Topic 5: Time Series, Moving Averages, and Smoothing Methods

ECON 4753 — University of Arkansas

Prof. Kyle Butts

Fall 2024

Roadmap

- Introduction to Time-Series
- Time-series Statistics
- Learning from Time-Series
- Smoothing Methods for Inference
  - Trends and Seasonality
- Smoothing Methods for Forecasting
  - Exponential Smoothing
  - Trends and Seasonality with SES

Time-series

Time-series data is a set of observations y_t that occur for a single unit measured over the course of time

- You observe a set of observations x_t for $t \in \{1, \ldots, T\}$
- In general, we call t the 'period'

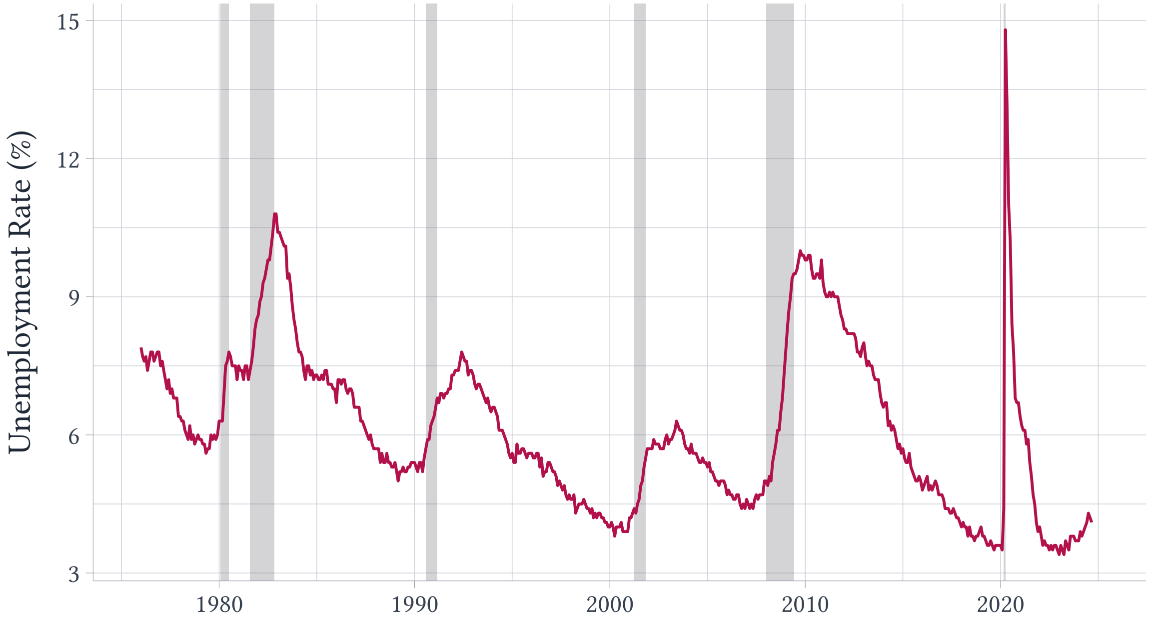
Time-series

Time-series data is a set of observations y_t that occur for a single unit measured over the course of time

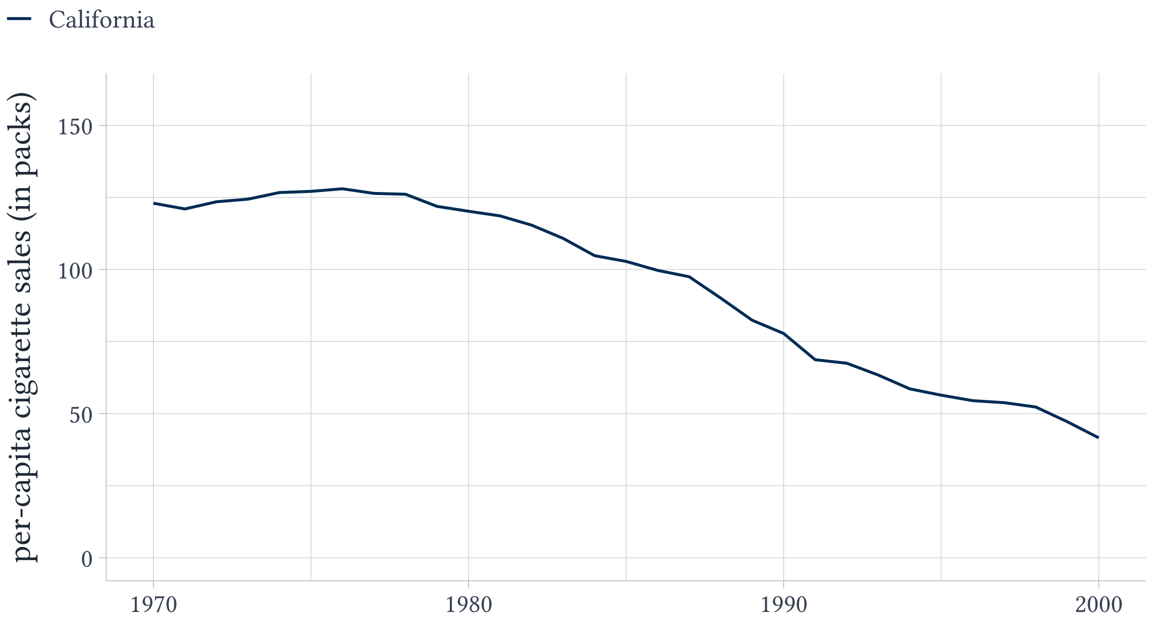
- You observe a set of observations x_t for $t \in \{1, \ldots, T\}$
- In general, we call t the 'period'

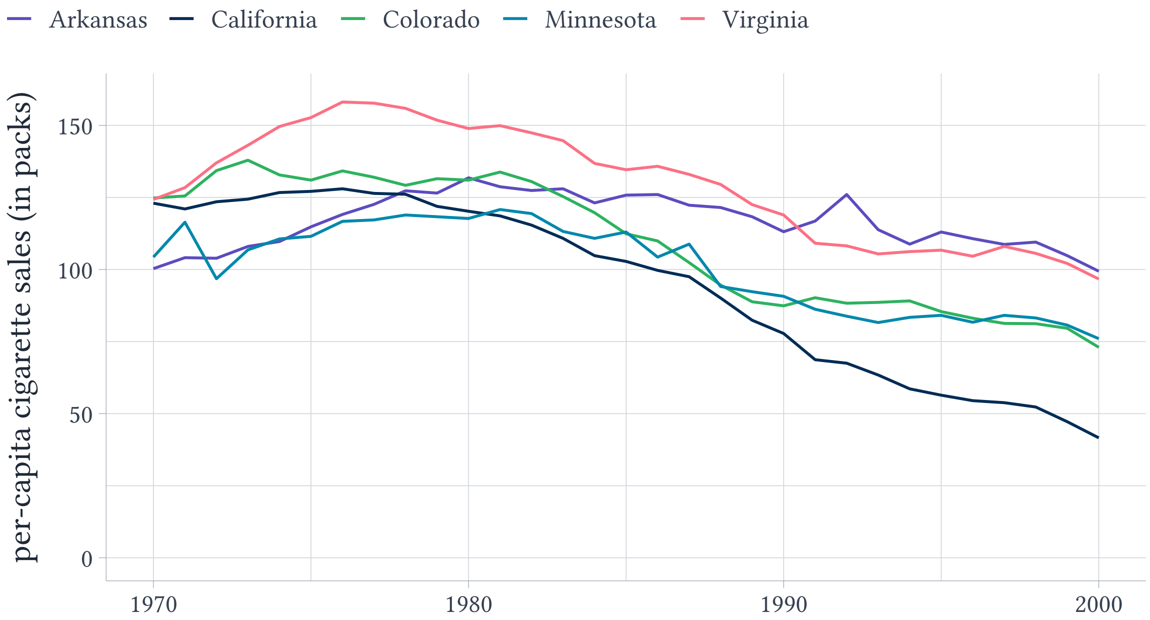
Examples include:

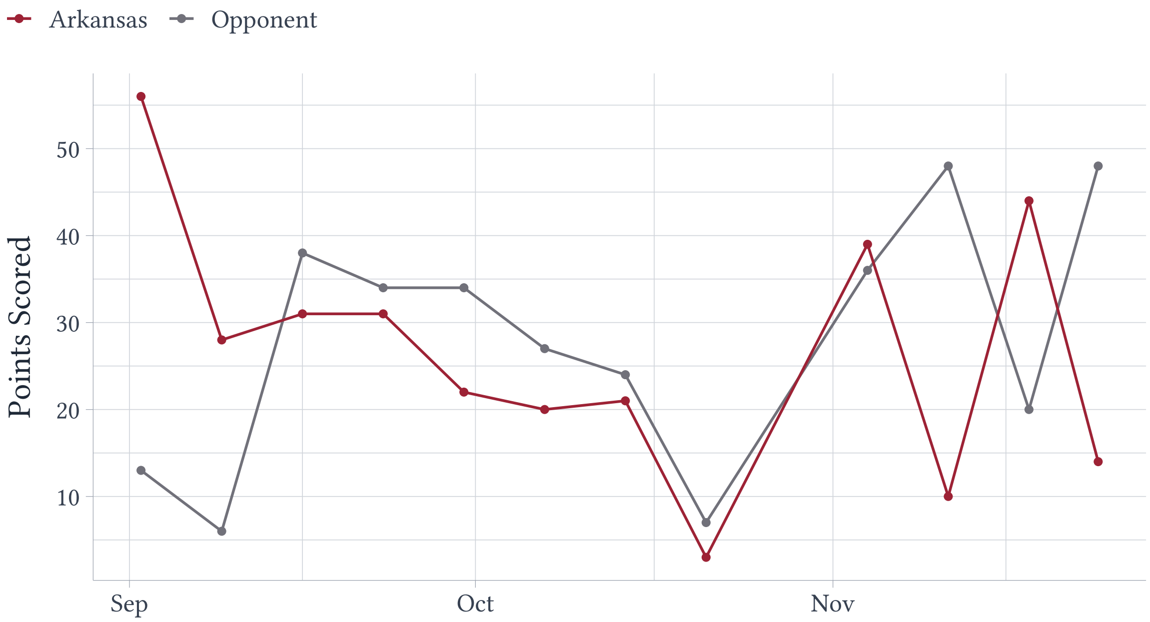
- Annual data on the DGP of a country
- Hourly stock price for a company
- Annual data on cigarette consumption per capita in a state
- A sport's teams number of points scored in games (unequally spaced)


Unemployment Rate (%)
15
12
9
6
3
1980
1990
2000
2010
2020

Apple Stock Price ($)
0
50
100
150
200
2010
2015
2020
2025


California
per-capita cigarette sales (in packs)
150
100
50
0
1970
1980
1990
2000


Arkansas
California
Colorado
Minnesota
Virginia
per-capita cigarette sales (in packs)
0
50
100
150
1970
1980
1990
2000


Arkansas
Opponent
Points Scored
50
40
30
20
10
Sep
Oct
Nov

What is special about time-series?

In our previous topics, we have been thinking about cross-sectional data

- That is, we view a set of individuals viewed at a point in time

In cross-sectional data, knowing about one indivdiual does not really tell me much information about another

- This is not *entirely true*; e.g. worker's in same firm have common experiences, kids in same school have same teacher quality, etc.

What is special about time-series?

