SIMPLE REGRESSION MODEL II: Incorporating Nonlinearities, Unbiasedness of OLS Estimators

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

Econometrics I

Incorporating Nonlinearities in Simple Regression

- Linear relationships may not be appropriate in some cases.
- By appropriately redefining variables we can easily incorporate nonlinearities into the simple regression.
- Our model will still be **linear in parameters**. We do not use nonlinear transformations of parameters.
- In practice natural logarithmic transformations are widely used. $(\log(y) = \ln(y))$. (also recall: $\log(e) = 1, \log(1) = 0, \log(z^a) = a\log(z), \log(z^a x^b) = a\log(z) + b\log(x), \log(e^b) = b$)
- Other transformations may also be used, e.g., adding quadratic or cubic terms, inverse form, etc.

Linearity in Parameters

- The linearity of the regression model is determined by the linearity of βs not x and y.
- We can still use nonlinear transformations of x and y such as $\log x$, $\log y$, x^2, $\sqrt{x}$, $1/x$, $y^{1/4}$. The model is still linear in parameters.
- But the models that include nonlinear transformations of βs are not linear in parameters and cannot be analyzed using OLS framework.
- For example the following models are not linear in parameters:

$$consumption = \frac{1}{\beta_0 + \beta_1 income} + u$$

$$y = \beta_0 + \beta_1^2 x + u$$

$$y = \beta_0 + e^{\beta_1 x} + u$$

Functional Forms using Natural Logarithms

Log-Level

$$\log y = \beta_0 + \beta_1 x + u$$

$$\begin{aligned} \Delta \log y &= \beta_1 \Delta x \\ \%\Delta y &= (100\beta_1)\Delta x \end{aligned}$$

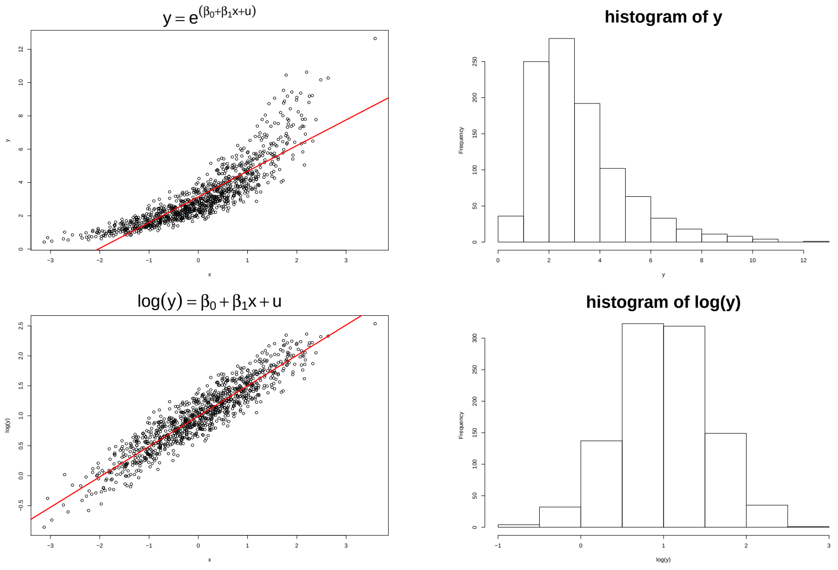
Interpretation: For a one-unit change in x, y changes by $\%(100\beta_1)$. Note: $100\Delta\log y = \%\Delta y$

The relationship between x and y, **before the (natural) logarithmic transformation** can be written as

$$y = \exp(\beta_0 + \beta_1 x + u) \equiv e^{\beta_0 + \beta_1 x + u}$$

Recall that $ln(e^z) = z$ (or equivalent notation: $\log(e^z) = z$). Applying the logarithmic transformation we obtain the log-level regression model.

