REVIEW: REGRESSION AND CAUSALITY (SKIPPED)

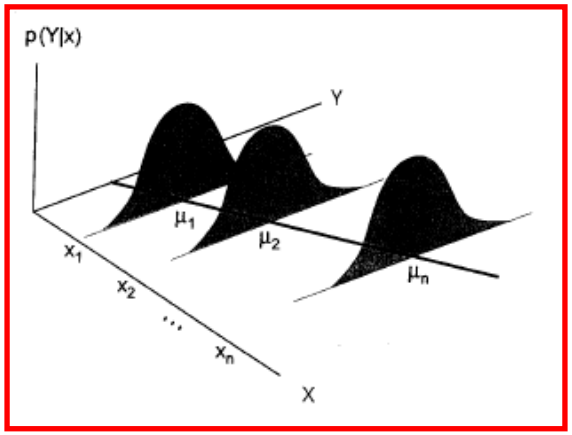
- Def. causality: The change in a ***cause*** *induces in a predictable* manner a particular change in an ***effect***.
    - The role of endogenous and exogenous variable depends on the context. E.g.: The relationship between income and education can either be:
      [a] the parent's income determines the child's education
      [b] A person's earlier education level determines a person income
- Time ordering: The cause *X* ***precedes*** the effect *Y*. This is the closest we get in ***empirical research*** to causality)
    - Note: in the income-education example the exogenous variable always precedes the endogenous variable.
- Co-variation: The variables *X* and *Y* co-vary jointly in a systematic manner (***not causality***).
    - No statement can be made about the ***influence*** that one variable has on the variation of another variable. Thus there is *no cause and effect* relationship.
    - Both variables seem to just correlate in their variation for ***what-ever reasons***.
- Spurious relationships: The empirical co-variation between *X* and *Y* can be induced by their ***joint relationship*** to another variable *Z* (or a set of variables).

- The variable *Z* is called ***confounder***.
- Bivariate regression analysis cannot control for confounding effects because only the two variables *X* and *Y* define the regression system.

- Theory based research (or confirmatory research) makes ***predictive statements*** about the [a] strength, [b] direction, [c] shape and [d] conditional distributions of $pdf(Y|X)$.

- Regression traces the conditional distribution of *Y* given a particular level of *X* by means of the conditional expectation.
  FOX Fig. 6.1 assumes positive linearity and equal conditional distributions
  We get the conditional expectations
  $\mu_{y_i|x_{i1}} \equiv E(y_i \mid x_{i1}) = \beta_0 + \beta_1 \cdot x_{i1}$

FOX Fig 6.1



REVIEW: NOTATIONAL CONSIDERATIONS (SKIPPED)

- For the *i*-th observation the ***population model*** is $y_i = \beta_0 + \beta_1 \cdot x_{i1} + \varepsilon_i$.
  - None of the population parameters β_0 and β_1 nor the error term ε_i are directly observable using an empirical sample.

- The parameter β_0 is called the ***intercept*** and the parameter β_1 is the ***slope***.
- The parameters β_0 and β_1 are *not indexed* by the individual observations i. Thus they are constant across all observations.
- In the population the error term ε_i, also called ***disturbance***, is directly associated to individual observation by the index i.

- The predicted value of the estimated model is $\hat{y}_i = b_0 + b_1 \cdot x_{i1}$ with the ***residual*** $e_i = y_i - \hat{y}_i$.
  - Some books use α instead of β_0 for the population intercept and a instead of b_0 for the estimated intercept.
  - Some people also write $\hat{\beta}_0$ and $\hat{\beta}_1$ for the estimated parameters and $\hat{\varepsilon}_i$ for the residuals.
  - Some books exclude the intercept when counting the number of variables in the equation; some books include the intercept in the count.
    Hamilton uses the symbol *K* for the number of parameters *including the intercept*. Thus for bivariate regression analysis the number of estimated parameters is $K = 2$.

REVIEW: BASIC INTERPRETATION

- A sample of only ***two data-points*** determines the regression line. Consequently, we will lose two degrees of freedom when performing bivariate regression analysis.
- The linear regression model separates

- the ***conditional expectations*** (conditional on the observation x_{i1})
  $\mu_{y_i|x_{i1}} \equiv E(y_i \mid x_{i1}) = \beta_0 + \beta_1 \cdot x_{i1}$ *(systematic component)*
- from the unobserved ***disturbances*** ε_i *(random component)* that are unique for each observations.

