Time Series Analysis

Discussion Section 01

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Organizational Issues

Please log in to **ISIS** (password: Zeit1718) and **download** the following files:

- USpop.dta

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Deterministic Models

- **Components of a Time Series**
- **Additive and Multiplicative Models**
- **Simple Trend Models**
- Smoothing Techniques
- Seasonal Adjustment

Components and Models

Four components

L_t long-term trend

C_t cyclical component

S_t 'seasonal' component

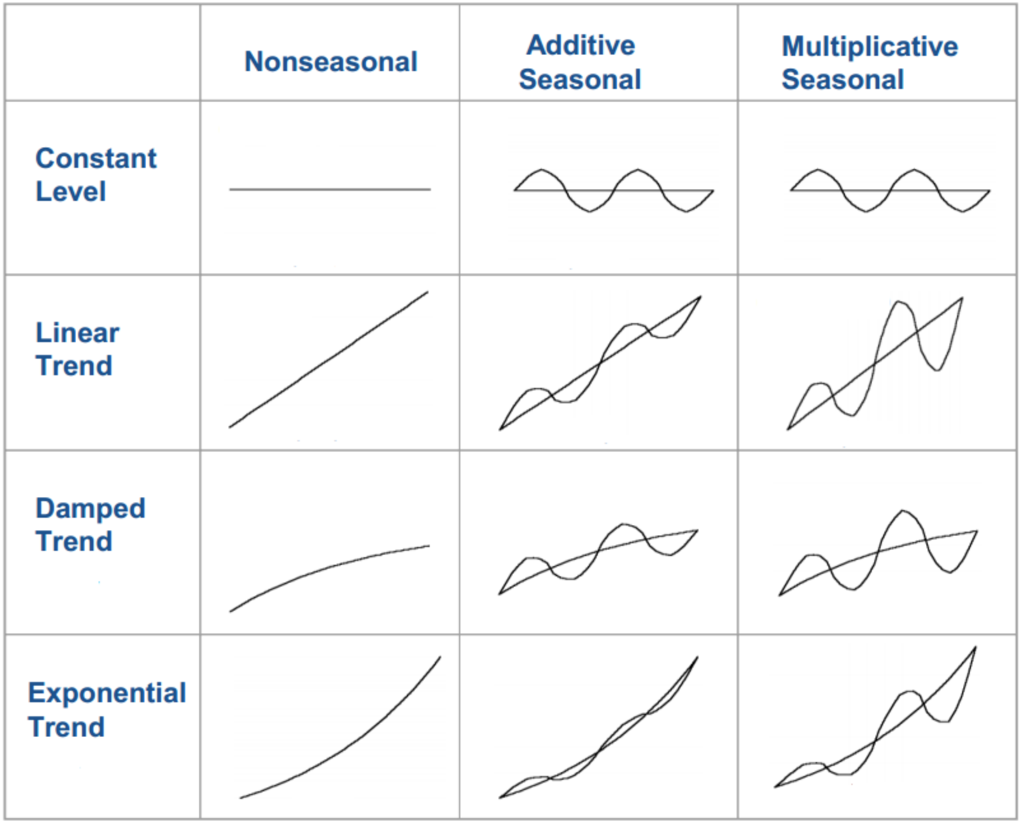
I_t irregular component

Additive Model

$$y_t = L_t + C_t + S_t + I_t \quad t = 1, \ldots, T$$

Multiplicative Model

$$y_t = L_t \cdot C_t \cdot S_t \cdot I_t \quad t = 1, \ldots, T$$



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.1:

- Generate a time variable (named “time”) for the “USpop.dta” dataset.

Recall:

- Creating New Variables: `generate newvar = exp`
- **System variables** (`_variables`)

  `_n` contains the number of the current observation. It is useful for indexing observations or generating sequences of numbers and can be used with mathematical operators.

- Time series dates:

Format (%fmt)	Description	Coding
%td	daily	0 = 01jan1960, 1 = 02jan1960
%tw	weekly	0 = 1960w1, 1 = 1960w2
%tm	monthly	0 = 1960m1, 1 = 1960m2
%tq	quarterly	0 = 1960q1, 1 = 1960q2
%th	halfyearly	0 = 1960h1, 1 = 1960h2
%ty	yearly	1960 = 1960, 1961 = 1961

  `format varlist %ty`

- Label the variable time “Time”.
- Plot the time series and describe its pattern. Which of the simple trend models do you think is appropriate?

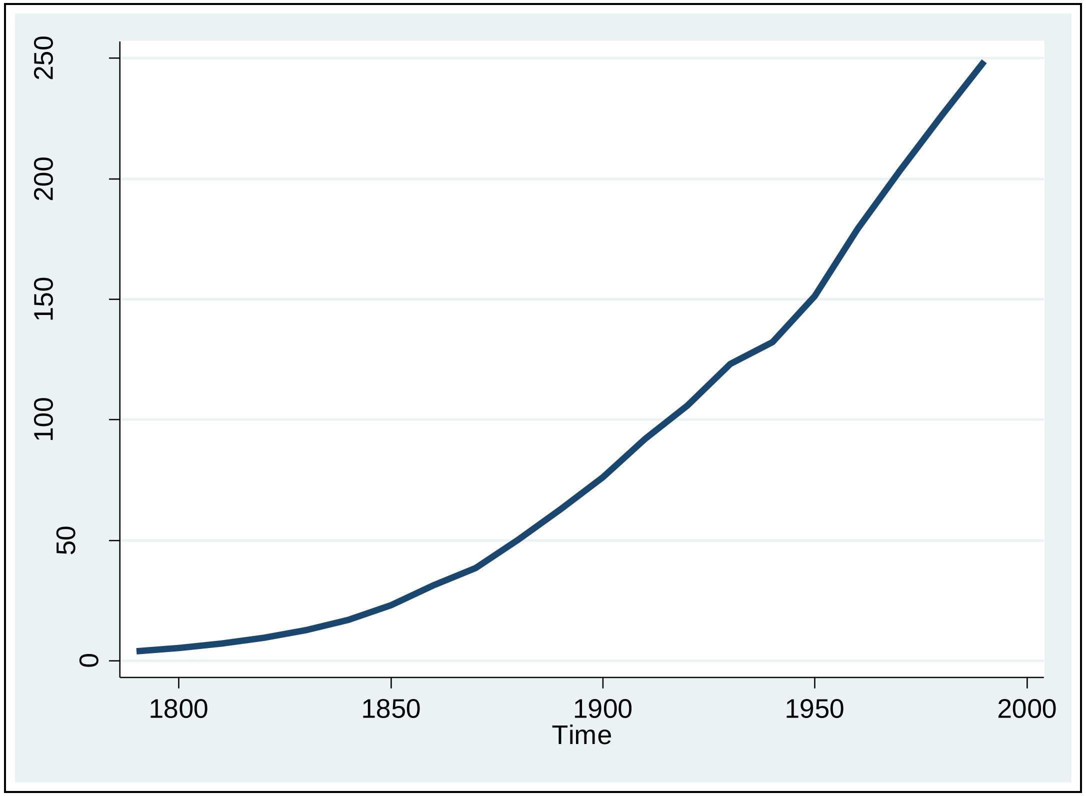
Solution 1.1-1:

```
. describe
[...]
                 storage  display     value
variable name      type   format      label      variable label
-------------------------------------------------------------------------------
uspop            double  %12.0g                 US-Population at ten-year
                                                  intervals, 1790-1990
[...]

. gen time = 1780 + 10*_n

. format time %ty

. tsset time
        time variable:  time, 1790 to 2010, but with gaps

. label variable time "Time"

. tsline uspop, lwidth(thick)
```

We want to start at 1790 and follow a ten-year interval ➔ there will be „gaps“ in the time!

Solution 1.1-2:

```
. tsline uspop, lwidth(thick)
```



Solution 1.1-3: Simple trend models

- Linear Trend Model

  $$L_t = c_1 + c_2 t$$

- Quadratic Trend Model

  $$L_t = c_1 + c_2 t + c_3 t^2$$

- Cubic Trend Model

  $$L_t = c_1 + c_2 t + c_3 t^2 + c_4 t^3$$

- Logarithmic Linear Trend Model (Exponential Growth)

  $$L_t = A \cdot e^{rt} \Rightarrow \ln(L_t) = \ln(A) + rt$$

- Logistic Curve

  $$L_t = \frac{1}{k + ab^t} \qquad b > 0$$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.2:

- Fit a linear trend model to the "USpop.dta" dataset. Calculate c_1 and c_2 and write down the estimated equation.
- Plot the original series and the fitted values. Do you think the linear model is appropriate?
- Predict the residuals. If the linear trend model would be the right one how should the residuals behave? Plot the residuals. What can you conclude?

Note: $L_t = c_1 + c_2 t$

`regress` *`depvar`* `[`*`indepvars`*`]` fits a model of *`depvar`* on *`indepvars`* using linear regression

`predict` *`newvar`* `, xb` predicts the fitted values from the last estimation

`predict` *`newvar`* `, residuals` predicts the residuals from the last estimation

Solution 1.2-1:

```
. regress uspop time

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  1,    19) =  224.33
       Model |  113745.587     1  113745.587           Prob > F      =  0.0000
    Residual |  9634.07504    19  507.056581           R-squared     =  0.9219
-------------+------------------------------           Adj R-squared =  0.9178
       Total |  123379.662    20  6168.9831            Root MSE      =  22.518

------------------------------------------------------------------------------
       uspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   1.215408   .0811489    14.98   0.000     1.045561    1.385254
       _cons |  -2211.338   153.4502   -14.41   0.000    -2532.513   -1890.163
------------------------------------------------------------------------------
```

Estimated equation: $\hat{L}_t = -221134 + 122 \cdot t$

Calculate the fitted values:

```
. gen L_lin = -2211.338+1.215408*time
```

or

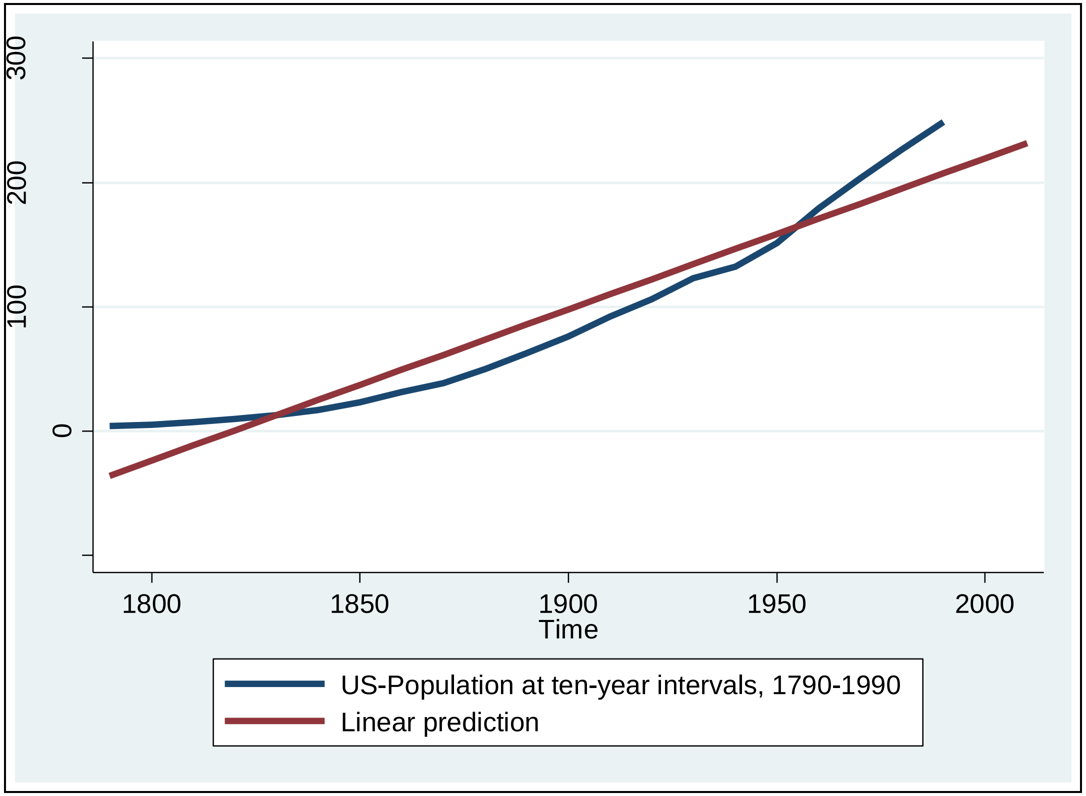
```
. predict L_lin, xb
```

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.2-2:

```
. tsline uspop || tsline L_lin
```

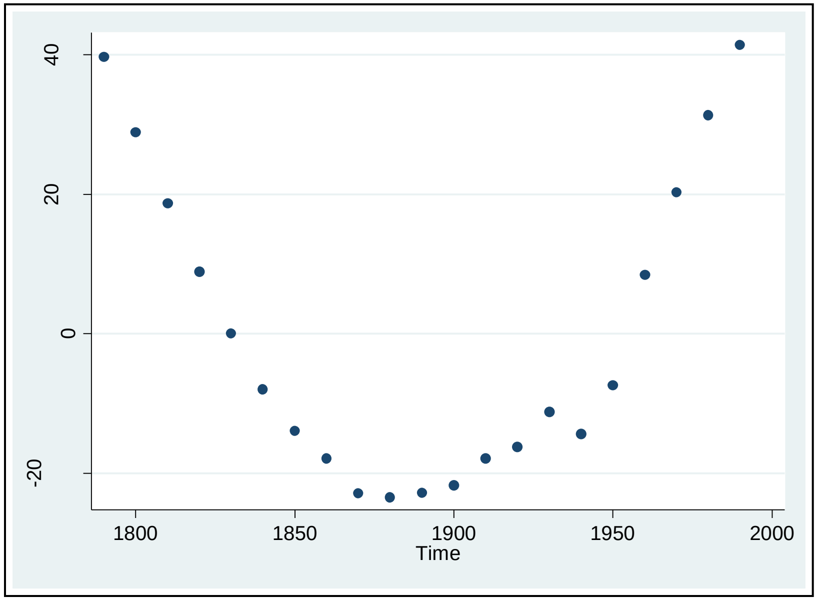


Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.2-3:

Calculate the residuals:

```
. gen res_lin = uspop - L_lin
```
or
```
. predict res_lin, residuals
. twoway scatter res_lin time
```



What do we want?
Residuals should fluctuate randomly around zero!
Here it is not the case ➔ clearly see a U-shape
➔ Indicates that our model is not appropriate

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.3:

- Plot “log(uspop)” against time.
- Fit a logarithmic linear trend model to the “USpop.dta” dataset and write down the estimated equation for L_t.

Note: $L_t = f(t) = A \cdot e^{rt}$

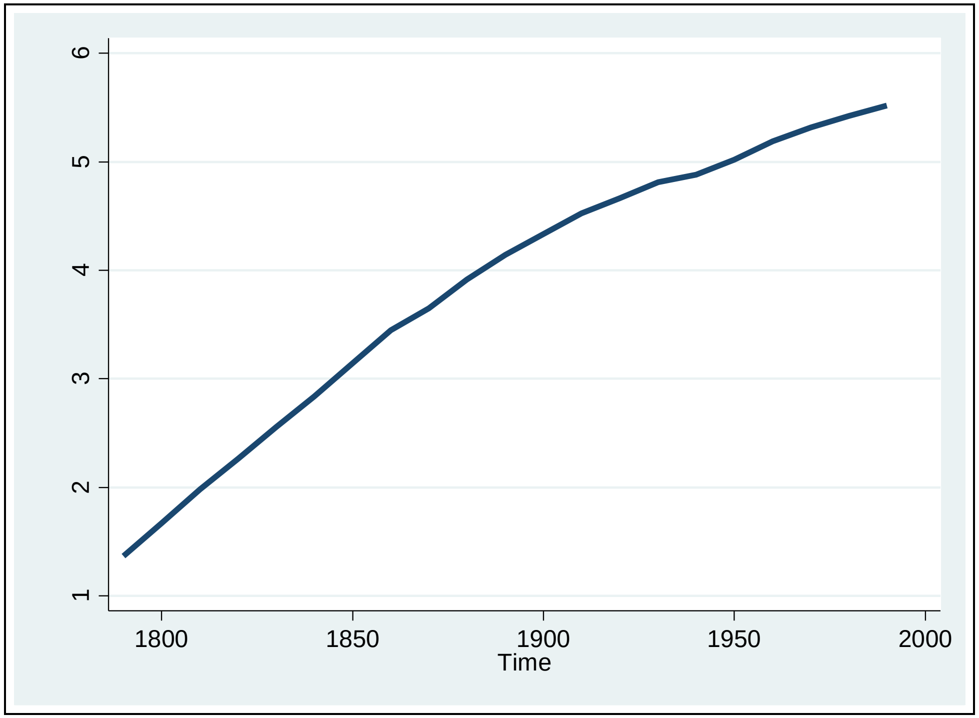
Estimation by taking the logarithms of both sides and fitting the log-linear regression equation:

$\log(L_t) = \log(A) + rt$

- Plot the original series together with the fitted values against time, and the residuals against time. What can you conclude?

