Regression analysis for trend and smooth component

PD Dr. Ralf Stecking and **Abigail Opokua Asare**

Department of Business Adminstration,
Economics and Law

Institute of Economics
Carl von Ossietzky University Oldenburg

28 October, 2025

Overview

1. Linear trends and linear regression
2. Coefficient of Determination
3. Higher-order polynomials
4. Exponential trend

Basic Steps in Forecasting

- Plot the series to determine it components
- Remove any trend or seasonal pattern
- Develop a forecasting model
- Validate the forecasting model
- Bonus tips:
  - Scale of original time series
  - Prediction intervals
  - Monitor forecast

Data Transformations

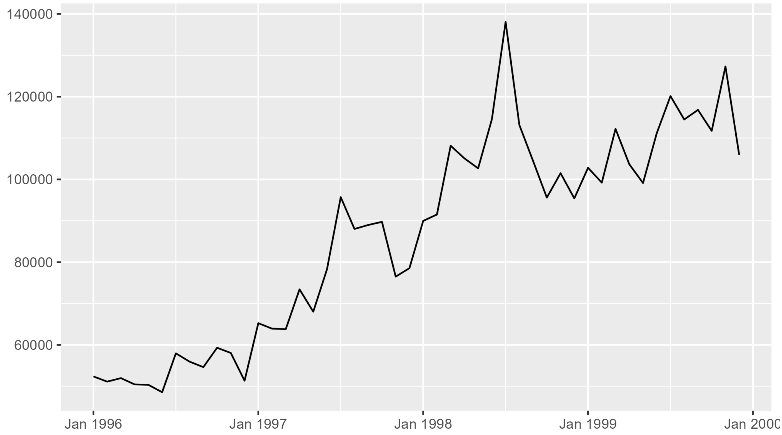
Data transformations are normally done to stabilize the variance of the data.

- Transformations: e.g. taking logarithms, square roots, or Box-Cox transformations to stabilize variance.
- Trend and Seasonal Adjustments
  - Differencing: subtracting the previous value to remove trend or seasonality.
  - Regression model

Linear Trend

- For many economical time series an obvious steady increase or decrease over time (**linear orientation**) can be observed.
- Assumption: the time series can be modeled by a straight line (**trend line**).
  - **Goal 0:** Describe the time series.
  - **Goal 1:** Predict the long term development of the time series.
  - **Goal 2:** Trend adjustment, i.e. elimination of the trend component $\rightarrow$ stationarity!
- Identification of the trend line: **using linear regression**.

Empirical example: DAX 1996-2000



Trend line I

- Assumption: We use linear regression model to capture a time series with a linear trend. The variable is described by a deterministic function m_t of time.
- Observation value x_t deviates from m_t just by random influence or measurement error u_t.
- x_t is composed of m_t and u_t additively, i.e.

$$x_t = m_t + u_t \quad t = 1, 2, ..., N$$

- The random influences u_t are realizations of independent random variables,
  - expectation of the mean **equals zero** and
  - the variance is **constant** through time.

Trend line II

- The equation of the trend line is a linear function:

$$m_t = b_0 + b_1 t$$

- The coefficients b_0 and b_1 are determined by **method of least squares**.
- Based on the differences

$$u_t = x_t - m_t \quad = x_t - b_0 - b_1 t \quad t = 1, 2, ..., N$$

a straight line can be chosen, where the sum of squared errors (S) is minimum, i.e.

$$S = \sum_{t=1}^{N} (x_t - b_0 - b_1 t)^2 \rightarrow MIN!$$

Minimizing the objective function I

- The **objective function** reads:

$$S = \sum_{t=1}^{N} (x_t - b_0 - b_1 t)^2 \rightarrow MIN!$$

- The first **partial derivatives** of S by coefficients b_0 and b_1 equal

$$\frac{\partial S}{\partial b_0} = 2 \sum_{t=1}^{N} (x_t - b_0 - b_1 t)(-1) \quad and$$