- A disturbance ***deviates from the systematic component***. The set of n disturbances has only $n - K$ degrees of freedom.
- Gaining ***further understanding*** of any potentially inherent pattern in the residuals (*unexplained part*) enhances our ***knowledge***.
  $\Rightarrow$ This gives rise to extensive ***residual analysis*** in regression analysis (see Chapter 4).

REVIEW: KEY ASSUMPTIONS OF REGRESSION ANALYSIS

- Key assumptions:
  1. ***Relationship*** among the independent variable and the dependent variable is ***linear*** (or can be transformed to linearity).
  2. The independent variable X is ***fixed*** (that is, it is ***deterministic variables*** and not influenced by random effects).
     Note: should it for some reason be influenced by randomness, then we need at least assume

that it is uncorrelated with the population disturbances. There is a strong risk of obtaining biased estimates for the regression coefficients.

3. Disturbances at any level of x_i have ***identical distributions***, with zero mean and constant variance.
4. Disturbances are in general assumed to be ***independent*** (uncorrelated) among each other.
5. The independence and identical distribution assumption is abbreviated by statement that the disturbances are i.i.d. (**i**ndependently **i**dentically **d**istributed)
6. The additional *i.i.d.* ***normality*** assumption associated with the distribution of the disturbances allows statistical significance testing of the estimated regression model.

- Notes:

If sample size is very small, then we need central limit theory

  - Arguing purely statistically, only the disturbances are required to be normal *i.i.d*. Neither *Y* nor *X* need to follow a normal distribution.
  - <u>However:</u> A joint bivariate normal distribution of the variables *X* and *Y* is highly desirable because it guarantees ***linearity*** and a ***balanced distribution*** of all data points in the scatterplot; thus minimizing any impact of potential outliers.

REVIEW: ORDINARY LEAST SQUARES ESTIMATION AND VARIANCE DECOMPOSITION

- The method of ordinary least squares fits a straight line through the scatterplot point cloud that minimizes the ***sum of the squared regression residuals***.
- As long as the model has an intercept the regression line always goes through means of X and Y, i.e., the point $(\bar{x}, \bar{y})$ will be on the regression line.

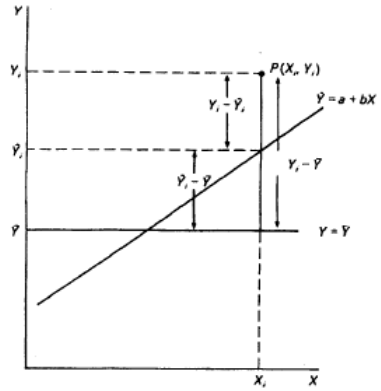


FIGURE 13-11. Decomposition of the total variation.

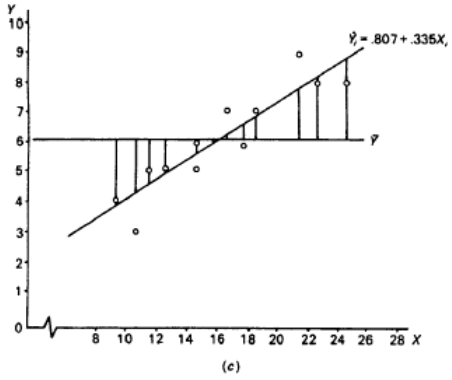
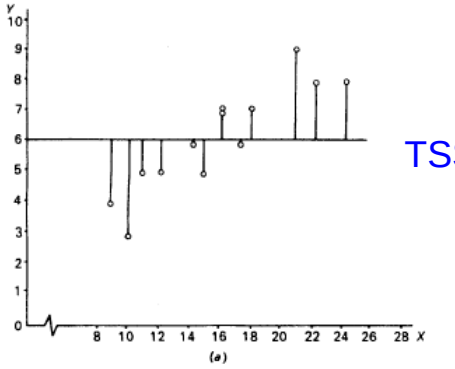
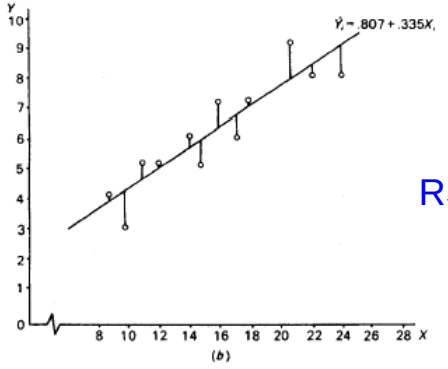


FIGURE 13-12 (continued).



TSS



RSS

FIGURE 13-12. Geometrical representation of (a) total variation, (b) residual variation, and (c) explained variation.