In time-series data, knowing last period's value of y_{t-1} is often very useful for this period's value of y_t

- This property is essential in forecasting; following a variable over time might let us predict futrue values

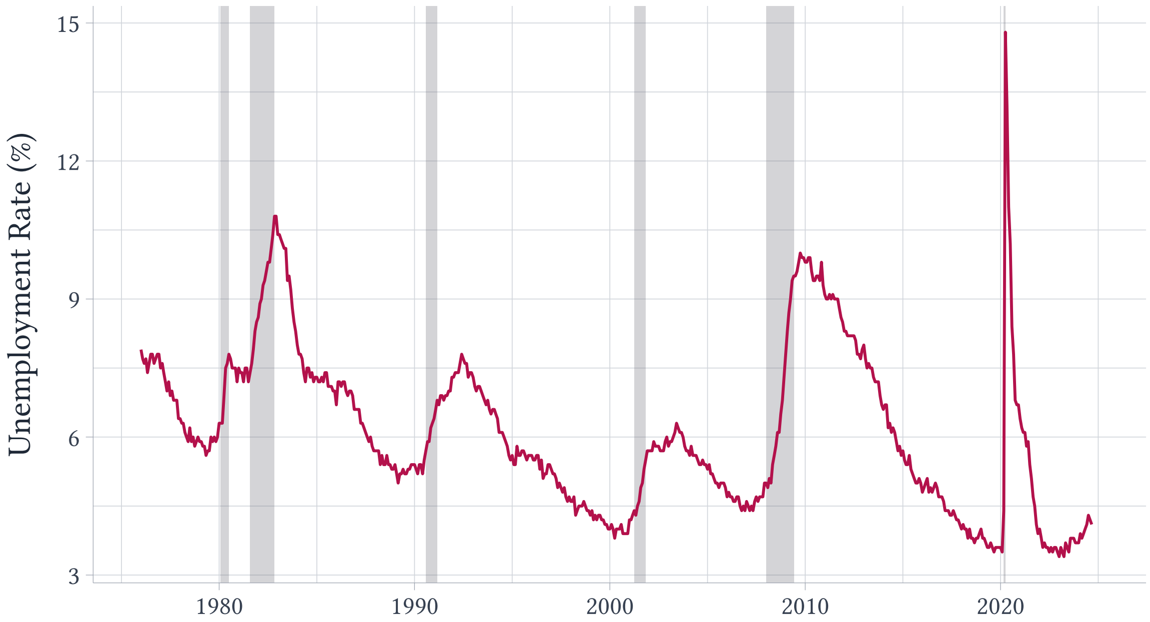
What is special about time-series?

In time-series data, knowing last period's value of y_{t-1} is often very useful for this period's value of y_t

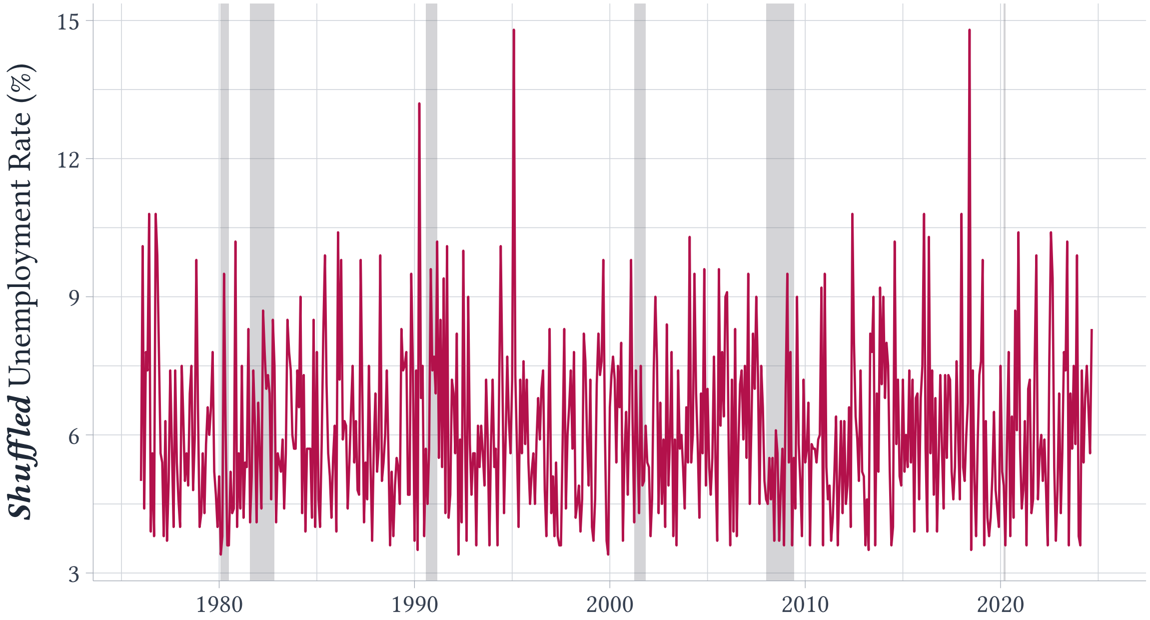
- This property is essential in forecasting; following a variable over time might let us predict futrue values

Another way of saying this, is if we randomly shuffled time-series data, we would lose information!

- This is not true of a cross-sectional dataset; we can reshuffle rows without problem



Unemployment Rate (%)
15
12
9
6
3
1980
1990
2000
2010
2020


Shuffled Unemployment Rate (%)
15
12
9
6
3
1980
1990
2000
2010
2020

What we can gain from using time-series

Time-series forecasting can be useful to:

- Predict future values based on past data
- Inform decision-making by anticipating changes over time
- Identify patterns like trends or seasonality

Roadmap

Introduction to Time-Series

Time-series Statistics

Learning from Time-Series

Smoothing Methods for Inference

- Trends and Seasonality

Smoothing Methods for Forecasting

- Exponential Smoothing
- Trends and Seasonality with SES

Statistics of Time-series

For the next few slides, we will discuss some statistics of time-series data that we might be interested in

To review, in cross-sectional data, we mainly cared about:

- the mean and the variance of a single variable, and
- the correlation between two variables

Autocovariance

Autocovariance measures the covariance between a variable and a lagged version of itself over successive time periods.

In formal terms, the autocovariance at lag k is defined as:

$$\gamma_k = \operatorname{cov}(y_t, y_{t-k}) = \mathbb{E}((y_t - \mu)(y_{t-k} - \mu))$$

where:

- μ is the mean of y_t,
- $\operatorname{cov}(y_t, y_{t-k})$ is the covariance between y_t and y_{t-k}.

Autocovariance

$$\gamma_k = \mathrm{cov}\left(y_t, y_{t-k}\right) = \mathbb{E}((y_t - \mu)(y_{t-k} - \mu))$$

Intuition: Autocovariance helps quantify how much the past values of y move together with its current value.

- When y_{t-k} was above the mean, was y_t typically above it's mean?

Autocovariance

$$\gamma_k = \operatorname{cov}(y_t, y_{t-k}) = \mathbb{E}((y_t - \mu)(y_{t-k} - \mu))$$

Intuition: Autocovariance helps quantify how much the past values of y move together with its current value.

- When y_{t-k} was above the mean, was y_t typically above it's mean?

In most settings, it is likely that $\gamma_1 \geq \gamma_2 \geq \dots$

- More-recent 'shocks' (in say $t-1$) tend to persist for a little and then fade-out

Autocovariance

$$\gamma_k = \operatorname{cov}(y_t, y_{t-k}) = \mathbb{E}((y_t - \mu)(y_{t-k} - \mu))$$

As an aside, note that when $k = 0$,

$$\gamma_0 = \operatorname{cov}(y_t, y_t) = \operatorname{var}(y_t)$$

Autocorrelation

Autocorrelation is the normalized version of autocovariance. It measures the correlation of a variable with its lagged values.

The autocorrelation at lag k is defined as:

$$\rho_k = \frac{\gamma_k}{\operatorname{var}(y_t)} = \frac{\operatorname{cov}(y_t, y_{t-k})}{\operatorname{var}(y_t)}$$

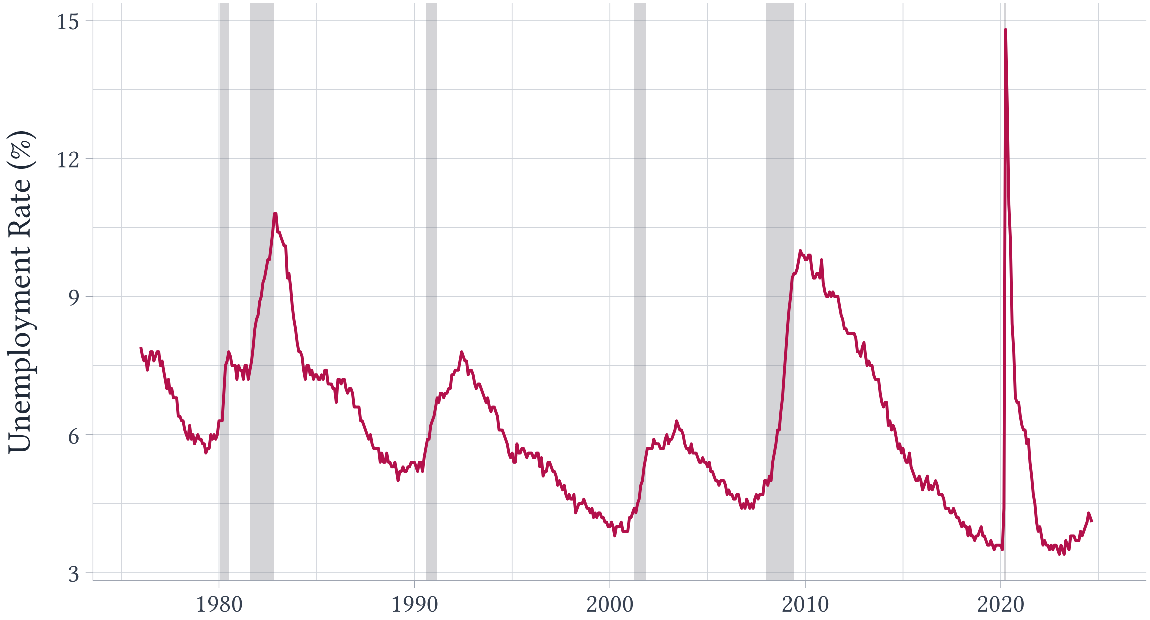
where:

- γ_k is the autocovariance at lag k,
- γ_0 is the variance of y_t (i.e., autocovariance at lag 0).

Autocorrelation

$$\rho_k = \frac{\gamma_k}{\text{var}\left(y_t\right)} = \frac{\text{cov}\left(y_t, y_{t-k}\right)}{\text{var}\left(y_t\right)}$$

Intuition: Autocorrelation tells us the strength of the relationship between y_t and its past values. It ranges between -1 and 1.



Unemployment Rate (%)
15
12
9
6
3
1980
1990
2000
2010
2020

Unemployment Example

In the unemployment example, the time-series

$$\hat{\gamma}_1 = \text{cov}\left(y_t, y_{t-1}\right) = 2.968 \quad \text{and} \quad \hat{\rho}_1 = 0.961$$

- Unsurprisingly the correlation of unemployment from 1-month to the next is very strong

Unemployment Example

In the unemployment example, the time-series

$$\hat{\gamma}_1 = \text{cov}\left(y_t, y_{t-1}\right) = 2.968 \quad \text{and} \quad \hat{\rho}_1 = 0.961$$

- Unsurprisingly the correlation of unemployment from 1-month to the next is very strong

This is useful for forecasting; a very strong autocorrelation tells us that recent values of y should be useful for predicting future values of y

Unemployment Example

Let's look at the correlation unemployment over 12 periods (year to year)

$$\hat{\rho}_{12} = 0.659$$

- Shocks to last year's unemployment seem to 'persist' into the current period

Unemployment Example

If we use the reshuffled gdp data, what do we think the autocorrelation may be?

Unemployment Example

If we use the reshuffled gdp data, what do we think the autocorrelation may be?

$$\hat{\rho}_{1,\texttt{reshuffled}} = -0.03081183$$

- When we completely randomly shuffled the data, we have destroyed any autocorrelation!

Unemployment Example

If we use the reshuffled gdp data, what do we think the autocorrelation may be?

$$\hat{\rho}_{1,\texttt{reshuffled}} = -0.03081183$$

- When we completely randomly shuffled the data, we have destroyed any autocorrelation!

