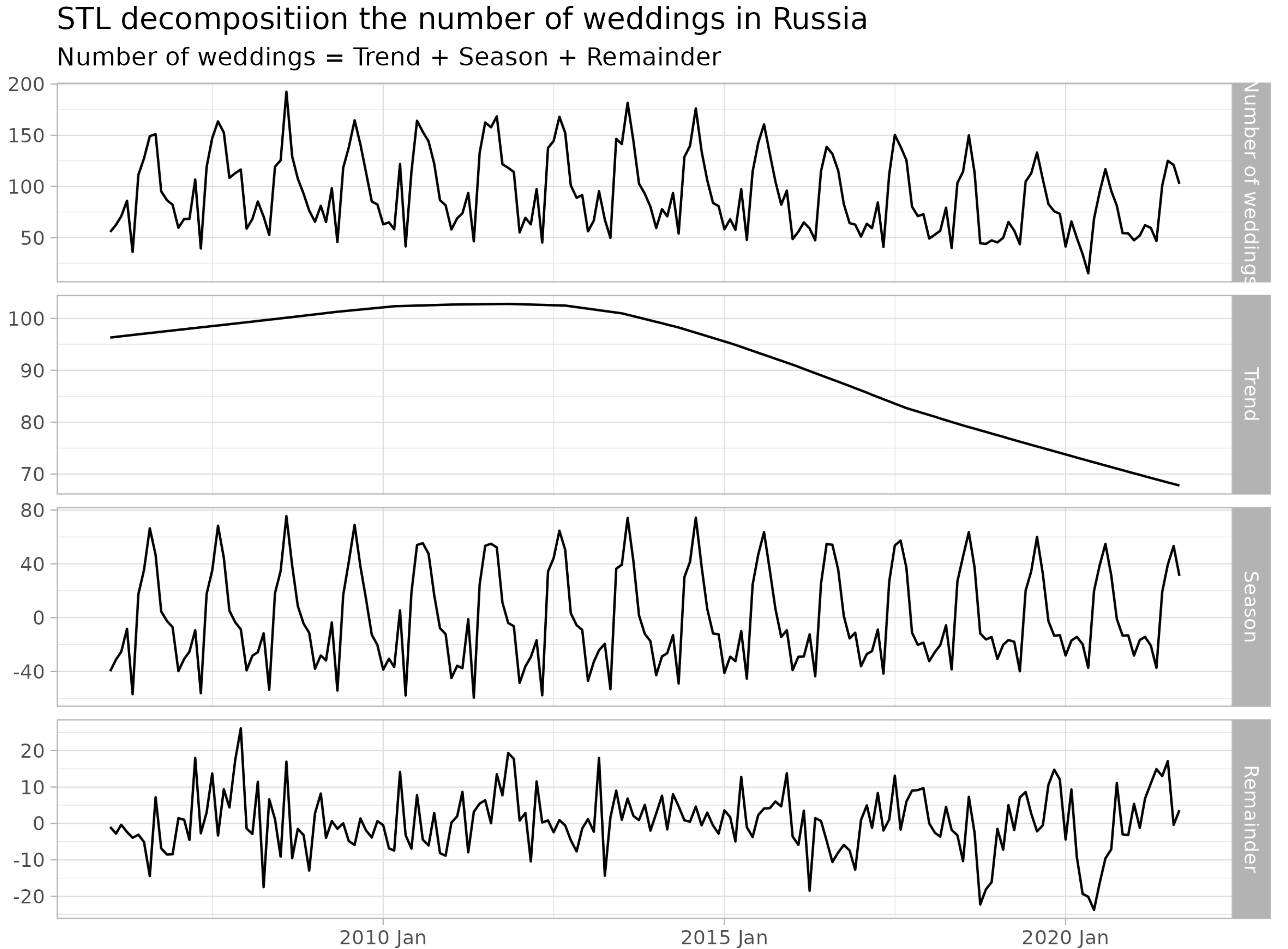
$$y_t = trend_t + seas_t + remainder_t$$

Trend — smoothly changing component of the series

Seasonal component — a component with a clear frequency and stable intensity

Random component (remainder) — everything else

Trend, seasonality and residual



Strict definition?

There will be no single strict definition for the components!

Strict definition?

There will be no single strict definition for the components!

Some models and algorithms formally define these components

Cyclical component

Sometimes the series can be decomposed further

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Cyclical component

Sometimes the series can be decomposed further

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Cyclical component — component with floating frequency and unstable intensity

Cyclical component

Sometimes the series can be decomposed further

$$y_t = trend_t + cycle_t + seas_t + remainder_t$$

Cyclical component — component with floating frequency and unstable intensity

Trend (in the narrow sense) — a smoothly changing monotonous component of a series

Additive and multiplicative decomposition

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Additive and multiplicative decomposition

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Multiplicative series decomposition:

$$y_t = trend_t \cdot seas_t \cdot remainder_t$$

Additive and multiplicative decomposition

Additive series decomposition:

$$y_t = trend_t + seas_t + remainder_t$$

Multiplicative series decomposition:

$$y_t = trend_t \cdot seas_t \cdot remainder_t$$

Let's transform one into another:

$$\ln y_t = \ln trend_t + \ln seas_t + \ln remainder_t$$

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Logarithm: $y_t \to \ln y_t$

Box-Cox transformation: $y_t \to bc_\lambda(y_t)$

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Logarithm: $y_t \to \ln y_t$

Box-Cox transformation: $y_t \to bc_\lambda(y_t)$

(Generalized) Box-Cox transformation:

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ if } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ if } \lambda \neq 0 \end{cases}$$

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Logarithm: $y_t \to \ln y_t$

Box-Cox transformation: $y_t \to bc_\lambda(y_t)$

(Generalized) Box-Cox transformation:

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ if } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ if } \lambda \neq 0 \end{cases}$$

How to select the parameter λ?

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Logarithm: $y_t \to \ln y_t$

Box-Cox transformation: $y_t \to bc_\lambda(y_t)$

(Generalized) Box-Cox transformation:

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ if } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ if } \lambda \neq 0 \end{cases}$$

How to select the parameter λ?

- Some models contain it inside and estimate λ within themselves

Box-Cox Transformation

For y_t, whose fluctuations increases with y_t, it's reasonable to try the logarithm or the Box-Cox transformation.

Logarithm: $y_t \to \ln y_t$

Box-Cox transformation: $y_t \to bc_\lambda(y_t)$

(Generalized) Box-Cox transformation:

$$bc_\lambda(y_t) = \begin{cases} \ln y_t, \text{ if } \lambda = 0, \\ \operatorname{sign}(y_t)(|y_t|^\lambda - 1)/\lambda, \text{ if } \lambda \neq 0 \end{cases}$$

How to select the parameter λ?

- Some models contain it inside and estimate λ within themselves
- You can choose λ by yourself to stabilize the amplitude of oscillations of the series

What to choose?

The formal definition of depends on the model

What to choose?

The formal definition of depends on the model

STL algorithm: one decomposition

$$y_t = trend_t + seas_t + remainder_t$$

What to choose?

The formal definition of depends on the model

STL algorithm: one decomposition

$$y_t = trend_t + seas_t + remainder_t$$

ETS(AAA): different decomposition

$$y_t = trend_t + seas_t + remainder_t$$

What to choose?

The formal definition of depends on the model

STL algorithm: one decomposition

$$y_t = trend_t + seas_t + remainder_t$$

ETS(AAA): different decomposition

$$y_t = trend_t + seas_t + remainder_t$$

It is important to understand the purpose of constructing the decomposition

Why decompose?

- Interesting by itself

Why decompose?

- Interesting by itself
- For predicting a series using component prediction

Why decompose?

- Interesting by itself
- For predicting a series using component prediction
- To get characteristics of the series

Why decompose?

- Interesting by itself
- For predicting a series using component prediction
- To get characteristics of the series

Why decompose?

- Interesting by itself
- For predicting a series using component prediction
- To get characteristics of the series

Why characteristics?

- To classify the new series into one of the given classes

Why decompose?

- Interesting by itself
- For predicting a series using component prediction
- To get characteristics of the series

Why characteristics?

- To classify the new series into one of the given classes
- To identify unknown clusters in series

Series Components: Summary

- Trend smoothly changes and includes a cyclical component

Series Components: Summary

- Trend smoothly changes and includes a cyclical component
- The seasonal component has clear periodicity and stable amplitude

Series Components: Summary

- Trend smoothly changes and includes a cyclical component
- The seasonal component has clear periodicity and stable amplitude
- The exact formalization of the components depends on the model

Naive Models

Naive Models: Plan

- White noise

Naive Models: Plan

- White noise
- Independent observations

Naive Models: Plan

- White noise
- Independent observations
- Random walk

White noise

White noise

Time series u_t is white noise if:

- $\mathbb{E}(u_t) = 0$;
- $\mathrm{Var}(u_t) = \sigma^2$;
- $\mathrm{Cov}(u_s, u_t) = 0$ for $s \neq t$

White noise

White noise

Time series u_t is white noise if:

- $\mathbb{E}(u_t) = 0$;
- $\mathrm{Var}(u_t) = \sigma^2$;
- $\mathrm{Cov}(u_s, u_t) = 0$ for $s \neq t$

- An integral part of all models; most often, white noise is not modelled explicitly

White noise

White noise

Time series u_t is white noise if:

- $\mathbb{E}(u_t) = 0$;
- $\mathrm{Var}(u_t) = \sigma^2$;
- $\mathrm{Cov}(u_s, u_t) = 0$ for $s \neq t$

- An integral part of all models; most often, white noise is not modelled explicitly
- Often independence and normality are assumed
  ARCH, GARCH volatility models are based on the fact that u_t and u_s can be dependent!

Independent observations

Model

$$y_t = \mu + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$

Independent observations

Model

$$y_t = \mu + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$

Estimators:

$$\hat{\mu}_{ML} = \bar{y}, \quad \hat{\sigma}^2_{ML} = \frac{\sum (y_i - \bar{y})^2}{T}$$

Independent observations

Model

$$y_t = \mu + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$

Estimators:

$$\hat{\mu}_{ML} = \bar{y}, \quad \hat{\sigma}^2_{ML} = \frac{\sum (y_i - \bar{y})^2}{T}$$

Interval forecast h steps ahead:

$$[\bar{y} - 1.96\hat{\sigma}; \bar{y} + 1.96\hat{\sigma}]$$

Random Walk

Naive model

$$y_t = y_{t-1} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, starting y_1 is given

Random Walk

Naive model

$$y_t = y_{t-1} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, starting y_1 is given

Let's reformulate: $y_t - y_{t-1} = \Delta y_t = u_t$

Random Walk

Naive model

$$y_t = y_{t-1} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, starting y_1 is given

Let's reformulate: $y_t - y_{t-1} = \Delta y_t = u_t$

Estimators:

$$\hat{\sigma}^2_{ML} = \frac{\sum(\Delta y_i - \overline{\Delta y})^2}{T-1}$$

Random Walk

Naive model

$$y_t = y_{t-1} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, starting y_1 is given