Log-level Model Illustration



Functional Forms using Natural Logarithms

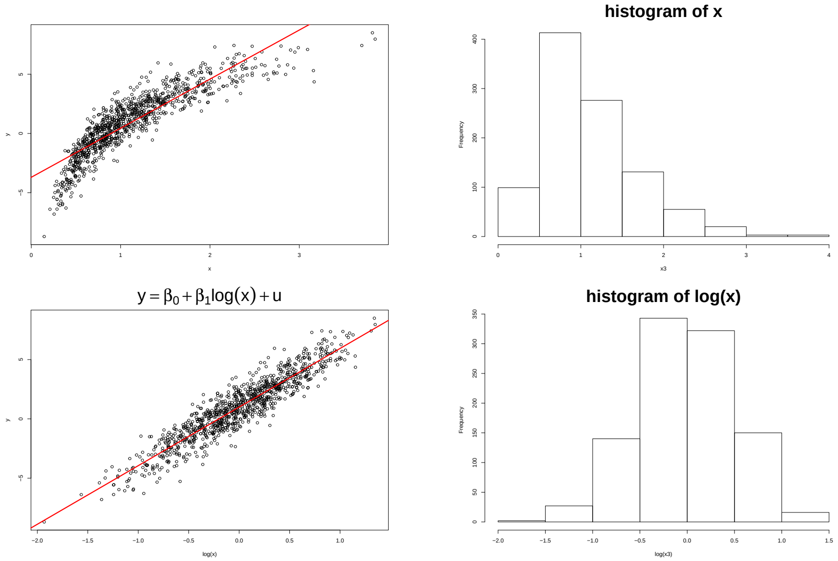
Level-Log

$$y = \beta_0 + \beta_1 \log x + u$$

$$\begin{aligned}\Delta y &= \beta_1 \Delta \log x \\ &= \left(\frac{\beta_1}{100}\right) \underbrace{100\Delta \log x}_{\%\Delta x}\end{aligned}$$

Interpretation: For a %1 change in x, y changes by $(\beta_1/100)$ (in its own units of measurement).

Level-Log Model Illustration



Functional Forms using Natural Logarithms

Log-Log (Constant Elasticity Model)

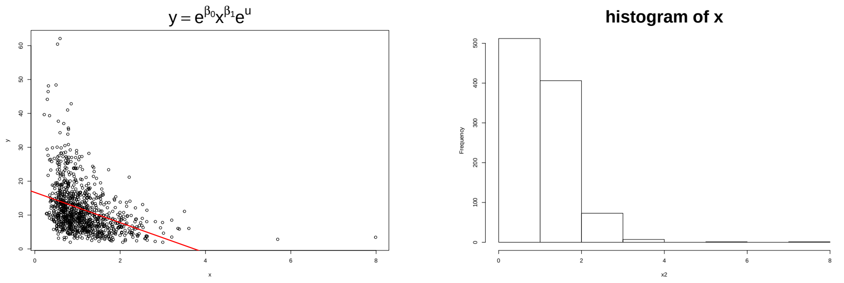
$$\log y = \beta_0 + \beta_1 \log x + u$$

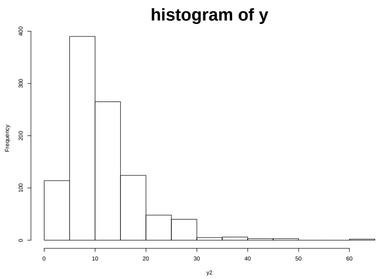
$$\begin{aligned}\Delta \log y &= \beta_1 \Delta \log x \\ \%\Delta y &= \beta_1 \%\Delta x\end{aligned}$$

Interpretation: β_1 is the elasticity of y with respect to x. It gives the percentage change in y for a %1 change in x.

$$\frac{\%\Delta y}{\%\Delta x} = \beta_1$$

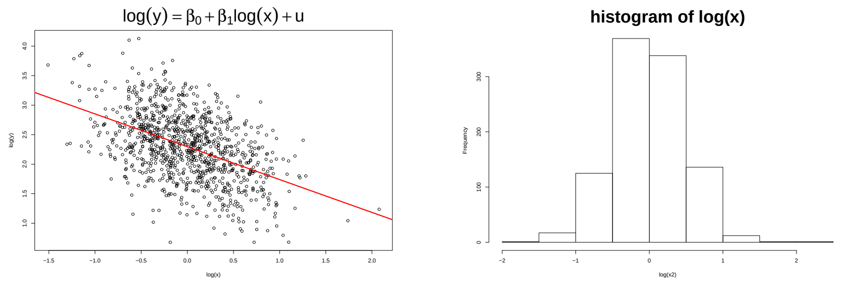
Log-Log Model Illustration

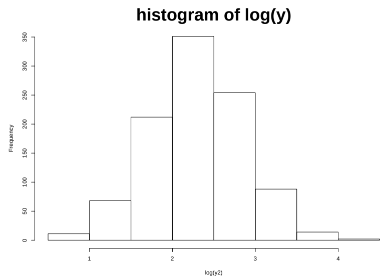




In this illustration $\beta_1 < 0$, x and y are negatively correlated. The scatter diagram resembles a constant elasticity demand curve. But the linear fit (without log transformation) is not good. Note that both x and y are skewed to right.