This makes sense. If I reshuffled the data, knowing last month's (reshuffled) unemployment is no longer useful for predicting this month's (reshuffled) unemployment rate

Roadmap

Introduction to Time-Series

Time-series Statistics

Learning from Time-Series

Smoothing Methods for Inference

- Trends and Seasonality

Smoothing Methods for Forecasting

- Exponential Smoothing
- Trends and Seasonality with SES

Two goals of time-series

There are two possible goals that we can tackle when working with time-series data:

1. Learn about *persistent* patterns in how y evolves over time while ignoring random fluctuations (inference)
   - $\rightarrow$ E.g. learn about seasonality, trends, etc.
2. Predict future values of y_t (forecasting)
   - $\rightarrow$ The above step might be useful in predicting future y, but not necessary (only care about prediction)

Will try to clarify when we are discussing forecasting vs. describing time-series patterns (inference)

Learning from time-series

We observe a set of time-series observations y_t. Think of the observed y as being generated by

$$y_t = \mu_t + \varepsilon_t$$

- μ_t is the 'typical' or 'systematic' value of y at time t
- ε_t is a random fluctuation

Of course, we do not know which fluctuations are due to μ_t changing over time or ε_t changing over time

- Without any more structure, this is is an impossible task

Learning from time-series

$$y_t = \mu_t + \varepsilon_t$$

All hope is not lost though. Looking at the previous time-series graphs, it is clear we can learn *something*

$$y_t = \mu_t + \varepsilon_t$$

Here are some examples of what we can hope to learn using time-series data:

1. Identify seasonality in data
   - $\rightarrow$ Does the change in μ_t over the year follow a standard pattern?
   - $\rightarrow$ E.g. retail sales increasing in December
2. Detect long-term trends
   - $\rightarrow$ How does μ_t change over time?
   - $\rightarrow$ E.g. trend in GDP over time
3. Assess how strongly autocorrelated the data is
   - $\rightarrow$ How 'sticky' shocks are from past periods are

Key insight in time-series forecasting

Key Insight: By analyzing the changes across time, we reveal structure and patterns that help in making better predictions. For example:

- Does yesterday's sales help us learn about what products people will buy today?
- Do we see an up-swing in jacket sales every October?

Key insight in time-series forecasting

Key Insight: By analyzing the changes across time, we reveal structure and patterns that help in making better predictions. For example:

- Does yesterday's sales help us learn about what products people will buy today?
- Do we see an up-swing in jacket sales every October?

Of course, this can fail if the underlying structure of the world changes over time

- If we are using data from early 2000s on homes, we will surely fail at forecasting during the Great Recession

Evaluating forecasting methods

As usual, we can use the mean-squared prediction error to evaluate our models:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

- Typically, will evaluate on the time-series data you do observe

Evaluating forecasting methods

Time-series forecasting is particularly difficult to evaluate

- Our training data is past-values up until today
- Our testing data is values in the future

If the structure of the world changes over time, then our testing data *can* look fundamentally different over time

- Consumer preferences change over time can make predicting future sales hard

Over-fitting

For this reason, we have to be *very* careful when using forecasting methods on time-series

- Over-fitting the past data makes us learn ‘false’ time-series relationships

Roadmap

Introduction to Time-Series

Time-series Statistics

Learning from Time-Series

Smoothing Methods for Inference

- Trends and Seasonality

Smoothing Methods for Forecasting

- Exponential Smoothing
- Trends and Seasonality with SES

Smoothing Methods

Recall we said y_t was generated by

$$y_t = \mu_t + \varepsilon_t$$

- μ_t is the 'typical' or 'systematic' value of y at time t

The idea of smoothing methods is to use time periods right around period t to estimate a smoothed value at period t

- Want to "smooth" over random fluctuations

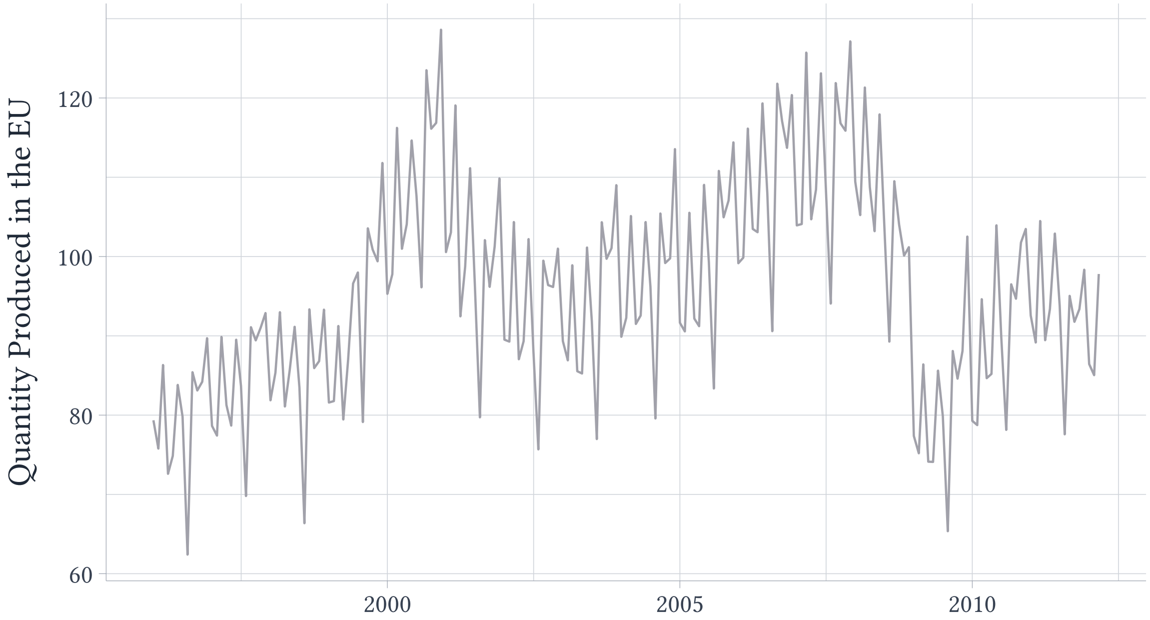
Example Electrical Manufacturing in the EU

On the following slide I'm going to show you production figures across the European Union

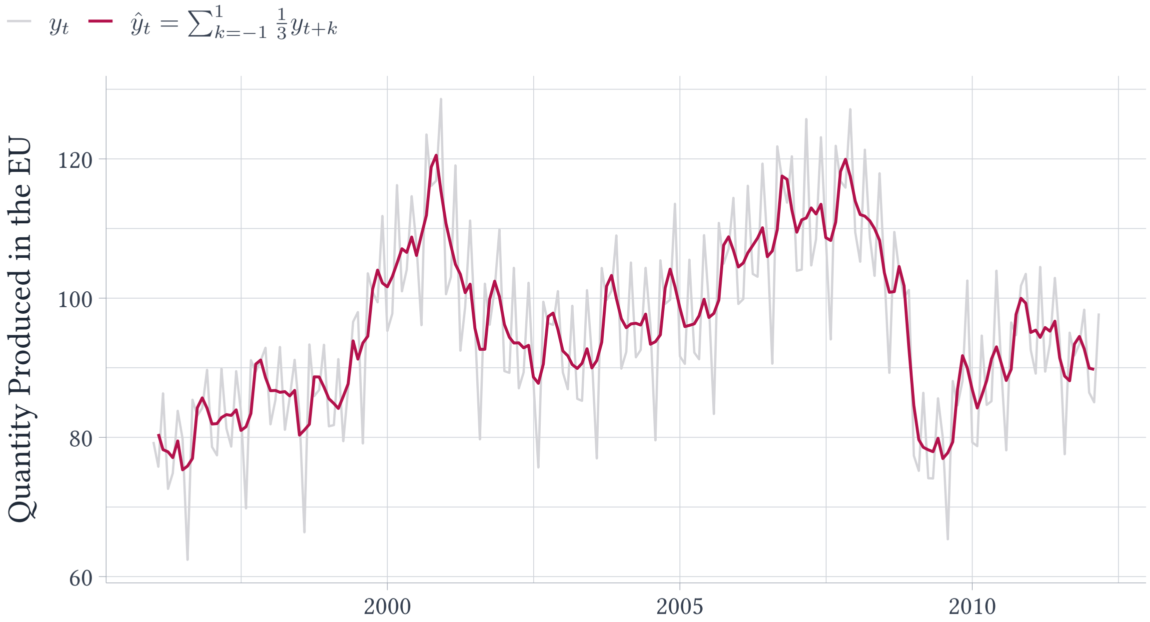
- Time-series data is on electircal manufactuing (computers and other technology) and is from EUROSTAT

When looking at this figure, try to imagine the 'systematic' component versus the random fluctuation ε_t

$$y_t = \mu_t + \varepsilon_t$$



Quantity Produced in the EU
120
100
80
60
2000
2005
2010


y_t
$\hat{y}_t = \sum_{k=-1}^{1} \frac{1}{3} y_{t+k}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Smoothing methods

In the previous figure, I created a moving average where I estimated the μ_t as being an average of $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}$

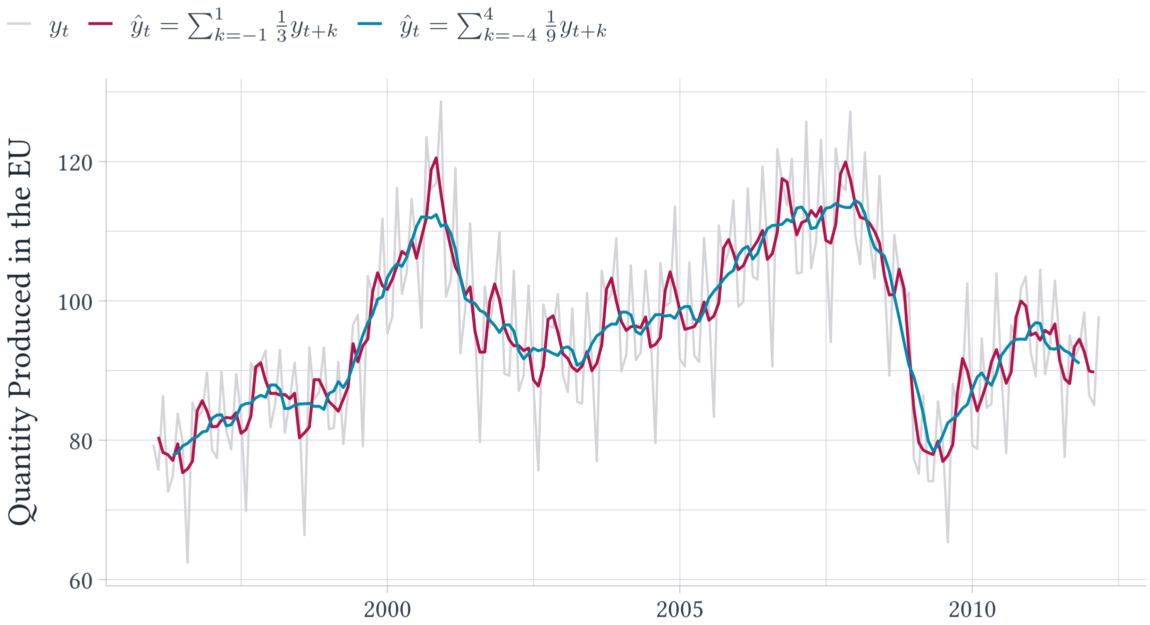
- This helped to smooth out some of the random fluctuations, perhaps better isolating systematic trends in y_t