Let's reformulate: $y_t - y_{t-1} = \Delta y_t = u_t$

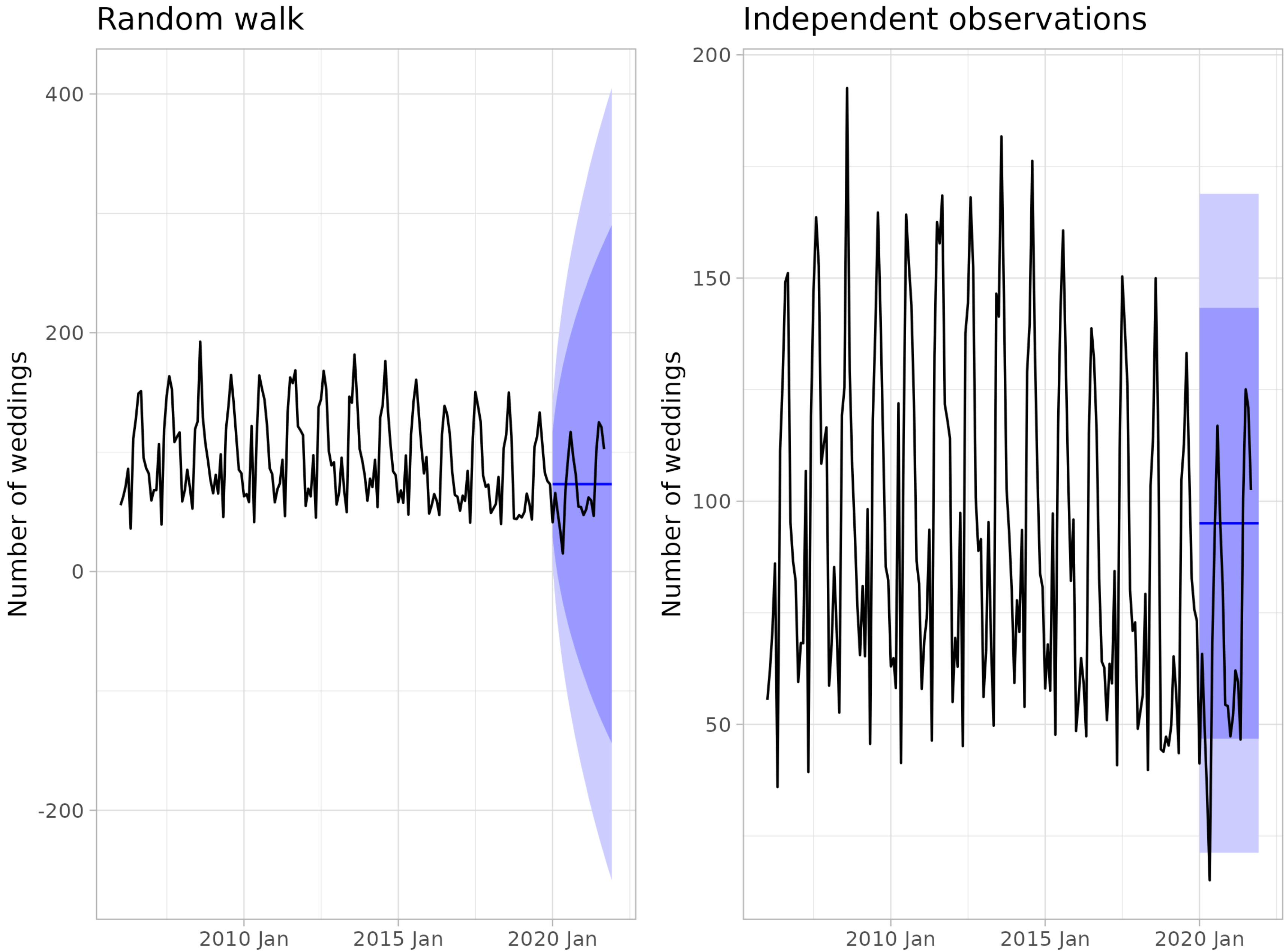
Estimators:

$$\hat{\sigma}^2_{ML} = \frac{\sum(\Delta y_i - \overline{\Delta y})^2}{T-1}$$

Interval forecast h steps ahead:

$$[y_T - 1.96\hat{\sigma}\sqrt{h}; y_T + 1.96\hat{\sigma}\sqrt{h}]$$

First predictions!



Seasonal random walk

Seasonal naive model

$$y_t = y_{t-12} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, y_1, …, y_{11} are given

Seasonal random walk

Seasonal naive model

$$y_t = y_{t-12} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, y_1, …, y_{11} are given

Let's reformulate: $y_t - y_{t-12} = \Delta_{12} y_t = u_t$

Seasonal random walk

Seasonal naive model

$$y_t = y_{t-12} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, y_1, …, y_{11} are given

Let's reformulate: $y_t - y_{t-12} = \Delta_{12} y_t = u_t$

Estimators:

$$\hat{\sigma}^2_{ML} = \frac{\sum(\Delta_{12} y_i - \overline{\Delta_{12} y})^2}{T-12}$$

Seasonal random walk

Seasonal naive model

$$y_t = y_{t-12} + u_t,$$

where u_t is white noise, $u_t \sim \mathcal{N}(0; \sigma^2)$, $y_1, \ldots, y_{11}$ are given

Let's reformulate: $y_t - y_{t-12} = \Delta_{12} y_t = u_t$

Estimators:

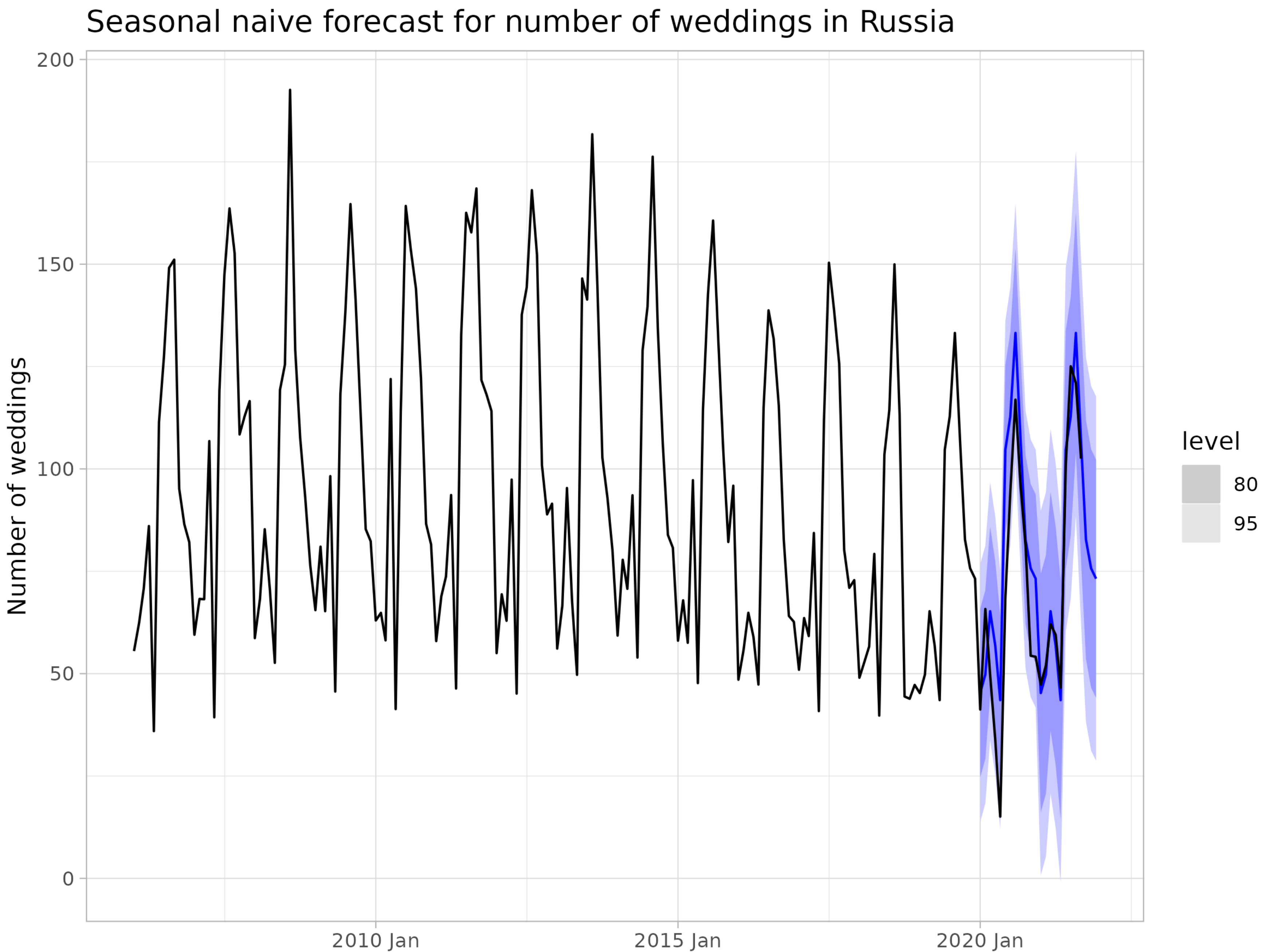
$$\hat{\sigma}^2_{ML} = \frac{\sum (\Delta_{12} y_i - \overline{\Delta_{12} y})^2}{T - 12}$$

Interval forecast for h steps ahead:

$$\left[y_{T-12+h\%12} - 1.96\hat{\sigma}\sqrt{\left\lceil \frac{h}{12} \right\rceil}; y_{T-12+h\%12} + 1.96\hat{\sigma}\sqrt{\left\lceil \frac{h}{12} \right\rceil} \right]$$

where $\%$ - remainder from the division, $\lceil x \rceil$ - ceiling function

Not bad already!



Why do we need naive models?

- Ideas for complex model:
  the stationary series models are similar to the independent observations model;
  non-stationary series models are similar to a random walk

Why do we need naive models?

- Ideas for complex model:
  the stationary series models are similar to the independent observations model;
  non-stationary series models are similar to a random walk
- Benchmark for comparison:
  when evaluating a complex model, it is very important to have a base of comparison

Why do we need naive models?

- Ideas for complex model:
  the stationary series models are similar to the independent observations model;
  non-stationary series models are similar to a random walk
- Benchmark for comparison:
  when evaluating a complex model, it is very important to have a base of comparison
- Averaging with other models' forecasts:
  you can average forecasts of a complex model and a naive seasonal one!

Naive Models: Summary

- White noise is what you don't want to simulate

Naive Models: Summary

- White noise is what you don't want to simulate
- Independent observations and random walk

Naive Models: Summary

- White noise is what you don't want to simulate
- Independent observations and random walk
- Ideas, parts, and helpers for other models

Naive Models: Summary

- White noise is what you don't want to simulate
- Independent observations and random walk
- Ideas, parts, and helpers for other models
- Base for comparison

STL algorithm

STL algorithm: Plan

- Local regression

STL algorithm: Plan

- Local regression
- STL outer loop

STL algorithm: Plan

- Local regression
- STL outer loop
- STL inner loop

STL algorithm: Plan

- Local regression
- STL outer loop
- STL inner loop
- STL options

STL

STL — Seasonal Trend decompositon with LOESS

STL — seasonality and trend decomposition using LOESS

STL

STL — Seasonal Trend decompositon with LOESS

STL — seasonality and trend decomposition using LOESS

LOESS — LOcal regrESSion

LOESS — local linear regression

STL as a black box

Input:

Row Y_t

Algorithm parameters n_p, n_i, n_o, n_l, n_s, n_t

STL as a black box

Input:

Row Y_t

Algorithm parameters n_p, n_i, n_o, n_l, n_s, n_t

Output:

Decomposition $Y_t = T_t + S_t + R_t$

STL as a black box

Input:

Row Y_t

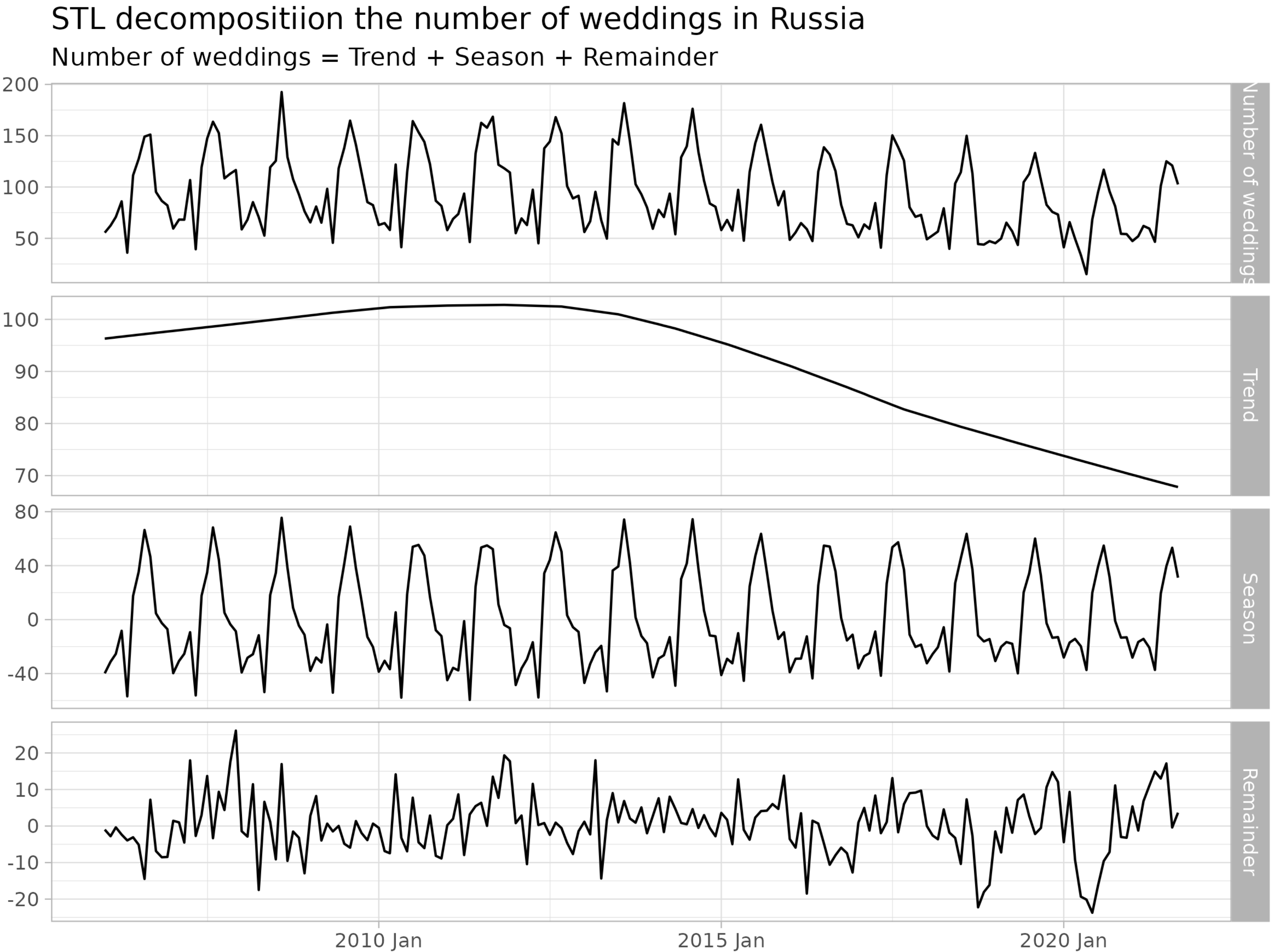
Algorithm parameters n_p, n_i, n_o, n_l, n_s, n_t

Output:

Decomposition $Y_t = T_t + S_t + R_t$

Black box set up

STL: result



STL decompositiion the number of weddings in Russia
Number of weddings = Trend + Season + Remainder
200
150
100
50
Number of weddings
100
90
80
70
Trend
80
40
0
-40
Season
20
10
0
-10
-20
Remainder
2010 Jan
2015 Jan
2020 Jan

LOESS

- We want to build a forecast for the point x
- Find local estimates $\hat{\beta}_1(x)$, $\hat{\beta}_2(x)$

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Predicting:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x) x.$$

LOESS

- We want to build a forecast for the point x
- Find local estimates $\hat{\beta}_1(x)$, $\hat{\beta}_2(x)$

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Predicting:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

Kernel function

- The function $K_h(x_i - x)$ decreases with increasing distance $|x_i - x|$;
- The h parameter controls the width of the smoothing window

LOESS

- We want to build a forecast for the point x
- Find local estimates $\hat{\beta}_1(x)$, $\hat{\beta}_2(x)$

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$$

- Predicting:

$$\hat{y} = \hat{\beta}_1(x) + \hat{\beta}_2(x)x.$$

Kernel function

- The function $K_h(x_i - x)$ decreases with increasing distance $|x_i - x|$;
- The h parameter controls the width of the smoothing window

For example, h is the number of points x_i next to x that we take into account

Nuances of local regression

- Select degrees of the polynomial

$$\min \sum_i K_h(x_i - x)(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i - \hat{\beta}_3 x_i^2)^2$$

- Select kernel function

$$K_h(d) = \frac{1}{\sqrt{2\pi h}} \exp\left(-d^2/2h^2\right)$$

- Select window width h

STL algorithm

Purpose: decomposition of $Y_t = T_t + S_t + R_t$

The algorithm contains two loops: outer and inner loop

1. Initialize $T_t = 0$, $R_t = 0$

   Outer loop:

STL algorithm

Purpose: decomposition of $Y_t = T_t + S_t + R_t$

The algorithm contains two loops: outer and inner loop

1. Initialize $T_t = 0$, $R_t = 0$

   Outer loop:

2. Calculate the weight of each observation, ρ_t:

   on the first pass, $\rho_t = 1$ for each observation;

   on subsequent passes, ρ_t depends negatively on the new value of R_t

STL algorithm

Purpose: decomposition of $Y_t = T_t + S_t + R_t$

The algorithm contains two loops: outer and inner loop

1. Initialize $T_t = 0$, $R_t = 0$

   Outer loop:

2. Calculate the weight of each observation, ρ_t:

   on the first pass, $\rho_t = 1$ for each observation;

   on subsequent passes, ρ_t depends negatively on the new value of R_t

3. Update the current decomposition $Y_t = T_t + S_t + R_t$ taking into account new weights ρ_t

STL: inner loop

Goal: update the decomposition $Y_t = T_t + S_t + R_t$.

1. Remove the previously calculated trend from the series:

$$Y_t^{det} = Y_t - T_t.$$

STL: inner loop

Goal: update the decomposition $Y_t = T_t + S_t + R_t$.

1. Remove the previously calculated trend from the series:
$$Y_t^{det} = Y_t - T_t.$$
2. Divide the detrended series into 12 series (one for each season)

STL: inner loop

Goal: update the decomposition $Y_t = T_t + S_t + R_t$.

1. Remove the previously calculated trend from the series:

$$Y_t^{det} = Y_t - T_t.$$

2. Divide the detrended series into 12 series (one for each season)
3. Smooth each of the series individually with LOESS:

$$C^{jan} = LOESS_\rho(Y_{jan}^{det}), C^{feb} = LOESS_\rho(Y_{feb}^{det}), \ldots$$

STL: inner loop

Goal: update the decomposition $Y_t = T_t + S_t + R_t$.

1. Remove the previously calculated trend from the series:

$$Y_t^{det} = Y_t - T_t.$$

2. Divide the detrended series into 12 series (one for each season)

3. Smooth each of the series individually with LOESS:

$$C^{jan} = LOESS_\rho(Y_{jan}^{det}), C^{feb} = LOESS_\rho(Y_{feb}^{det}), \ldots$$

4. Extract the low-frequency component (double moving average + LOESS):

$$L_t = LOESS(MA(MA(C_t)))$$

STL: inner loop

1-3. Remove the previously calculated trend from the series, break it down into 12 series and smooth each of them with LOESS.
4. Extract the low-frequency component L_t.
5. Get new seasonal component by removing the low-frequency component:

$$S_t^{new} = C_t - L_t$$

STL: inner loop

1-3. Remove the previously calculated trend from the series, break it down into 12 series and smooth each of them with LOESS.

4. Extract the low-frequency component L_t.

5. Get new seasonal component by removing the low-frequency component:

$$S_t^{new} = C_t - L_t$$

6. Get new trend component by removing new seasonal component from the original series and smoothing with LOESS:

$$T_t^{new} = LOESS_\rho(Y_t - S_t^{new})$$

STL parameters

- n_p — periodicity of seasonality, for example, $n_p = 12$

STL parameters

- n_p — periodicity of seasonality, for example, $n_p = 12$
- n_o is the number of iterations of the outer loop:
  the larger the number n_o, the weaker the impact of outliers;
  $n_o = 1$ is often sufficient

STL parameters

- n_p — periodicity of seasonality, for example, $n_p = 12$
- n_o is the number of iterations of the outer loop:
  the larger the number n_o, the weaker the impact of outliers;
  $n_o = 1$ is often sufficient
- n_i — number of passes of the inner loop:
  $n_i = 2$ is often enough to achieve convergence.

STL smoothing parameters

- n_l — low pass filter smoothing strength

STL smoothing parameters

- n_l — low pass filter smoothing strength
- n_s — seasonal contract smoothing strength

STL smoothing parameters

- n_l — low pass filter smoothing strength
- n_s — seasonal contract smoothing strength
- n_t — smoothing strength when highlighting a trend at the last step

  What to configure?

STL smoothing parameters

- n_l — low pass filter smoothing strength
- n_s — seasonal contract smoothing strength
- n_t — smoothing strength when highlighting a trend at the last step

  What to configure?

1. Be sure to specify the periodicity n_p

STL smoothing parameters

- n_l — low pass filter smoothing strength
- n_s — seasonal contract smoothing strength
- n_t — smoothing strength when highlighting a trend at the last step

  What to configure?

1. Be sure to specify the periodicity n_p
2. Maybe play around with n_s

STL algorithm: Summary

- LOESS — local regression

STL algorithm: Summary

- LOESS — local regression
- STL is a well-proven algorithm without an underlying model

STL algorithm: Summary

- LOESS — local regression
- STL is a well-proven algorithm without an underlying model
- If you wish, you can play around with the smoothing parameters

Series Characteristics

Series Characteristics: Plan

- Sample autocorrelation

Series Characteristics: Plan

- Sample autocorrelation
- Sample partial autocorrelation

Series Characteristics: Plan

- Sample autocorrelation
- Sample partial autocorrelation
- STL features

Tasks for multiple series

- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

Tasks for multiple series

- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

How to solve?

1. Generate features for each series

Tasks for multiple series

- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

How to solve?

1. Generate features for each series
2. Apply the algorithms for cross data to the obtained features

Tasks for multiple series

- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

How to solve?

1. Generate features for each series
2. Apply the algorithms for cross data to the obtained features

Tasks for multiple series

- Classify the new series into one of the existing classes
- Understand which series are close to each other
- Cluster series into an unknown set of clusters

How to solve?

1. Generate features for each series
2. Apply the algorithms for cross data to the obtained features

Classify using random forest;

Measure distance using the Mahalanobis metric;

Cluster using hierarchical clustering

Creating features

Two sets of features:

- Sample ACF (autocorrelation function, AutoCorrelation Function)

Creating features

Two sets of features:

- Sample ACF (autocorrelation function, AutoCorrelation Function)
- Sample PACF (partial autocorrelation function, Partial ACF)

Creating features

Two sets of features:

- Sample ACF (autocorrelation function, AutoCorrelation Function)
- Sample PACF (partial autocorrelation function, Partial ACF)

Creating features

Two sets of features:

- Sample ACF (autocorrelation function, AutoCorrelation Function)
- Sample PACF (partial autocorrelation function, Partial ACF)

From one series we get:

$ACF_1, ACF_2, ACF_3, \ldots$

$PACF_1, PACF_2, PACF_3, \ldots$

ACF

Sample ACF

Let's evaluate a set of paired regressions:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

ACF

Sample ACF

Let's evaluate a set of paired regressions:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

ACF

Sample ACF

Let's evaluate a set of paired regressions:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-k}, \quad ACF_k = \hat{\beta}_2$$

ACF

Sample ACF

Let's evaluate a set of paired regressions:

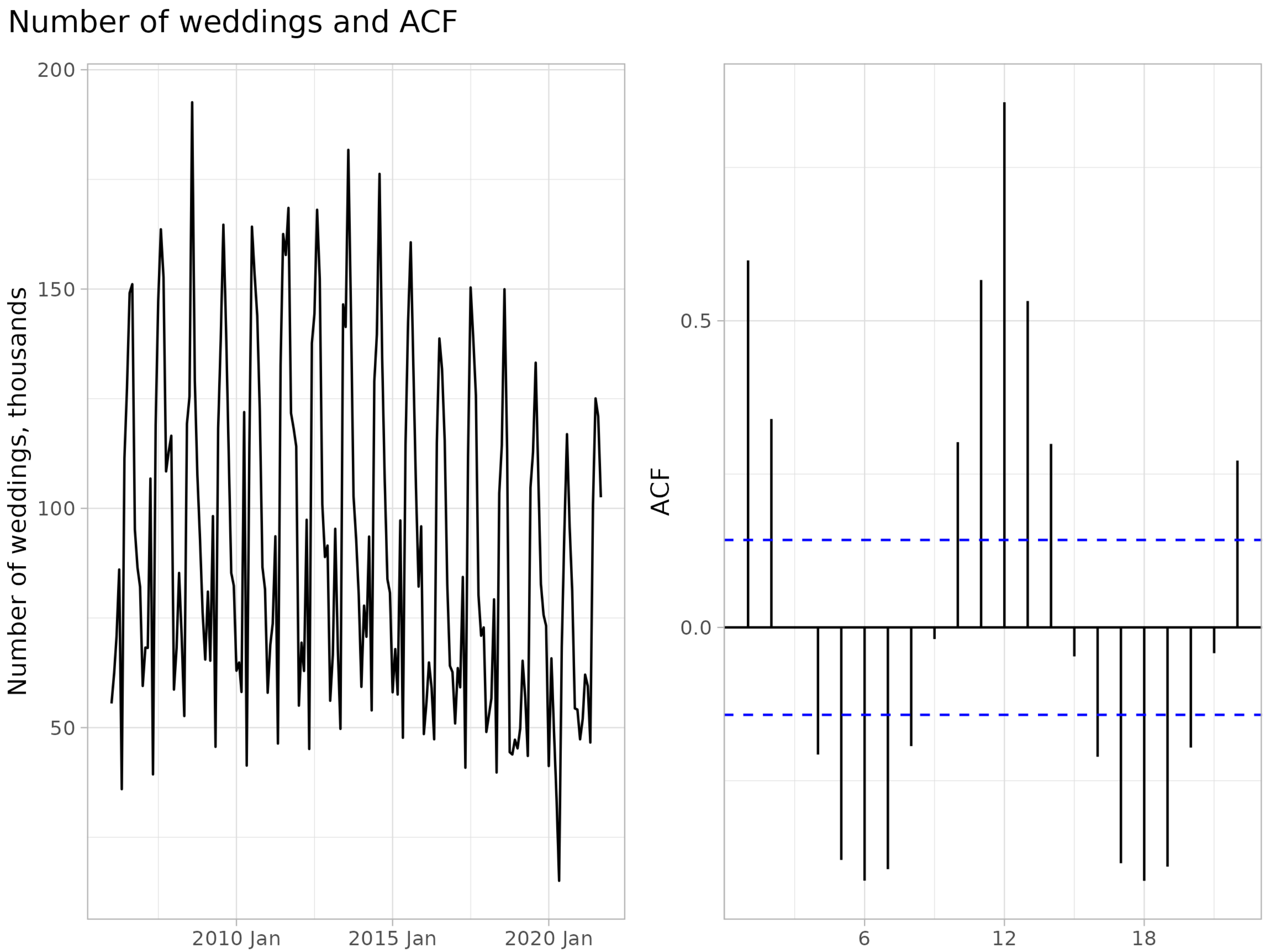
$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-1}, \quad ACF_1 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-2}, \quad ACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 y_{t-k}, \quad ACF_k = \hat{\beta}_2$$

Meaning ACF_2: How many units is y_t above average on average if y_{t-2} is one unit above average.

Series and its ACF



Why is ACF a correlation?

Classic definition

Sample ACF

ACF_k — sample correlation between series y_t and series y_{t-k}

Why is ACF a correlation?

Classic definition

Sample ACF

ACF_k — sample correlation between series y_t and series y_{t-k}

The difference between the definitions is small

PACF

Sample PACF

Let's evaluate a set of multiple regressions:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

PACF

Sample PACF

Let's evaluate a set of multiple regressions:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

PACF

Sample PACF

Let's evaluate a set of multiple regressions:

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \ldots + \hat{\beta}_k y_{t-k}, \quad PACF_k = \hat{\beta}_k$$

PACF

Sample PACF

Let's evaluate a set of multiple regressions:

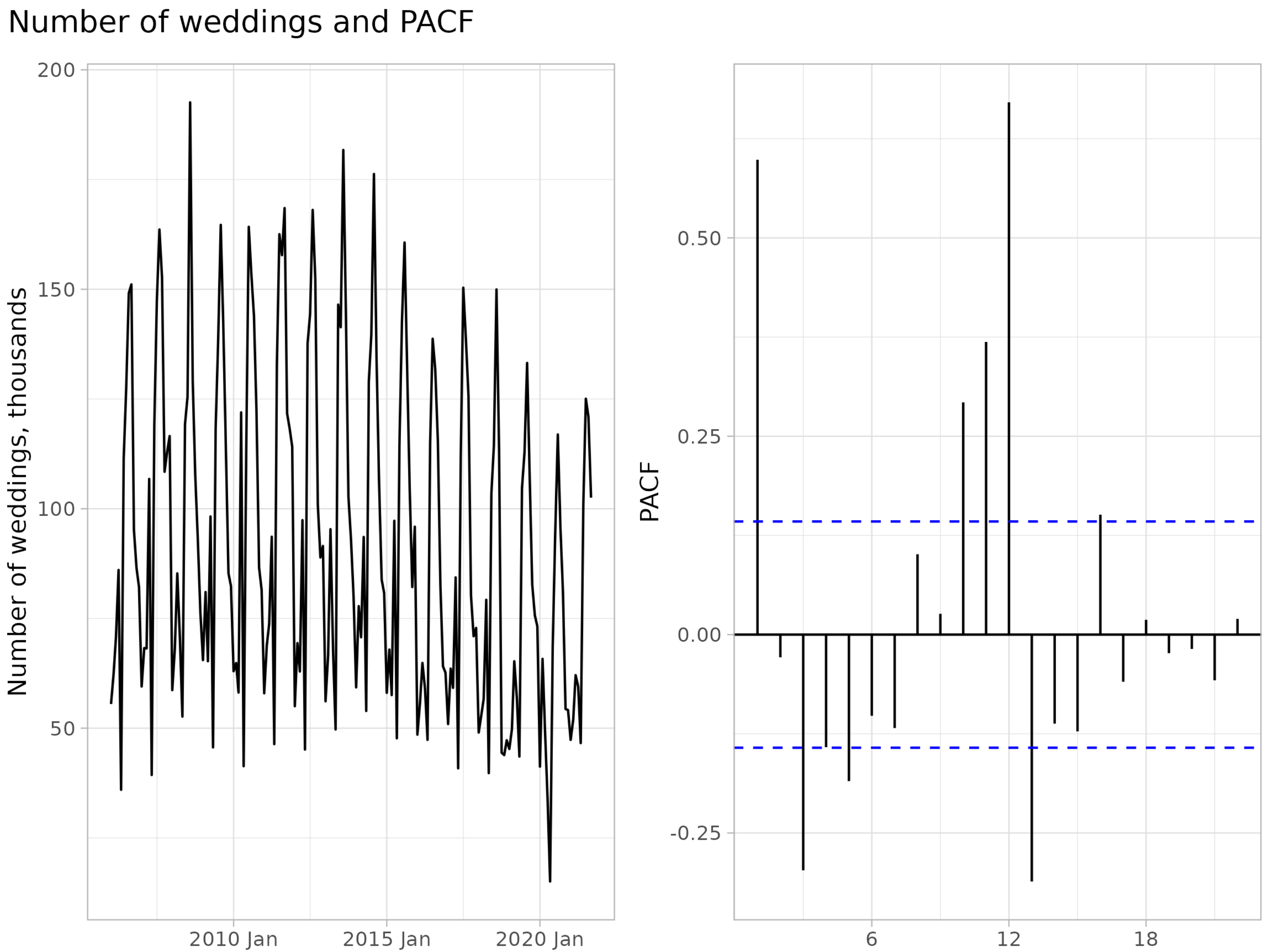
$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1}, \quad PACF_1 = \hat{\beta}_1;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2}, \quad PACF_2 = \hat{\beta}_2;$$

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}_1 y_{t-1} + \ldots + \hat{\beta}_k y_{t-k}, \quad PACF_k = \hat{\beta}_k$$

Meaning $PACF_2$: how many units is y_t above average on average if y_{t-2} is one unit above average, and y_{t-1} is at the middle level

Series and its PACF



Why is PACF a correlation?

Classic definition

Custom PACF

$PACF_4$ — sample correlation between a_t residuals and b_t residuals:

a_t — regression residuals

$$y_t \mid 1, y_{t-1}, y_{t-2}, y_{t-3};$$

b_t — regression residuals

$$y_{t-4} \mid 1, y_{t-1}, y_{t-2}, y_{t-3}$$

Why is PACF a correlation?

Classic definition

Custom PACF

$PACF_4$ — sample correlation between a_t residuals and b_t residuals:

a_t — regression residuals

$$y_t \mid 1, y_{t-1}, y_{t-2}, y_{t-3};$$

b_t — regression residuals

$$y_{t-4} \mid 1, y_{t-1}, y_{t-2}, y_{t-3}$$

The difference between the definitions of small

STL features

Output:

$$y_t = T_t + S_t + R_t$$

STL features

Output:

$$y_t = T_t + S_t + R_t$$

Let's measure:

- Strength of trend F_{trend}
- Strength of seasonality F_{seas}

Strength of trend and seasonality

We got the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

Strength of trend and seasonality

We got the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

Definition idea:

For an ideal decomposition with uncorrelated components:

$$F_{trend} = \frac{\mathrm{sVar}(trend)}{\mathrm{sVar}(trend) + \mathrm{sVar}(remainder)},$$

Strength of trend and seasonality

We got the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

Definition idea:

For an ideal decomposition with uncorrelated components:

$$F_{trend} = \frac{\mathrm{sVar}(trend)}{\mathrm{sVar}(trend) + \mathrm{sVar}(remainder)},$$

$$F_{seas} = \frac{\mathrm{sVar}(seas)}{\mathrm{sVar}(seas) + \mathrm{sVar}(remainder)},$$

Strength of trend and seasonality

We have the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

Strength of trend and seasonality

We have the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

In practice:

- Trend strength:

$$F_{trend} = \max\left\{1 - \frac{\mathrm{sVar}(remainder)}{\mathrm{sVar}(trend + remainder)}, 0\right\}.$$

Strength of trend and seasonality

We have the decomposition:

$$y_t = trend_t + seas_t + remainder_t.$$

In practice:

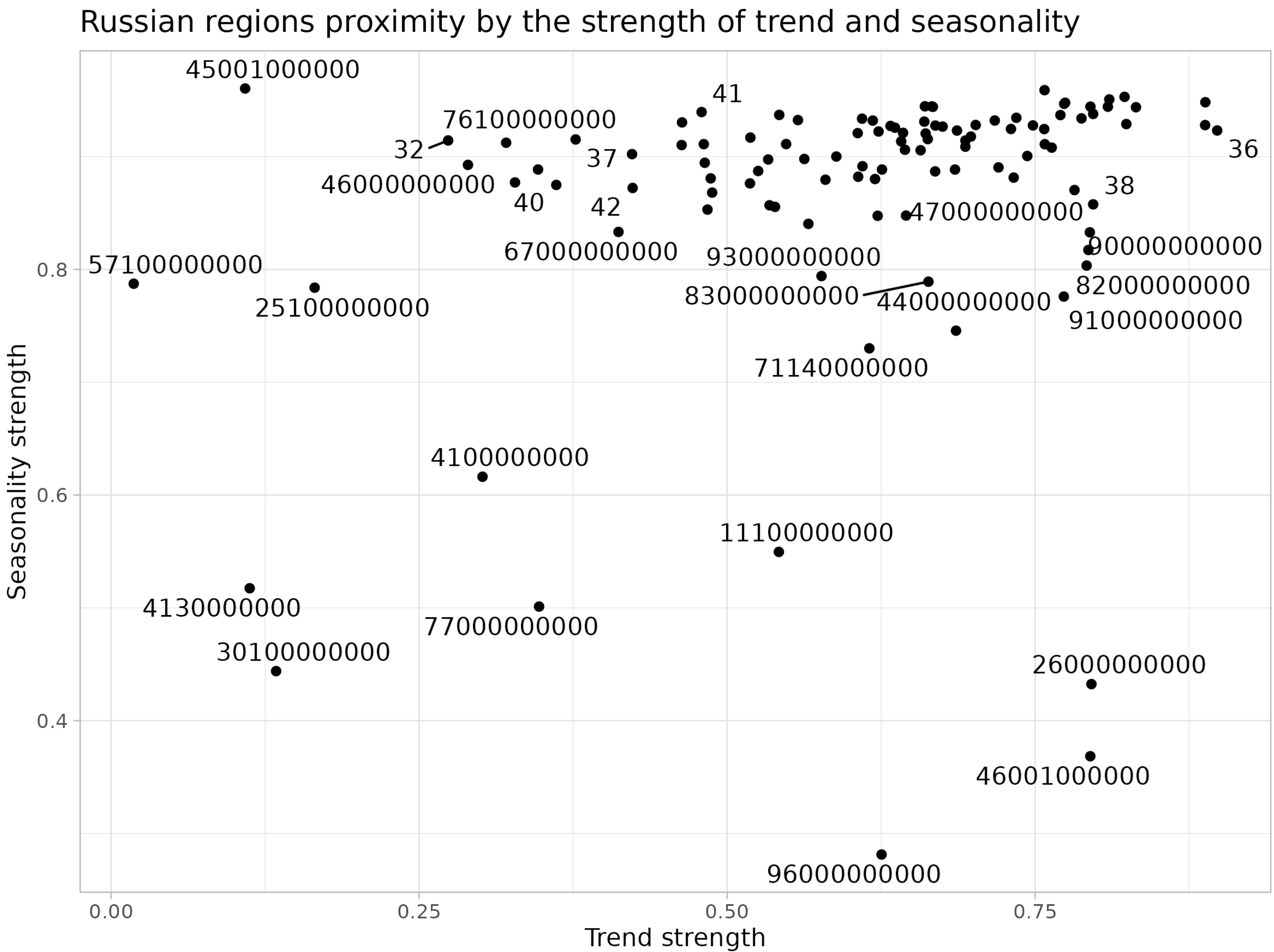
- Trend strength:

$$F_{trend} = \max\left\{1 - \frac{\mathrm{sVar}(remainder)}{\mathrm{sVar}(trend + remainder)}, 0\right\}.$$

- Seasonality strength:

$$F_{seas} = \max\left\{1 - \frac{\mathrm{sVar}(remainder)}{\mathrm{sVar}(seas + remainder)}, 0\right\}.$$

Strength of trend and seasonality



Series Characteristics: Summary

- ACF — coefficients in paired regressions or correlations

Series Characteristics: Summary

- ACF — coefficients in paired regressions or correlations
- PACF — coefficients in multiple regressions or correlations

Series Characteristics: Summary

- ACF — coefficients in paired regressions or correlations
- PACF — coefficients in multiple regressions or correlations
- STL allows you to measure strength of trend and seasonality in comparison to the residual component

ETS Model (Part I)

ETS Model: Plan

- ETS as a model

ETS Model: Plan

- ETS as a model
- Formulas for predictions

ETS Model: Plan

- ETS as a model
- Formulas for predictions
- Adding a trend

ETS Model: Plan

- ETS as a model
- Formulas for predictions
- Adding a trend
- Idea of damped trend

How many ETS models in total?

ETS — Error, Trend, Seasonality (error, trend, seasonality)

How many ETS models in total?

ETS — Error, Trend, Seasonality (error, trend, seasonality)

Error: A, M

Trend: N, A, Ad, M, Md

Seasonality: N, A, M

How many ETS models in total?

ETS — Error, Trend, Seasonality (error, trend, seasonality)

Error: A, M

Trend: N, A, Ad, M, Md

Seasonality: N, A, M

A — additive component

M — multiplicative component

N — no component

d — damping for the trend

How many ETS models in total?

ETS — Error, Trend, Seasonality (error, trend, seasonality)

Error: A, M

Trend: N, A, Ad, M, Md

Seasonality: N, A, M

A — additive component

M — multiplicative component

N — no component

d — damping for the trend

Formally: 30 options

Historical names

ETS(ANN) — simple exponential smoothing

ETS(AAA) — an additive Holt-Winters method

ETS(AAM) — the multiplicative Holt-Winters method

ETS(AAdM) — Holt-Winters method with a fading trend

ETS(ANN) terminology

y_t — the observed series;

ℓ_t — trend, cleaned series;

u_t — a random error

ETS(ANN) terminology

y_t — the observed series;

ℓ_t — trend, cleaned series;

u_t — a random error

$$y_t = \ell_{t-1} + u_t;$$

ETS(ANN) terminology

y_t — the observed series;

ℓ_t — trend, cleaned series;

u_t — a random error

$$y_t = \ell_{t-1} + u_t;$$

$$\ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0;$$

ETS(ANN) terminology

y_t — the observed series;

ℓ_t — trend, cleaned series;

u_t — a random error

$$y_t = \ell_{t-1} + u_t;$$

$$\ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0;$$

$$u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent}$$

ETS(ANN) terminology

y_t — the observed series;

ℓ_t — trend, cleaned series;

u_t — a random error

$$y_t = \ell_{t-1} + u_t;$$

$$\ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0;$$

$$u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent}$$

Parameters: α, σ^2, ℓ_0

Recognized?

ETS(ANN) is a generalization of random walk

Recognized?

ETS(ANN) is a generalization of random walk

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

Recognized?

ETS(ANN) is a generalization of random walk

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

Substitute $\alpha = 1$:

$$y_t = \ell_t = \ell_{t-1} + u_t$$

Estimation

Maximum likelihood method is used for estimation

Estimation

Maximum likelihood method is used for estimation

Main idea: decompose the likelihood into a sum

$$\begin{aligned}\ln L(y \mid \theta) = \ln L(y_1 \mid \theta) + \ln L(y_2 \mid y_1, \theta) + \ldots + \\ + \ln L(y_T \mid y_{T-1}, \ldots, y_1, \theta),\end{aligned}$$

where $\theta = (\alpha, \ell_0, \sigma^2)$

Estimation

Maximum likelihood method is used for estimation

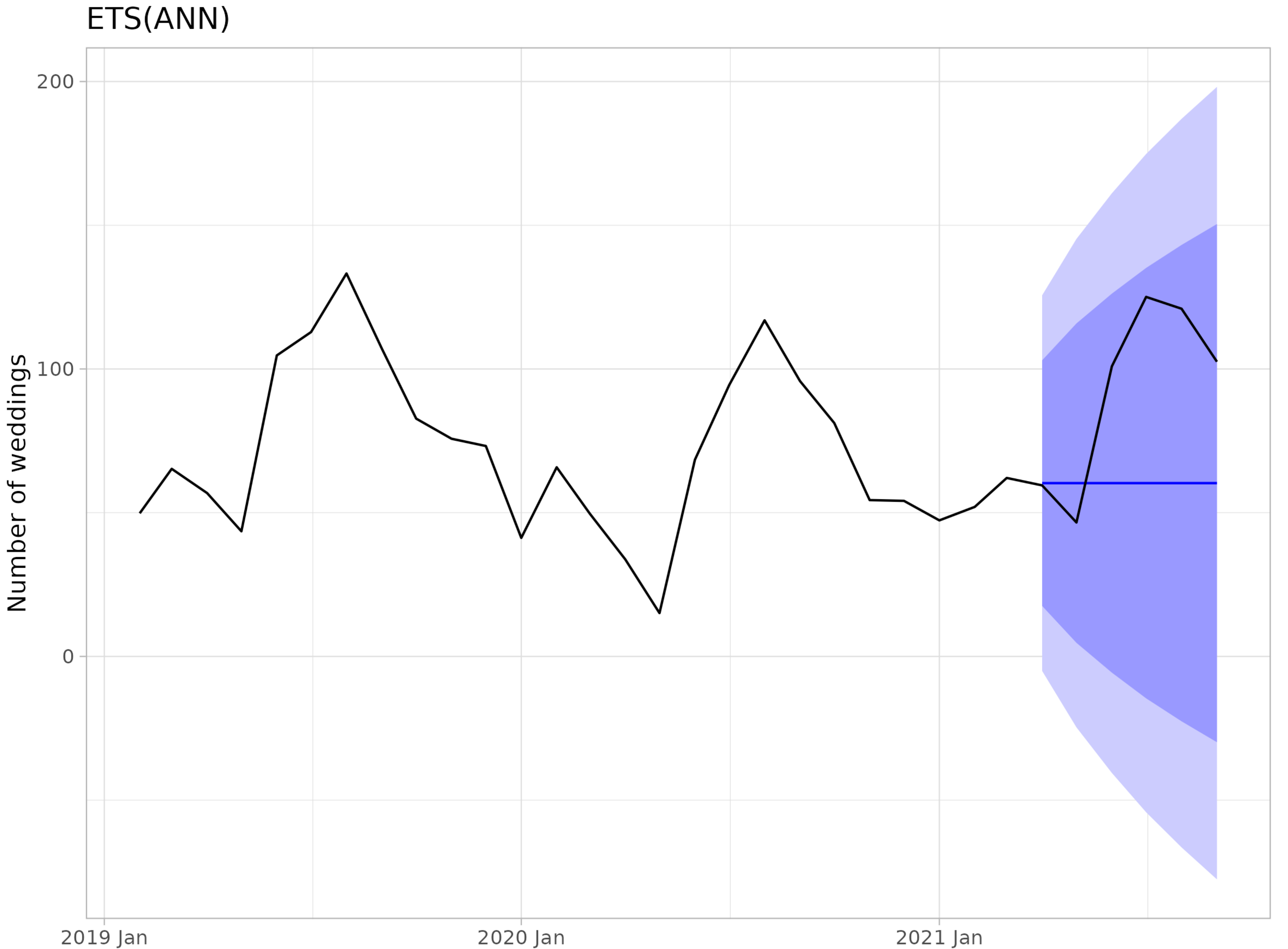
Main idea: decompose the likelihood into a sum

$$\ln L(y \mid \theta) = \ln L(y_1 \mid \theta) + \ln L(y_2 \mid y_1, \theta) + \ldots + \\ + \ln L(y_T \mid y_{T-1}, \ldots, y_1, \theta),$$

where $\theta = (\alpha, \ell_0, \sigma^2)$

Unfortunately, there are no explicit formulas for the estimators

Forecasting



Forecast 1 step ahead

Luckily, there are recurrent formulas for predictions

Forecast 1 step ahead

Luckily, there are recurrent formulas for predictions

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

Forecast 1 step ahead

Luckily, there are recurrent formulas for predictions

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

$$y_{T+1} = \ell_T + u_{T+1}$$

Forecast 1 step ahead

Luckily, there are recurrent formulas for predictions

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

$$y_{T+1} = \ell_T + u_{T+1}$$

$$(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2)$$

Forecast 2 steps ahead

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

Forecast 2 steps ahead

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

$$y_{T+2} = \ell_{T+1} + u_{T+2} = \ell_T + \alpha u_{T+1} + u_{T+2}$$

Forecast 2 steps ahead

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

$$y_{T+2} = \ell_{T+1} + u_{T+2} = \ell_T + \alpha u_{T+1} + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2(\alpha^2 + 1))$$

Predictive intervals

From distribution law

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2(\alpha^2 + 1))$$

we can derive a

Predictive intervals

From distribution law

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T; \sigma^2(\alpha^2 + 1))$$

we can derive a predictive interval

$$[\hat{\ell}_T - 1.96\hat{\sigma}\sqrt{\hat{\alpha}^2 + 1}; \hat{\ell}_T + 1.96\hat{\sigma}\sqrt{\hat{\alpha}^2 + 1}].$$

What was discovered in the 1950s?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

What was discovered in the 1950s?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

Let's rewrite the second equation:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \alpha y_t + (1-\alpha)\ell_{t-1}$$

What was discovered in the 1950s?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

Let's rewrite the second equation:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \alpha y_t + (1-\alpha)\ell_{t-1}$$

Simple exponential smoothing:

$$\hat{\ell}_1 = y_1$$

What was discovered in the 1950s?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

Let's rewrite the second equation:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \alpha y_t + (1-\alpha)\ell_{t-1}$$