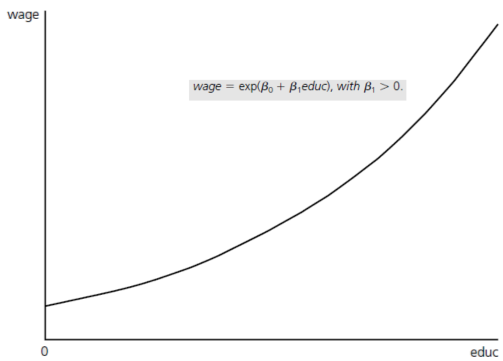
Log-Log Model Illustration





After the log transformation we can see that both x and y are symmetrically distributed. Now the linear fit is much better. Notice that the slope is negative (see the previous slide).

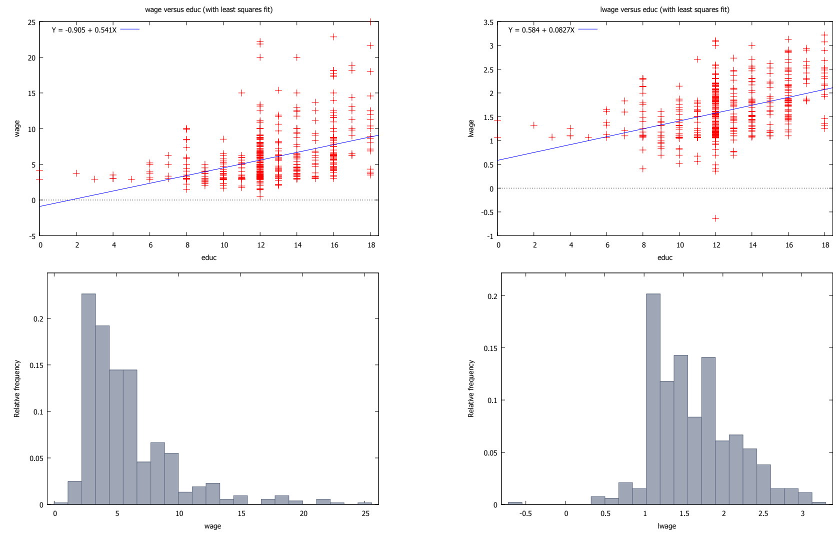
Example: Wage-Education Relationship, $\log(wage) = \beta_0 + \beta_1 educ + u$



Note that an additional year of education has a larger marginal impact on wage as the level of education increases. This is reasonable because the impact of an additional year of education on top of a 4-year college degree, say, is expected to be much larger than the impact of an additional year at the primary school level.

12

Wage-Education Relationship



Log-Level Simple Wage Equation

$$\widehat{\text{logwage}} = \underset{(0.097)}{0.584} + \underset{(0.008)}{0.083}\,\text{educ}$$

$$n = 526 \quad R^2 = 0.186$$

(standard errors in parentheses)

- After multiplying the slope estimate by 100 it can be interpreted as %; $100 \times 0.083 = 8.3$
- An additional year of education is predicted to increase average wages by %8.3. This is called *return to another year of education*.
- **WRONG**: *an additional year of education increases logwage by %8.3*. Here, wage increases by %8.3 not logwage.
- $R^2 = 0.186$: Education explains about %18.6 of the variation in $logwage$.

Log-Log Example: CEO Salaries (ceosal1.gdt)

Model:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u$$

Estimation results:

$$\widehat{\log(\text{salary})} = \underset{(0.288)}{4.822} + \underset{(0.035)}{0.257}\,\log(\text{sales})$$

$$n = 209 \quad R^2 = 0.211$$

(standard errors in parentheses)

- Interpretation: %1 increase in firm sales increases CEO salary by %0.257. In other words, the elasticity of CEO salary with respect to sales is 0.257. About %4 increase in firm sales will increase CEO salary by %1.
- $R^2 = 0.211$: logsales can explain about %21.1 of variation in logsalary.

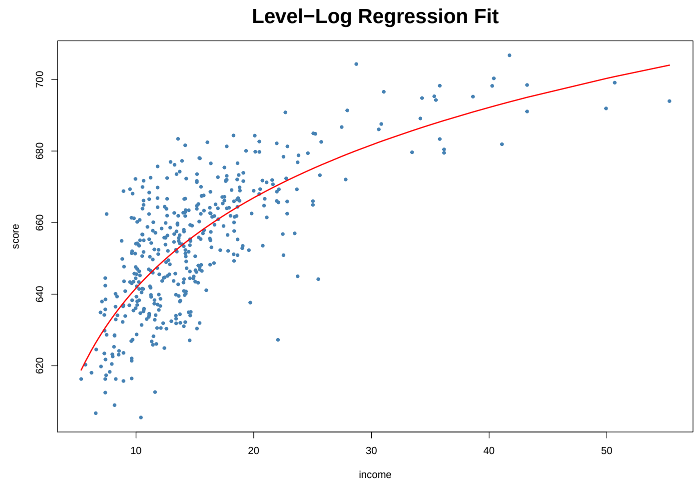
Level-Log Example

$$TestScore_i = \beta_0 + \beta_1\ income_i + u_i,$$

The straight line represents the linear fit.