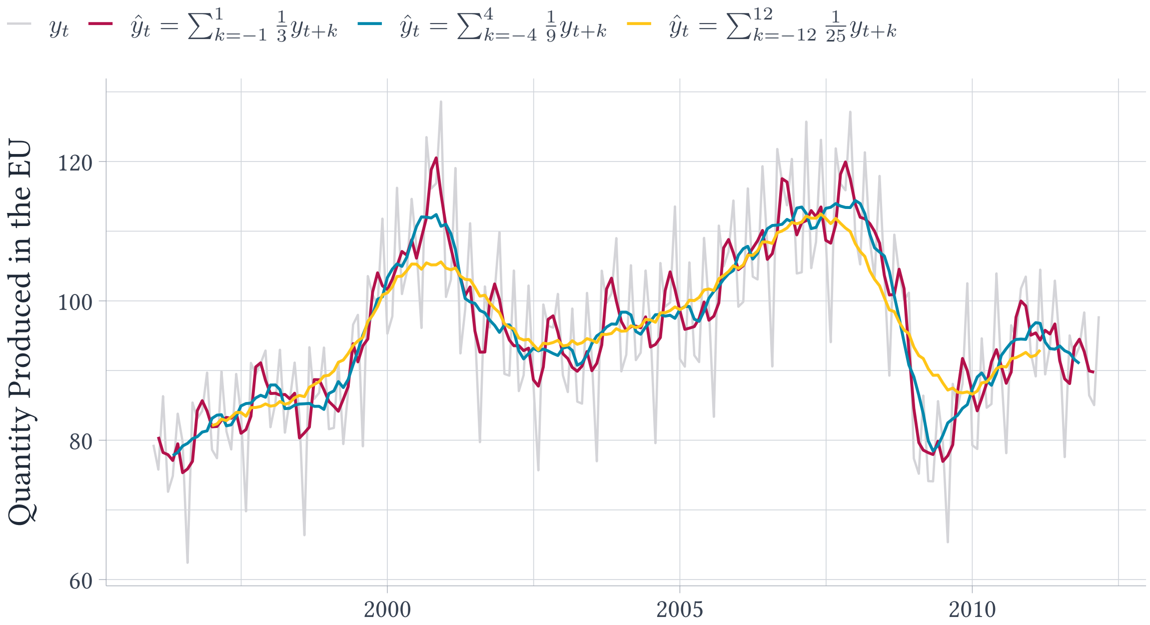
Smoothing methods

In the previous figure, I created a moving average where I estimated the μ_t as being an average of $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}$

- This helped to smooth out some of the random fluctuations, perhaps better isolating systematic trends in y_t

What happens if I average a bit more over time?


y_t
$\hat{y}_t = \sum_{k=-1}^{1} \frac{1}{3} y_{t+k}$
$\hat{y}_t = \sum_{k=-4}^{4} \frac{1}{9} y_{t+k}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010


y_t
$\hat{y}_t = \sum_{k=-1}^{1} \frac{1}{3} y_{t+k}$
$\hat{y}_t = \sum_{k=-4}^{4} \frac{1}{9} y_{t+k}$
$\hat{y}_t = \sum_{k=-12}^{12} \frac{1}{25} y_{t+k}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Moving average

In general, our moving average can be calculated as follows:

$$\hat{y}_t = \sum_{k=-K}^{K} \frac{1}{2K+1} y_{t+k}$$

This is just the sample mean using observations within $\pm K$ periods of t

- K is the number of observations on each side of y_t we include
- $2K+1$ is the number of observations. Note $+1$ because we include y_t

Moving average

In general, our moving average can be calculated as follows:

$$\hat{y}_t = \sum_{k=-K}^{K} \frac{1}{2K+1} y_{t+k}$$

This is just the sample mean using observations within $\pm K$ periods of t

- K is the number of observations on each side of y_t we include
- $2K+1$ is the number of observations. Note $+1$ because we include y_t

I will show you how to do this using the `slider` package in `R`

What happened to the end points?

$$\hat{y}_t = \sum_{k=-K}^{K} \frac{1}{2K+1} y_{t+k}$$

Note when calculating a rolling-average, we will face problems on either end of our observed time-series

- E.g. for my first observation, I do not have the y from the period before

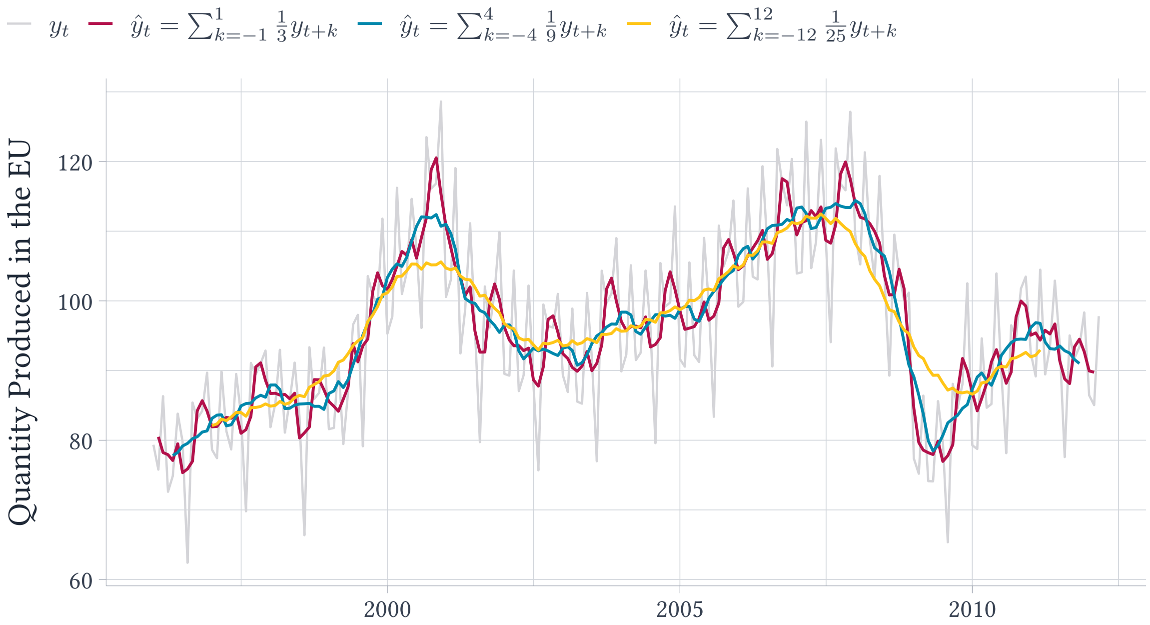
What happened to the end points?

$$\hat{y}_t = \sum_{k=-K}^{K} \frac{1}{2K+1} y_{t+k}$$

Note when calculating a rolling-average, we will face problems on either end of our observed time-series

- E.g. for my first observation, I do not have the y from the period before

That is what causes the truncated ends of the smoothed time-series graph



y_t
$\hat{y}_t = \sum_{k=-1}^{1} \frac{1}{3} y_{t+k}$
$\hat{y}_t = \sum_{k=-4}^{4} \frac{1}{9} y_{t+k}$
$\hat{y}_t = \sum_{k=-12}^{12} \frac{1}{25} y_{t+k}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Problems with moving averages

“Over-smoothing”

When K is large, we are using observations quite far away from the current period (e.g. using data from 12 months ago)

- This prevents $\hat{y}_t$ from being driven too much by the current period’s observation (for better or worse!)

Problems with moving averages

"Over-smoothing"

When K is large, we are using observations quite far away from the current period (e.g. using data from 12 months ago)

- This prevents $\hat{y}_t$ from being driven too much by the current period's observation (for better or worse!)

When we have a high-degree of smoothing, our smoothed time-series misses out on true shocks to μ_t that are short-lived

- In our previous example, the overly-smoothed version misses the short jump in manufacturing in the early 2000s

Problems with moving averages

Seasonality

Say you had time-series data on candy sales over the course of the last decade

- You would see a bump every October for Halloween (i.e. it is part of μ_t)

Even a moderately small $K = 1$ would make $\hat{y}_t$ be too small in October

- Temporary seasonal swings in y (i.e. last only a period or two) are going to be lost

Selecting K

There is a trade-off at play

- Using a small K only uses the most recent information (perhaps better picking up on recent shocks)
- Using a larger K helps average over non-persistant random noise

Selecting K

There is a trade-off at play

- Using a small K only uses the most recent information (perhaps better picking up on recent shocks)
- Using a larger K helps average over non-persistant random noise

This is an example of a *bias-variance tradeoff*

- Smaller K lowers bias, but increases variance

Mean-squared prediction error

Say we wanted to use data to tell us the ‘best’ K to use for forming $\hat{y}_t$

We could search over $K = 0, 1, 2, 3, \ldots$ and see which gives us the smallest mean-squared prediction error:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

Mean-squared prediction error

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

For smoothing averages, when $K = 0$, we just use $\hat{y}_t = y_t$ and we have MSE of 0

- As K increases, the MSE necessarily grows

Trying to select K this way fails utterly because we are using our training data as our testing data!

Roadmap

Seasonality, Trends, and Shocks

It is often desirable to break up μ_t into two components:

$$y_t = T_t + S_t + \varepsilon_t$$

- S_t is the seasonality term (e.g. year over year)
- T_t is the trend-term
- and ε_t is the remaining noise (random fluctuations)

Let's look into how we can try to separate trends from seasonality

- This section will cover the 'classical' decomposition (see 3.4 in Forecasting: Principles and Practices)

Moving average to remove seasonality, S_t

It turns out, there is a particular moving average that can remove seasonality from the data

- For this example, we will think of monthly data and try to remove annual trend (you can similarly do this with quarterly data)

Remember we can write our general moving average as

$$\hat{y}_t = \sum_{k=-K}^{K} w_k y_{t+k}$$

- If we choose K and w_k right, we will try to remove seasonality

Moving average to remove seasonality, S_t

$$
\begin{aligned}
\hat{y}_t = \frac{1}{24}y_{t-6} + \frac{1}{12}y_{t-5} + \frac{1}{12}y_{t-4} + \frac{1}{12}y_{t-3} + \frac{1}{12}y_{t-2} + \frac{1}{12}y_{t-1} \\
+ \frac{1}{12}y_t \\
+ \frac{1}{12}y_{t+1} + \frac{1}{12}y_{t+2} + \frac{1}{12}y_{t3+3}\frac{1}{12}y_{t+4} + \frac{1}{12}y_{t+5} + \frac{1}{24}y_{t+6}
\end{aligned}
$$

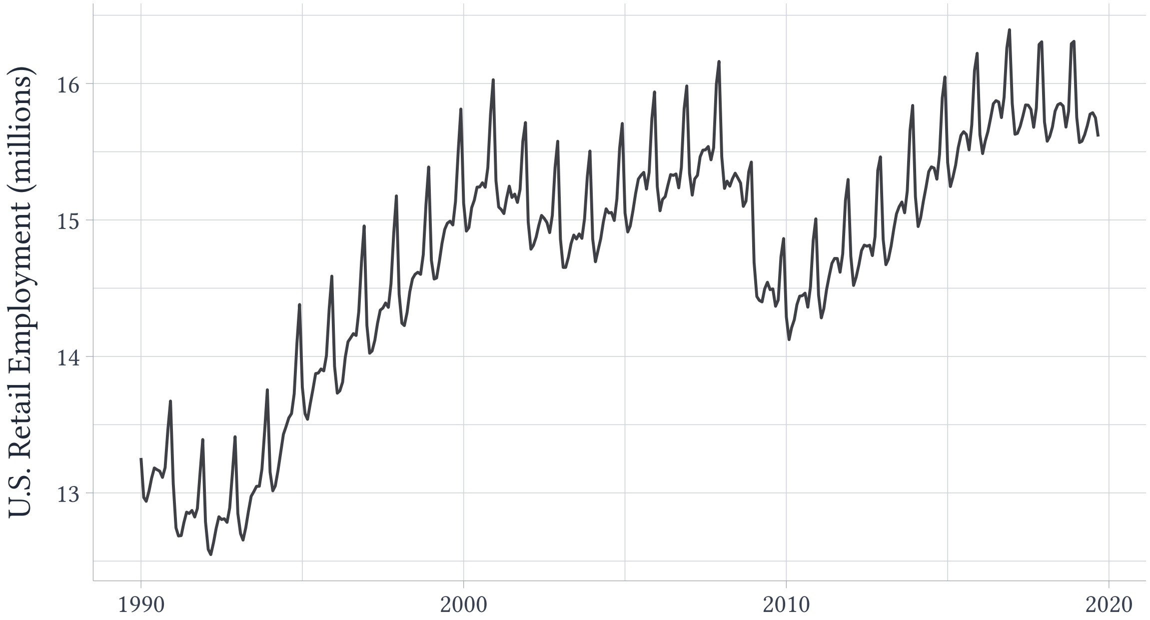
Basically a $\pm K$ smoothing average, but first and last get a half the weight