Simple exponential smoothing:

$$\hat{\ell}_1 = y_1$$

$$\hat{\ell}_t = \alpha y_t + (1-\alpha)\hat{\ell}_{t-1}$$

What was discovered in the 1950s?

$$\begin{cases} y_t = \ell_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

Let's rewrite the second equation:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) = \alpha y_t + (1-\alpha)\ell_{t-1}$$

Simple exponential smoothing:

$$\hat{\ell}_1 = y_1$$

$$\hat{\ell}_t = \alpha y_t + (1-\alpha)\hat{\ell}_{t-1}$$

$$\min_{\alpha} \sum (y_t - \hat{\ell}_t)^2$$

Adding trend!

y_t — the observed series;

ℓ_t — trend, cleaned series;

b_t — current growth rate of the cleaned series;

u_t — a random error

Adding trend!

y_t — the observed series;

ℓ_t — trend, cleaned series;

b_t — current growth rate of the cleaned series;

u_t — a random error

ETS(AAN):

A — additive error;

A — additive trend;

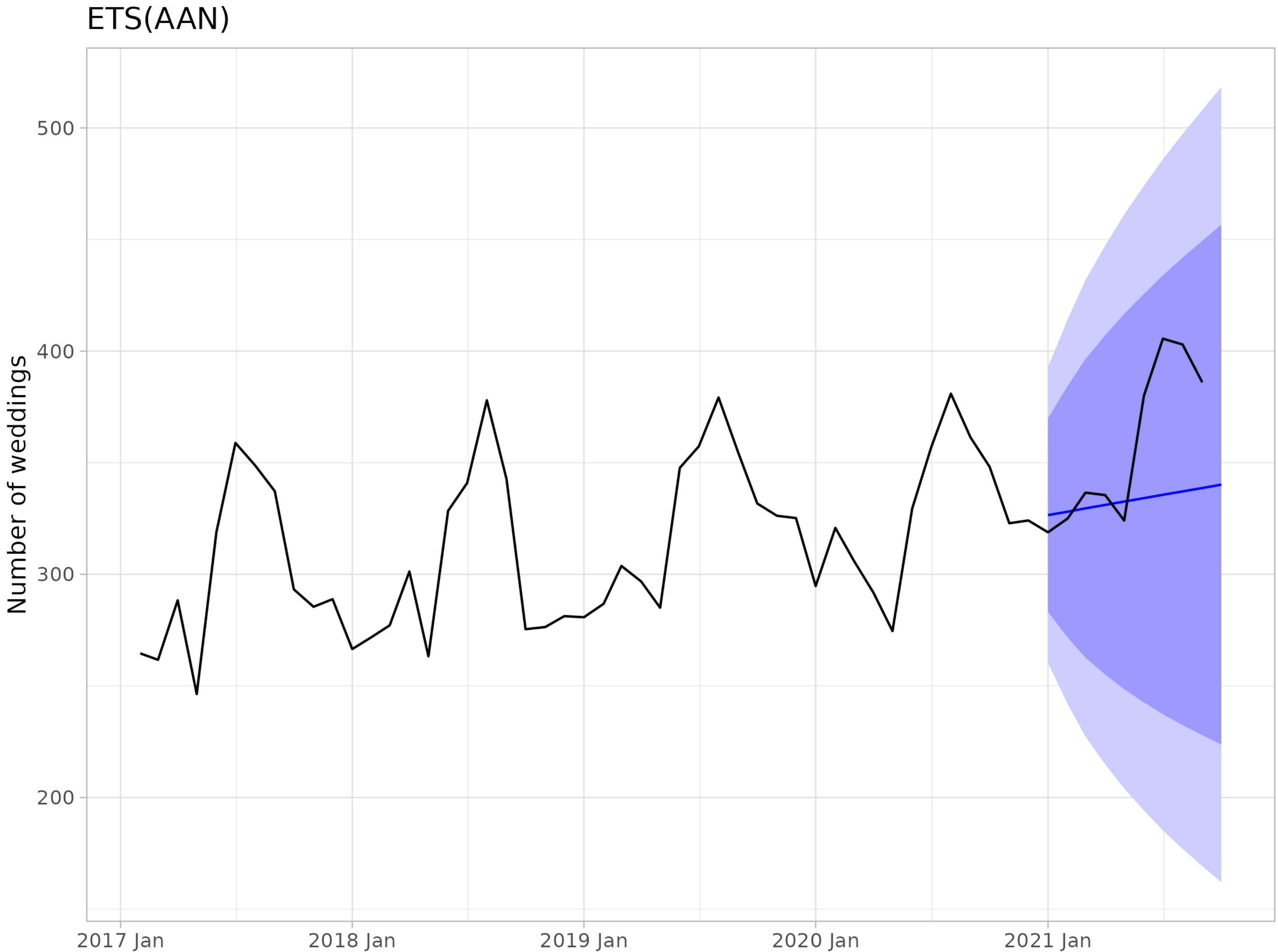
N — no seasonality

ETS(AAN): equations

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \end{cases}$$

Parameters: α, β, σ^2, ℓ_0, b_0

ETS(AAN): Forecasting



Forecast 1 step ahead

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent.} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0 \end{cases}$$

$$y_{T+1} = \ell_T + b_T + u_{T+1}$$

$$(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + b_T; \sigma^2)$$

Forecast 2 steps ahead

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0 \end{cases}$$

$$y_{T+2} = \ell_{T+1} + b_{T+1} + u_{T+2} = (\ell_T + b_T + \alpha u_{T+1}) + \\ + (b_T + \beta u_{T+1}) + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + 2b_T; \sigma^2((\alpha + \beta)^2 + 1))$$

Problem with trend in ETS(AAN)

In the ETS(AAN) model growth rate of the ℓ_t trend is defined by the formula

$$b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0$$

Problem with trend in ETS(AAN)

In the ETS(AAN) model growth rate of the ℓ_t trend is defined by the formula

$$b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0$$

Consequently,

$$\mathbb{E}(b_t) = \mathbb{E}(b_{t-1}), \quad \mathbb{E}(b_{T+h} \mid b_T) = b_T$$

Problem with trend in ETS(AAN)

In the ETS(AAN) model growth rate of the ℓ_t trend is defined by the formula

$$b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0$$

Consequently,

$$\mathbb{E}(b_t) = \mathbb{E}(b_{t-1}), \quad \mathbb{E}(b_{T+h} \mid b_T) = b_T$$

Long-term forecast of a positive indicator at $b_T < 0$ will become negative

Contradiction

In short-term we expect a change in the indicator:

we want a trend in the model.

Contradiction

In short-term we expect a change in the indicator:

we want a trend in the model.

In long-term negative values are impossible:

we don't want a trend in the model.

Contradiction

In short-term we expect a change in the indicator:

we want a trend in the model.

In long-term negative values are impossible:

we don't want a trend in the model.

Solution: damped or fading trend.

Extra parameters are expensive!

We want richer trend dynamics — we need additional parameters.

Extra parameters are expensive!

We want richer trend dynamics — we need additional parameters.

Additional parameters — risk overfitting of the model, wider confidence intervals for the remaining parameters.

Extra parameters are expensive!

We want richer trend dynamics — we need additional parameters.

Additional parameters — risk overfitting of the model, wider confidence intervals for the remaining parameters.

Let's solve the problem with only one new parameter!

Damped trend

We introduce the trend damping parameter $\phi \in (0; 1)$ into the slope equation:

$$b_t = \phi b_{t-1} + \beta u_t, \text{ starts at } b_0$$

Damped trend

We introduce the trend damping parameter $\phi \in (0; 1)$ into the slope equation:

$$b_t = \phi b_{t-1} + \beta u_t, \text{ starts at } b_0$$

And for the rest of the equations:

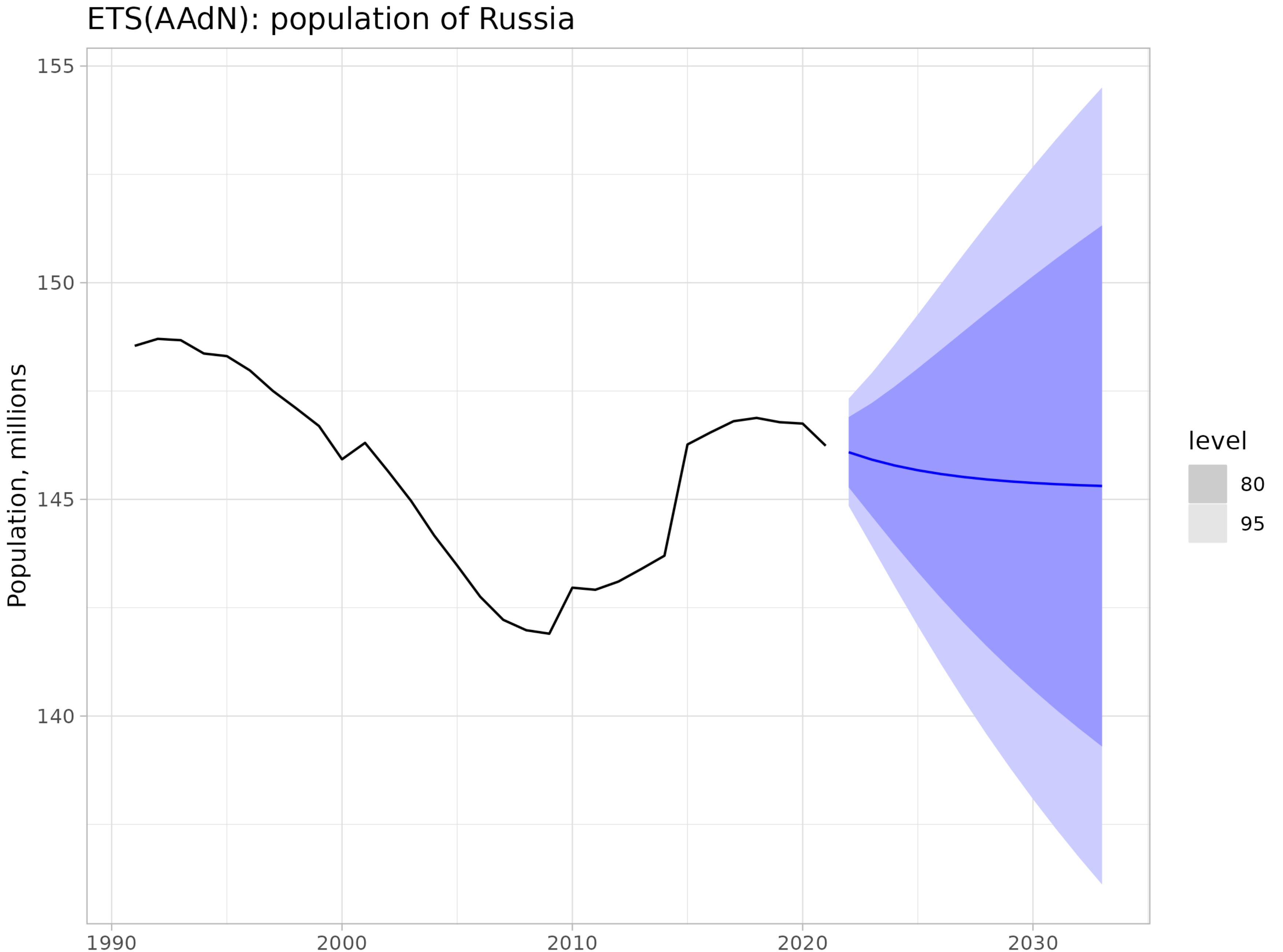
$$\begin{cases} y_t = \ell_{t-1} + \phi b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha u_t, \text{ starts at } \ell_0 \end{cases}$$

General form of ETS(AAdN)

$$
\begin{cases}
y_t = \ell_{t-1} + \phi b_{t-1} + u_t; \\
\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\
b_t = \phi b_{t-1} + \beta u_t, \text{ starts at } b_0; \\
u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent}
\end{cases}
$$

Parameters: α, σ^2, ℓ_0, b_0, β, ϕ

ETS(AAdN): Forecasting



ETS Model: Summary

- Formulas for exponential smoothing have been around for a long time

ETS Model: Summary

- Formulas for exponential smoothing have been around for a long time
- ETS — a wide class of modern models