Solution 1.3-1:

```
. gen luspop = log(uspop)
. tsline luspop
```



Solution 1.3-2:

```
. regress luspop time

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  1,    19) =  570.37
       Model |  33.2140643     1  33.2140643           Prob > F      =  0.0000
    Residual |  1.10641977    19  .058232619           R-squared     =  0.9678
-------------+------------------------------           Adj R-squared =  0.9661
       Total |  34.3204841    20   1.7160242           Root MSE      =  .24131

------------------------------------------------------------------------------
      luspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    .020769   .0008696    23.88   0.000     .0189488    .0225892
       _cons |  -35.41252   1.644456   -21.53   0.000    -38.85441   -31.97063
------------------------------------------------------------------------------
```

Estimated equation:

$$\log\left(\hat{L}_t\right) = \log\left(\hat{A}\right) + \hat{r}t = -35.413 + .021 \cdot t$$

$$\hat{L}_t = \hat{A} \cdot e^{\hat{r}t} = e^{-35.413} \cdot e^{.021 \cdot t}$$

Solution 1.3-3:

Estimated equation:

$$\log\left(\hat{L}_t\right) = \log\left(\hat{A}\right) + \hat{r}t = -35.413 + .021 \cdot t$$

$$\hat{L}_t = \hat{A} \cdot e^{\hat{r}t} = e^{-35.413} \cdot e^{.021 \cdot t}$$

Prediction of the fitted values of “luspop”:

```
. predict lL_exp, xb
```

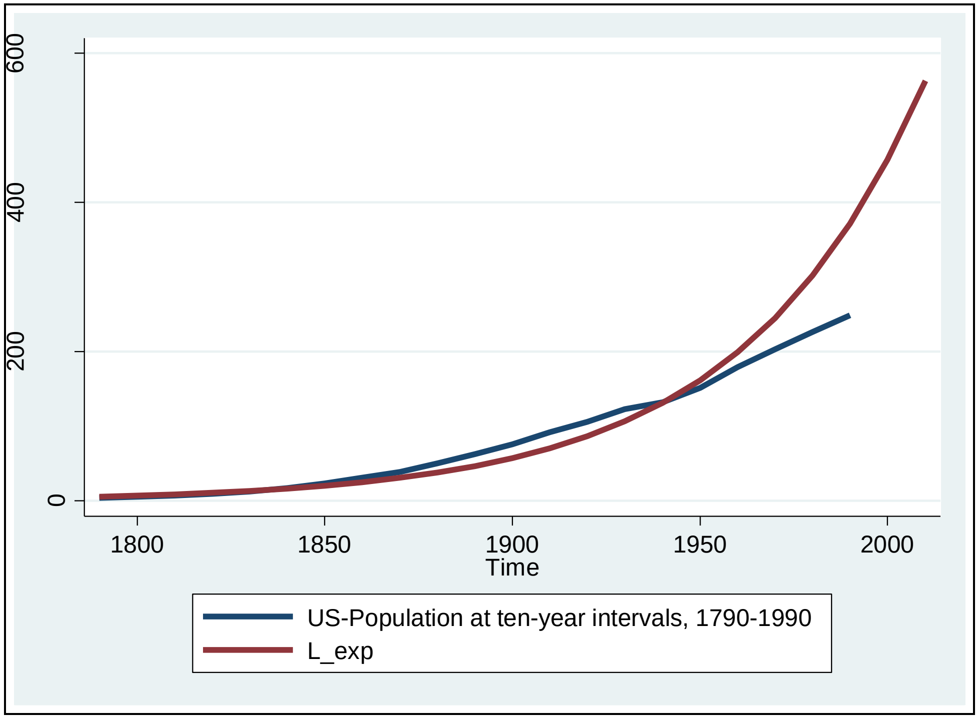
Prediction of the fitted values of “uspop”:

```
. gen L_exp = exp(_b[_cons])*exp(_b[time]*time)
```
or
```
. gen L_exp = exp(lL_exp)
```

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

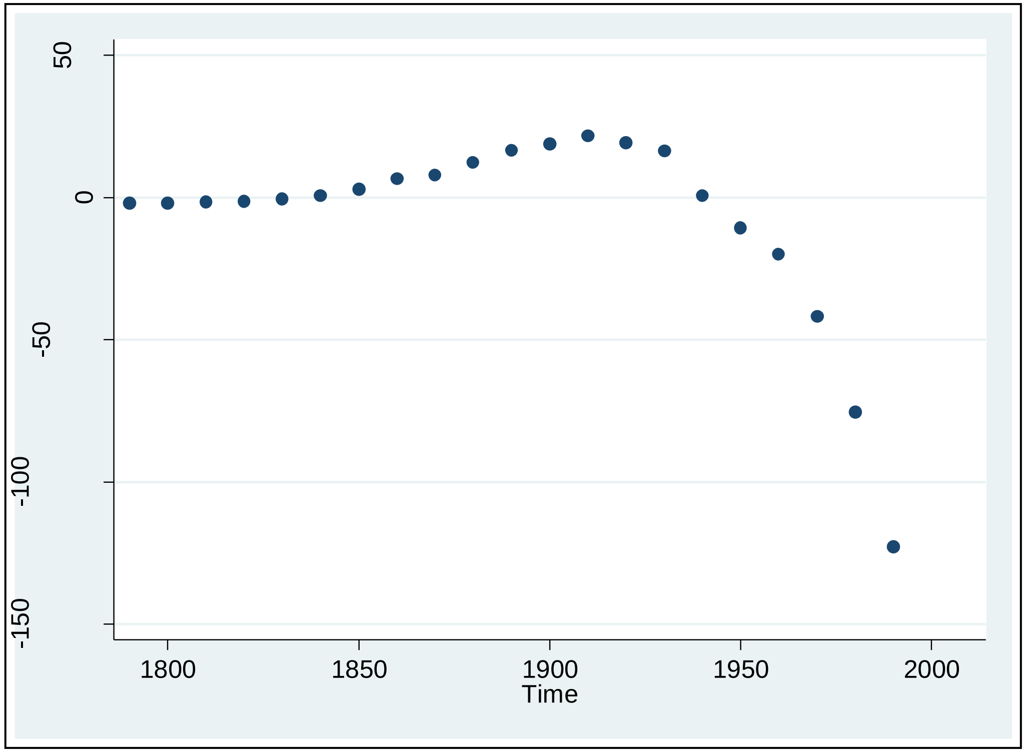
Solution 1.3-4:

```
. gen L_exp = exp(_b[_cons])*exp(_b[time]*time)
. tsline uspop L_exp
```



Solution 1.3-5:

```
. generate res_exp = uspop-L_exp
. twoway scatter res_exp time
```



What do we want?
Residuals should fluctuate randomly around zero!
Here it is not the case ➔ clearly see a systematic pattern
➔ Indicates that our model is not appropriate

Exercise 1.4:

- Calculate the US population growth rate between 1790 and 1800.
- Calculate the same growth rate using the log operator.

$$\log(x_1) - \log(x_0) \approx \frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0} \quad x_1, x_0 > 0 \quad \text{for small changes in } x$$

Note: With the command `display` you can use Stata as a substitute for a hand calculator.
Example: `display uspop[1]` displays the first observation of the variable “uspop”

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.4:

```
. di (uspop[2]-uspop[1])/uspop[1]
.35102924

. di ln(uspop[2])-ln(uspop[1])
.3008667
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.5:

- Split the time series into two appropriate periods and fit a logarithmic linear trend model to each of them.
- Plot the original time series together with the fitted values. What can you conclude?
- If necessary try to split into two different periods.

Solution 1.5-1:

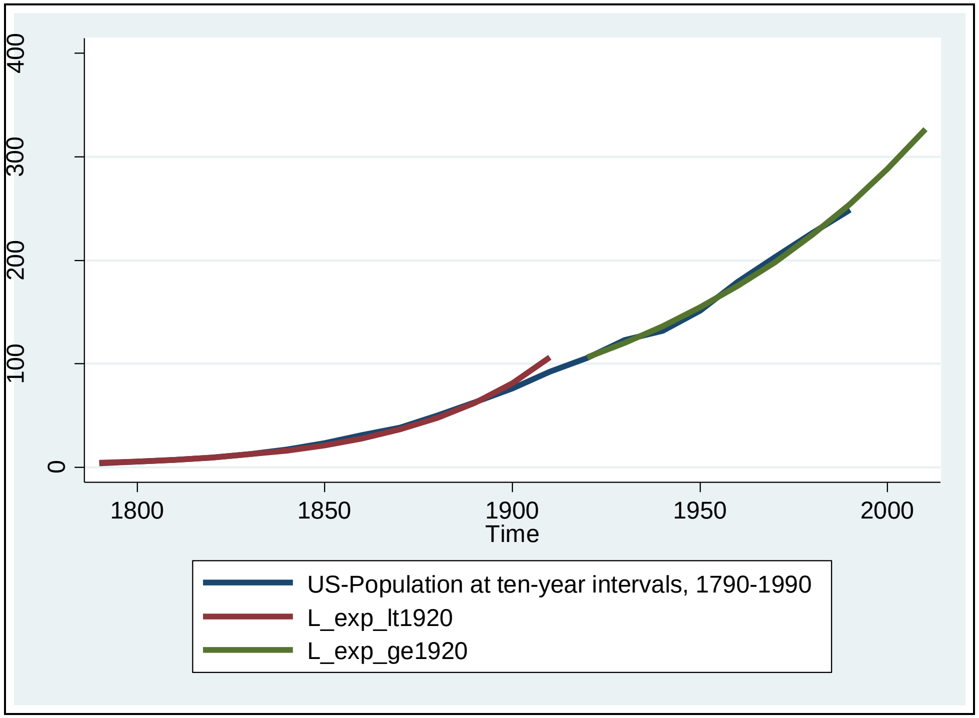
```
. regress luspop time if time < 1920
[...]
      luspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0267397   .0005692    46.98   0.000      .025487    .0279924
       _cons |  -46.40423   1.053148   -44.06   0.000    -48.72219   -44.08627
------------------------------------------------------------------------------

. predict lL_exp_lt1920, xb

. regress luspop time if time >= 1920
[...]
      luspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .0124711   .0003865    32.26   0.000     .0115253    .0134169
       _cons |  -19.27795   .7557339   -25.51   0.000    -21.12716   -17.42873
------------------------------------------------------------------------------

. predict lL_exp_ge1920, xb

. gen L_exp_lt1920 = exp(ly_exp_lt1920)

. gen L_exp_ge1920 = exp(ly_exp_ge1920)
```

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.5-2:

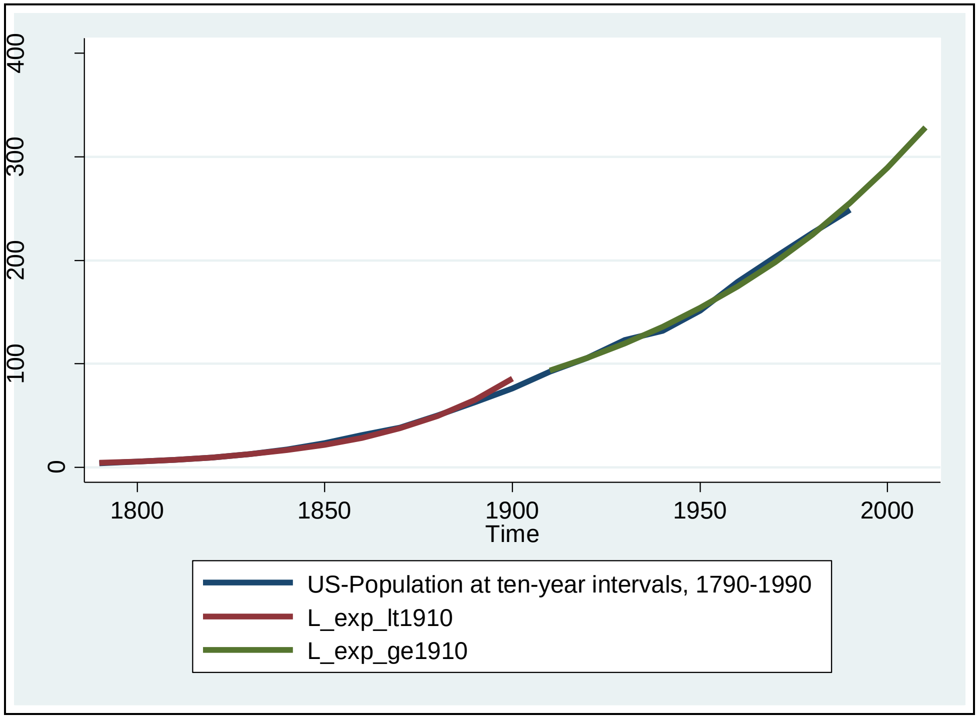
```
. tsline uspop || tsline L_exp_lt1920 if time < 1920 || tsline
L_exp_ge1920 if time >= 1920
```



Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.5-3:

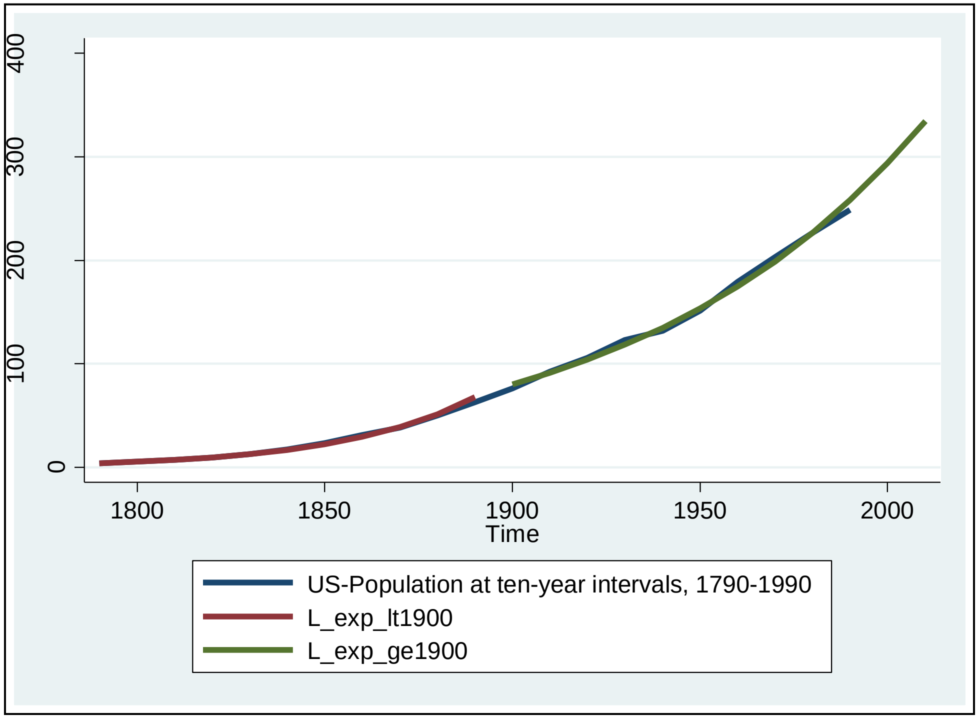
```
. tsline uspop || tsline L_exp_lt1910 if time < 1910 || tsline
L_exp_ge1910 if time >= 1910
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.5-4:

```
. tsline uspop || tsline L_exp_lt1900 if time < 1900 || tsline
L_exp_ge1900 if time >= 1900
```



Exercise 1.6:

- Fit a quadratic trend model to the “USpop.dta” dataset. Write down the estimated equation.

Note: $L_t = c_1 + c_2 t + c_3 t^2$

- Plot the residuals. What can you conclude?

- Fit a cubic trend model to the “USpop.dta” dataset. Write down the estimated equation.

Note: $L_t = c_1 + c_2 t + c_3 t^2 + c_4 t^3$

- Plot the residuals. What can you conclude?

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.6-1:

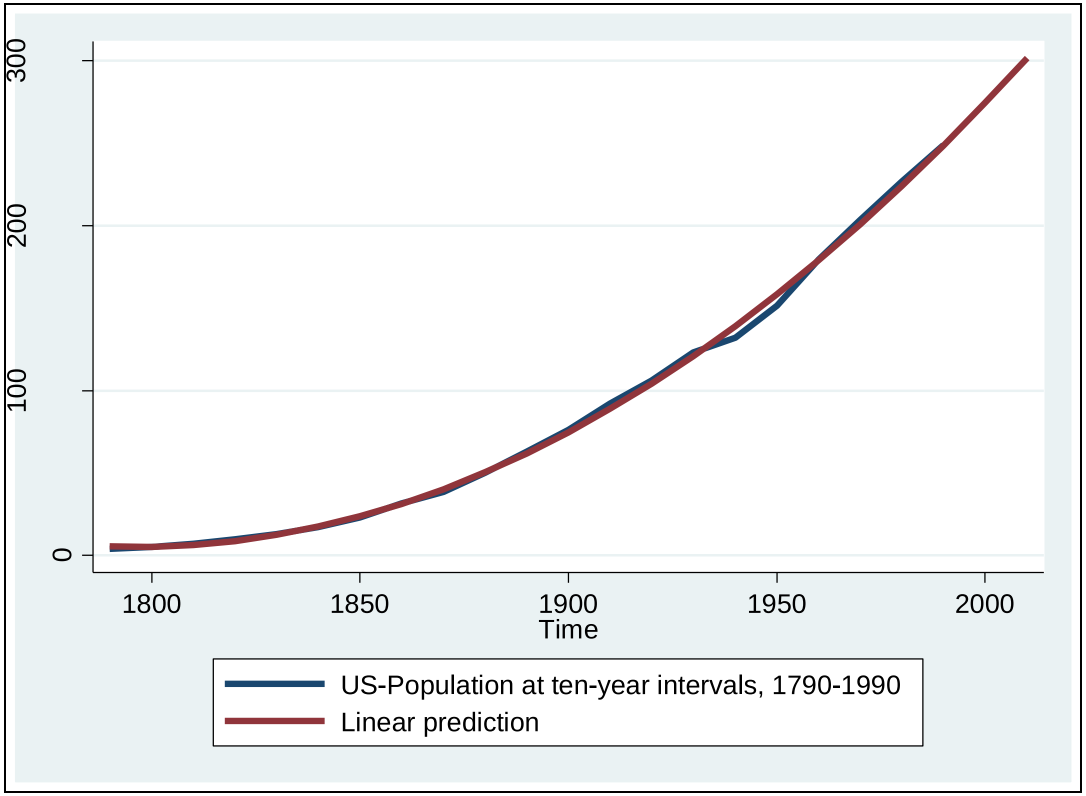
```
. gen time2 = time^2

. regress uspop time time2

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  2,    18) = 8050.39
       Model |  123241.883     2  61620.9414           Prob > F      =  0.0000
    Residual |  137.779342    18  7.65440787           R-squared     =  0.9989
-------------+------------------------------           Adj R-squared =  0.9988
       Total |  123379.662    20  6168.9831            Root MSE      =  2.7667

------------------------------------------------------------------------------
       uspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |  -23.37855   .6983151   -33.48   0.000    -24.84566   -21.91145
       time2 |   .0065063   .0001847    35.22   0.000     .0061183    .0068944
       _cons |    21006.1   659.4327    31.85   0.000     19620.68    22391.51
------------------------------------------------------------------------------

. predict L_qua, xb

. predict res_qua, residuals
```

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.6-2:

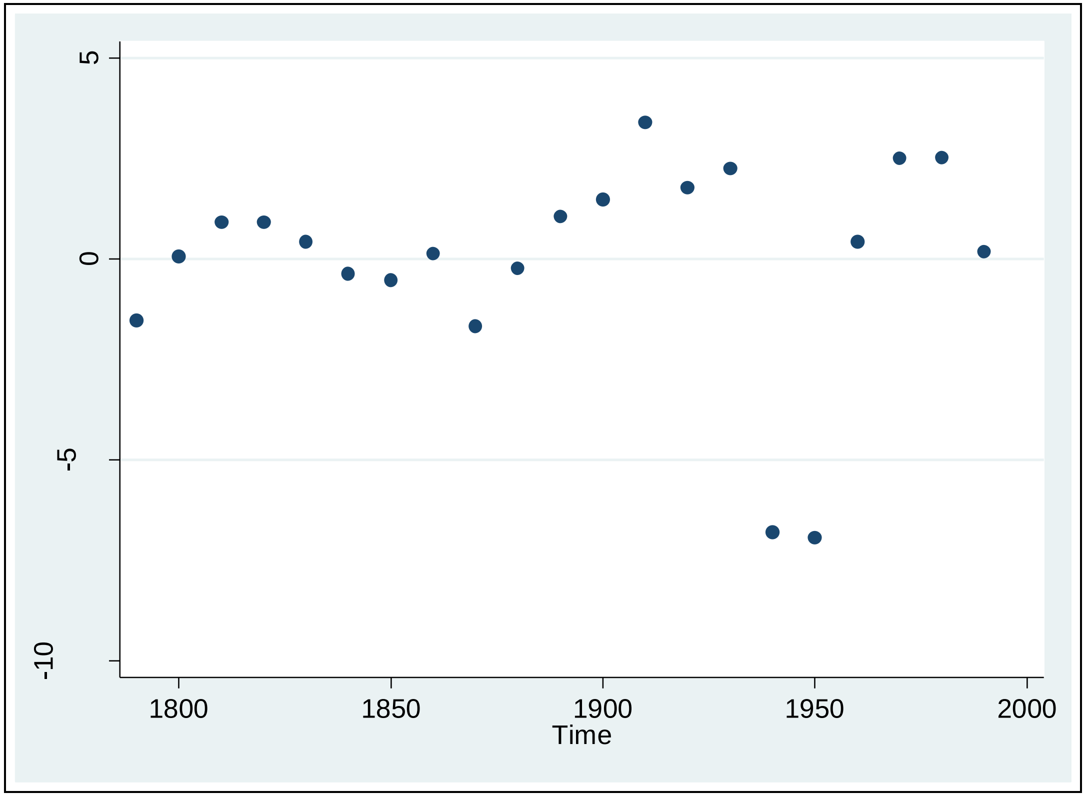
```
. tsline uspop L_qua
```



Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.6-3:

```
. scatter res_qua time
```



Franziska Plitzko

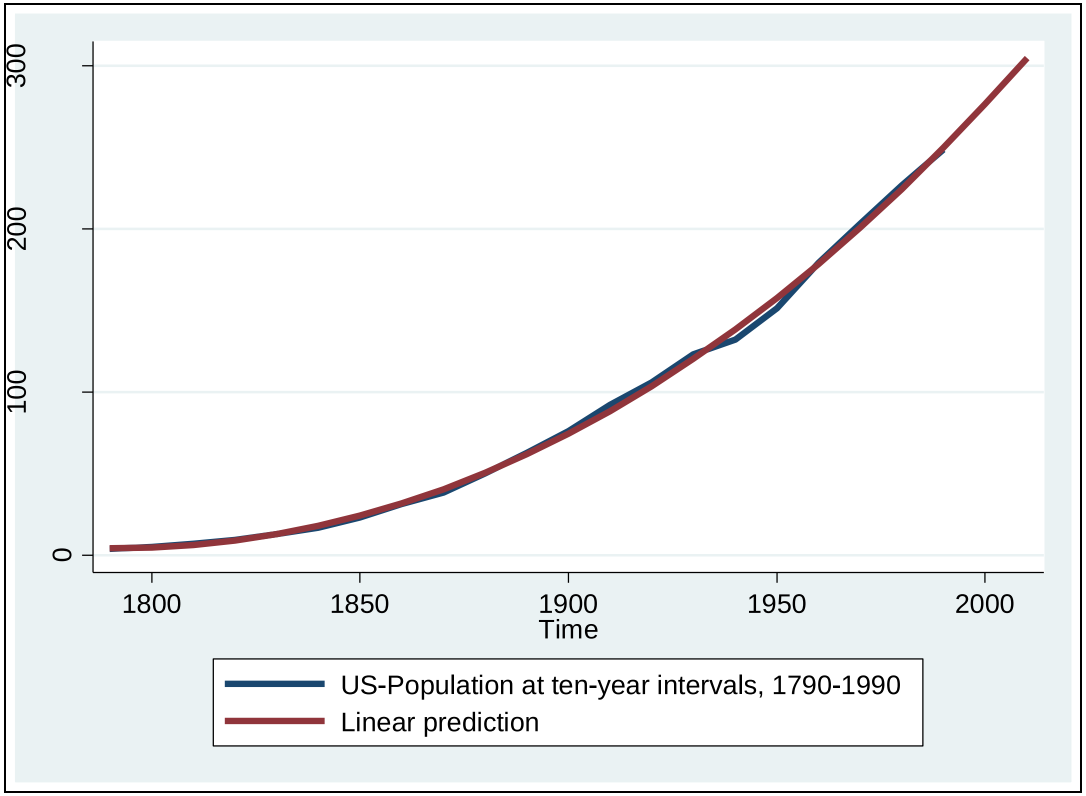
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.6-4:

```
. gen time3 =time^3

. regress uspop time time2 time3

      Source |       SS       df       MS              Number of obs =      21
-------------+------------------------------           F(  3,    17) = 5319.49
       Model |  123248.37     3  41082.79              Prob > F      =  0.0000
    Residual | 131.292208    17  7.72307107            R-squared     =  0.9989
-------------+------------------------------           Adj R-squared =  0.9987
       Total | 123379.662    20  6168.9831             Root MSE      =   2.779

------------------------------------------------------------------------------
       uspop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   11.18374    37.7178     0.30   0.770    -68.39385    90.76134
       time2 |  -.0117918   .0199662    -0.59   0.563    -.0539167    .0303331
       time3 |   3.23e-06   3.52e-06     0.92   0.372    -4.20e-06    .0000107
       _cons |  -741.4073   23738.17    -0.03   0.975    -50824.56    49341.75
------------------------------------------------------------------------------

. predict L_cub, xb

. predict res_cub, residuals
```

Solution 1.6-5:

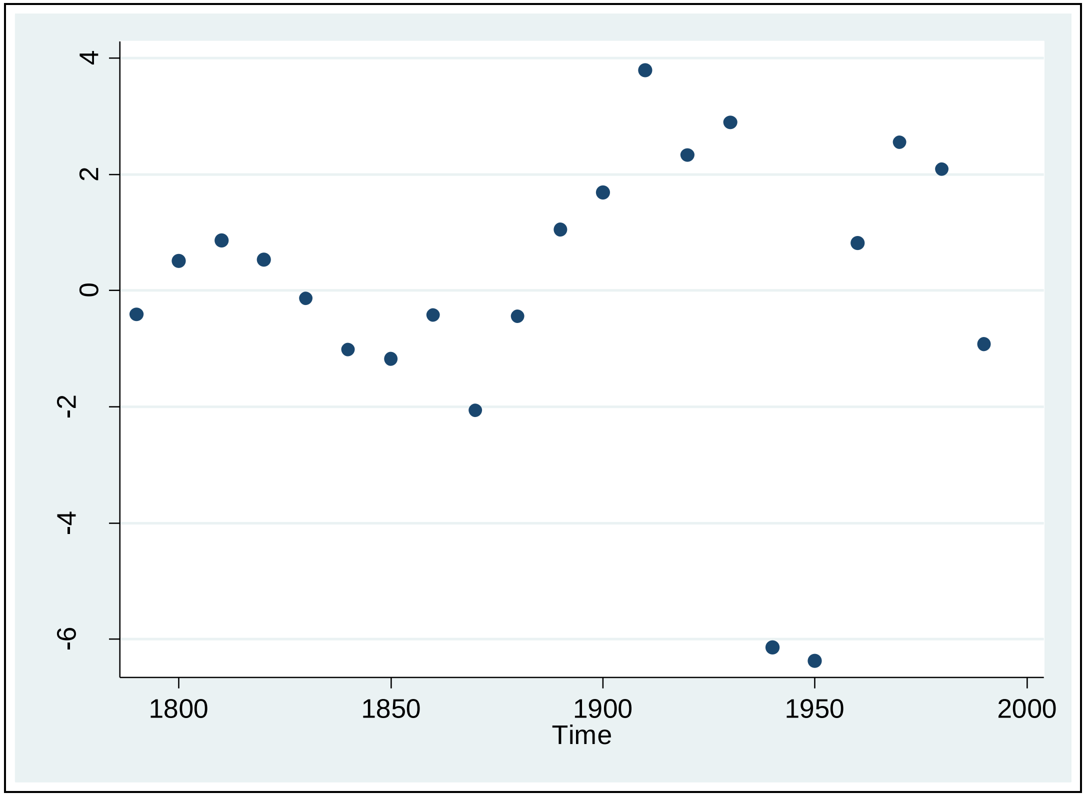
```
. tsline uspop L_cub
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.6-6:

```
. scatter res_cub time
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.7:

- Compare the (at least) five models you have fitted to the dataset. Which of these models fits best to the series?
- Calculate for each model the forecast for the year 2000 and compare them with the US population in 2000 (281.55 million) and in 2010 (310.3 million).

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.7-1:

Calculate for each model the forecast for the year 2000 and compare them with the US population in 2000 (281.55 million) and in 2010 (310.3 million).

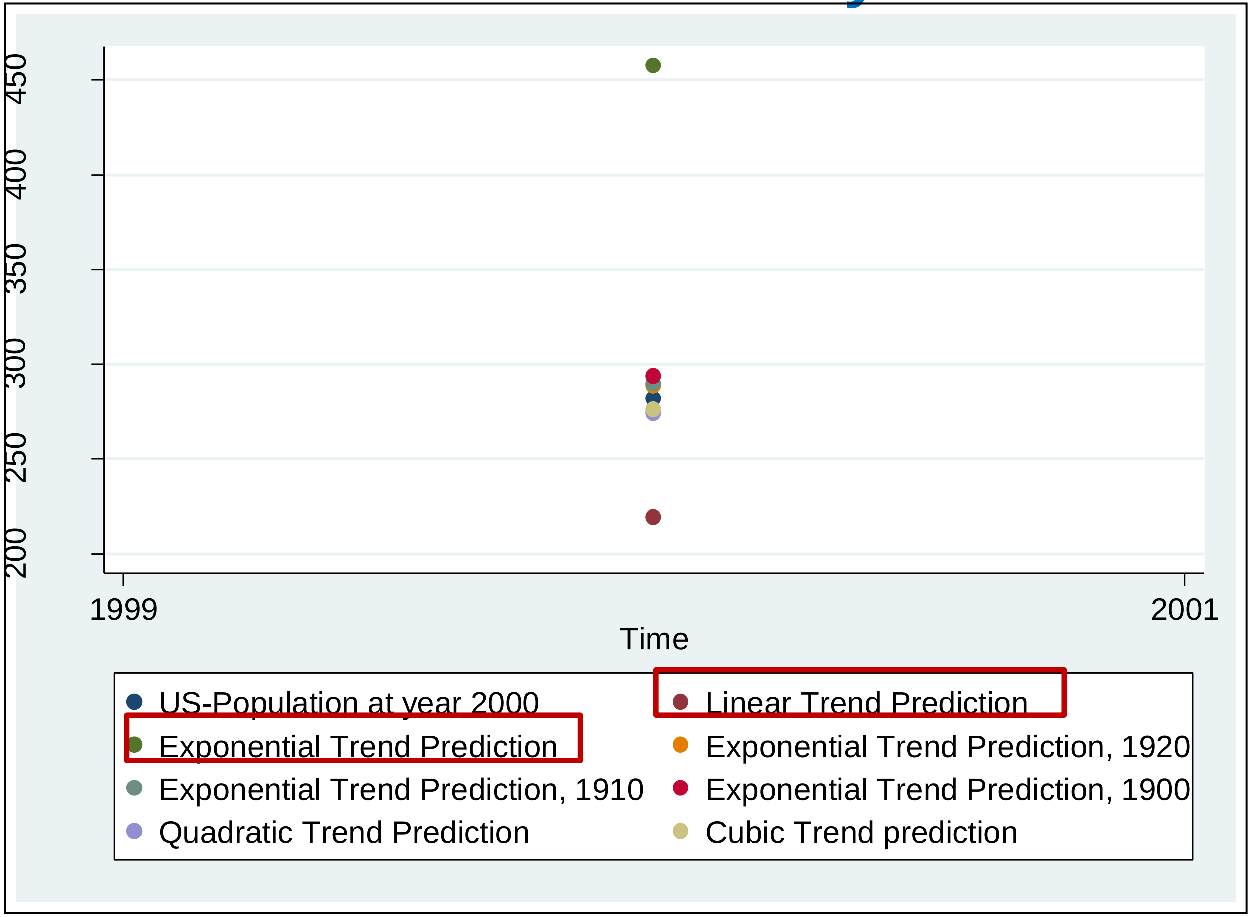
. list time L_lin L_exp L_exp_ge1920 L_exp_ge1910 L_exp_ge1900 L_qua L_cub if time >= 2000

```
     +---------------------------------------------------------------------------------------+
     | time      L_lin       L_exp    L_~e1920    L_~e1910    L_~e1900       L_qua       L_cub |
     |---------------------------------------------------------------------------------------|
 22. | 2000   219.4774    457.3676    288.3588    289.4849    293.9057    274.3476    276.3069 |
 23. | 2010   231.6315    562.9426    326.6588    328.3181    334.7128    301.4662    304.4946 |
     +---------------------------------------------------------------------------------------+
```

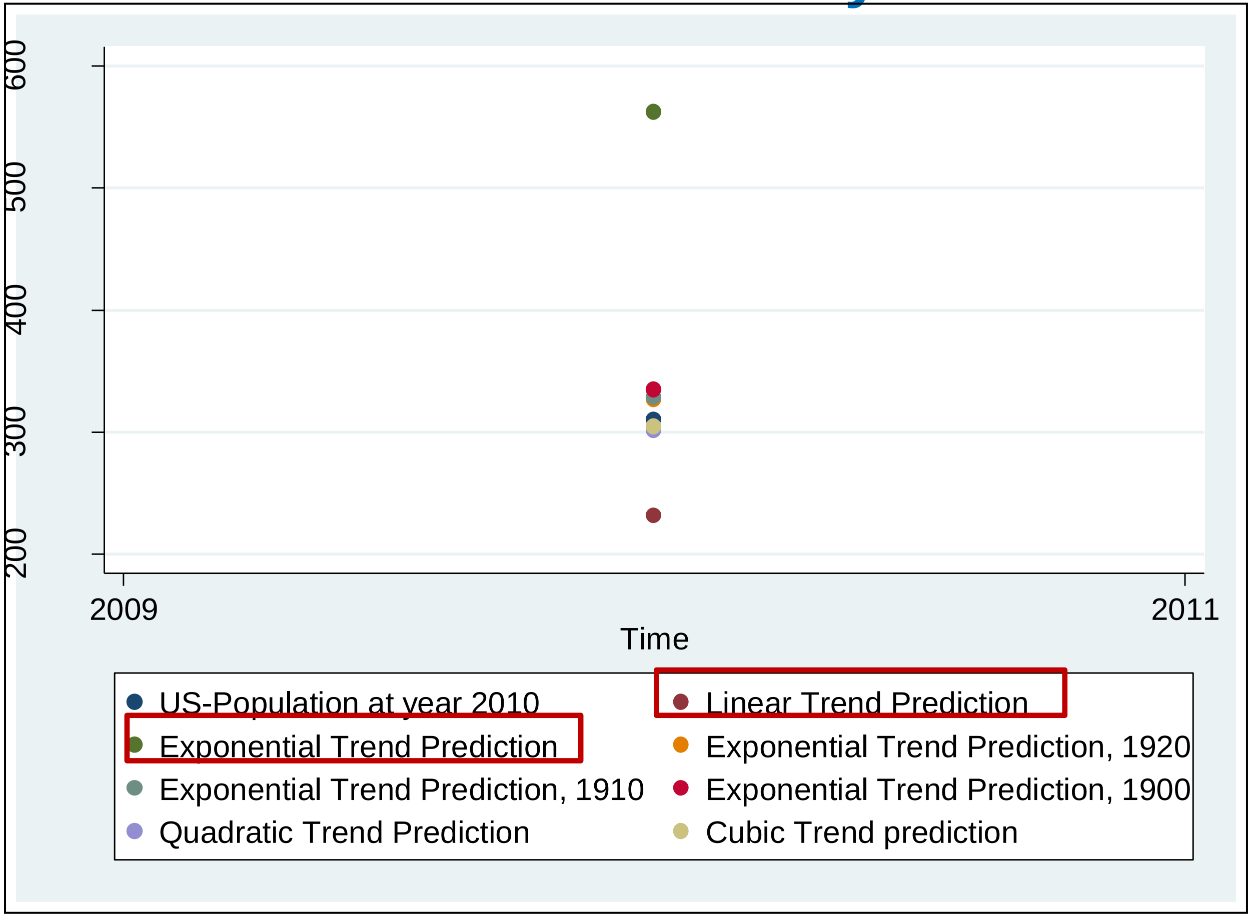
. list time p_lin p_exp p_exp_ge1920 p_exp_ge1910 p_exp_ge1900 p_qua p_cub if time >= 2000 ("forecast error")

```
     +------------------------------------------------------------------------------------------+
     | time       p_lin       p_exp    p_e~1920    p_e~1910    p_e~1900       p_qua       p_cub |
     |------------------------------------------------------------------------------------------|
 22. | 2000   -.2204672    .6244633    .0241833     .028183    .0438846   -.0255814   -.0186224 |
 23. | 2010   -.2535239    .8141881    .0527193    .0580668    .0786749   -.0284685    -.018709 |
     +------------------------------------------------------------------------------------------+
```

Solution 1.7-2: Forecast at year 2000



Solution 1.7-3: Forecast at year 2010



Organizational Issues

Please log in to **ISIS** (password: Zeit1718) and **download** the following files:

- oil.dta
- Siemens.dta
- prod.dta

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Deterministic Models

- Components of a Time Series
- Additive and Multiplicative Models
- Simple Trend Models
- **Smoothing Techniques**
- Seasonal Adjustment

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Smoothing Techniques...