Level-Log Example

$$TestScore_i = \beta_0 + \beta_1\ \log(income_i) + u_i,$$



Level-Log Example: Test Scores and Income

Model:

$$TestScore = \beta_0 + \beta_1 \ \log(income) + u$$

Data set: CASchools (From Stock and Watson's text) Estimation results:

$$\widehat{\text{TestScore}} = \underset{(3.84)}{557.83} + \underset{(1.40)}{36.42}\,\text{log(income)}$$

$$n = 420 \quad R^2 = 0.56$$

(standard errors in parentheses)

- Interpretation: 1% increase in income is associated with $0.01 \times 36.42 = 0.364$ point increase in test scores. A 3% increase in income is associated with about $3 \times 0.364 = 1.09$ point increase in test scores.
- $R^2 = 0.56$: $\log(income)$ can explain about 56% of the variation in test scores.

Functional Forms using Natural Logarithms: Summary

Table: Functional Forms

Model	Dependent	Explanatory	Interpretation
Level-Level	y	x	$\Delta y = \beta_1 \Delta x$
Log-Level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Level-Log	y	$\log(x)$	$\Delta y = \frac{\beta_1}{100}\%\Delta x$
Log-Log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Exercise: Match the following interpretations with the models above

- In response to a one unit change in x, y is predicted to change by β_1 units. Model = ...
- In response to a one per cent change in x, y is predicted to change by $0.01\beta_1$ units. Model = ...
- In response to a 10 per cent change in x, y is predicted to change by $10\beta_1$ per cent. Model = ...
- In response to a 10 unit change in x, y is predicted to change by $1000\beta_1$ per cent. Model = ...

Statistical Properties of OLS Estimators, $\hat{\beta}_0, \hat{\beta}_1$

- What are the properties of the distributions of $\hat{\beta}_0, \hat{\beta}_1$ over different random samples from the population?
- What are the expected values and variances of OLS estimators?
- We will first examine finite sample properties: unbiasedness and efficiency. These are valid for any sample size n.
- Recall that unbiasedness means that the mean of the sampling distribution of an estimator is equal to the unknown parameter value.
- Efficiency is related to the variance of the estimators. An estimator is said to be efficient if its variance is the smallest among a set of unbiased estimators.

Unbiasedness of OLS Estimators

We need the following assumptions for unbiasedness:

1. SLR.1: Model is linear in parameters: $y = \beta_0 + \beta_1 x + u$
2. SLR.2: Random sampling: we have a random sample from the target population.
3. SLR.3: Sample variation in x. The variance of x must not be zero:
$$\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$
4. SLR.4: Zero conditional mean: $\mathsf{E}(u|x) = 0$. Since we have a random sample we can write:
$$\mathsf{E}(u_i|x_i) = 0, \quad \forall\ i = 1, 2, \ldots, n$$

Unbiasedness of OLS Estimators

THEOREM: Unbiasedness of OLS

If all SLR.1-SLR.4 assumptions hold then OLS estimators are unbiased:

$$\mathsf{E}(\hat{\beta}_0) = \beta_0, \quad \mathsf{E}(\hat{\beta}_1) = \beta_1$$

PROOF:

This theorem says that the centers of the sampling distributions of OLS estimators (i.e. their expectations) are equal to the unknown population parameter.

Notes on Unbiasedness

- Unbiasedness is feature of the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ that are obtained via repeated random sampling.
- As such, it does not say anything about the estimate that we obtain for a given sample. It is possible that we could obtain an estimate which is far from the true value.
- Unbiasedness generally fails if any of the SLR assumptions fail.
- SLR.2 needs to be relaxed for time series data. But there are ways that it cannot hold in cross-sectional data as well.
- If SLR.4 fails then the OLS estimators will generally be biased. This is the most important issue in nonexperimental data.
- If x and u are correlated then we have **biased estimators**.
- **Spurious correlation**: we find a relationship between y and x that is really due to other unobserved factors that affect y.