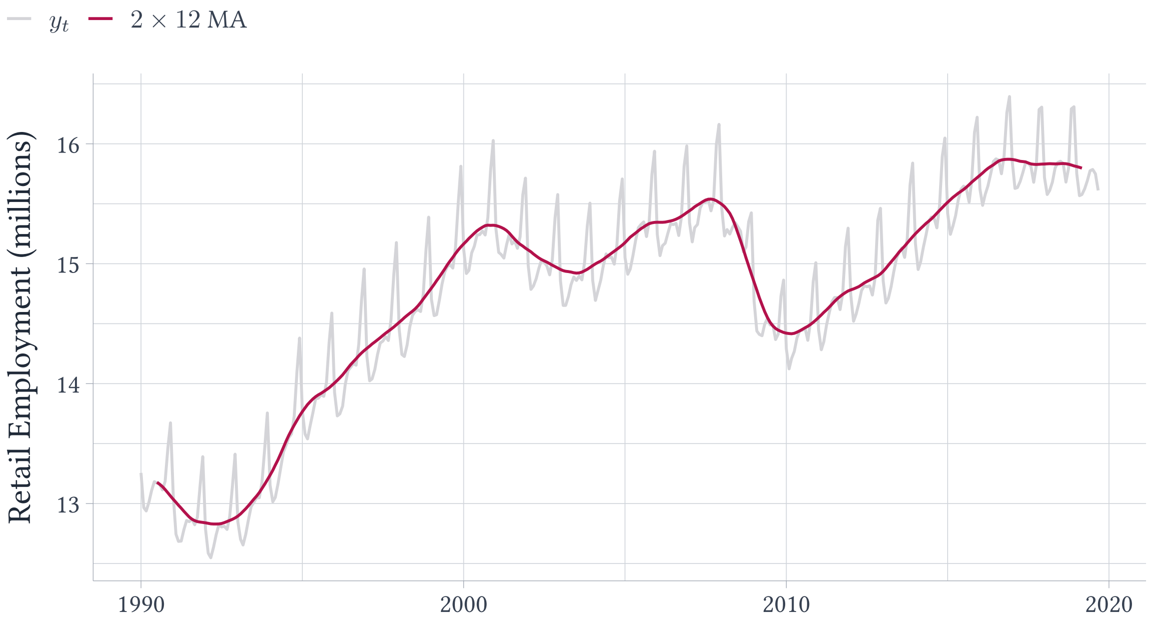
Moving average to remove seasonality, S_t

Or, can do the following:

- 12-month rolling average
- 2-month rolling average of the 12-month rolling average

Sometimes called the 2×12 MA


U.S. Retail Employment (millions)
13
14
15
16
1990
2000
2010
2020



y_t
2×12 MA
Retail Employment (millions)
16
15
14
13
1990
2000
2010
2020

2×12 MA

Our 2×12 moving-average serves as the classical estimate of $\hat{T}_t$, i.e. the time-trend

2×12 MA

Our 2×12 moving-average serves as the classical estimate of $\hat{T}_t$, i.e. the time-trend

For quarterly data, you would do

$$\hat{y}_t = \frac{1}{8}y_{t-2} + \frac{1}{4}y_{t-1} + \frac{1}{4}y_t + \frac{1}{4}y_{t+1} + \frac{1}{8}y_{t+2}$$

De-trending our data

Now, we can "de-trend" our data by forming $y_t - \hat{T}_t$

- $\hat{T}_t$ is our 2×12 moving average estimate

What remains is

$$y_t - \hat{T}_t \approx S_t + \varepsilon_t$$

Estimating seasonality

We want to know how does $y_t - \hat{T}_t$ cycle throughout the year

- E.g. is retail employment systematically higher in November and December?

The classical way to estimate this is take the average of $y_t - \hat{T}_t$ separately for each month

- Each month's average serves as the estimated month's "seasonal trend", $\hat{S}_t$
  - $\rightarrow$ Takes the same value year over year

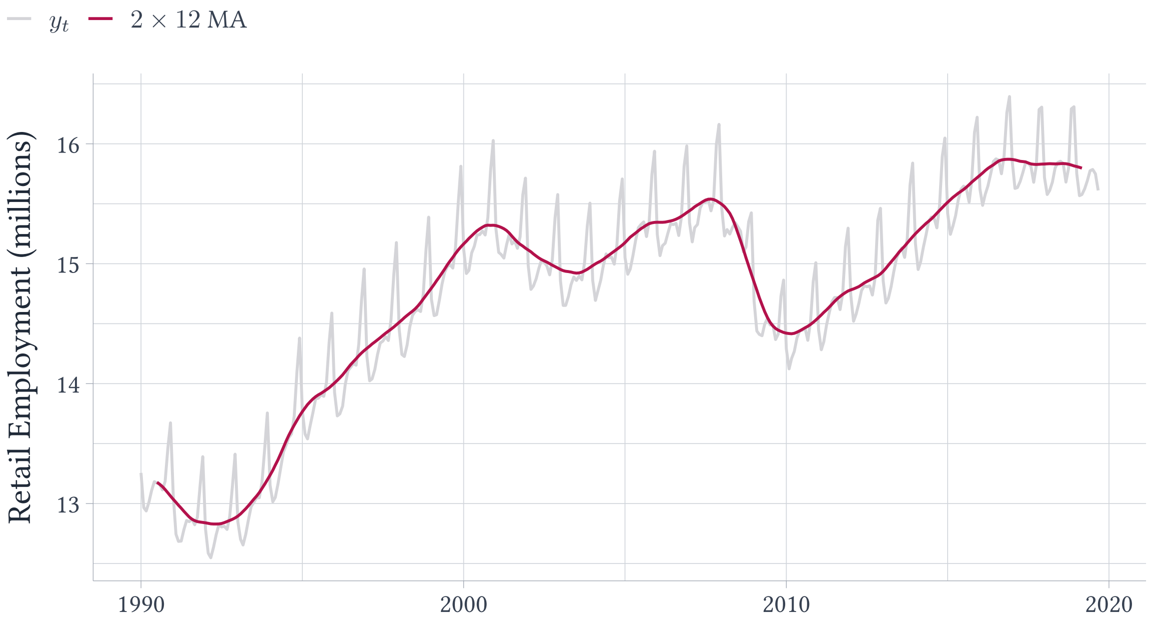
Seasonality estimation in R

The classical way to estimate this is take the average of $y_t - \hat{T}_t$ separately for each month

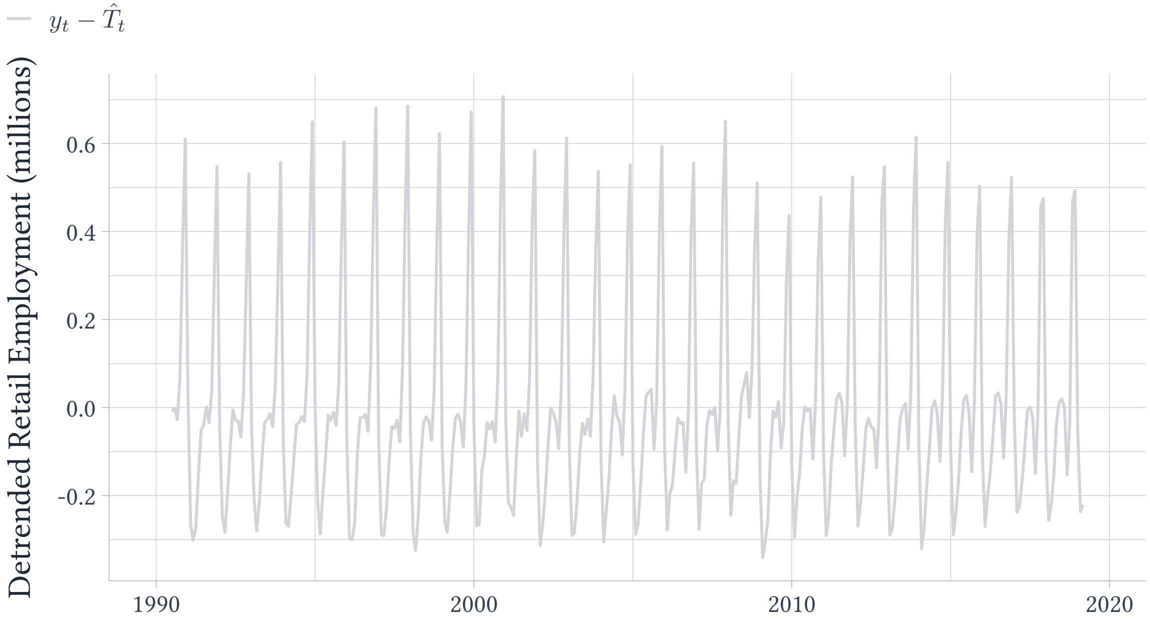
- This can be done by regression $y_t - \hat{T}_t$ on a set of month indicators (and no intercept)

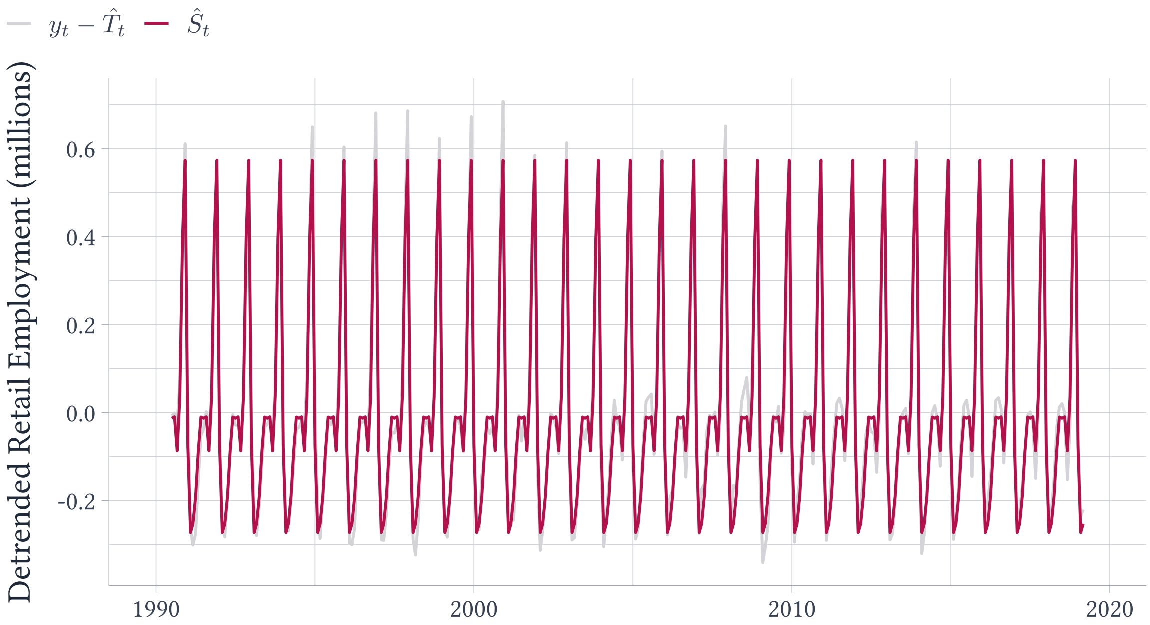
In `R`, this can be done with

```
feols(y_minus_trend ~ 0 + i(month(date)), data = df)
```



y_t
2×12 MA
Retail Employment (millions)
16
15
14
13
1990
2000
2010
2020


$y_t - \hat{T}_t$
Detrended Retail Employment (millions)
0.6
0.4
0.2
0.0
-0.2
1990
2000
2010
2020


$y_t - \hat{T}_t$
$\hat{S}_t$
Detrended Retail Employment (millions)
0.6
0.4
0.2
0.0
-0.2
1990
2000
2010
2020