- Simple Moving Average
- Centered Moving Average
- Exponentially Weighted Moving Average
- Holt-Winter's two parameter exponential smoothing
- Hodrick-Prescott filter

... are useful for

- filtering/smoothing
- forecasting

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Crude oil import price index

Jan. 2000 - Sep. 2008

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.8:

Calculate the exponentially weighted moving average for the first four observations of the oil variable (import price index 2000 = 100). Use $\tilde{y}_0 = 82.6$ as the initial value and $\alpha = .3$ as the smoothing parameter.

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

Data:

time	oil
Jan. 2000	82.6
Feb. 2000	91.2
Mar. 2000	94.3
Apr. 2000	82.9
May 2000	94.9
Jun. 2000	100.7
...	...

Statistisches Bundesamt (2008) “Daten zur Energiepreisentwicklung”

Solution 1.8:

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

$$\alpha = 0.3 \quad \tilde{y}_0 = 82.6$$

$$\tilde{y}_1 = 0.3 \cdot 82.6 + 0.7 \cdot 82.6 = 82.6$$

$$\tilde{y}_2 = 0.3 \cdot 91.2 + 0.7 \cdot 82.6 = 85.18$$

$$\tilde{y}_3 = 0.3 \cdot 94.3 + 0.7 \cdot 85.18 = 87.916$$

$$\tilde{y}_4 = 0.3 \cdot 82.9 + 0.7 \cdot 87.916 = 86.4112$$

time	oil
Jan. 2000	82.6
Feb. 2000	91.2
Mar. 2000	94.3
Apr. 2000	82.9
May 2000	94.9
Jun. 2000	100.7
...	...

Exercise 1.9:

- Upload the “oil.dta” dataset.
- Generate the appropriate `time` variable (labeled “Time”).
- Plot the series.

Solution 1.9:

Generate the time variable:

```
. describe
 obs:            105
 vars:             1                          4 Nov 2008 09:05
 size:           840 (99.9% of memory free)
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
oil             float  %9.0g                  Crude oil import price index, Jan.
                                                     2000 - Sep. 2008
--------------------------------------------------------------------------------
. di 12*40-1
479
. generate time = 479+_n
. format time %tmm_Y
```

Declare data to be time series data:

```
. tsset time
```

Plot the series:

```
. tsline oil
```

Alternative:

```
generate time2=tm(2000m1)+_n-1
format time2 %tmm_Y
tsset time2
. list oil time time2 in 1/1

     +-------------------------+
     |  oil      time     time2 |
     |-------------------------|
  1. | 82.6    Jan 00    Jan 00 |
     +-------------------------+
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Crude oil import price index

Jan. 2000 - Sep. 2008

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.10:

- Calculate the exponentially weighted moving average. Use $\tilde{y}_0 = 82.6$ as the initial value and $\alpha = .3$ as the smoothing parameter.

Stata commands:

Macro definition

```
. local lclname = exp
```

Example:

```
. local num = 5
. di `num'
5
```

Replace contents of existing variable

```
replace oldvar = exp
```

Loop over consecutive values

```
forvalues lname = range {
  commands referring to `lname'
}
```

where *range* is #1/#2 meaning #1 to #2 in steps of 1

Solution 1.10-1:

Generate a new variable for the smoothed values:

```
. gen oil_exp3 = oil in 1
```

Or

```
. gen oil_exp3 = oil
```

Generate a local for the value of *α*

```
. local alpha = 0.3
```

Replace the values of `oil_exp3` using the exponential smoothing formula

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

```
. forvalues num = 2/105 {
. replace oil_exp3=`alpha'*oil+(1-`alpha')*oil_exp3[`num'-1] in `num'
. }
```

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.10-2:

$$\tilde{y}_1 = 0.3 \cdot 82.6 + 0.7 \cdot 82.6 = 82.6$$

$$\tilde{y}_2 = 0.3 \cdot 91.2 + 0.7 \cdot 82.6 = 85.18$$

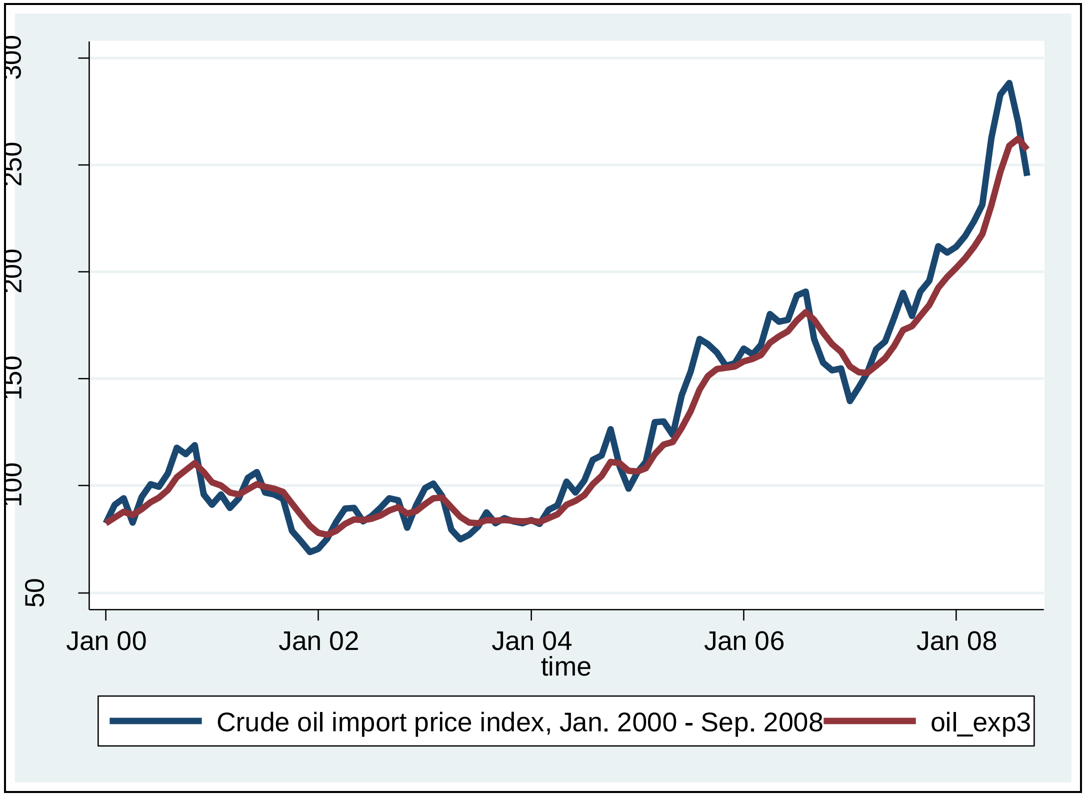
$$\tilde{y}_3 = 0.3 \cdot 94.3 + 0.7 \cdot 85.18 = 87.916$$

$$\tilde{y}_4 = 0.3 \cdot 82.9 + 0.7 \cdot 87.916 = 86.4112$$

```
. list time oil oil_exp3 in 1/4

     +---------------------------+
     |   time    oil    oil_exp3 |
     |---------------------------|
  1. | Jan 00   82.6        82.6 |
  2. | Feb 00   91.2       85.18 |
  3. | Mar 00   94.3      87.916 |
  4. | Apr 00   82.9     86.4112 |
     +---------------------------+
```

Solution 1.10-3:

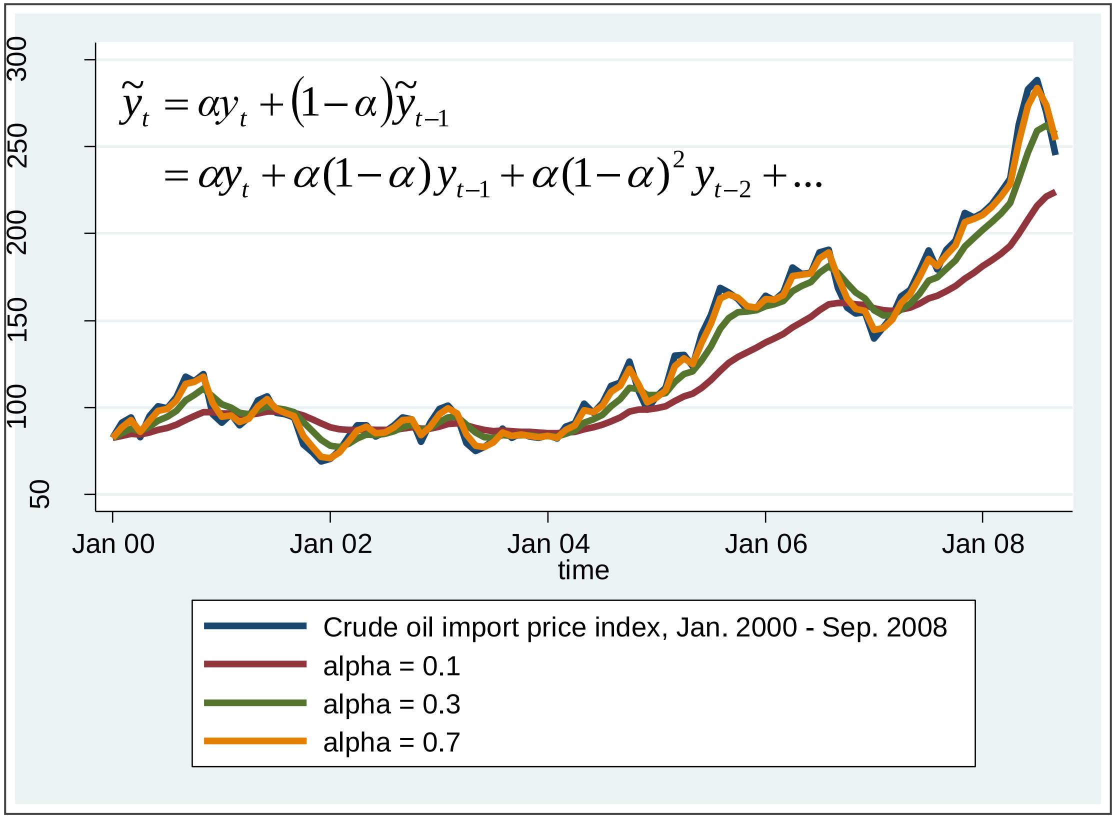
```
. tsline oil oil_exp3
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.10-4:

```
. tsline oil oil_exp1 oil_exp3 oil_exp7
```



Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.11:

- Calculate the exponentially weighted moving average. Use $\tilde{y}_0 = 82.6$ as the initial value and $\alpha = .3$ as the smoothing parameter.

Stata command:

Exponential smoothing

`tssmooth exponential`

`parms(#a)`	specifies the parameter alpha for the exponential smoother; 0 < #a < 1
`s0(#)`	specifies the initial value to be used

Solution 1.11:

Stata's command for exponential smoothing (forecasting version)

$\alpha = 0.3 \quad \tilde{y}_0 = 82.6$

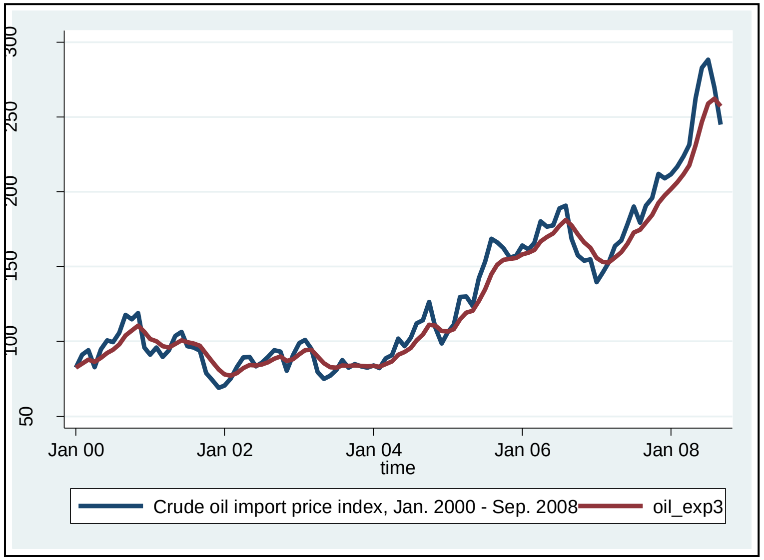
```
. tssmooth exponential oil_exp3_stata = oil, parms(.3) s0(82.6) replace
exponential coefficient  =       0.3000
sum-of-squared residuals =        22922
root mean squared error  =       14.775

. list time oil oil_exp3 oil_exp3_stata in 1/4
     +--------------------------------------+
     |   time      oil   oil_exp3   oil_ex~a |
     |--------------------------------------|
  0. |                       82.6            |
  1. | Jan 00     82.6       82.6       82.6 |
  2. | Feb 00     91.2      85.18       82.6 |
  3. | Mar 00     94.3     87.916      85.18 |
  4. | Apr 00     82.9    86.4112     87.916 |
```

For your information:
We calculate by hand using the initial value at t=0! STATA uses the initial value at t=1. STATA does not calculate $\tilde{y}_1$, it sets the initial value at this index.

Not in our data, but used by us!

Exercise 1.12:

- True or false and explain: The above EWMA model is appropriate for forecasting the oil price.
- For large values of α, will the response to changes in the mean of the unfiltered series be slow or fast?



Solution 1.12:

- If the time series has an **upward (downward) trend**, the EWMA model will **underpredict (overpredict)** future values of y_t.

  → remove any trend from the data before using EWMA

  → the trend term can be added to the untrended initial forecast to obtain the final forecast

- "The choice of the smoothing constant α determines how quickly the smoothed series or forecast will adjust to changes in the mean of the unfiltered series. For small values of α, the response will be slow because more weight is placed on the previous estimate of the mean of the unfiltered series, whereas larger values of α will put more emphasis on the most recently observed value of the unfiltered series."

Stata Time Series Preference Manuel, Release 11, p. 336

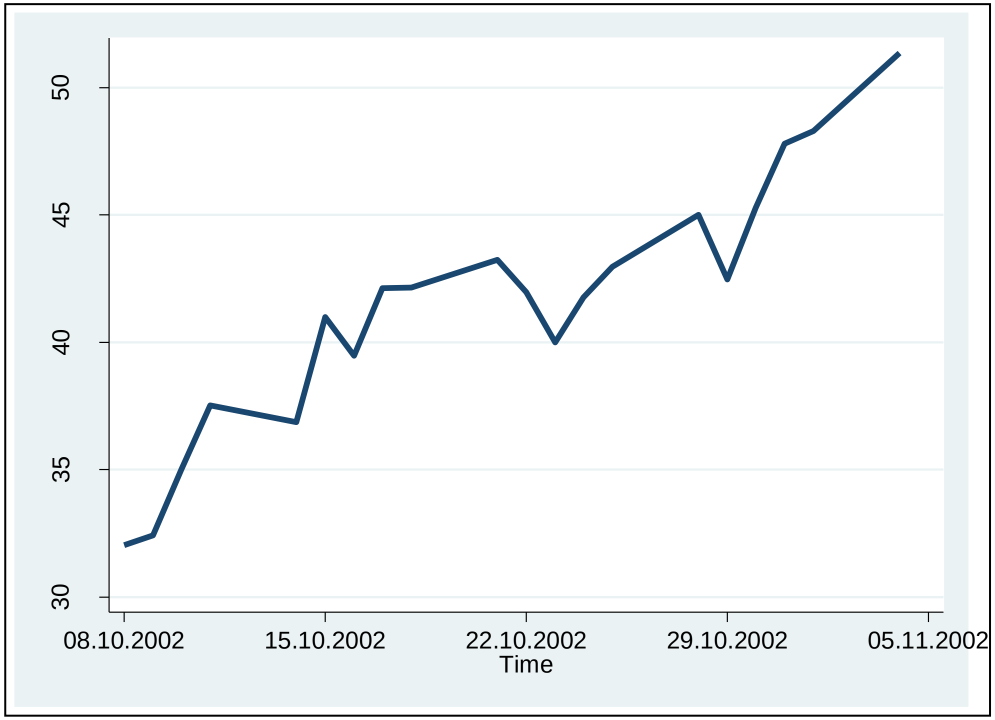
Exercise 1.13:

- Upload the “Siemens.dta” dataset.
- Plot the series.
- Calculate the EWMA. Use the first observation as the initial value and α=.3 as the smoothing parameter.
- Plot the original series and the smoothed series.