Unbiasedness of OLS: A Simple Monte Carlo Experiment

- Population model (DGP - Data Generating Process):

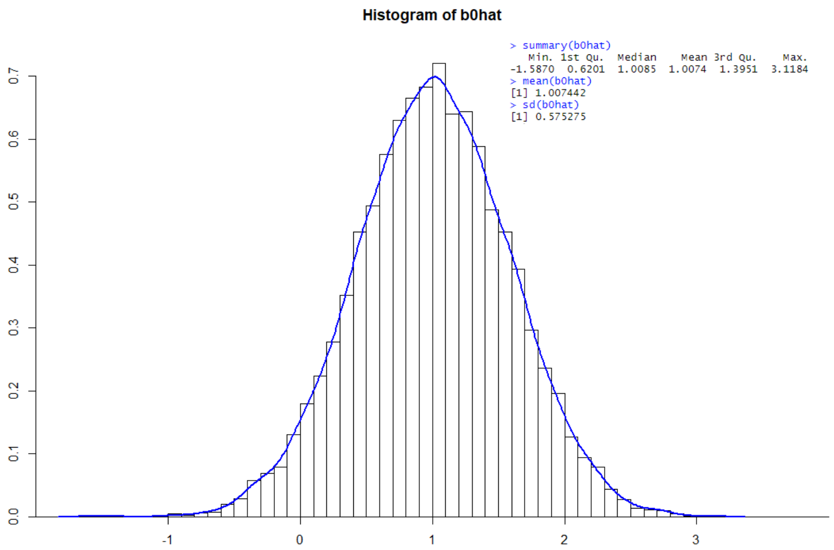
$$y = 1 + 0.5x + 2 \times N(0, 1)$$

- True parameter values are known: $\beta_0 = 1$, $\beta_1 = 0.5$, $u = 2 \times N(0, 1)$ (what is the variance of u?). $N(0, 1)$ represents a random draw from the standard normal distribution.
- The values of x are drawn from the Uniform distribution: $x = 10 \times Unif(0, 1)$
- Using random numbers we can generate artificial data sets. Then, for each data set we can apply the OLS method to find estimates.
- After repeating these steps many times, say 1000, we would obtain 1000 slope and intercept estimates. Then we can analyze the sampling distribution of these estimates.
- This is a simple example of Monte Carlo simulation experiment. These experiments may be useful in analyzing properties of estimators.

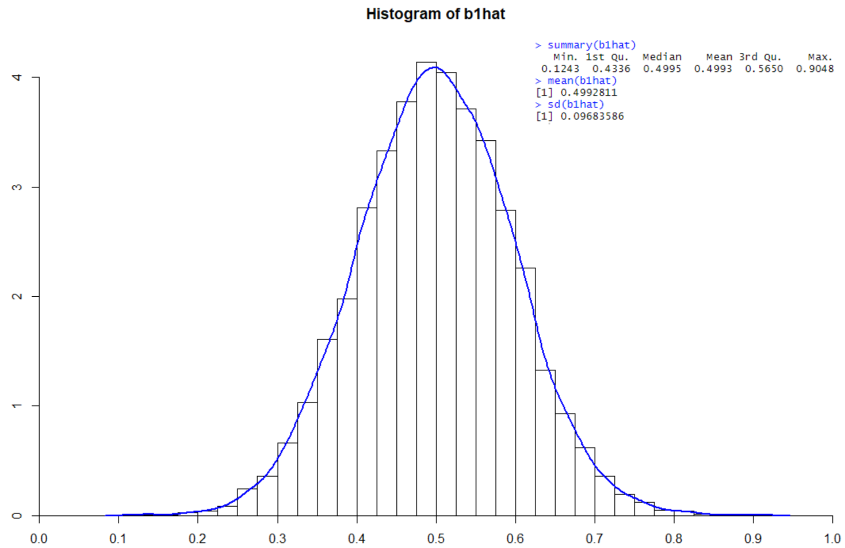
R Implementation

```r
# Set the random seed
# So that we will obtain the same results
# Otherwise, simulation results will change
set.seed(1234567)

# set sample size
n <- 50
# the number of simulations
MCreps <- 10000

# set true parameters: betas and sd of u
beta0 <- 1
beta1 <- 0.5
su <- 2

# initialize b0hat and b1hat to store results later:
b0hat <- numeric(MCreps)
b1hat <- numeric(MCreps)

# Draw a sample of x
# this is going to be fixed in repeated samples
x <- 10*runif(n,0,1)

# repeat MCreps times:
for(i in 1:MCreps) {
  print(i)
  # Draw a sample of y:
  u <- rnorm(n,0,su)
  y <- beta0 + beta1*x + u
  # estimate parameters by OLS and store them in the vectors
  bhat <- coefficients( lm(y~x) )
  b0hat[i] <- bhat["(Intercept)"]
  b1hat[i] <- bhat["x"]
}
# draw histogram and summary statistics
```