ETS Model: Summary

- Formulas for exponential smoothing have been around for a long time
- ETS — a wide class of modern models
- The slope of the trend line can change

ETS Model: Summary

- Formulas for exponential smoothing have been around for a long time
- ETS — a wide class of modern models
- The slope of the trend line can change
- Damped trend: on a small forecasting horizon there is a trend, on a large horizon — none

ETS Model (Part II)

ETS Model: Plan

- Adding seasonality

ETS Model: Plan

- Adding seasonality
- Formulas for predictions

ETS Model: Plan

- Adding seasonality
- Formulas for predictions
- Decomposition into components

ETS Model: Plan

- Adding seasonality
- Formulas for predictions
- Decomposition into components
- Multiplicative components

Adding seasonality!

y_t — the observed series;

ℓ_t — trend, cleaned series;

b_t — current growth rate of the cleaned series;

s_t — seasonal component;

u_t — a random error

Adding seasonality!

y_t — the observed series;

ℓ_t — trend, cleaned series;

b_t — current growth rate of the cleaned series;

s_t — seasonal component;

u_t — a random error

ETS(AAA):

A — additive error;

A — additive trend;

A — additive seasonality

ETS(AAA): equations

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ s_t = s_{t-12} + \gamma u_t; \text{ starts at} s_0, s_{-1}, \ldots, s_{-11} \end{cases}$$

Parameters: α, β, γ, σ^2, ℓ_0, b_0, s_0, s_{-1}, …, s_{-11}

Restriction: $s_0 + s_{-1} + \ldots + s_{-11} = 0$

ETS(AAA): equations

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ s_t = s_{t-12} + \gamma u_t; \text{ starts at} s_0, s_{-1}, \ldots, s_{-11} \end{cases}$$

Parameters: α, β, γ, σ^2, ℓ_0, b_0, s_0, s_{-1}, ..., s_{-11}

Restriction: $s_0 + s_{-1} + \ldots + s_{-11} = 0$

How many independent parameters are we estimating?

ETS(AAA): equations

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ s_t = s_{t-12} + \gamma u_t; \text{ starts at} s_0, s_{-1}, \ldots, s_{-11} \end{cases}$$

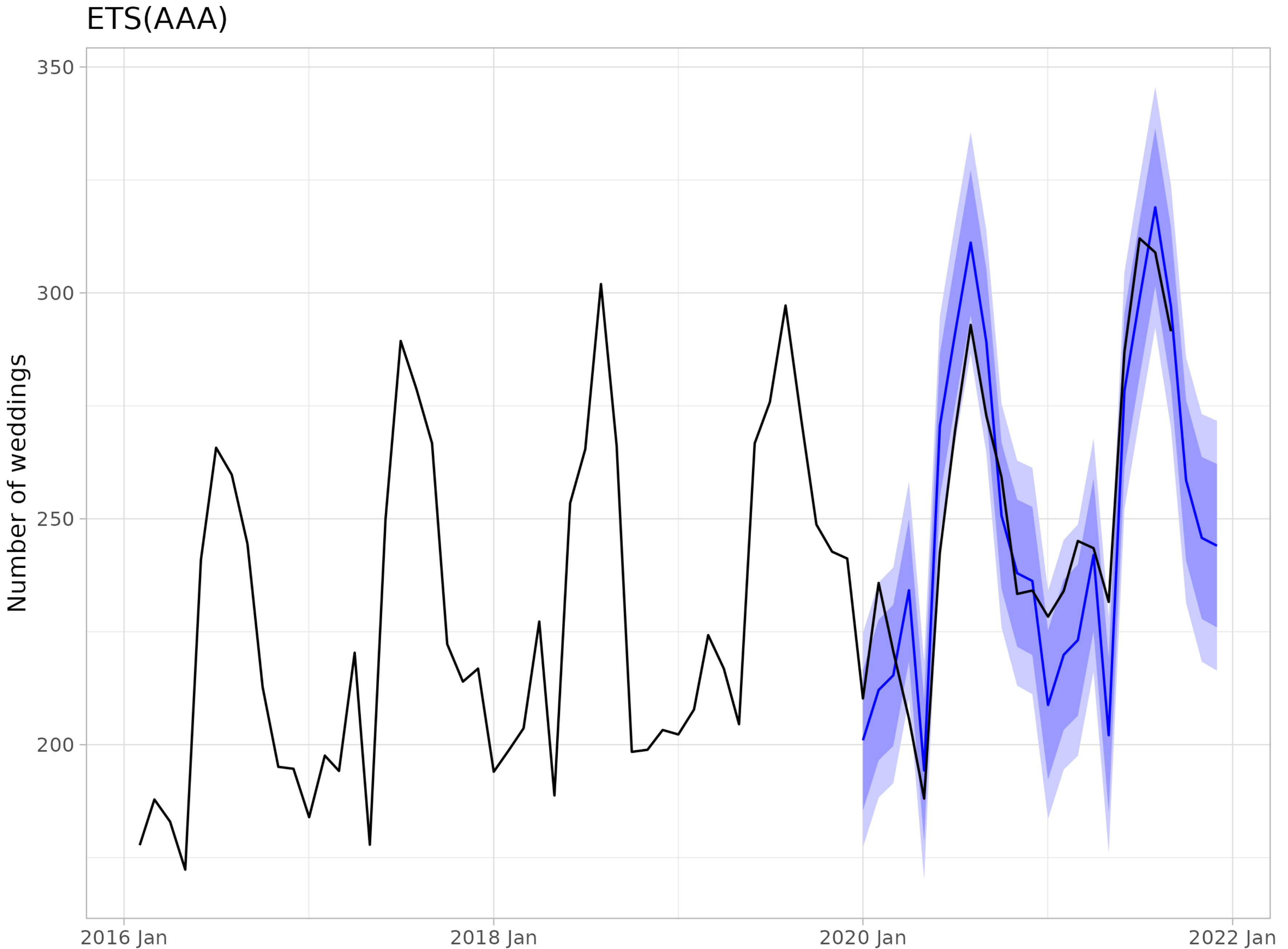
Parameters: α, β, γ, σ^2, ℓ_0, b_0, s_0, s_{-1}, …, s_{-11}

Restriction: $s_0 + s_{-1} + \ldots + s_{-11} = 0$

How many independent parameters are we estimating?

Correct answer: 17

ETS(AAA): Forecasting



Forecast 1 step ahead

$$
\begin{cases}
y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\
\ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\
u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent.} \\
b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\
s_t = s_{t-12} + \gamma u_t
\end{cases}
$$

$$
y_{T+1} = \ell_T + b_T + s_{T-11} + u_{T+1}
$$

$$
(y_{T+1} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + b_T + s_{T-11}; \sigma^2)
$$

Forecast 2 steps ahead

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ s_t = s_{t-12} + \gamma u_t \end{cases}$$

$$y_{T+2} = \ell_{T+1} + b_{T+1} + s_{T-10} + u_{T+2} = (\ell_T + b_T + \alpha u_{T+1}) + \\ + (b_T + \beta u_{T+1}) + s_{T-10} + u_{T+2}$$

$$(y_{T+2} \mid \mathcal{F}_T) \sim \mathcal{N}(\ell_T + 2b_T + s_{T-10}; \sigma^2((\alpha + \beta)^2 + 1))$$

Decomposition for free!

Consider the output of ETS(AAA):

Parameter estimates: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\ell}_0$, $\hat{b}_0$, $\hat{s}_0$, $\hat{s}_{-1}$, ..., $\hat{s}_{-11}$.

Constraints: $\hat{s}_0 + \hat{s}_{-1} + \ldots + \hat{s}_{-11} = 0$.

Decomposition for free!

Consider the output of ETS(AAA):

Parameter estimates: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\ell}_0$, $\hat{b}_0$, $\hat{s}_0$, $\hat{s}_{-1}$, ..., $\hat{s}_{-11}$.

Constraints: $\hat{s}_0 + \hat{s}_{-1} + \ldots + \hat{s}_{-11} = 0$.

Estimated component values: $\hat{\ell}_t$, $\hat{b}_t$, $\hat{s}_t$.

Decomposition for free!

Consider the output of ETS(AAA):

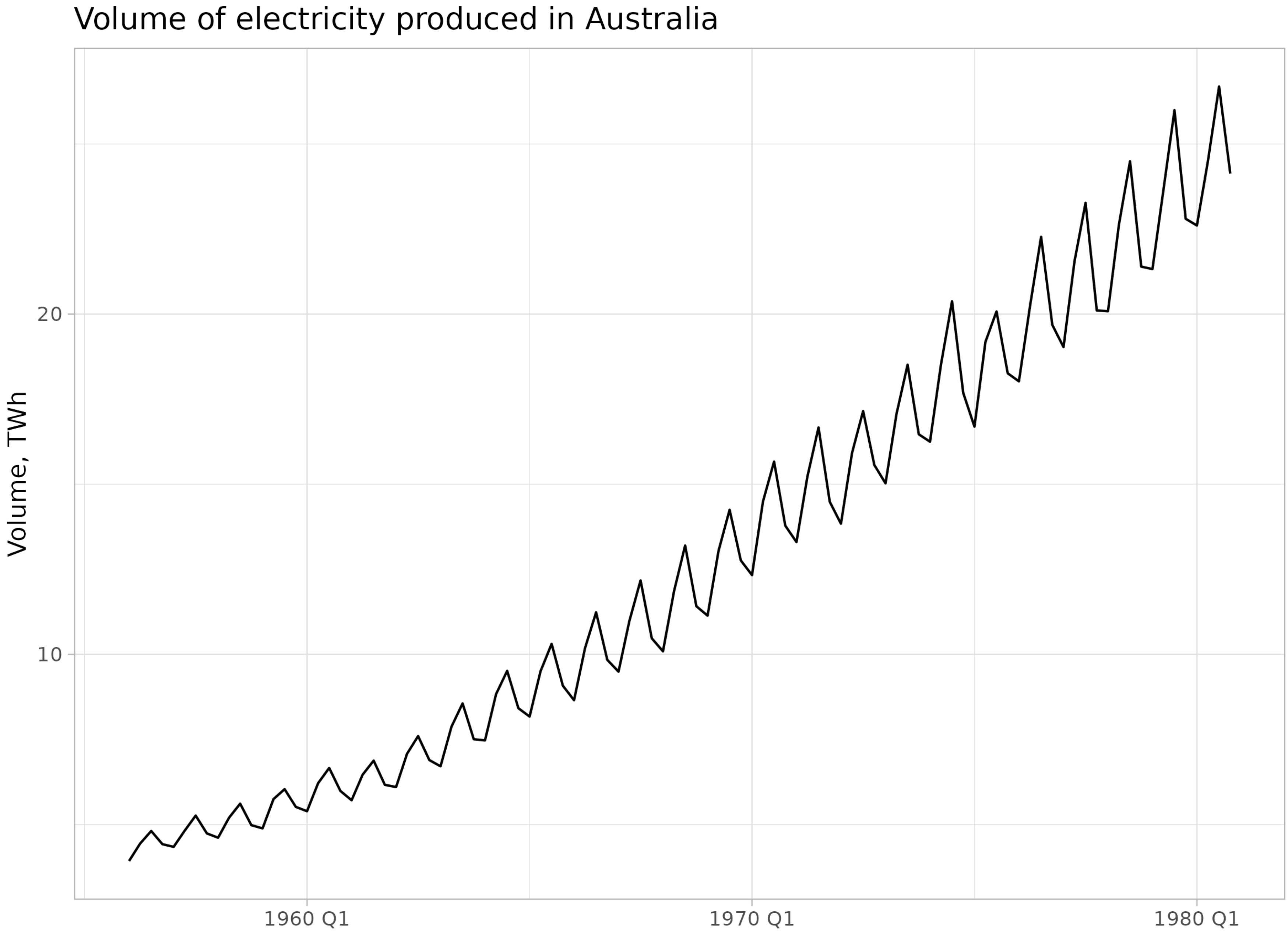
Parameter estimates: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, $\hat{\sigma}^2$, $\hat{\ell}_0$, $\hat{b}_0$, $\hat{s}_0$, $\hat{s}_{-1}$, ..., $\hat{s}_{-11}$.