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.13-1:

```
. tsset time

. tsline SIE
```



Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.13-2:

Generate a new variable for the smoothed values:

```
. gen SIE_exp3 = SIE
```

Generate a local for the value of α

```
. local alpha = 0.3
```

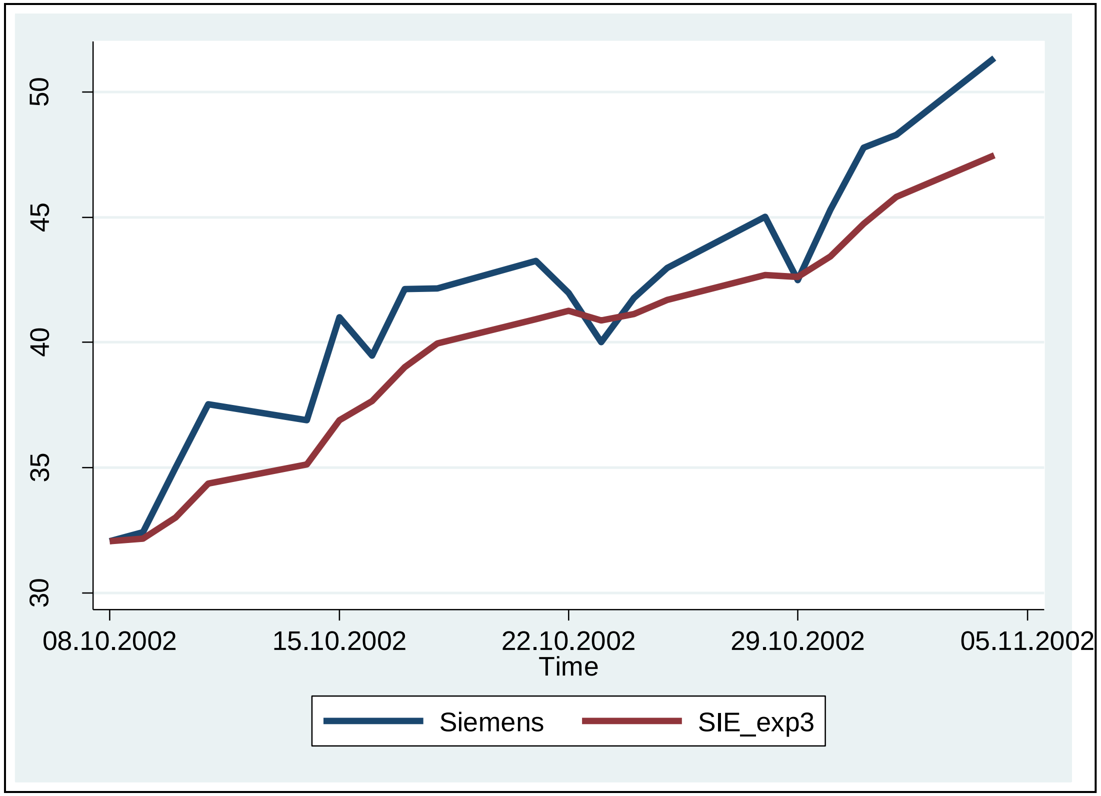
Replace the values of SIE_exp3 using the exponential smoothing formula

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

```
. forvalues num = 2/20 {
. replace SIE_exp3 = `alpha'*SIE +(1-`alpha')*SIE_exp3[`num'-1] in `num'
. }
```

Solution 1.13-3:

```
. tsline SIE SIE_exp3
```



Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercises (Siemens.dta)

```
. tssmooth exponential SIE_exp3_stata=SIE, parms(.3) s0(32.05) replace

. list time SIE SIE_exp3 SIE_exp3_stata in 1/10
```

	time	SIE	SIE_exp3	SIE_ex~a
1.	08.10.2002	32.05	32.05	32.05
2.	09.10.2002	32.42	32.161	32.05
3.	10.10.2002	35	33.0127	32.161
4.	11.10.2002	37.53	34.36789	33.0127
5.	12.10.2002	.	.	**34.36789**
6.	13.10.2002	.	.	**34.36789**
7.	14.10.2002	36.88	35.12152	**34.36789**
8.	15.10.2002	41	36.88506	35.12152
9.	16.10.2002	39.48	37.66354	36.88506
10.	17.10.2002	42.13	39.00348	37.66354

Careful: The data consists of daily data without weekends. The `tssmooth` command will create weekend dates and include missings as values for SIE. The calculations for 1.14-1.16 are done before this `tssmooth` command to avoid those missings in those calculations.

Because the data for 12.10.2002 and 13.10.2002 (Weekend) are missing, the estimations are calculated in STATA as followed:

$$\tilde{y}_{13.10.2002} = \alpha \cdot \tilde{y}_{12.10.2002} + (1-\alpha)\cdot \tilde{y}_{12.10.2002} = \tilde{y}_{12.10.2002}$$

$$\tilde{y}_{14.10.2002} = \alpha \cdot \tilde{y}_{13.10.2002} + (1-\alpha)\cdot \tilde{y}_{13.10.2002} = \tilde{y}_{13.10.2002}$$

Because this is a single-exponential procedure, the loss will not be noticed several periods later.

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.14:

- Detrend the original series (assume a linear trend). $L_t = c_1 + c_2 t$
- Smooth the detrended series using EWMA. Use the first observed value as the initial value and $\alpha = .3$ as the smoothing parameter.
- Add back the trend.
- Plot the result together with the original series and the smoothed original series.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.14-1:

Detrend the original series (assume a linear trend): $L_t = c_1 + c_2 t$

```
. regress SIE time
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  1,    18) =  108.76
       Model |  407.151469     1  407.151469           Prob > F      =  0.0000
    Residual |  67.3841375    18  3.74356319           R-squared     =  0.8580
-------------+------------------------------           Adj R-squared =  0.8501
       Total |  474.535606    19  24.9755582           Root MSE      =  1.9348

------------------------------------------------------------------------------
         SIE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   .5575393   .0534614    10.43   0.000     .4452212    .6698575
       _cons |  -8675.072   835.8098   -10.38   0.000    -10431.04     -6919.1
------------------------------------------------------------------------------
```

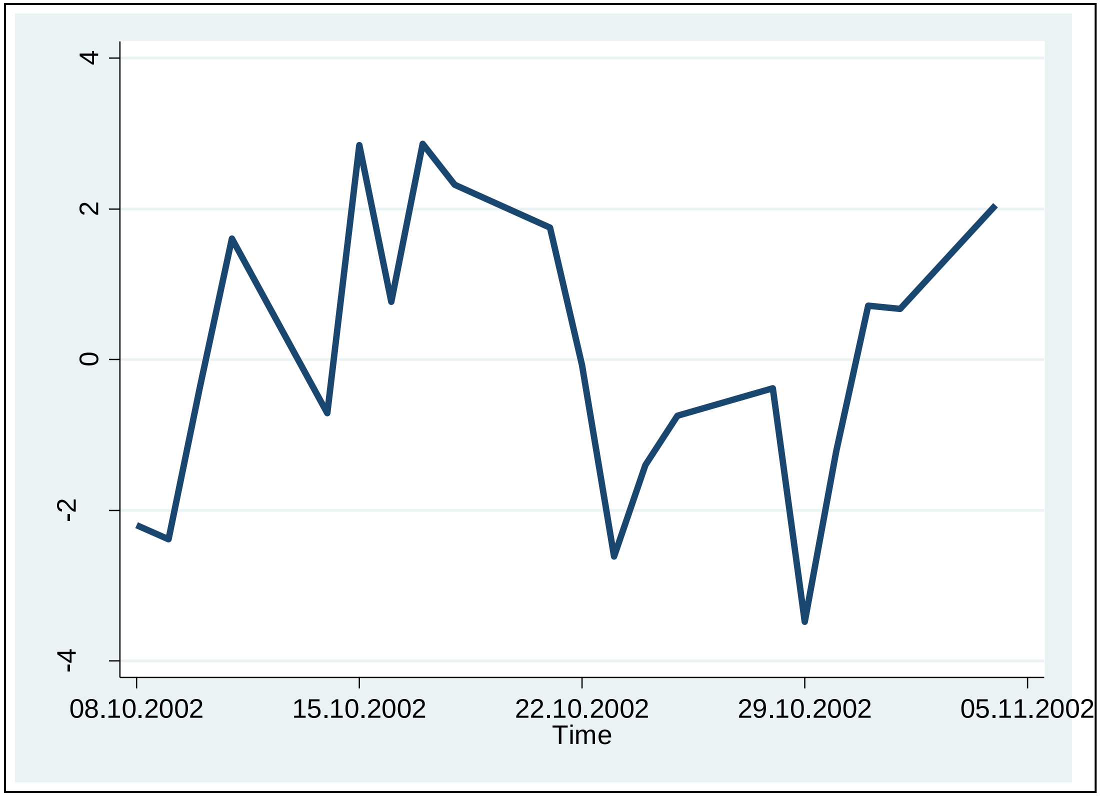
$$\hat{L}_t = -8675.072 + 0.5575393 \cdot t$$

Remove the trend: $\hat{\varepsilon}_t = y_t - \hat{L}_t = y_t - (\hat{c}_1 + \hat{c}_2 t)$

```
. predict res, residuals
```

Solution 1.14-2:

```
. tsline res
```



Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

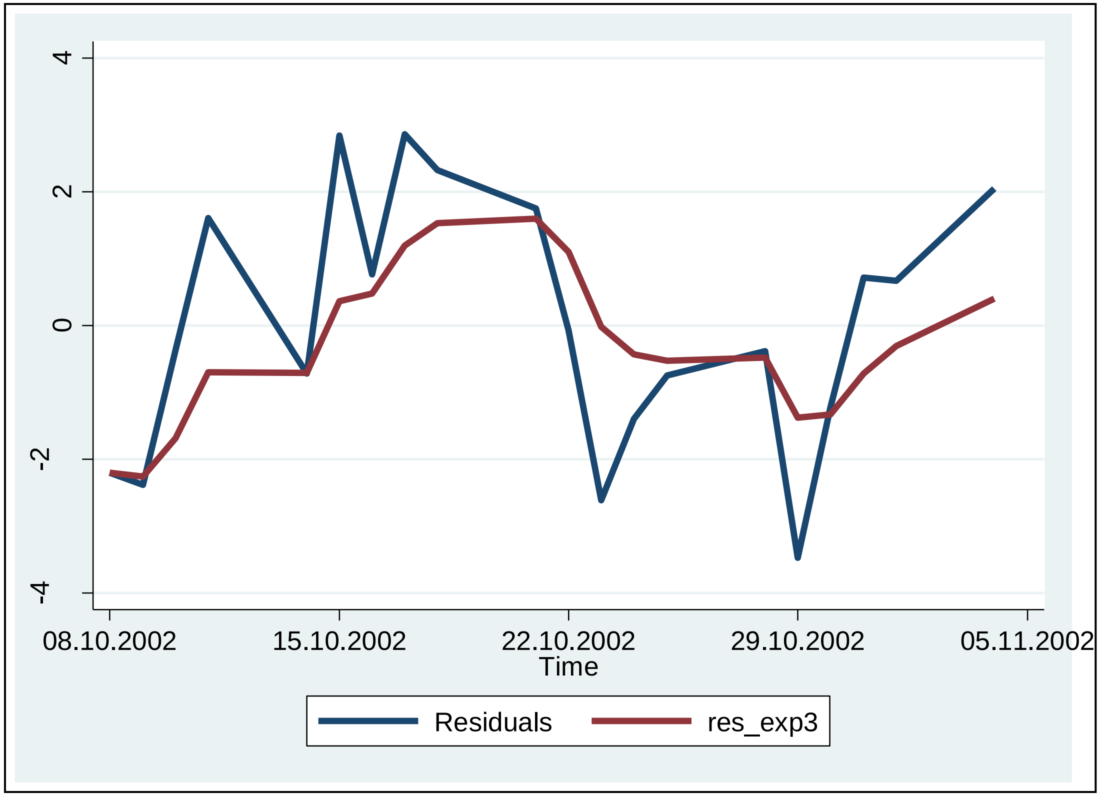
Solution 1.14-3:

Smooth the detrended series. Use the first observed residual as the initial value and $\alpha = .3$ as the smoothing parameter.

```
. gen res_exp3 = res in 1

. local alpha = 0.3

. forvalues num = 2/20 {

. replace res_exp3 = `alpha'*res + (1-alpha') * res_exp3[`num'-1] in
`num'

}
```

Solution 1.14-4:

```
. tsline res res_exp3
```



Solution 1.14-5:

$$\hat{\varepsilon}_t = y_t - \hat{L}_t = y_t - (\hat{c}_1 + \hat{c}_2 t)$$

$$y_t^{EWMA} = \hat{\varepsilon}_t^{EWMA} + \hat{L}_t = \hat{\varepsilon}_t^{EWMA} + \hat{c}_1 + \hat{c}_2 t$$

SIE	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
time	.5575393	.0534614	10.43	0.000	.4452212	.6698575
_cons	-8675.072	835.8098	-10.38	0.000	-10431.04	-6919.1

Add back the trend.

```
. gen SIE_exp3_trend = res_exp3 + .5575393 * time -8675.072
```

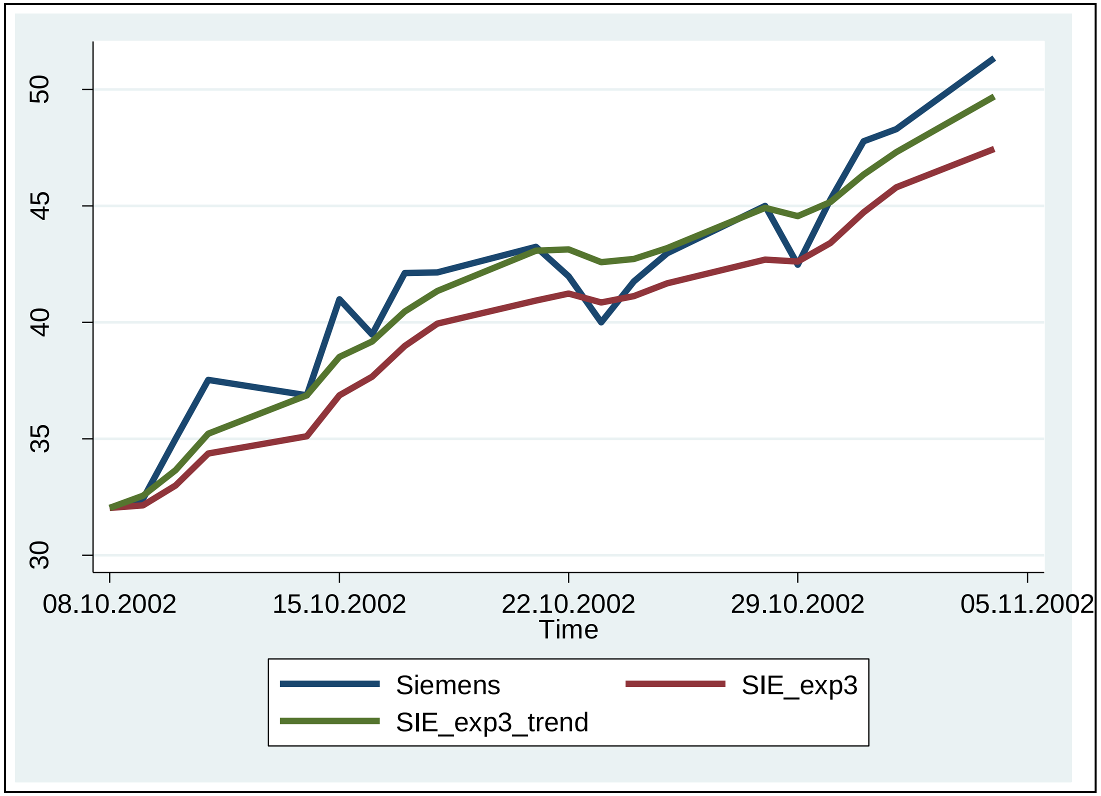
or

```
. gen SIE_exp3_trend = res_exp3+_b[time]* time+_b[_cons]
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.14-6:

```
. tsline SIE SIE_exp3 SIE_exp3_trend
```



Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.15:

Calculate the first three smoothed values of the Siemens time series using the Holt-Winter's two parameter exponential smoothing. Smoothing parameters and initial values: $\alpha = 0.3 \quad \gamma = 0.6 \quad \tilde{y}_0 = 32.05 \quad r_0 = 0$

time	SIE
08.10.2002	32.05
09.10.2002	32.42
10.10.2002	35
11.10.2002	37.53
...	...

Notice: $\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$ and $r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$

Here r_t is a smoothed series representing the trend, for example the average rate of increase.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.15:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1}) \text{ and } r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma)r_{t-1}$$

$$\alpha = 0.3 \quad \gamma = 0.6 \quad \tilde{y}_0 = 32.05 \quad r_0 = 0$$

$$\tilde{y}_1 = 0.3 \cdot y_1 + (1-0.3)(\tilde{y}_0 + r_0)$$
$$= 0.3 \cdot 32.05 + 0.7 \cdot 32.05 = 32.05$$

$$r_1 = 0.6 \cdot (\tilde{y}_1 - \tilde{y}_0) + (1-0.6) \cdot r_0$$
$$= 0.6 \cdot (32.05 - 32.05) + 0.4 \cdot 0 = 0$$

$$\tilde{y}_2 = 0.3 \cdot y_2 + (1-0.3)(\tilde{y}_1 + r_1)$$
$$= 0.3 \cdot 32.42 + 0.7 \cdot 32.05 = 32.161$$

$$r_2 = 0.6 \cdot (\tilde{y}_2 - \tilde{y}_1) + (1-0.6) \cdot r_1$$
$$= 0.6 \cdot (32.161 - 32.05) + 0.4 \cdot 0 = 0.0666$$

$$\tilde{y}_3 = 0.3 \cdot y_3 + (1-0.3)(\tilde{y}_2 + r_2)$$
$$= 0.3 \cdot 35 + 0.7 \cdot (32.161 + 0.0666) = 33.05932$$

Exercise 1.16:

- Smooth the Siemens.dta series using the Holt-Winter's two parameter exponential smoothing. Use the following smoothing parameters and initial values: $\alpha = 0.3 \quad \gamma = 0.6 \quad \tilde{y}_0 = 32.05 \quad r_0 = 0$

Notice: $\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$

$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$

- Plot the result together with the original series and the "simple" smoothed series.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.16-1:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$$
$$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$$

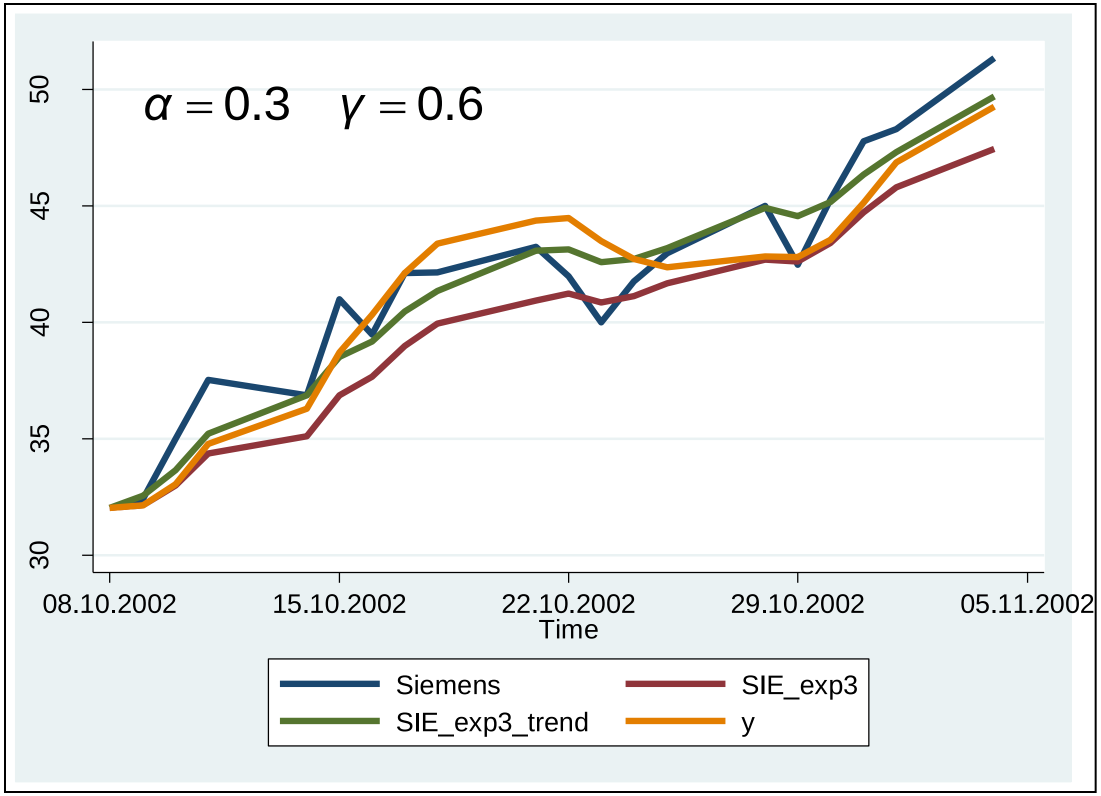
```
. local alpha = 0.3