Sampling Distribution of $\hat{\beta}_0$



Sampling Distribution of $\hat{\beta}_1$



Variances of the OLS Estimators

- Unbiasedness of OLS estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$ is a feature about the center of the sampling distributions.
- We should also know how far we can expect $\hat{\beta}_1$ to be away from β_1 on average.
- In other words, we should know the sampling variation in OLS estimators in order to establish efficiency and to calculate standard errors.
- SLR.5: Homoscedasticity (constant variance assumption): This says that the variance of u conditional on x is constant:

$$\text{Var}(u|x) = \sigma^2$$

- This is also the unconditional variance: $\text{Var}(u) = \sigma^2$
- Using this assumption we can say that u and x are statistically independent: $\text{E}(u|x) = \text{E}(u) = 0$ and $\text{Var}(u|x) = \text{Var}(u) = \sigma^2$

Variances of the OLS Estimators

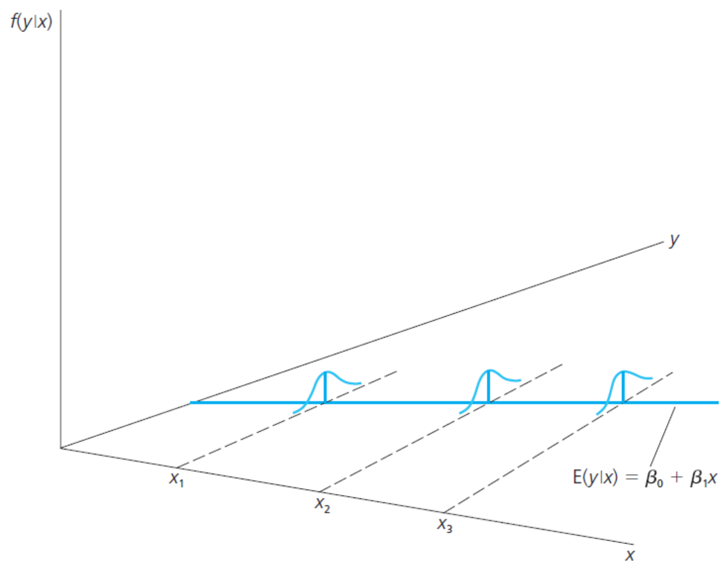
- Assumptions SLR.4 and SLR.5 can be rewritten in terms of the conditional mean and variance of y:

$$\text{E}(y|x) = \beta_0 + \beta_1 x$$

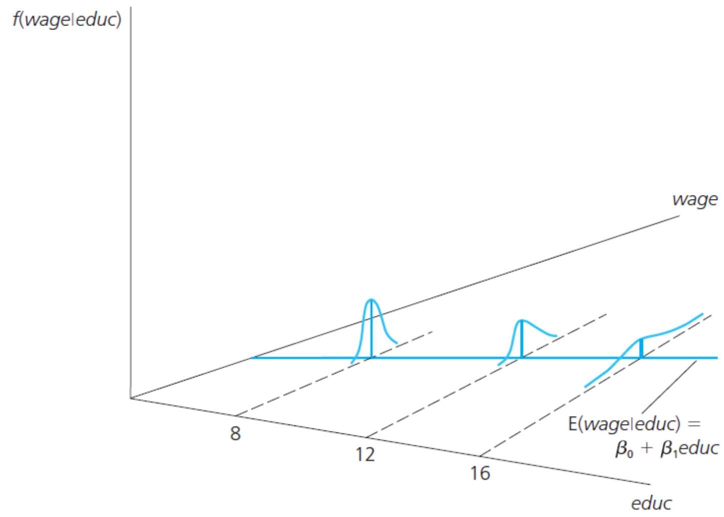
$$\text{Var}(y|x) = \sigma^2$$

- Conditional expectation of y given x is linear in x.
- Conditional variance of y given x is constant and equal to the error variance, σ^2.

Simple Regression Model under Homoscedasticity



An example of Heteroskedasticity



31

Sampling Variances of the OLS Estimators

Under assumptions SLR.1 through SLR.5:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{s_x^2}$$

and

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- These formulas are not valid under heteroscedasticity (if SLR.5 does not hold).
- Sampling variances of OLS estimators increase with the error variance and decrease with the sampling variation in x.

Error Terms and Residuals

- Error terms and residuals are not the same.
- Error terms are not observable:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- Residuals can be calculated after the model is estimated:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

- Residuals can be rewritten as a function of error terms:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i = \beta_0 + \beta_1 x_i + u_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) x_i$$

- From unbiasedness: $\text{E}(\hat{u}) = \text{E}(u) = 0$.