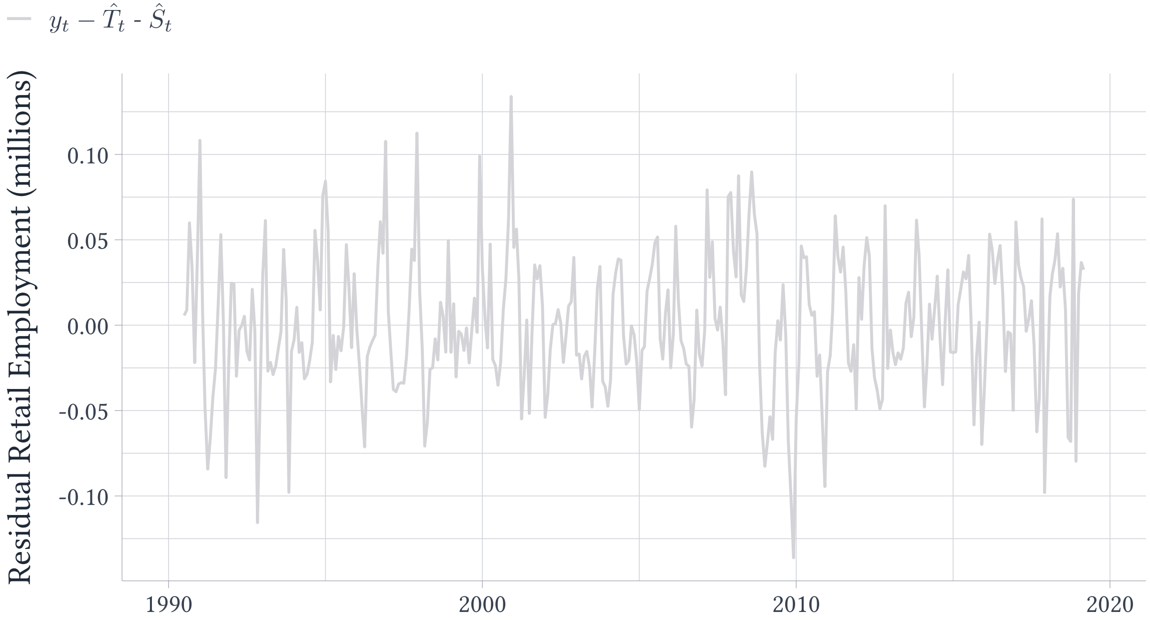
Residual

$$y_t - \hat{T}_t - \hat{S}_t \approx \varepsilon_t$$

- $\hat{T}_t$ is the 2×12 moving average estimate of trends
- $\hat{S}_t$ is the monthly average of $y_t - \hat{T}_t$

What remains after this is a de-trended and de-seasoned data, i.e. random fluctuations

- Should visually inspect this to see how good we did at removing trends and seasonality



$y_t - \hat{T}_t$ - $\hat{S}_t$
Residual Retail Employment (millions)
0.10
0.05
0.00
-0.05
-0.10
1990
2000
2010
2020

Roadmap

- Introduction to Time-Series
- Time-series Statistics
- Learning from Time-Series
- Smoothing Methods for Inference
  - Trends and Seasonality
- Smoothing Methods for Forecasting
  - Exponential Smoothing
  - Trends and Seasonality with SES

Smoothing Methods for Forecasting

Our previous goal was to learn the ‘systematic’ part of the time-series $y_t = \mu_t + \varepsilon_t$

“Rules” for forecasting

Typically, we will want to only use data from period t or prior in our model

- E.g. I can’t use y_{t+2} to predict tomorrow’s y_{t+1}

When predicting the future, I can’t view use data from the future

- So the model I learn can only use past data

Simplest forecasting method

The simplest forecasting method is to use y_{t-1}, the previous period's value, as the forecast for y_t:

$$\hat{y}_t = y_{t-1}$$

This method is going to use only information from the most recent observations

- Maybe the most recent observation is the most-relevant for predicting today

  $\rightarrow$ I.e. autocorrleation is high

- If μ_t is really wild (i.e. no trends/seasonality), then we should only use recent information

Cons of using y_{t-1} as a forecast

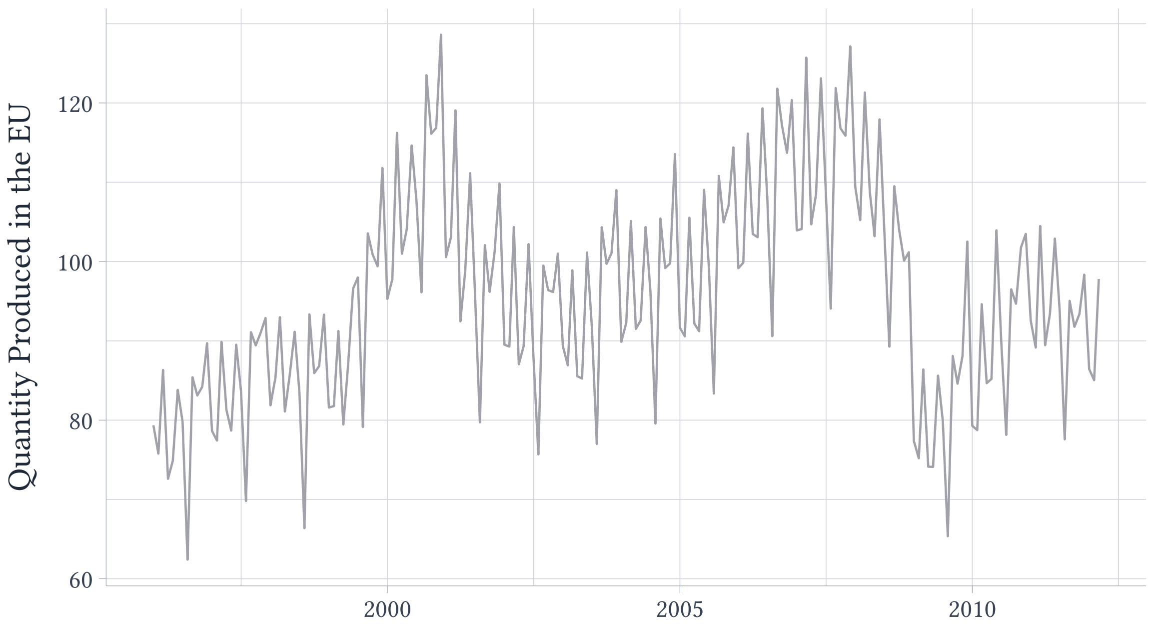
Using y_{t-1} could fail when:

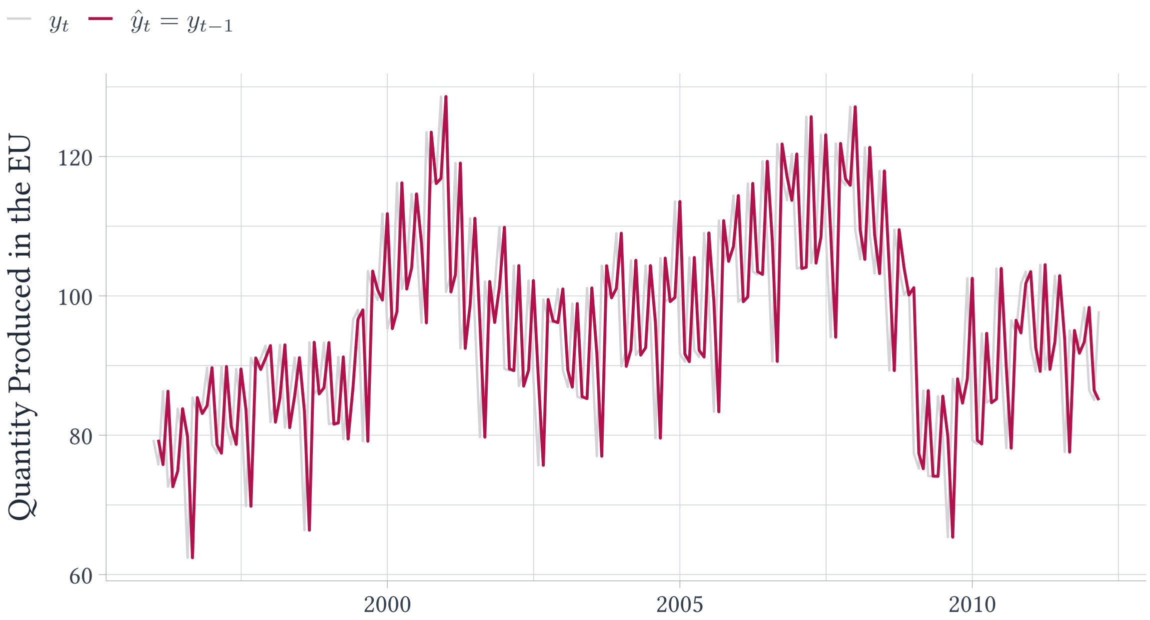
- The data has trends or seasonality that y_{t-1} doesn't capture
  - $\rightarrow$ Using August's jacket sales to predict September's jacket sales will not do well

Cons of using y_{t-1} as a forecast

Using y_{t-1} could fail when:

- The data has trends or seasonality that y_{t-1} doesn't capture
  - $\rightarrow$ Using August's jacket sales to predict September's jacket sales will not do well
- y_{t-1} can be quite *noisy*
  - $\rightarrow$ Maybe yesterday's value of y was weird because of a bad news story that turned out to not be a big deal


Quantity Produced in the EU
120
100
80
60
2000
2005
2010


y_t
$\hat{y}_t = y_{t-1}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Predicition Error

On the last slide, it's hard to see, but the $\hat{y}_t = y_{t-1}$ does a bad job at predicting y_t

- The data jumps around too much, so yesterday's value is only weakly predictive of today's value

Smoothing Methods

We can try to improve on the simple method by smoothing over the last K periods:

$$\hat{y}_t = \sum_{k=1}^{K} w_k y_{t-k},$$

where

- K is the number of lags to smooth over
- w_k is the weights put on the k-th lagged value of y

Smoothing Methods

$$\hat{y}_t = \sum_{k=1}^{K} w_k y_{t-k},$$

For example:

- $K = 1$ and $w_1 = 1$ is the simple method $\hat{y}_t = y_{t-1}$

Smoothing Methods

$$\hat{y}_t = \sum_{k=1}^{K} w_k y_{t-k},$$

For example:

- $K = 1$ and $w_1 = 1$ is the simple method $\hat{y}_t = y_{t-1}$
- $K = 3$ and $w_3 = \frac{1}{3}$ is the average of three-previous periods

Average of previous ys

Say we use an average of the K most recent observations:

$$\hat{y}_t = \sum_{k=1}^{K} \frac{1}{K} y_{t-k},$$

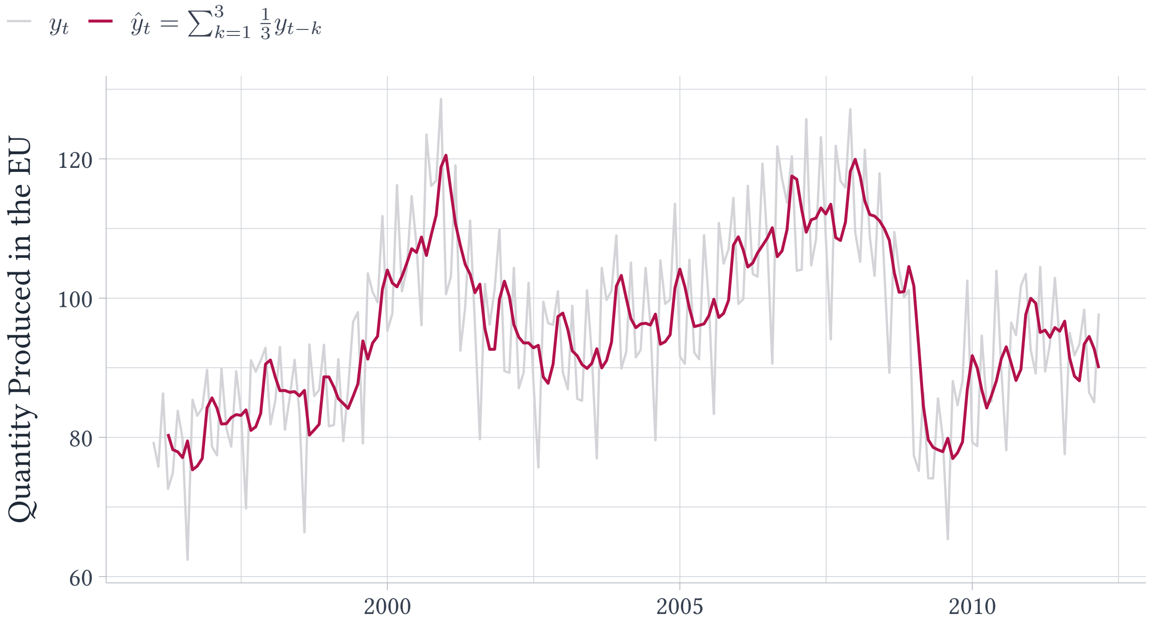
Average of previous ys

Say we use an average of the K most recent observations:

$$\hat{y}_t = \sum_{k=1}^{K} \frac{1}{K} y_{t-k},$$

As you move ahead one time period, you lose 1 observation $(t - K)$ and gain one observation t