. local gamma = 0.6

. gen r = 0

. gen y = 32.05

. forvalues num = 2/20 {

. replace y = `alpha'*SIE+(1-`alpha')*(y[`num'-1] +r[`num'-1]) in `num'

. replace r = `gamma'*(y-y[`num'-1])+((1-`gamma') *r[`num'-1]) in `num'

. }
```

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.16-2:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$$
$$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$$

```
. tsline SIE SIE_exp3 SIE_exp3_trend y
```

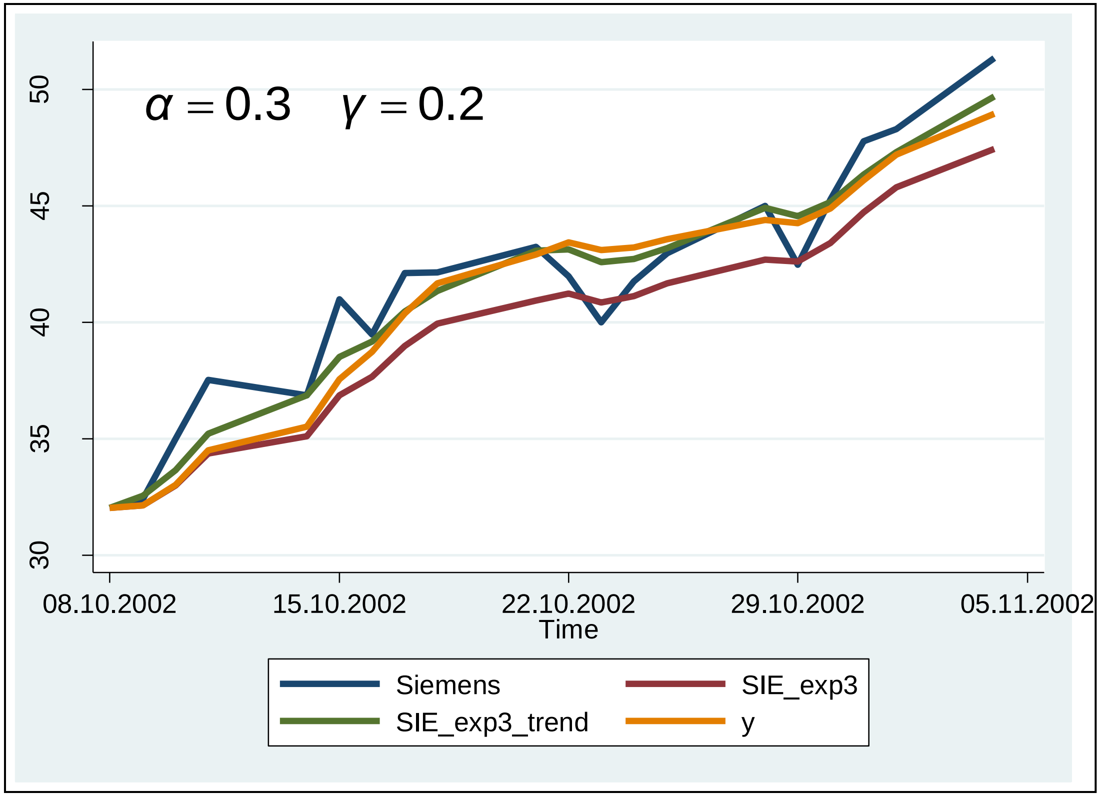


Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.16-3:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$$
$$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$$

```
. tsline SIE SIE_exp3 SIE_exp3_trend y
```

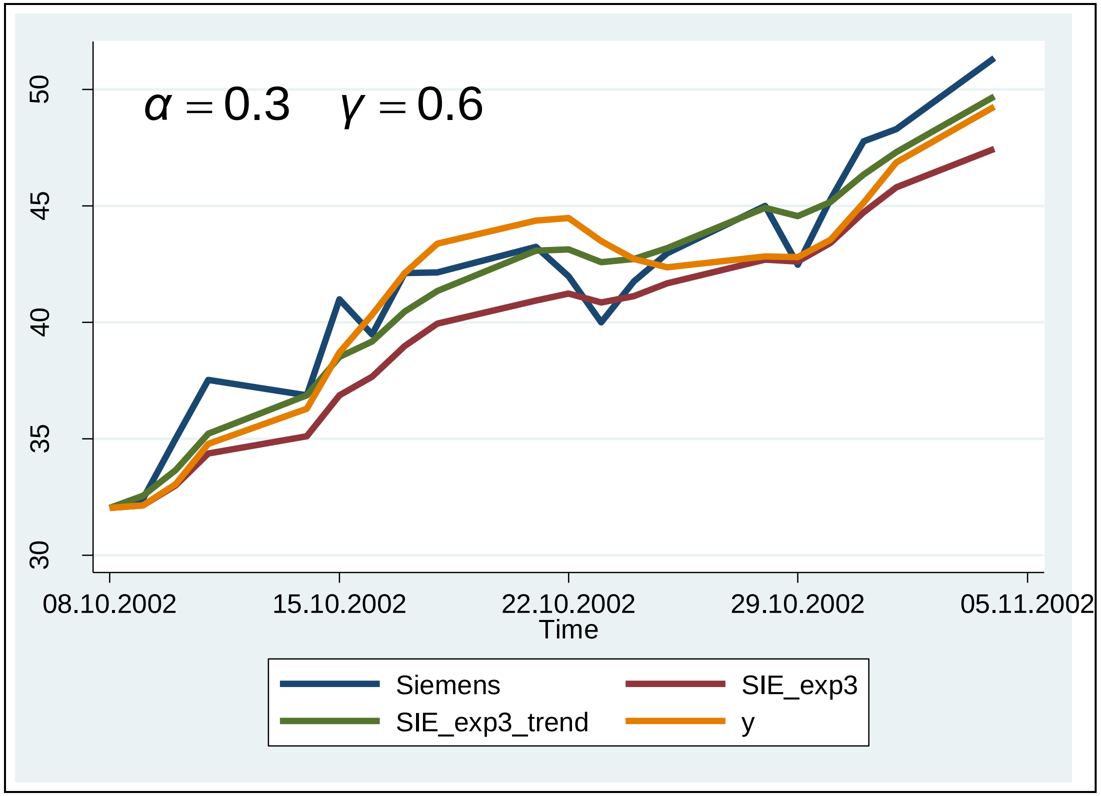


Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.16-4:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$$
$$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$$

```
. tsline SIE SIE_exp3 SIE_exp3_trend y
```

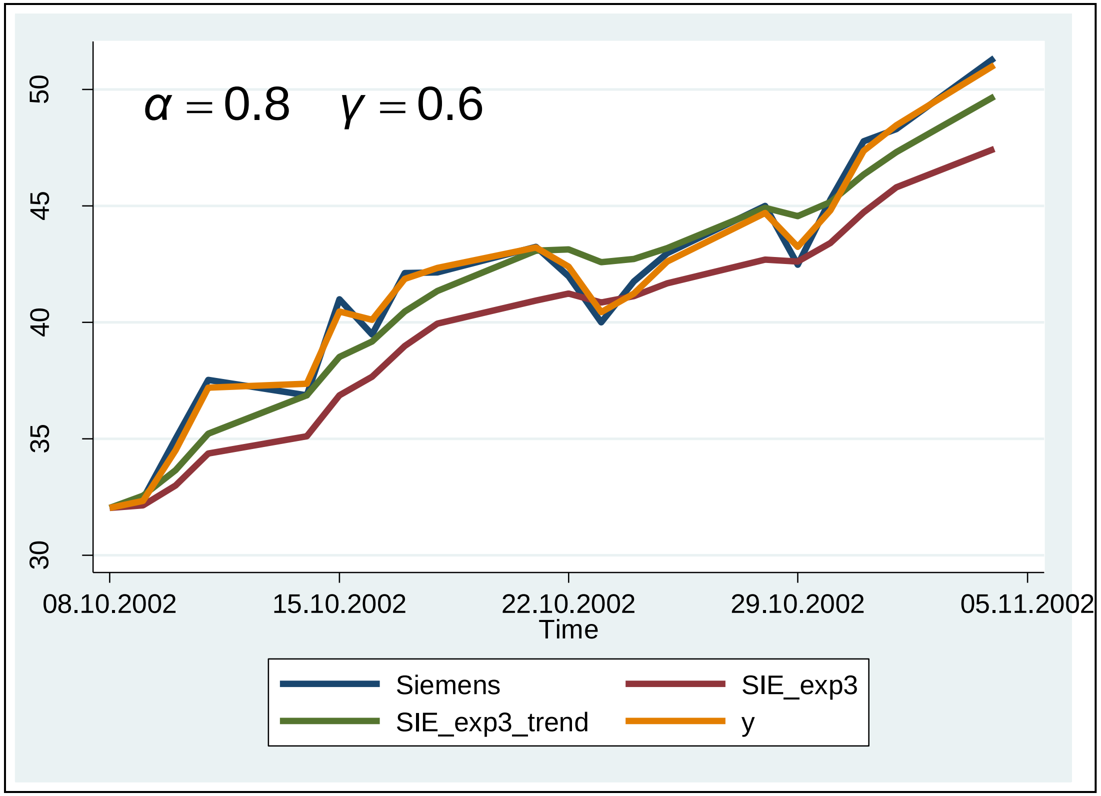


Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.16-5:

$$\tilde{y}_t = \alpha y_t + (1-\alpha)(\tilde{y}_{t-1} + r_{t-1})$$
$$r_t = \gamma(\tilde{y}_t - \tilde{y}_{t-1}) + (1-\gamma) r_{t-1}$$

```
. tsline SIE SIE_exp3 SIE_exp3_trend y
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.17:

Hodrick-Prescott filter

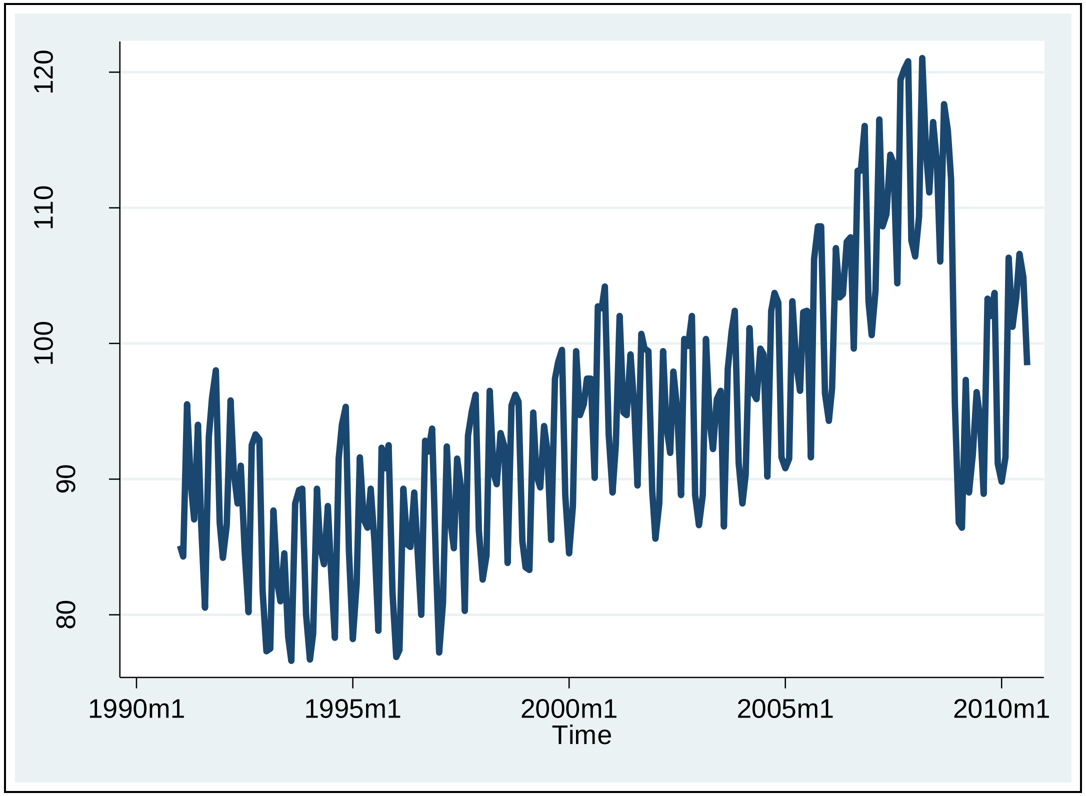
- Compare the "smoothness" of the filtered output series (prod.dta) for different values of λ.

Notice: $$\min \sum_{t=1}^{T} \underbrace{\left(y_t - \hat{G}_t\right)^2}_{\text{Fit}} + \underbrace{\lambda \sum_{t=2}^{T-1} \left[\left(\hat{G}_{t+1} - \hat{G}_t\right) - \left(\hat{G}_t - \hat{G}_{t-1}\right)\right]^2}_{\text{Penalty for Non-Smoothness}}$$

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.17-1:

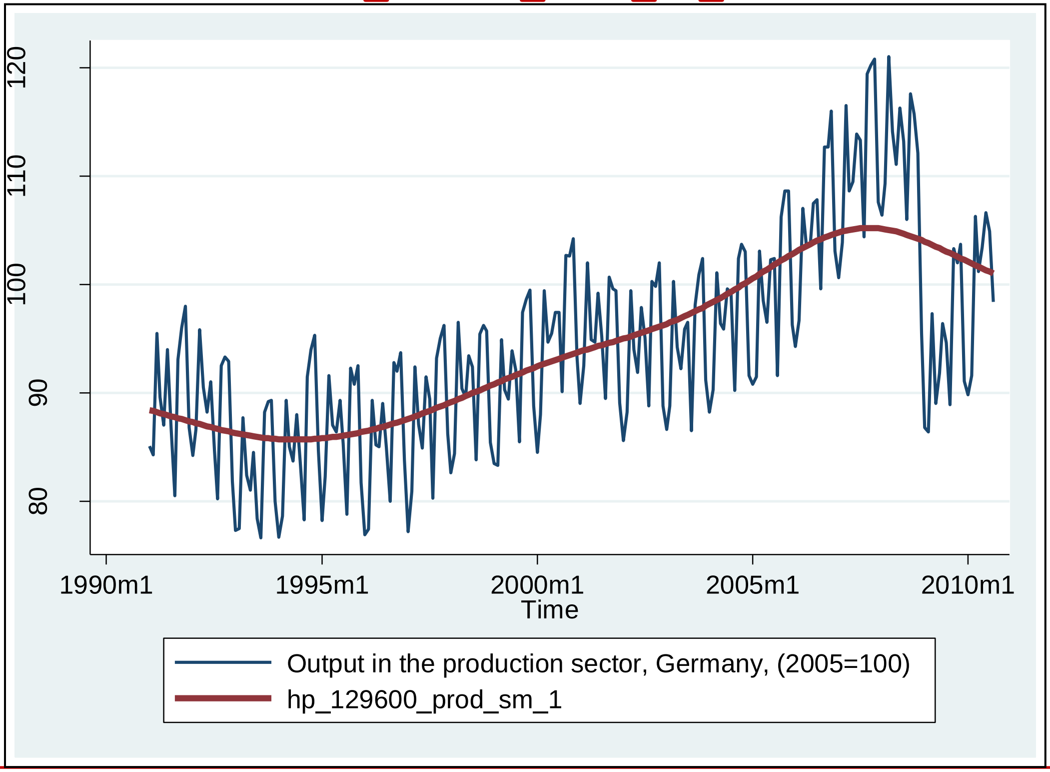
```
. tsline prod
```



Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.17-2:

```
. hprescott prod, stub(hp_129600) smooth(129600)

. tsline prod hp_129600_prod_sm_1
```



The Hodrick-Prescott filter is not implemented in STATA. However you can download an ADO-File for the Hodrick-Prescott filter from STATA. Write the following two commands to download it:

```
. ssc describe hprescott
. ssc install hprescott
```

Afterwards you can use:

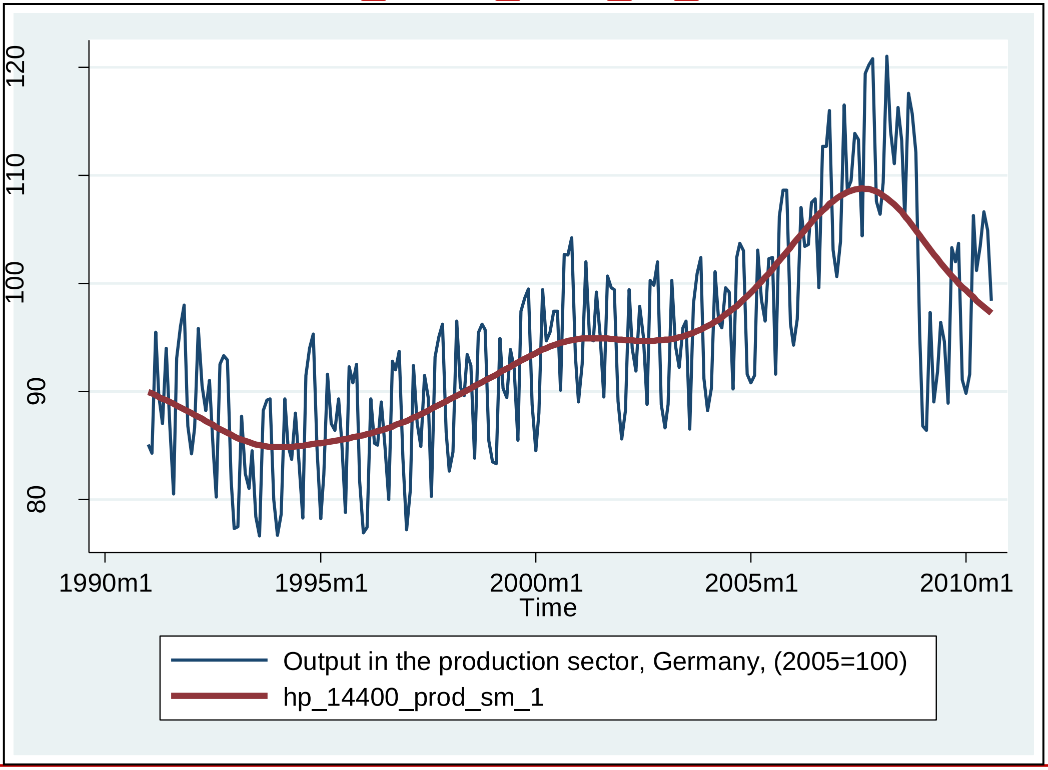
```
. hprescott prod,
stub(hp_129600)
smooth(129600)
```