Mean of (x,y) would always on the regression line

- Definition of variance terms: Total Sum of Squares (TSS), Residual Sum of Squares (RSS) and Explained Sum of Squares (ESS)

$$TSS = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 \text{ with } df = n - 1$$

$$RSS = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \text{ with } df = n - K$$

$$ESS = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y} \right)^2 \text{ with } df = K - 1$$

- The validity of decomposition $TSS = ESS + RSS$ relies on the fact that the product $\sum(y_i - \hat{y}_i) \cdot (\hat{y}_i - \bar{y}) = 0$, which can be easily shown later using matrix algebra.
- **Key consequence:** $Cov(\hat{y}, e) = 0 \Rightarrow Cov(X, e) = 0$
  - Because $Cov(\hat{y}, e) = Cov(b_0 + b_1 \cdot X, e) = b_1 \cdot \underbrace{Cov(X, e)}_{=0} = 0.$
  - Therefore, the independent variables ***cannot explain any remaining variation*** in the regression residuals.
    That is, the independent variable has exhausted all information in the regression model.

REVIEW: ESTIMATION OF SLOPE AND INTERCEPT

- We want to ***minimize the sum of squared residuals*** RSS in ordinary least squares regression with the residuals being defined by $e_i = y_i - \underbrace{(b_0 + b_1 \cdot x_{i1})}_{\hat{y}_i}$ (see HAM p 33), that is,

$$\min_{b_0, b_1} RSS = \min_{b_0, b_1} \sum (y_i - \underbrace{(b_0 + b_1 \cdot x_i)}_{=\hat{y}_i})^2$$

- Minimizing this function is done by setting its ***first derivatives*** equal to ***zero*** and solving for the unknown parameters b_0 and b_1:

$$\frac{\partial RSS}{\partial b_0} = -\sum y_i + n b_0 + b_1 \sum x_{i1} \equiv 0$$

$$\frac{\partial RSS}{\partial b_1} = -\sum x_i \cdot y_i + b_0 \sum x_{i1} + b_1 \sum x_{i1}^2 \equiv 0$$

- The first equation shows that the ***regression line must go through the means*** $(\bar{x}, \bar{y})$ of the independent and the dependent variables because,

$$-\sum y_i + nb_0 + b_1 \sum x_{i1} \equiv 0$$
$$\Rightarrow \sum y_i = nb_0 + b_1 \sum x_{i1} \quad | \div n$$
$$\Rightarrow \bar{y} = b_0 + b_1 \cdot \bar{x}$$

Therefore, $b_0 = \bar{y} - b_1 \cdot \bar{x}$.

Inserting this expression for b_0 into $\frac{\partial RSS}{\partial b_1} = 0$ and solving for b_1 gives $b_1 = \frac{n \cdot \sum y_i \cdot x_{i1} - \sum x_{i1} \cdot \sum y_i}{n \cdot \sum x_{i1}^2 - (\sum x_{i1})^2}$.

- Equivalent expressions for the slope parameter b_1 in bivariate regression are

$$b_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{s_{YX}}{s_X^2} = r \cdot \frac{s_Y}{s_X}$$ Slope of regression line

or $b_1 = \frac{Cov[Y,X]}{Var[X]}$ (see Hamiliton p 294).

Question: What would the slope be if we regress X on Y. Answer: $b_{X|Y} = r \cdot \frac{s_X}{s_y}$

- It follows from $e_i = y_i - \underbrace{(b_0 + b_1 \cdot x_{i1})}_{\hat{y}_i}$ and that $(\bar{x}, \bar{y})$ is on the estimated regression line that the ***sum of the regression residuals is zero***, i.e., $\sum e_i = 0$.

Review: R^2 and Adjusted R^2_{adj}

- The goodness of fit measure (proportion of explained variance relative to the total variance) is defined by

$$R^2 \equiv \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- The adjusted goodness of fit takes the degrees of freedom into account because, the more variables are enter into the regression equation, the better the fit of the model will be (recall the perfect fit of the regression through two points)

$$R^2_{adj} \equiv 1 - \frac{RSS/(n-K)}{TSS/(n-1)}$$

For comparing models have different number of independent variables (Considering the degree of freedom)

When *n* is large relative to *K* then the difference between the adjusted and the ordinary R^2 is negligible.

Statistical Inference on the Unknown Regression Parameters

- Question: Why are estimated regression parameters b_0 and b_1 ***random variables*** that have their own distribution?
  Answer: Repeated samples from a population will yield different sets of observations.

Thus the estimated regression parameters will differ from sample to sample and therefore will have a distribution.