- The most recent observation y_t updates what we think the moving average is


y_t
$\hat{y}_t = \sum_{k=1}^{3} \frac{1}{3} y_{t-k}$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Prediction

Note for the next period y_{t+1}, we can form our out-of-sample forecast as:

$$\hat{y}_{t+1} = \sum_{k=1}^{K} \frac{1}{K} y_{t-k},$$

- This is not true of our mouving average

Roadmap

Introduction to Time-Series

Time-series Statistics

Learning from Time-Series

Smoothing Methods for Inference

- Trends and Seasonality

Smoothing Methods for Forecasting

- Exponential Smoothing
- Trends and Seasonality with SES

Exponential Smoothing

It turns out, there is a (typically) better method over the sample average of the previous K periods.

It is called Exponential Smoothing and is quite popular because it works well

- Also has some generalizations that allow it to be more flexible

Simple Exponential Smoothing

The simple exponential smoothing method forms predictions in a *recursive manner*:

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

- To predict y in period $t + 1$, take a weighted sum of the observed y_t and the prediction $\hat{y}_{t-1}$

We learn the true value of y_t and update our prediction for the next period

- When $y_t > \hat{y}_t$, we revise up our forecast
- When $y_t < \hat{y}_t$, we revise down our forecast

How much to update, α

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

α tells us how much to update

- $\alpha = 1$ means throw out old prediction and use y_t
- $\alpha = 0$ means do not update at all
- $1 > \alpha > 0$ means updating more (close to 1) or less strongly (close to 0)

Simple Exponential Smoothing

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

There's a problem here though, because how do we get $\hat{y}_t$? Well we get it using $\hat{y}_{t-1}$

- And to get $\hat{y}_{t-1}$ we need $\hat{y}_{t-2}$...
- and turtles all the way down ...

Simple Exponential Smoothing

Starting from period $t = 1$,

$$\hat{y}_2 = \alpha y_1 + (1 - \alpha)\ell_0$$

$$\hat{y}_3 = \alpha y_2 + (1 - \alpha)\hat{y}_2$$

$$\vdots$$

$$\hat{y}_T = \alpha y_{T-1} + (1 - \alpha)\hat{y}_{T-1}$$

$$\hat{y}_{T+1} = \alpha y_T + (1 - \alpha)\hat{y}_T.$$

So, we only need the starting point ℓ_0

Simple Exponential Smoothing

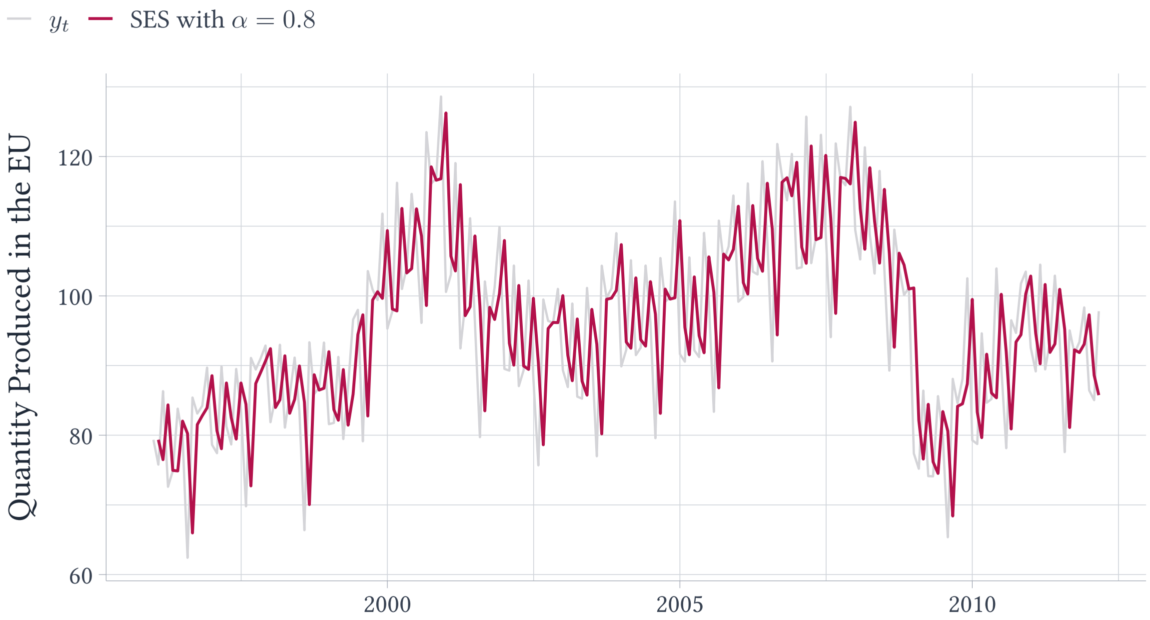
Since we usually do not care that much about predicting early period's y's, we can just pick ℓ_0 to be y_1

- It does not turn out to matter that much for forecasting if T is large

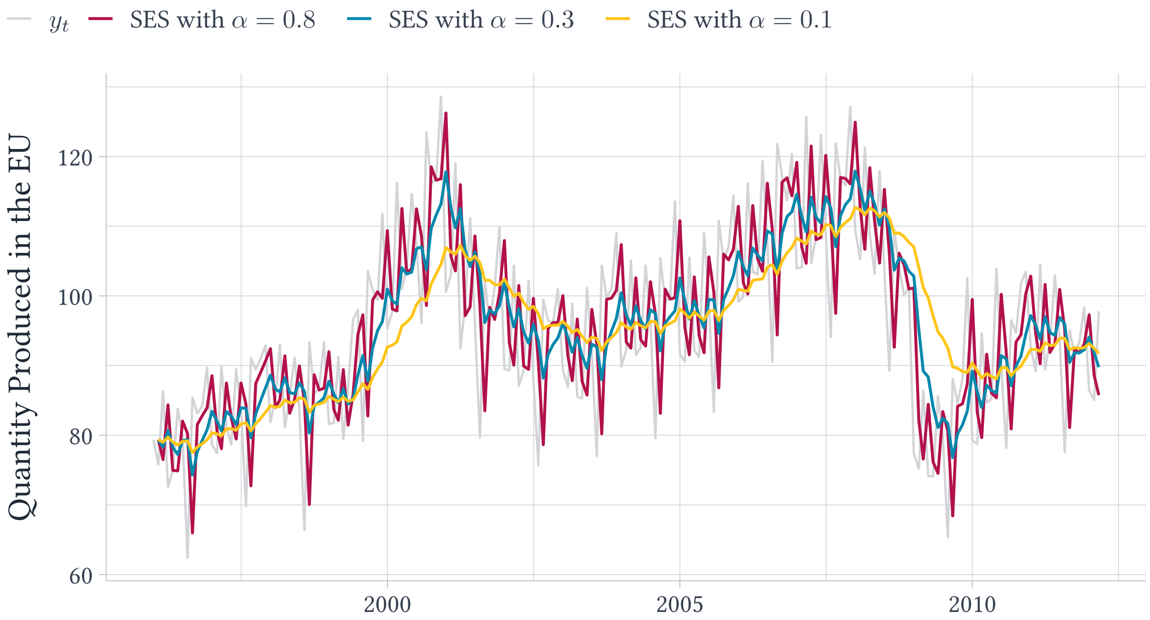
Updating parameter α

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

Let's look at a couple examples of α to build intuition on how this works



y_t
SES with $\alpha = 0.8$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010


y_t
SES with $\alpha = 0.8$
SES with $\alpha = 0.3$
SES with $\alpha = 0.1$
Quantity Produced in the EU
120
100
80
60
2000
2005
2010

Simple Exponential Smoothing and Recursion

We can trace out how $\hat{y}_t$ works as follows:

$$
\begin{aligned}
\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\hat{y}_t \\
&= \alpha y_t + (1-\alpha)\left(\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}\right)
\end{aligned}
$$

Simple Exponential Smoothing and Recursion

We can trace out how $\hat{y}_t$ works as follows:

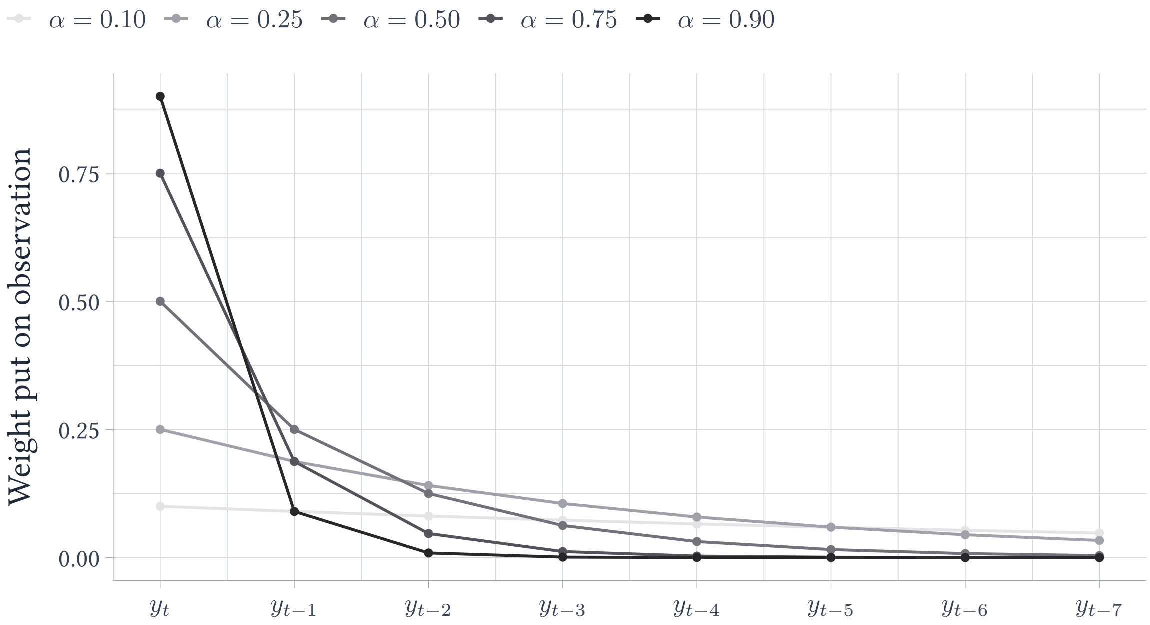
$$\begin{aligned}
\hat{y}_{t+1} &= \alpha y_t + (1-\alpha)\hat{y}_t \\
&= \alpha y_t + (1-\alpha)\left(\alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}\right) \\
&= \alpha y_t + \alpha(1-\alpha) y_{t-1} + (1-\alpha)^2 \hat{y}_{t-1}
\end{aligned}$$