$$\frac{\partial S}{\partial b_1} = 2 \sum_{t=1}^{N} (x_t - b_0 - b_1 t)(-t)$$

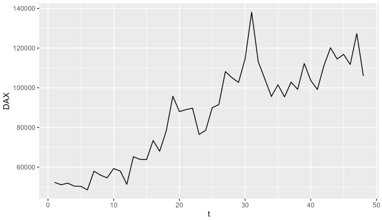
Minimizing the objective function II

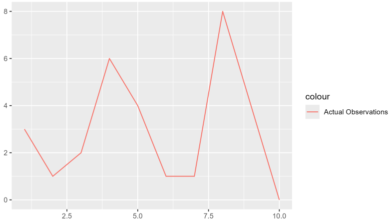
- The **method of least squares** used for estimation of the regression function minimizes S for both regression coefficients b_0 and b_1.
- To this end the first derivatives $\frac{\partial S}{\partial b_0}$ and $\frac{\partial S}{\partial b_1}$ are **set to zero** and solved for b_0 and b_1.
- Finally, the **derived regression coefficients** read:

$$b_1 = \frac{N \sum t x_t - (\sum t)(\sum x_t)}{N \sum t^2 - (\sum t)^2} = \frac{\overline{t x_t} - \bar{t} \cdot \bar{x}}{\overline{t^2} \cdot (\bar{t})^2}$$

$$b_0 = \bar{x} - b_1 \bar{t}$$

DAX 1996-2000 with trend line





Linear Regression – Dax data

	(1)
Intercept	46 975.060***
	(1983.780)
trend	1617.418***
	(86.509)
Num.Obs.	48
R2	0.826
R2 Adj.	0.822
AIC	1029.1
BIC	1034.7
RMSE	10 286.44

Decomposition of the total variation

For linear models estimated by method of least squares it applies:

Total variation = explained variation + not explained variation resp.

$$\sum_{t=1}^{N}(x_t-\bar{x})^2=\sum_{t=1}^{N}(\hat{x}_t-\bar{x})^2+\sum_{t=1}^{N}(x_t-\hat{x})^2$$

with $\hat{x}_t = x_t - u_t = m_t = b_0 + b_1 t$ as **estimated value** for x_t

Carl von Ossietzky Universität Oldenburg

Goodness-of-fit: coefficient of determination

$$r^2 = \frac{explained\ variation}{Total\ variation} = \frac{\sum_{t=1}^{N}(\hat{x}_t - \bar{x})^2}{\sum_{t=1}^{N}(x_t - \bar{x})^2}$$

- The determination coefficient r^2 measures the goodness-of-fit of the trend line towards the time series. R^2 is normalized between zero and one.
- Higher R^2 values indicate a better fit of the model to the data.
- The **adjusted** R^2 is R^2 corrected for degrees of freedom so that the total number of explanatory variables used is accounted for.

Higher-order polynomials for trend and smooth component

- In many cases a **linear** function is not very appropriate to describe the average movement of a time series.
- Example: Unemployment data 1990-2010
  - curved process
  - trend reversal
- Problem: The smooth component (trend + cyclical component) must **not** follow every single movement of the time series. It has to be smooth: the typical process of the curve shall be visible.

Higher-order polynomial I

- The general linear regression model includes k+1 well-known functions of time $m_0(t), m_1(t), m_2(t), ..., m_k(t)$ and assumes a linear combination in order to model the trend process:

$$m(t) = b_0 m_0(t) + b_1 m_1(t) + b_2 m_2(t) + ... + b_k m_k(t)$$

  with $b_0, b_1, ..., b_k$ chosen with the highest goodness-of-fit for the observed time series $\rightarrow$ method of **least squares!**

- The model of the linear trend is a special case, while $m_0(t) = 1$ and $m_1(t) = t$.

- If no ordinary linear trend can be assumed $\rightarrow$ **polynomials!**

Higher-order polynomial II

- A **polynomial** trend of order k

$$m_(t) = b_0 t + b_1 t + b_2 t^2 + ... + b_k t^k$$

arises from the linear regression model by choosing

$$m_0(t) = 1, m_1(t) = t, m_2(t) = t^2, ..., m_k(t) = t^k$$

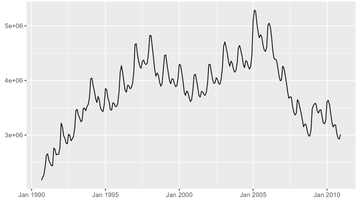
- Again the coefficients are estimated by the method of least squares, i.e. $b_0, b_1, ..., b_k$ so that

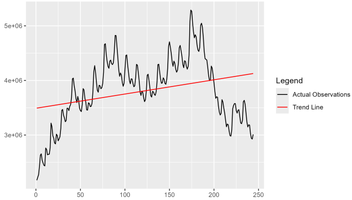
$$S = \sum_{t=1}^{N} (x_t - m(t))^2 = \sum_{t=1}^{N} (x_t - b_0 - b_1 t - b_2 t^2 - ... - b_k t^k)^2$$

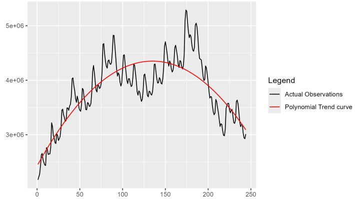
becomes minimal!