For STATA Version 13 or higher you can use:

```
. tsfilter hp prod,
smooth(129600)
```

Solution 1.17-3:

```
. hprescott prod, stub(hp_14400) smooth(14400)

. tsline prod hp_14400_prod_sm_1
```



Recommended for monthly data:

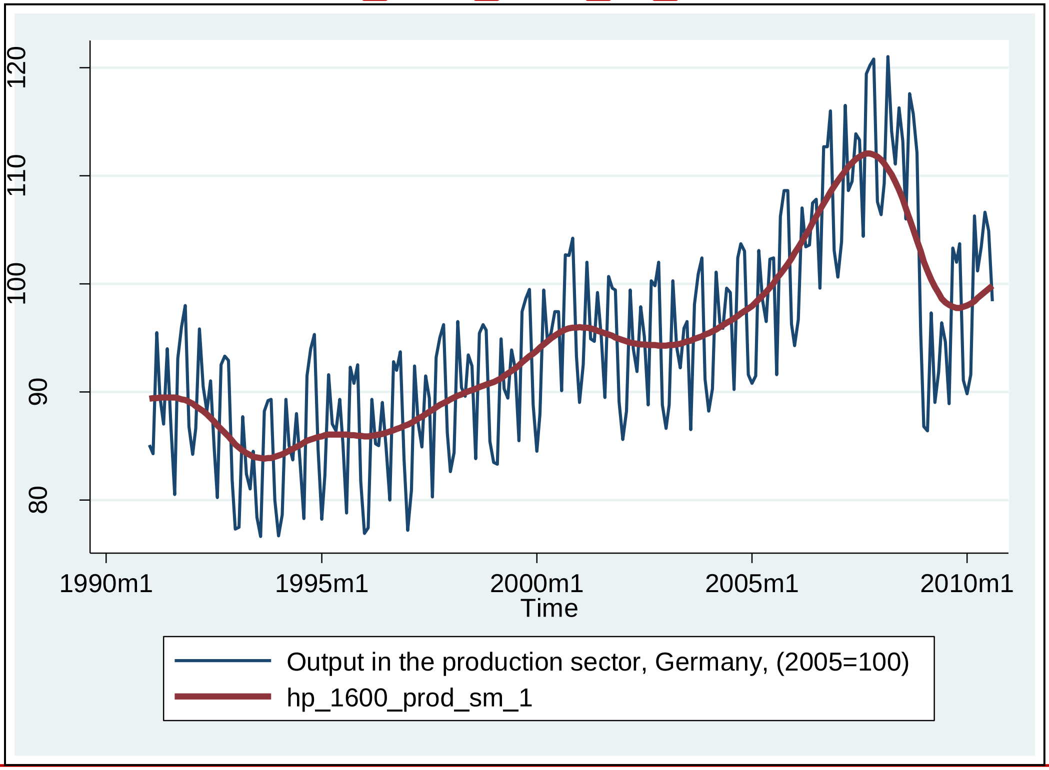
$\lambda = [129600, 86400, 14400]$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.17-4:

```
. hprescott prod, stub(hp_1600) smooth(1600)

. tsline prod hp_1600_prod_sm_1
```



Recommended for quarterly data:

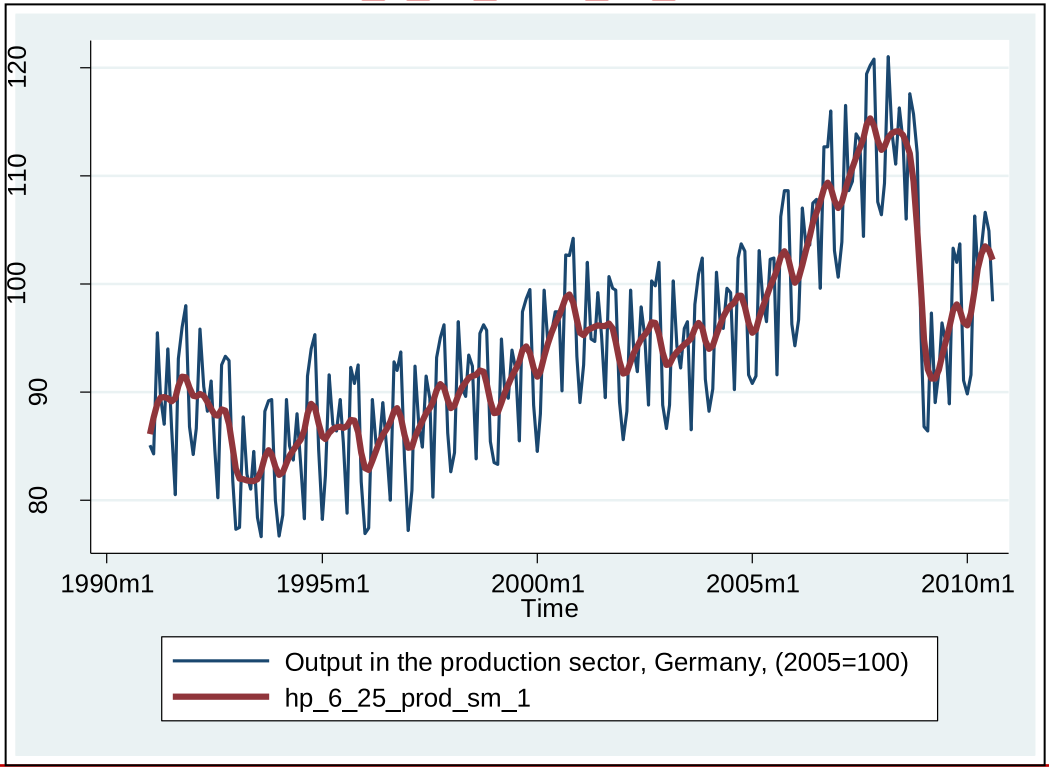
$$\lambda = 1600$$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.17-5:

```
. hprescott prod, stub(hp_6_25) smooth(6.25)

. tsline prod hp_6_25_prod_sm_1
```



Recommended for yearly data:

$$\lambda = 6.25$$

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Please log in to **ISIS** (password: Zeit1718) and **download** the following file:

- unemp.dta

Deterministic Models

- Components of a Time Series
- Additive and Multiplicative Models
- Simple Trend Models
- Smoothing Techniques
- **Seasonal Adjustment**

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.18:

- Show that a 12-month centered moving average is a weighted moving average over 13 periods.

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.18-1:

Centered moving average over 13 periods:

$$\underbrace{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}}_{\tilde{y}_7 = \frac{1}{13}(y_1 + y_2 + \ldots + y_{12} + y_{13})}, y_{14}, \ldots, y_t$$

$$y_1, \underbrace{y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}, y_{14}}_{\tilde{y}_8 = \frac{1}{13}(y_2 + y_3 + \ldots + y_{13} + y_{14})}, \ldots, y_t \qquad \cdots$$

Moving average over 12 periods for monthly data:

$$\underbrace{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}}_{\tilde{y}_{6.5} = \frac{1}{12}(y_1 + y_2 + \ldots + y_{11} + y_{12})}, y_{13}, \ldots, y_t$$

$$y_1, \underbrace{y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}}_{\tilde{y}_{7.5} = \frac{1}{12}(y_2 + y_3 + \ldots + y_{12} + y_{13})}, \ldots, y_t \qquad \cdots$$

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.18-2:

Moving average over 12 periods for monthly data:

$$\underbrace{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}}_{\tilde{y}_{6.5}=\frac{1}{12}(y_1+y_2+\ldots+y_{11}+y_{12})}, y_{13}, \ldots, y_t$$

$$y_1, \underbrace{y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}}_{\tilde{y}_{7.5}=\frac{1}{12}(y_2+y_3+\ldots+y_{12}+y_{13})}, \ldots, y_t \qquad \cdots$$

$$\tilde{y}_7 = \frac{1}{2}(\tilde{y}_{6.5} + \tilde{y}_{7.5}) = \frac{1}{2}\left[\frac{1}{12}(y_1 + y_2 + \ldots + y_{11} + y_{12}) + \frac{1}{12}(y_2 + y_3 + \ldots + y_{12} + y_{13})\right]$$

$$= \frac{1}{24}(y_1 + 2y_2 + 2y_3 + 2y_4 + 2y_5 + 2y_6 + 2y_7 + 2y_8 + 2y_9 + 2y_{10} + 2y_{11} + 2y_{12} + y_{13})$$

→ **Weighted moving average over 13 periods**

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Seasonal Adjustment (Multiplicative model)

The objective is to eliminate the seasonal component *S*:

1. Isolate the combined long-term trend and cyclical components ($G_t = L_t \cdot C_t$) by removing the combined seasonal and irregular components.
2. Divide the original data by the smoothed series to estimate the combined seasonal and irregular components:

$$y_t = L_t \cdot C_t \cdot S_t \cdot I_t \quad t = 1, \ldots, T \qquad \frac{y_t}{\tilde{y}_t} = \frac{L_t \cdot C_t \cdot S_t \cdot I_t}{L_t \cdot C_t} = S_t \cdot I_t$$

3. Eliminate the irregular component as completely as possible. Average the values of the combined seasonal and irregular components corresponding to the same period. These averages will then be estimates of the seasonal indices.
4. Deseasonalize the original series by dividing each value by its corresponding seasonal index.

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.19:

- Load the “unemp.dta” dataset. Plot the series. Describe the seasonal pattern.
- Compute a series which is assumed to be relatively free of seasonal and irregular fluctuations.

  Recall: `tssmooth ma newvar = var, weights([numlist_l]<#c>[numlist_f])`

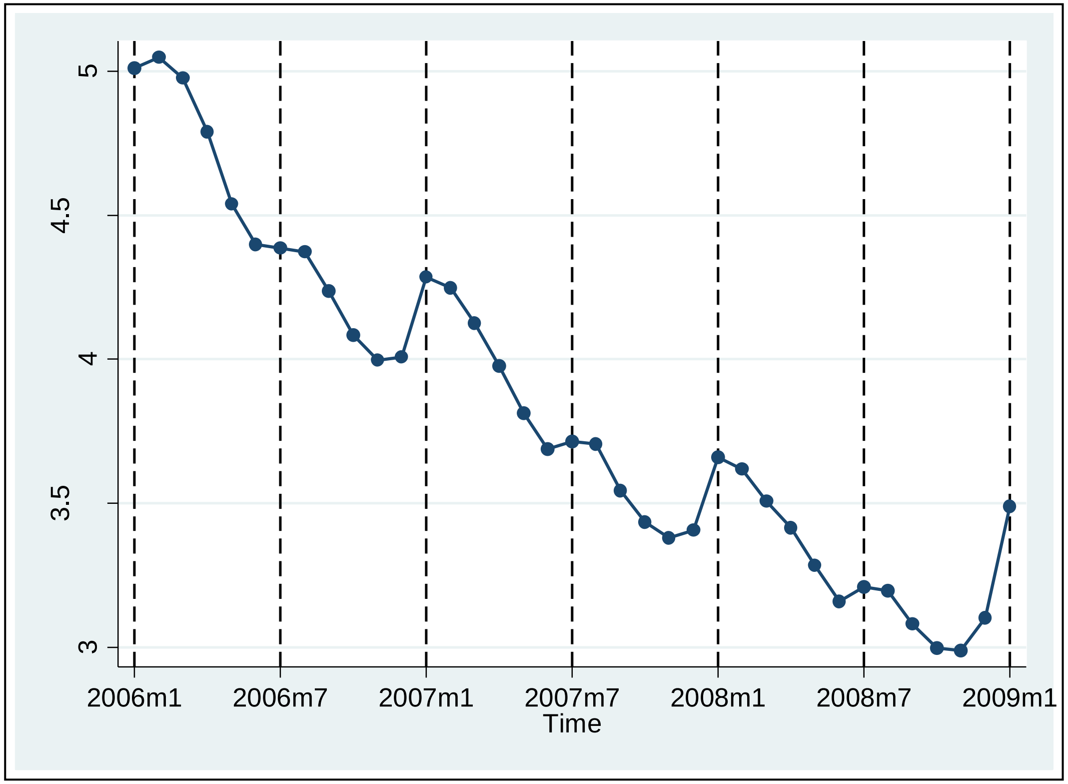
  The option is required for the weighted moving average and describes the span of the moving average, as well as the weights to be applied to each term in the average. The middle term literally is surrounded by < and >, so you might type weights `(1 2 <2> 2 1)`. *numlist_l* is optional and specifies the weights that are to be applied to the lagged terms when computing the moving average. *#_c* is required and specifies the weight to be applied to the current term. *numlist_f* is optional and specifies the weights to be applied to the forward terms when computing the moving average.
- Divide the original data by the smoothed series to estimate the combined seasonal and irregular components.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.19-1: Original data

```
. twoway connect unemp time, xline(552(6)588, lpattern(dash)
lcolor(black)) xlabel(552(6)588) lwidth(medthick)
```



Franziska Plitzko

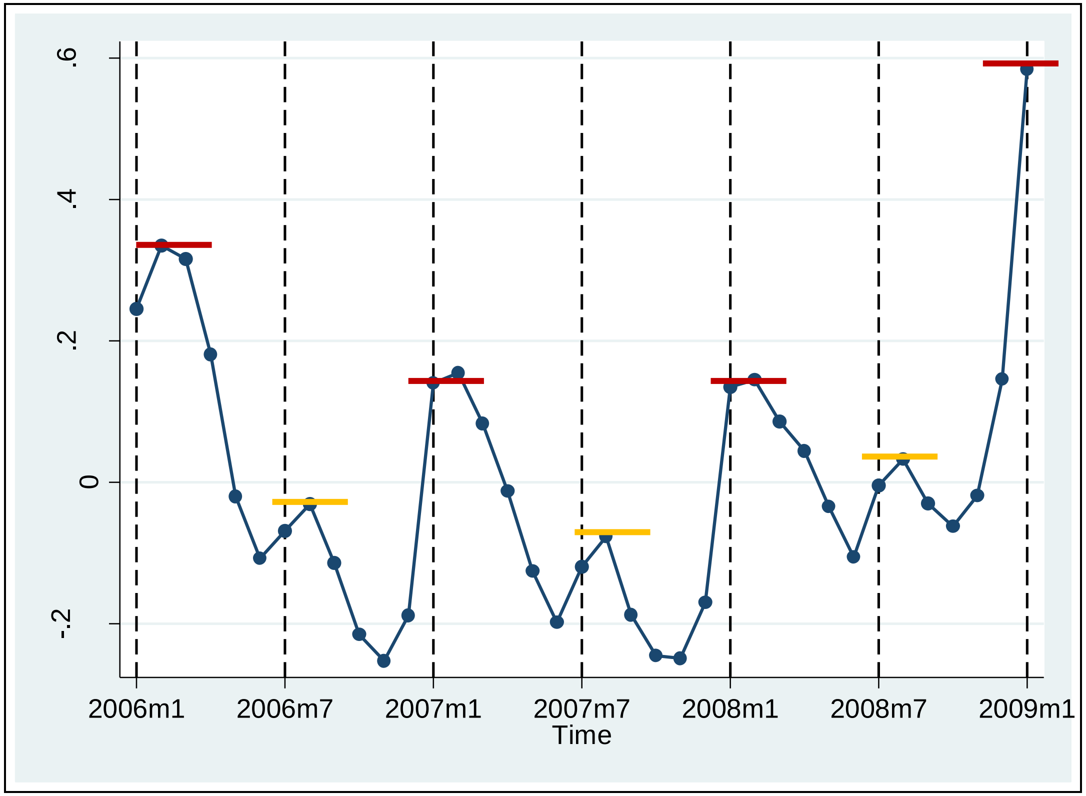
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.19-2:

```
. regress unemp time

      Source |       SS       df       MS              Number of obs =      37
-------------+------------------------------           F(  1,    35) =  322.32
       Model |  11.2661364     1  11.2661364           Prob > F      =  0.0000
    Residual |   1.2233632    35  .034953234           R-squared     =  0.9020
-------------+------------------------------           Adj R-squared =  0.8993
       Total |  12.4894996    36  .346930546           Root MSE      =  .18696

------------------------------------------------------------------------------
       unemp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |  -.0516814   .0028787   -17.95   0.000    -.0575254   -.0458374
       _cons |   33.29268   1.641124    20.29   0.000     29.96102    36.62434
------------------------------------------------------------------------------

. predict res, res

. twoway connect res time
```

Solution 1.19-3:

Residuals

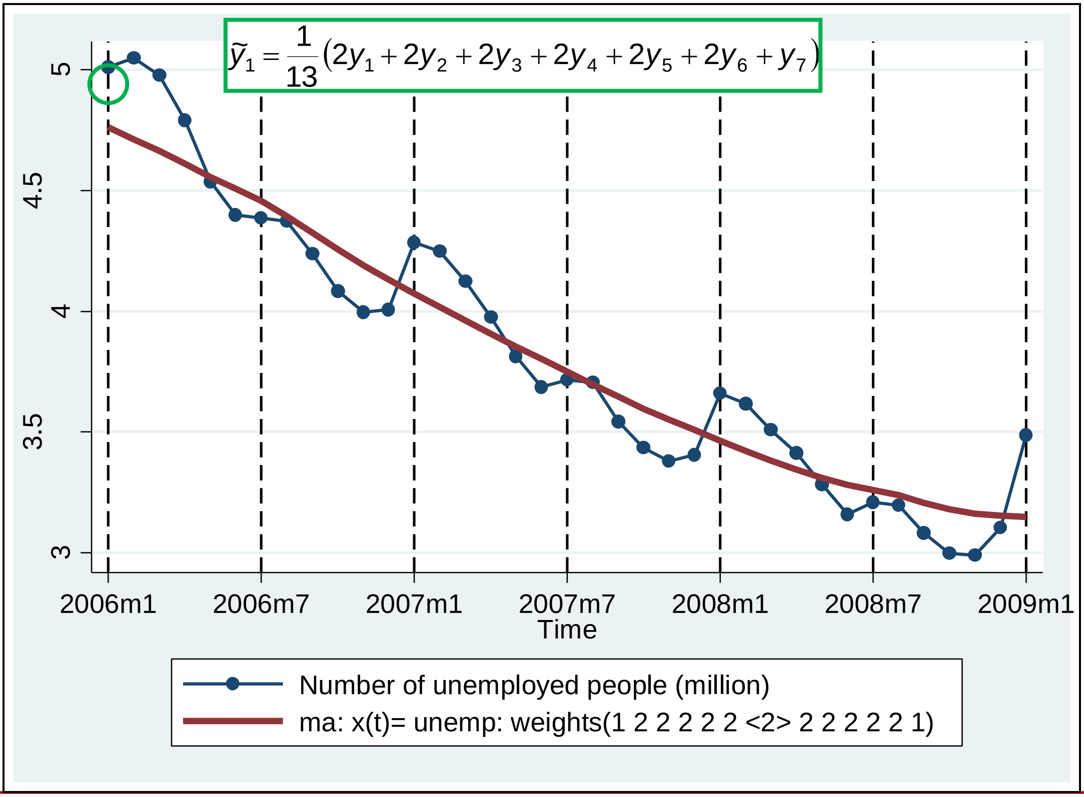


peaks Jan./Feb.
(construction sector)

peaks Aug.
(end of training/school)

Solution 1.19-4: Original data and 12-month centered moving average

```
. tssmooth ma unemp_ma12 = unemp, weights(1 2 2 2 2 2 <2> 2 2 2 2 2 1)
. Twoway connect unemp time || tsline unemp_ma12
```



The smoothed series (**red line**) only consists of the long-term and the cyclical component ($G_t = L_t \cdot C_t$).