Constraints: $\hat{s}_0 + \hat{s}_{-1} + \ldots + \hat{s}_{-11} = 0$.

Estimated component values: $\hat{\ell}_t$, $\hat{b}_t$, $\hat{s}_t$.

We automatically get decomposition: $y_t = \hat{\ell}_t + \hat{s}_t + remainder_t$.

Oscillation amplitude can vary



Various oscillation amplitude

Possible solutions:

- Switch to logarithms, $y_t \to \ln y_t$

Various oscillation amplitude

Possible solutions:

- Switch to logarithms, $y_t \to \ln y_t$
- Box-Cox transformation, $y_t \to bc(y_t, \lambda)$

Various oscillation amplitude

Possible solutions:

- Switch to logarithms, $y_t \to \ln y_t$
- Box-Cox transformation, $y_t \to bc(y_t, \lambda)$
- Multiplicative components

ETS(MNM): equations

ETS(MNM) for monthly data:

$$
\begin{cases}
y_t = \ell_{t-1} \cdot s_{t-12} \cdot (1 + u_t); \\
\ell_t = \ell_{t-1} \cdot (1 + \alpha u_t), \text{ starts at } \ell_0; \\
s_t = s_{t-12} \cdot (1 + \gamma u_t), \text{ starts at } s_0, \ldots, s_{-11}; \\
u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent}
\end{cases}
$$

ETS(MNM): equations

ETS(MNM) for monthly data:

$$\begin{cases} y_t = \ell_{t-1} \cdot s_{t-12} \cdot (1 + u_t); \\ \ell_t = \ell_{t-1} \cdot (1 + \alpha u_t), \text{ starts at } \ell_0; \\ s_t = s_{t-12} \cdot (1 + \gamma u_t), \text{ starts at } s_0, \ldots, s_{-11}; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent} \end{cases}$$

ETS(ANA):

$$\begin{cases} y_t = \ell_{t-1} + s_{t-12} + u_t; \\ \ell_t = \ell_{t-1} + \alpha u_t, \text{ starts at } \ell_0; \\ s_t = s_{t-12} + \gamma u_t, \text{ starts at } s_0, \ldots, s_{-11}; \end{cases}$$

Units

Series y_t, ℓ_t — initial units.

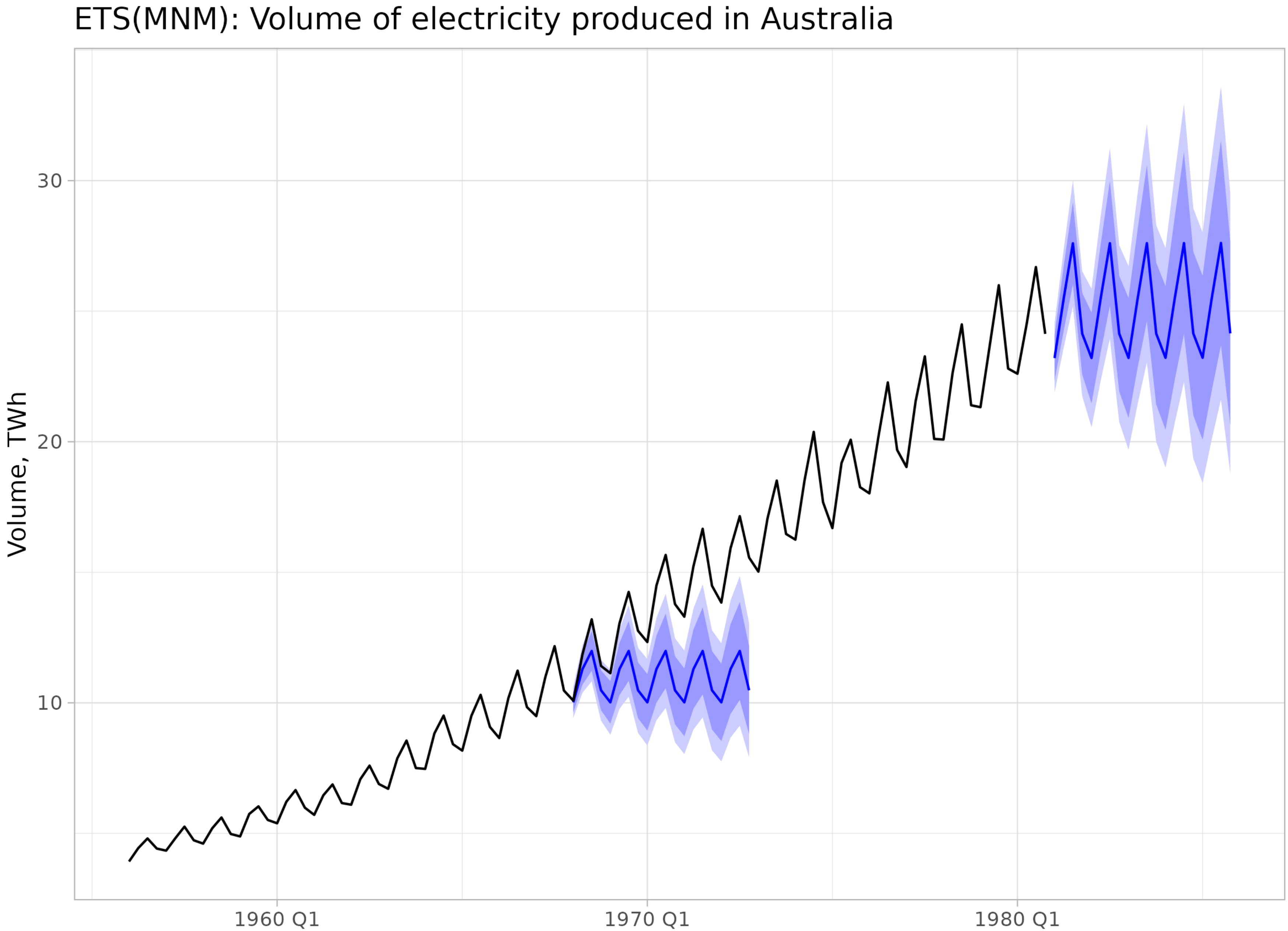
Units

Series y_t, ℓ_t — initial units.

The s_t series is measured relative to one, for example, $s_t = 0.9$ — 10% below the trend.

The u_t series is measured relative to zero, for example, $u_t = -0.1$ — a 10% drop.

ETS(MNM): Forecasting



Which option to choose?

Different amplitude of fluctuations: indication of multiplicative models.

Which option to choose?

Different amplitude of fluctuations: indication of multiplicative models.

Automatic selection based on the AIC criterion works.

Which option to choose?

Different amplitude of fluctuations: indication of multiplicative models.

Automatic selection based on the AIC criterion works.

You can get the ETS(AAdA) model with seasonality.

Which option to choose?

Different amplitude of fluctuations: indication of multiplicative models.

Automatic selection based on the AIC criterion works.

You can get the ETS(AAdA) model with seasonality.

Some of the multiplicative models can be numerically unstable or not implemented in the software.

ETS Model: Summary

- The slope of the trend and seasonality may change

ETS Model: Summary

- The slope of the trend and seasonality may change
- Automatic decomposition into components

ETS Model: Summary

- The slope of the trend and seasonality may change
- Automatic decomposition into components
- Multiplicative models take into account changing oscillation amplitudes

ETS Model: Summary

- The slope of the trend and seasonality may change
- Automatic decomposition into components
- Multiplicative models take into account changing oscillation amplitudes
- A lot of possible combinations

Theta method

Theta method: Plan

- An unexpected leader

Theta method: Plan

- An unexpected leader
- Author's version

Theta method: Plan

- An unexpected leader
- Author's version
- Special case of ETS

Theta method

Appeared in 2000 and became a sensation at the competition M3 for predicting series.

Theta method

Appeared in 2000 and became a sensation at the competition M3 for predicting series.

Works for non-seasonal series.

Theta method

Appeared in 2000 and became a sensation at the competition M3 for predicting series.

Works for non-seasonal series.

Initially suggested without a statistical model.

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)
2. Predict zero-line using linear regression

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)
2. Predict zero-line using linear regression
3. Predict the second line using ETS(ANN)

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)
2. Predict zero-line using linear regression
3. Predict the second line using ETS(ANN)
4. Average the forecasts

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)
2. Predict zero-line using linear regression
3. Predict the second line using ETS(ANN)
4. Average the forecasts

Author's version

1. Decompose the series into two theta lines ($\theta = 0$, $\theta = 2$)
2. Predict zero-line using linear regression
3. Predict the second line using ETS(ANN)
4. Average the forecasts

You can pre-delete seasonality and add it back in the end

What is a theta line?

Zero theta line — regression on time:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 t$$

What is a theta line?

Zero theta line — regression on time:

$$\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 t$$

Theta line for arbitrary theta:

$$\Delta^2 y_t^{new} = \theta \Delta^2 y_t$$

Intuition

- The zero theta line catches the long-term trend of the series

Intuition

- The zero theta line catches the long-term trend of the series
- Theta line ($\theta = 2$) catches the short-term trend: acceleration of theta line is θ times stronger than of the initial series

Intuition

- The zero theta line catches the long-term trend of the series
- Theta line ($\theta = 2$) catches the short-term trend: acceleration of theta line is θ times stronger than of the initial series
- Averaging reduces the variance of predictions

How is the theta line selected?

We take $\theta = 2$:

$$\Delta^2 y_t^{new} = 2\Delta^2 y_t$$

How is the theta line selected?

We take $\theta = 2$:

$$\Delta^2 y_t^{new} = 2\Delta^2 y_t$$

Or

$$y_t^{new} - 2y_{t-1}^{new} + y_{t-2}^{new} = 2(y_t - 2y_{t-1} + y_{t-2})$$

How is the theta line selected?

We take $\theta = 2$:

$$\Delta^2 y_t^{new} = 2\Delta^2 y_t$$

Or

$$y_t^{new} - 2y_{t-1}^{new} + y_{t-2}^{new} = 2(y_t - 2y_{t-1} + y_{t-2})$$

The new series y_t^{new} is completely determined by y_1^{new}, y_2^{new}

How is the theta line selected?

We take $\theta = 2$:

$$\Delta^2 y_t^{new} = 2\Delta^2 y_t$$

Or

$$y_t^{new} - 2y_{t-1}^{new} + y_{t-2}^{new} = 2(y_t - 2y_{t-1} + y_{t-2})$$

The new series y_t^{new} is completely determined by y_1^{new}, y_2^{new}

We solve the optimization problem:

$$\sum_{t=1}^{T} (y_t - y_t^{new})^2 \to \min$$

Statistical Model

Formal model appeared in 2003:

$$\begin{cases} y_t = \ell_t + b + u_t; \\ \ell_t = \ell_{t-1} + b + \alpha u_t; \\ \ell_1 = y_1 \end{cases}$$

Statistical Model

Formal model appeared in 2003:

$$\begin{cases} y_t = \ell_t + b + u_t; \\ \ell_t = \ell_{t-1} + b + \alpha u_t; \\ \ell_1 = y_1 \end{cases}$$

Or:

$$\Delta y_t = b + (\alpha - 1)u_{t-1} + u_t$$

Theta method — ETS variant

A special case of a more general model — ETS(AAN):

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_1; \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent.} \end{cases}$$

Theta method — ETS variant

A special case of a more general model — ETS(AAN):

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_1; \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent.} \end{cases}$$

Remove the trend stochasticity setting $\beta = 0$.

Theta method — ETS variant

A special case of a more general model — ETS(AAN):

$$\begin{cases} y_t = \ell_{t-1} + b_{t-1} + u_t; \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha u_t, \text{ starts at } \ell_1; \\ b_t = b_{t-1} + \beta u_t, \text{ starts at } b_0; \\ u_t \sim \mathcal{N}(0; \sigma^2) \text{ and are independent.} \end{cases}$$

Remove the trend stochasticity setting $\beta = 0$.

Nuances of initialization are possible.

Theta method: Summary

- Works well for non-seasonal data

Theta method: Summary

- Works well for non-seasonal data
- A special variation of the ETS model