- This type of trend models is well suited in order to describe a time series.
- Warning: Be careful with forecasts by trend extrapolation. Out of its adjustment range polynomials tend towards $\pm\infty$ rapidly.

Empirical example: Unemployment data 1990-2010







Regressions – Unemployment Data

	Linear	Polynomial
Intercept	3 490 640.303***	2 427 880.803***
	(83 986.186)	(44 379.343)
trend	2605.813***	28 526.776***
	(655.111)	(1057.007)
$trend^2$		−105.800***
		(4.155)
Num.Obs.	244	244
R2	0.089	0.675
R2 Adj.	0.086	0.672
AIC	7179.2	6929.8
BIC	7189.7	6943.8
RMSE	585 462.64	349 802.82

Exponential trend I

- **Exponential trend** is the main approach in order to model **non linear** trend or smooth components.
- The trend function is represented by an **exponential function**:

$$m(t) = a \cdot e^{bt}$$

- The slope of the trend function

$$\frac{\partial m}{\partial t} = b \cdot a \cdot e^{bt} = b \cdot m(t)$$

  is not constant, but proportional to the actual level.
- Coefficient b is a **constant rate** for the growth of the trend function.

Exponential trend II

- In the **multiplicative model** (excluding the seasonal S and cyclical component C) time series x_t can be represented by:

$$x_t = a \cdot e^{bt} \cdot u_t$$

- By taking the **logarithm** of:

$$ln\ x_t = ln\ a \cdot bt \cdot ln\ u_t$$

- **Substitution** by $ln\ a = b_0; b = b_1; ln\ u_t = u_t^*$:

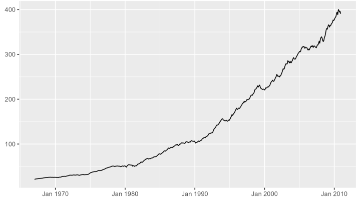
$$ln\ x_t = b_0 + b_1 t + u_t^* \;\; resp. \;\; m(t)^* = b_0 + b_1 t$$

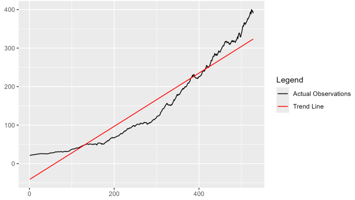
with $m(t)^* = in\ x_t - u_t^*$

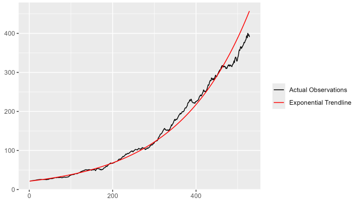
- The **calculation of coefficients** again can be carried out by simple linear regression.

Carl von Ossietzky Universität Oldenburg

REX 1967-2010







Regression – Rex Data

	Linear	Exponential
Intercept	-41.391^{***}	3.063^{***}
	(3.154)	(0.005)
trend	0.691^{***}	0.006^{***}
	(0.011)	(0.000)
Num.Obs.	528	528
R2	0.915	0.994
R2 Adj.	0.915	0.994
AIC	5164.3	6958.8
BIC	5177.1	6971.6
RMSE	32.00	175.06

Important Lessons

- **Identification stage:** the choice of model
- **Estimation stage:** estimate the parameters of the model
- **Diagnostic checking stage:** criterion to evaluate the relative goodness of fit of the models

- **Linear regression** is powerful for capturing linear trends in data.

- **Linear regression** can be used to fit global trend that applies to the entire series.
- R^2 helps assess the goodness-of-fit of the regression model..
- **Higher-order polynomials** can model more complex trends but require careful use to avoid overfitting.
- **Exponential trends** are useful in modeling growth patterns where the dependent variable grows at a constant percentage rate, fitting well with the multiplicative model.
- **Avoid** selecting overly **complicated trend patterns**.
- Always **evaluate** how the model performs on the **test data**.