Estimating the Error Variance

- We would like to find an unbiased estimator for σ^2.
- Since by assumption we have $\mathsf{E}(u^2) = \sigma^2$ an unbiased estimator is:
$$\frac{1}{n}\sum_{i=1}^{n} u_i^2$$
- But we cannot use this because we do not observe u. Replacing the errors with the residuals:
$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{SSR}{n}$$
- However, this estimator is **biased**. We need to make degrees of freedom adjustment. Thus, the unbiased estimator is:
$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{SSR}{n-2}$$
- degrees of freedom (dof) = number of observations - number of parameters = n-2

Standard Errors of OLS estimators

- The square root of the variance of the error term is called **the standard error of the regression (SER)**:
$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2} = \sqrt{\frac{SSR}{n-2}}$$
- $\hat{\sigma}$ is also called the *root mean squared error*.
- Standard error of the OLS slope estimate can be written as :
$$\mathsf{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{\hat{\sigma}}{s_x}$$
- Standard errors summarize the uncertainty surrounding the coefficient estimates.

t-test

- The null hypothesis: $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$. Note that the null and the alternative hypotheses always involve the **true but unknown parameters** (not the regression estimates).
- Standard statistical packages, including R, report statistics for that null and alternative hypothesis. The t-tests reported in the standard output is always two-sided.
- To conduct the test, we can use t-statistic
$$t - statistic = \frac{estimated\ value - hypothesizedvalue}{standard\ error}$$
Under the assumption of normality, the t-statistic follows the t-distribution with appropriate degrees of freedom. We will re-visit t-tests after we finish multiple linear regression analysis.

t-test

- If we assume that $H_0 : \beta_1 = 0$ is true then the t-test statistic becomes
$$t - statistic = \frac{estimated\ value}{standard\ error} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$
which follows t distribution with $n-k-1$ degrees of freedom where k is the number of explanatory variables. In the simple regression $k=1$ so that it follows $t(n-2)$.
- Decision rule: Reject H_0 if the absolute value of the t-statistic is larger than 1.96 in large samples. For small samples $n < 120$ we need to look up critical values (or use the p-value).

R Example: t-test

```
> data("wage1")
> res1 <- lm(wage1$wage ~ wage1$educ)
> summary(res1)
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
wage1$educ   0.54136    0.05325  10.167   <2e-16 ***

Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,Adjusted R-squared:  0.1632
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

$$\widehat{wage} = -0.9 + 0.54\ educ, \quad R^2 = 0.165$$

$se(\hat{\beta}_0) = 0.685$, $se(\hat{\beta}_1) = 0.0533$. T-test on the slope parameter:

$$t_{\beta_1} = \frac{0.54136}{0.05325} = 10.167 \sim \ t(524)$$

R Example: t-test

$$\widehat{wage} = -0.9 + 0.54\ educ, \quad R^2 = 0.165$$

$se(\hat{\beta}_0) = 0.685$, $se(\hat{\beta}_1) = 0.0533$. T-test on the slope parameter:

$$t_{\beta_1} = \frac{0.54136}{0.05325} = 10.167 \sim \ t(524)$$

Decision: Reject $H_0 : \beta_1 = 0$ in favor of the alternative because $|t_{\beta_1}| > 1.96$. We can say that education level has a statistically significant impact on wage. The slope parameter is significantly different from zero.

We can also use p-value to make a decision. But we compare p-value with a theoretical type-I error probability, α, say 5% or $\alpha = 0.05$. We reject the null hypothesis if $p - value < \alpha$. This is reported in the last column of the R output (column `Pr(>|t|)`).

Regression through the Origin

- In some rare cases we want $y = 0$ when $x = 0$. For example, tax revenue is zero whenever income is zero.
- We can redefine the simple regression model without the constant term as follows: $\tilde{y} = \tilde{\beta}_1 x$.
- Using OLS principle

$$\min \sum_{i=1}^{n} (\tilde{y} - \tilde{\beta}_1 x_i)^2$$

- First Order Condition:

$$\sum_{i=1}^{n} x_i(\tilde{y} - \tilde{\beta}_1 x_i) = 0$$

- Solving this we obtain the OLS estimator of the slope parameter:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Regression through the Origin

- A simple regression with both intercept and slope parameter:

$$salary = \beta_0 + \beta_1 roe + u$$

- No intercept

$$salary = \beta_1 roe + u$$

  Note that the intercept is forced to be zero (hence the regression through the origin)
- No slope

$$salary = \beta_0 + u$$

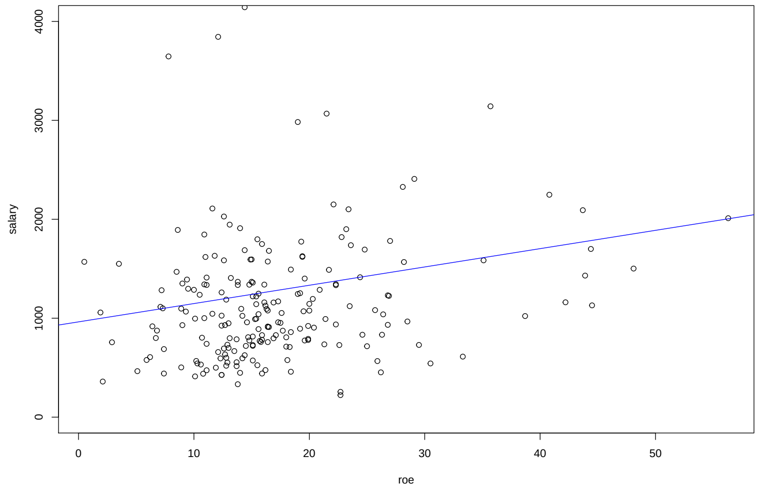
  In fact, there is no explanatory variable.
- Now, let's estimate each of these models and plot the sample regression functions.

Regression through the Origin