Simple Exponential Smoothing and Recursion

You can repeat this process many times

$$
\begin{aligned}
\hat{y}_{t+1} &= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2\hat{y}_{t-1} \\
&= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-1} + (1-\alpha)^3\hat{y}_{t-2} \\
&= \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-1} + \alpha(1-\alpha)^3 y_{t-2} + (1-\alpha)^4\hat{y}_{t-3}
\end{aligned}
$$

Simple Exponential Smoothing is actually taking a weighted average of past y values all the way back to the first-period


$\alpha = 0.10$ $\alpha = 0.25$ $\alpha = 0.50$ $\alpha = 0.75$ $\alpha = 0.90$
Weight put on observation
0.75
0.50
0.25
0.00
y_t y_{t-1} y_{t-2} y_{t-3} y_{t-4} y_{t-5} y_{t-6} y_{t-7}

Weights and ‘adaptability’

From the previous figure, it is clear that different values of α bput different weights on long-run values

- A large α is more ‘adaptable’ in that it can respond much more quickly to changes in y
- A small α puts weight more evenly

The different emphasis that these weights put have implications for how the SES method deals with trends and seasonality

Roadmap

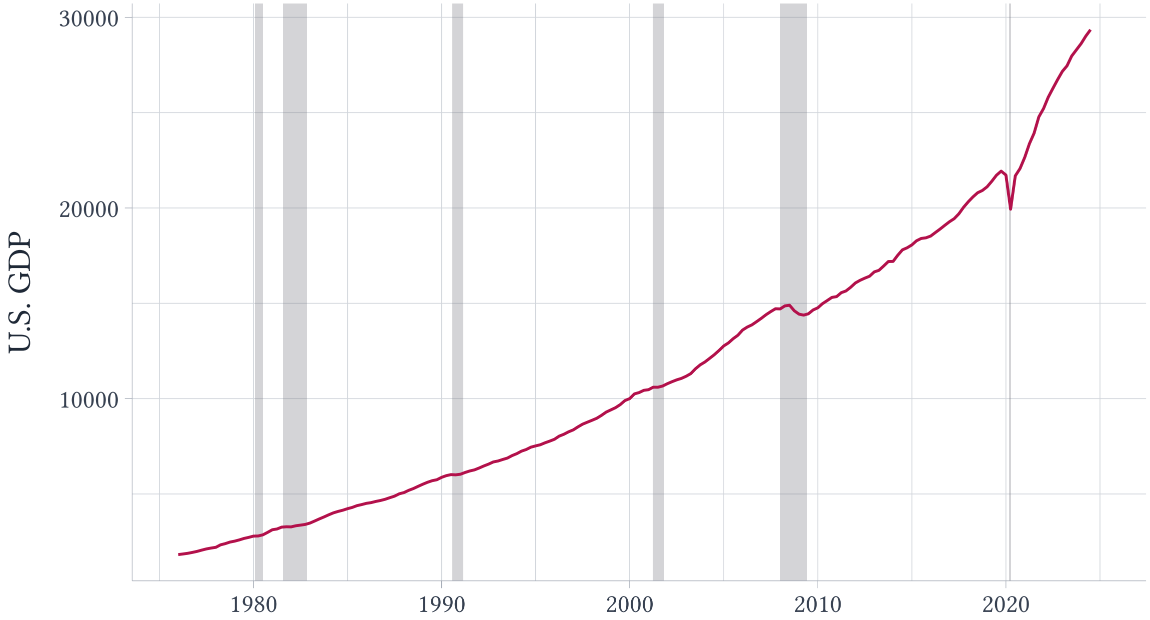
- Introduction to Time-Series
- Time-series Statistics
- Learning from Time-Series
- Smoothing Methods for Inference
  - Trends and Seasonality
- Smoothing Methods for Forecasting
  - Exponential Smoothing
  - Trends and Seasonality with SES

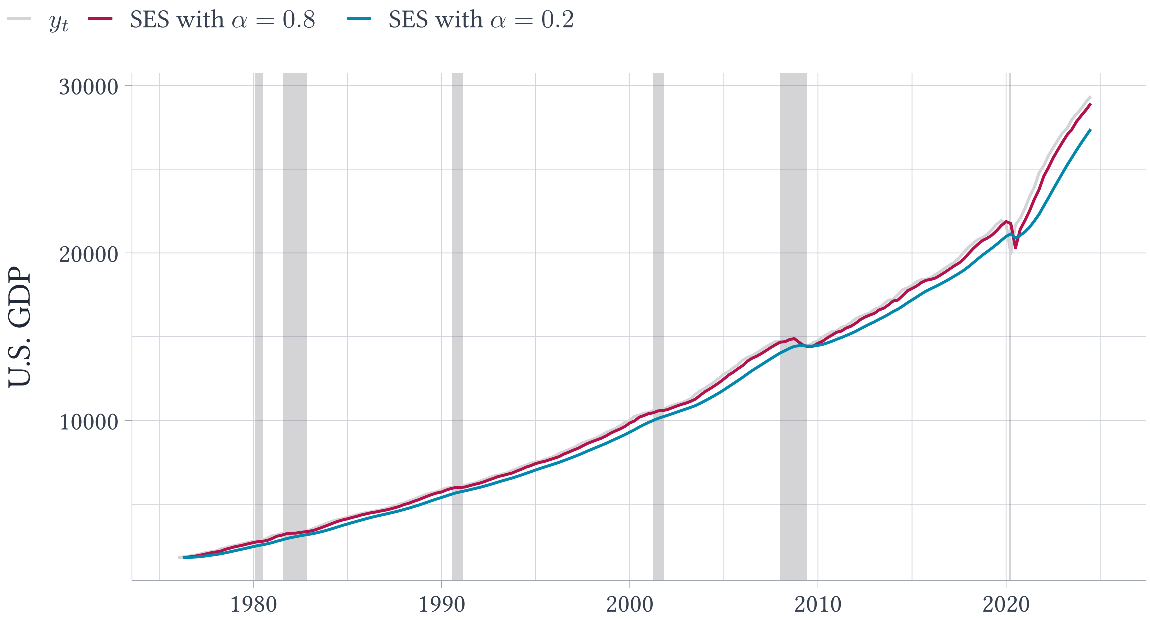
Trends

The simple exponential smoothing method can fail when the data has long-term trends

- When α is smaller, we lean more on observations from the past (when the trend was lower)

For example, let's look at smoothing US GDP estimates


U.S. GDP
30000
20000
10000
1980
1990
2000
2010
2020



y_t
SES with $\alpha = 0.8$
SES with $\alpha = 0.2$
U.S. GDP
30000
20000
10000
1980
1990
2000
2010
2020

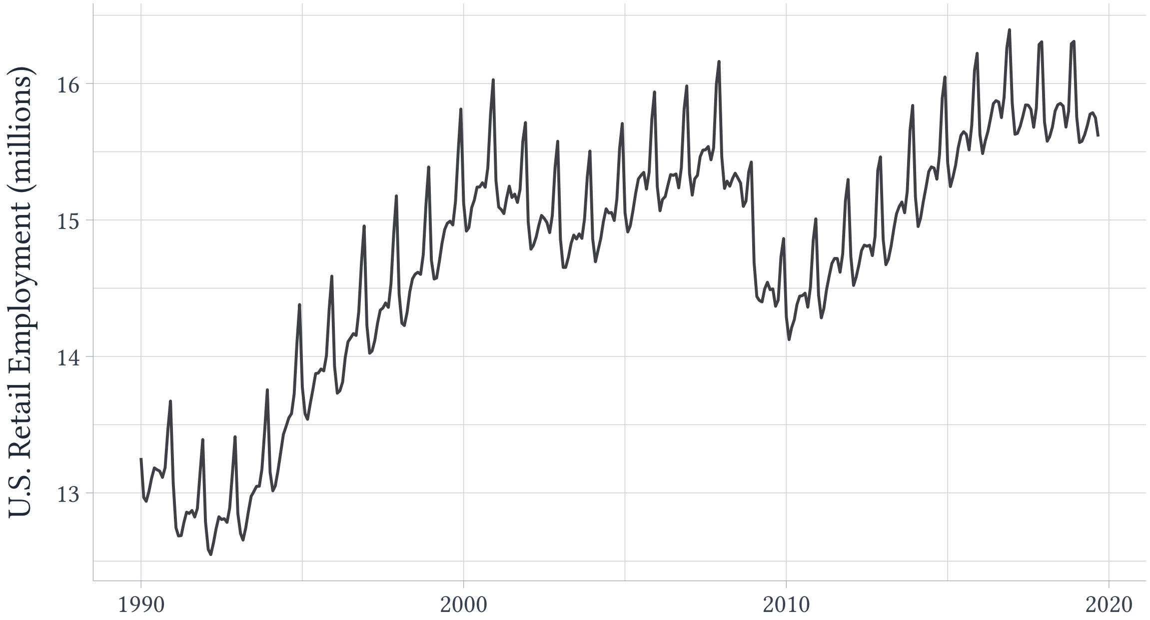
Simple Exponential Smoothing and Trends

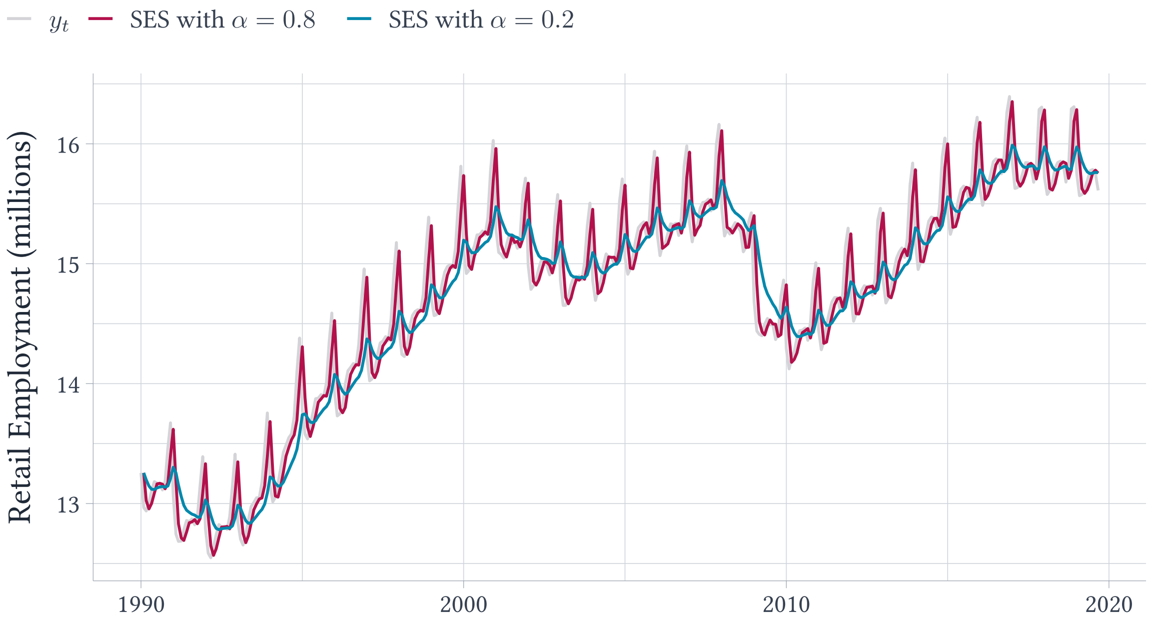
Since GDP is consistently trending upwards, old values of y_{t-k} are systematically lower

- Lower values of α will put more weight on older values, hence $\hat{y}_t$ being systematically too low

SES will also miss trends

- E.g. consider retail employment which is systematically higher around the end of the year


U.S. Retail Employment (millions)
16
15
14
13
1990
2000
2010
2020



y_t
SES with $\alpha = 0.8$
SES with $\alpha = 0.2$
Retail Employment (millions)
16
15
14
13
1990
2000
2010
2020

Holt-Winters method

Once again, we face this problem with these methods that we focus mainly on using recent observations in terms of t

- This ignores *trends* and *seasonality*

The Holt-Winters method is a more advanced method that allows for (1) repeated seasonality (year-over-year) and (2) smoothly-evolving trends

- Allows us to not ‘overfit’ short term fluctuations as much by learning something about general longer-run trends