STATA graph (detailed):

```
. twoway connect
unemp time,
xline(552(6)588,
lpattern(dash)
lcolor(black))
xlabel(552(6)588)
lwidth(medthick) ||
tsline unemp_ma12,
lwidth(thick)
legend(rows(2))
```

Solution 1.19-5:

```
. tssmooth ma unemp_ma12 = unemp, weights(1 2 2 2 2 2 <2> 2 2 2 2 2 1)
The smoother applied was
     (1/24)*[1*x(t-6) + 2*x(t-5) + 2*x(t-4) + 2*x(t-3) + 2*x(t-2) + 2*x(t-1) +
     2*x(t) + 2*x(t+1) + 2*x(t+2) + 2*x(t+3) + 2*x(t+4) + 2*x(t+5) + ...; x(t)= unemp

. tssmooth ma unemp_ma12_test = unemp, weights(0.5 1 1 1 1 1 <1> 1 1 1 1 1 0.5)
The smoother applied was
     (1/12)*[.5*x(t-6) + 1*x(t-5) + 1*x(t-4) + 1*x(t-3) + 1*x(t-2) + 1*x(t-1) +
     1*x(t) + 1*x(t+1) + 1*x(t+2) + 1*x(t+3) + 1*x(t+4) + 1*x(t+5) + ...; x(t)= unemp

. list  unemp_ma12  unemp_ma12_test in 1/5
```

	unemp~12	unemp_~t
1.	4.762484	4.762484
2.	4.711365	4.711365
3.	4.663529	4.663529
4.	4.610632	4.610632
5.	4.556282	4.556282

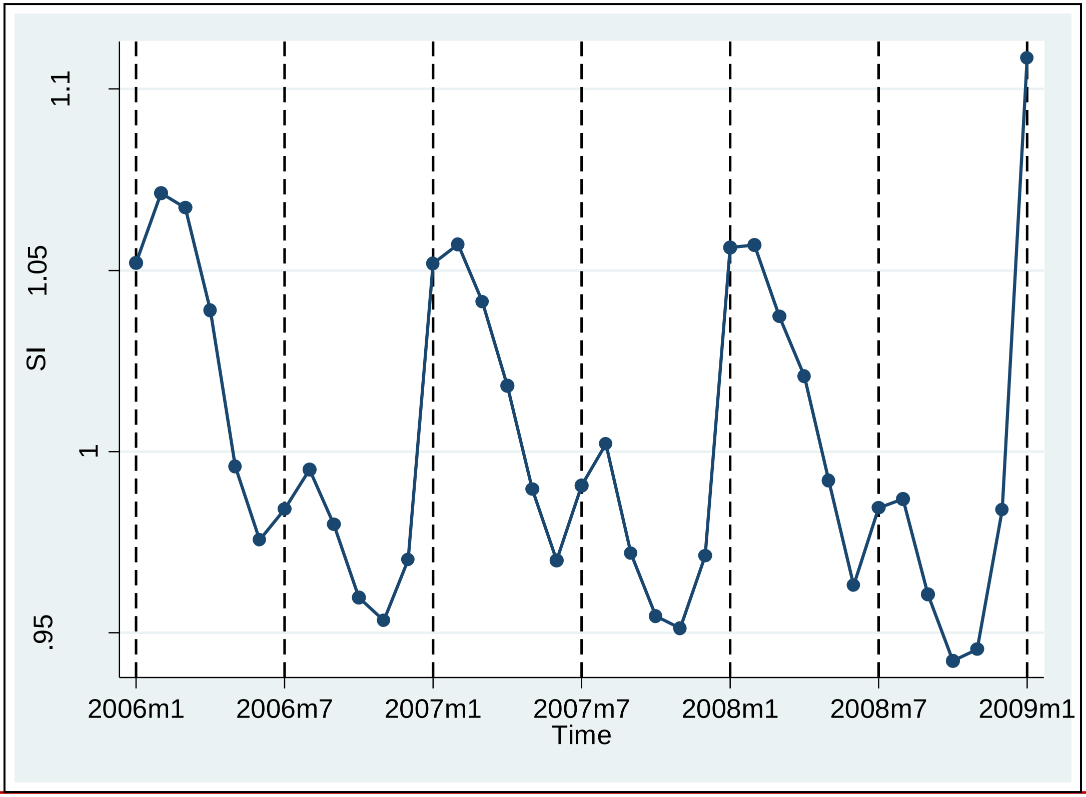
Both weighting options result in the same solution.

(1/24)=1/(sum of all used weights)
=1/(1+2+…+2+1)

(1/12)=1/(sum of all used weights)
=1/(0.5+1+…+1+0.5)

Solution 1.19-6: Original data divided by the smoothed series

```
. generate SI = unemp / unemp_ma12
. twoway connect SI time, xline(522(6)588, lpattern(dash)) xlabel(552(6)588)
```



This series consists of the combined seasonal and irregular component ($S_t \cdot I_t$).

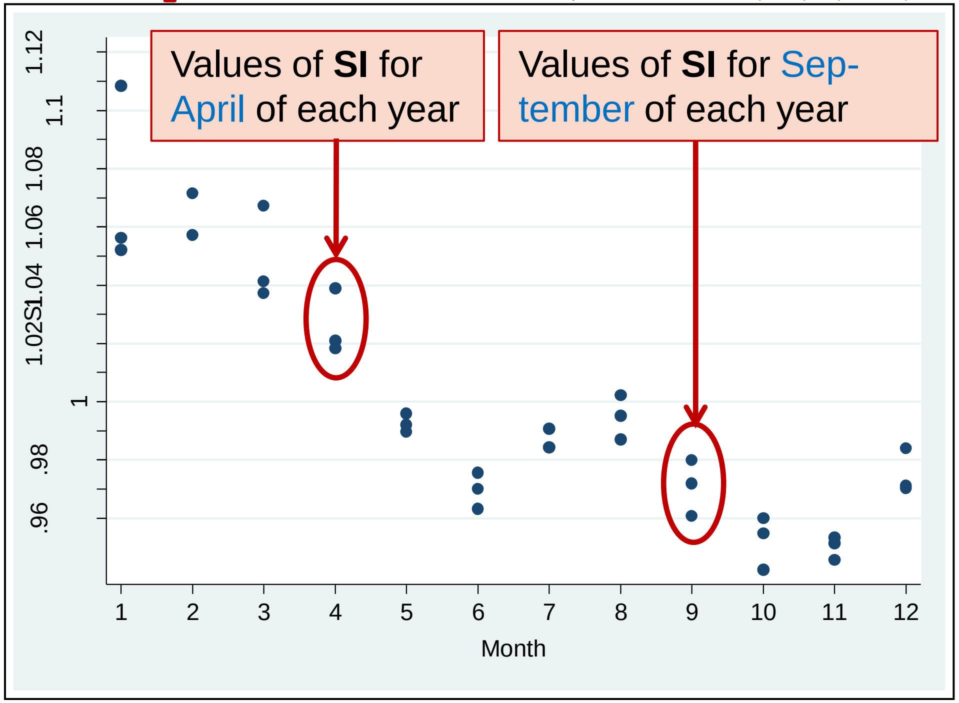
$$\frac{y_t}{\tilde{y}_t} = \frac{L_t \cdot C_t \cdot S_t \cdot I_t}{L_t \cdot C_t} = S_t \cdot I_t$$

STATA graph (detailed):

```
. twoway connect SI
time,
lwidth(medthick)
xline(552(6)588,
lpattern(dash)
lcolor(black))
xlabel(552(6)588)
```

Solution 1.19-7: Seasonal and irregular component plotted for each month

```
. generate month = month(dofm(time))
. label var month "Month"
. twoway scatter SI month, xlabel(1(1)12)
```



Generates a variable called **`month`** where each month is numbered in an orderly manner, i.e. 1=Jan., 2=Feb., …

FYI: `month()` needs time dates with day-month-year-information. We only have the information month-day. Therefore we convert our time variable beforehand with the **`dofm()`** command. Now STATA has the following Input:

```
. gen newtime=dofm(time)
. format newtime %td
. list time newtime in 1/2
     +-------------------+
     |   time    newtime |
     |-------------------|
  1. | 2006m1  01jan2006 |
  2. | 2006m2  01feb2006 |
     +-------------------+
```

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.19-8:

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Exercise 1.20:

- Average the values of the combined seasonal and irregular components corresponding to the same month. These averages will then be estimates of the seasonal indices.

  Note: Final seasonal indices are computed by multiplying these averages by a factor that brings their sum to 12.

- Deseasonalize the original series by dividing each value by its corresponding seasonal index.

Franziska Plitzko
Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.20-1: Averaging SI over months

```
. egen Q = mean(SI), by(month)
. list month Q in 1/14
     +------------------+
     | month          Q |
     |------------------|
  1. |     1   1.067169 |
  2. |     2   1.061885 |
  3. |     3   1.048618 |
  4. |     4   1.026062 |
  5. |     5   .9925745 |
     |------------------|
  6. |     6    .969637 |
  7. |     7    .986414 |
  8. |     8   .9947245 |
  9. |     9   .9708399 |
 10. |    10    .952306 |
     |------------------|
 11. |    11   .9500977 |
 12. |    12   .9751537 |
 13. |     1   1.067169 |
 14. |     2   1.061885 |
     +------------------+
```

The command **egen** generates variables (just as **generate()**), but it has some more functions, e.g. **mean()** function which we can use directly.

```
. list SI if month==1, mean

      +----------+
      |       SI |
      |----------|
   1. | 1.052074 |
  13. | 1.051846 |
  25. | 1.056263 |
  37. | 1.108493 |
      |----------|
 Mean | 1.067169 |
      +----------+
```

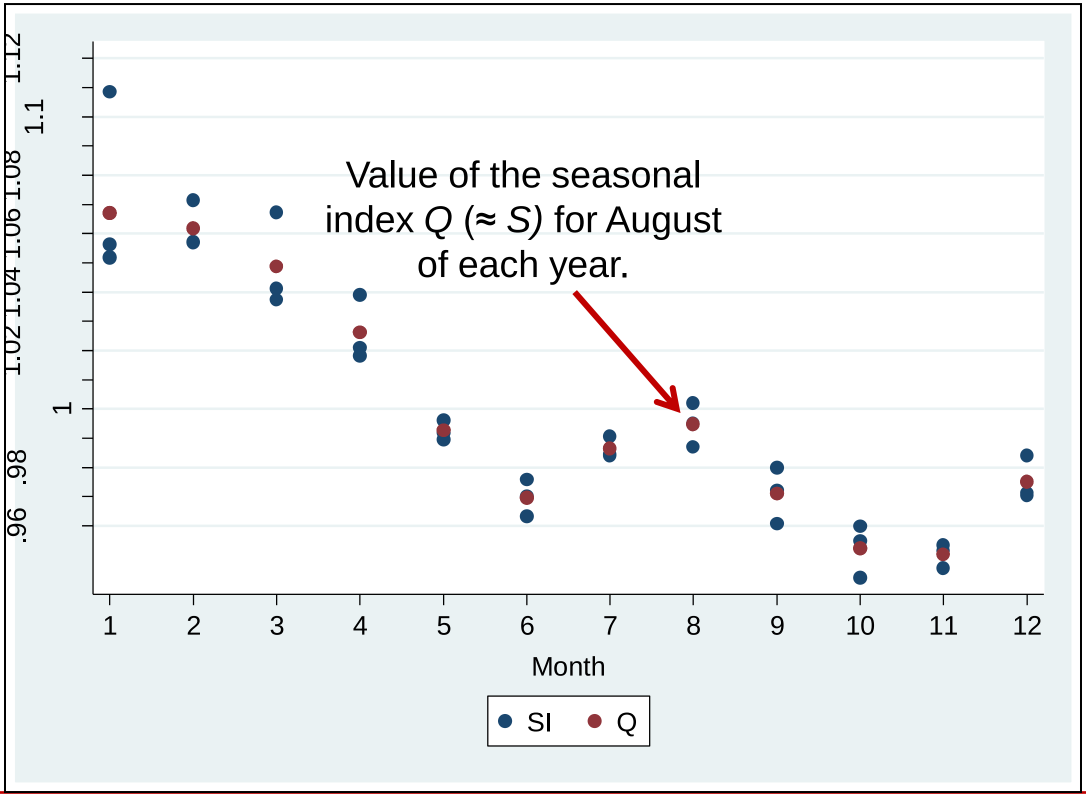
```
. list SI if month==2, mean

      +----------+
      |       SI |
      |----------|
   2. | 1.071384 |
  14. | 1.057185 |
  26. | 1.057084 |
      |----------|
 Mean | 1.061885 |
      +----------+
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.20-2:

Averages over the *SI* values for each quarter (red dots)

```
. twoway scatter SI month || scatter Q month, xlabel(1(1)12)
```



To eliminate the irregular fluctuations we have to average over the *SI* values for each month.

These averages (*Q*) are estimates of the seasonal indices (one for each month).

Final seasonal indices are computed by dividing these averages by their overall average.

Franziska Plitzko

Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.20-3:

```
. sum Q in 1/12

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           Q |        12    .9996233    .0413462    .9500977   1.067169
```

Generate the series S_t

```
. gen S=Q
```

Generate the seasonally adjusted time series

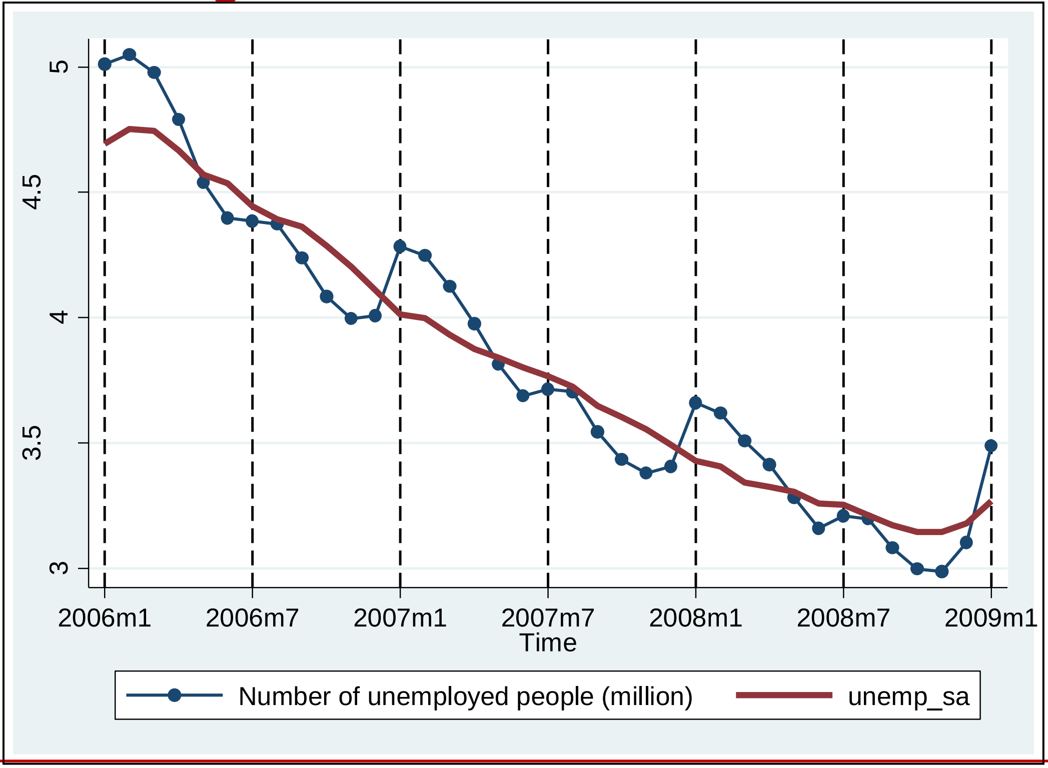
```
. gen unemp_sa = unemp/S
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de

Solution 1.20-4:

Original and seasonally adjusted series

```
. twoway connect unemp time, xline(552(6)588, lpattern(dash)) xlabel(552(6)588) ||
tsline unemp_sa
```



The seasonally adjusted series (red line) was obtained by dividing each value of the original series by its corresponding seasonal index (S).

STATA graph (detailed) :

```
. twoway connect
unemp time,
xline(552(6)588,
lpattern(dash)
lcolor(black))
xlabel(552(6)588)
lwidth(medthick) ||
tsline unemp_sa,
clwidth(thick)
```


Raum: H 5103 D, **Tel.:** 314-78734, **E-Mail:** franziska.plitzko@tu-berlin.de