```
> res1 <- lm(salary ~ roe, data = ceosal1)

# Regression through the origin:
> res2 <- lm(salary ~ 0 + roe, data = ceosal1)

# Regression on a constant
> res3 <- lm(salary ~ 1, data = ceosal1)

# Plot
> plot(roe, salary, ylim = c(0,4000))
> abline(res1,col="blue")
> abline(res2,col="red")
> abline(res3,col="black")
```

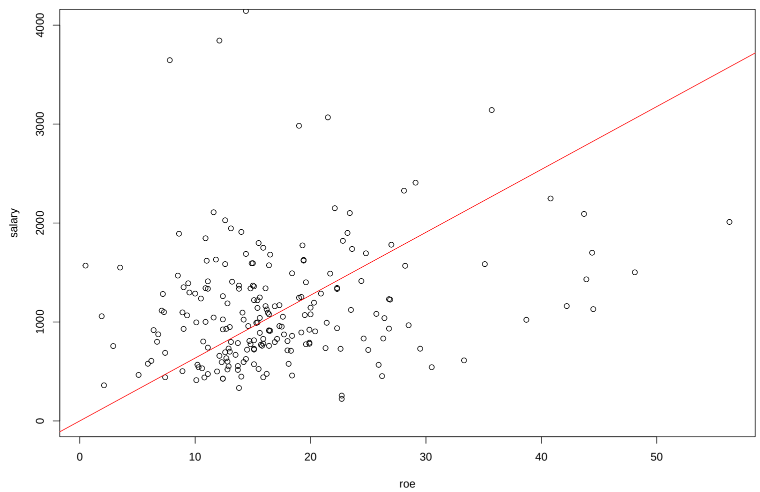
Full Regression with nonzero slope and intercept

$$\widehat{salary} = 963.2 + 18.5 roe$$



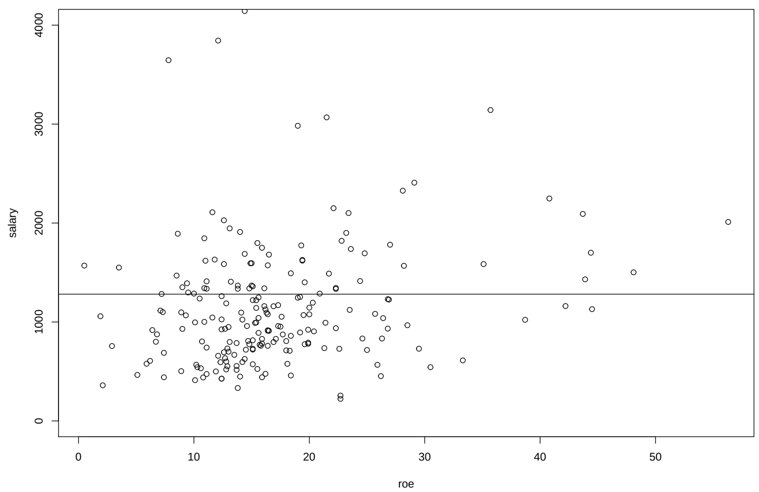
Regression through the origin

$$\widehat{salary} = 63.54 roe$$



Regression on a constant

$$\widehat{salary} = 1281$$



All in one graph