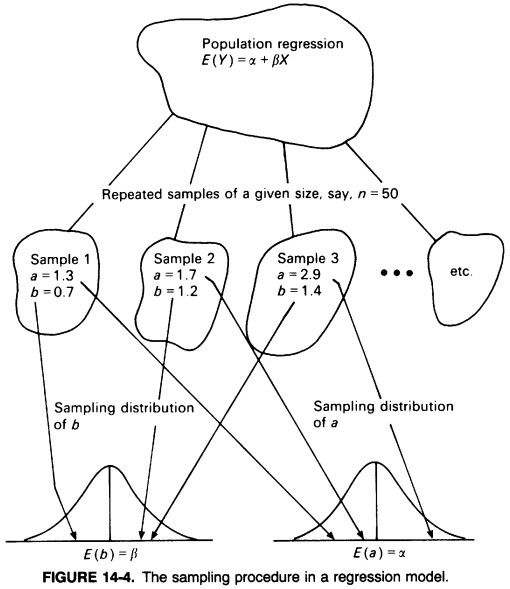


FIGURE 14-4. The sampling procedure in a regression model.

- Assuming *i.i.d.* disturbances ε_i, then the estimated regression parameters b_0 and b_1 will be ***asymptotically normal distributed*** with $b_0 \sim \mathcal{N}(\beta_0, \sigma_{b_0}^2)$ and $b_1 \sim \mathcal{N}(\beta_1, \sigma_{b_1}^2)$ due to the central limit theorem. Thus, if other assumptions are satisfied, the estimated parameters are ***unbiased*** with $E(b_0) = \beta_0$ and $E(b_1) = \beta_1$.
- The square roots of the estimated parameter variances $\hat{\sigma}_{b_0}^2$ and $\hat{\sigma}_{b_1}^2$ are called ***standard errors*** of the estimated regression coefficients.

- Several equations make use of the ***residual standard deviation*** $s_e = \sqrt{\frac{RSS}{n-K}}$. It is also called the ***Root Mean Square Error***. Standard deviation of regression model
- One can show that the standard errors of the regression parameters can be calculated by

$$\sqrt{Var(b_1)} = SE_{b_1} = \frac{s_e}{\sqrt{TSS_X}} \text{ and}$$

$$\sqrt{Var(b_0)} = SE_{b_0} = s_e \cdot \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{TSS_X}} \text{ with } TSS_X = \sum_{i=1}^{n}(X_i - \bar{X})^2$$

- Both standard errors depend on the spread TSS_X. Notice that with increasing spread the estimated regression lines become more precise (i.e., they have smaller standard errors).

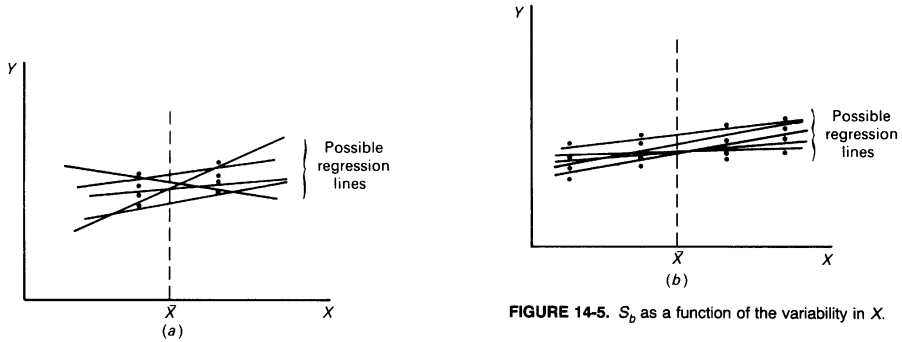


FIGURE 14-5. S_b as a function of the variability in X.

Low uncertainty and unbiasedness of any estimates are desirable properties.

Two-sided test

- Since we work with estimates of standard errors, the distribution of the test statistic $t = \frac{b_1 - \beta_1}{SE_{b_1}}$

  follows a *t-distribution* with $df = n - K$ rather than a standard normal distribution. This *t*-statistic resembles a *z*-transformed variable around a *hypothetical* value β_1.

K = 2 (Intercept & Slope)

- If the exogenous variable X does not explain any variation in the endogenous variable Y then the slope estimate b_1 does not differ significantly from zero.
  - The null hypothesis in this case is $H_0 : \beta_1 = 0$ and the alternative hypothesis becomes $H_1 : \beta_1 \neq 0$
  - Therefore, the test statistic reduces to $t_{obs} = (b_1 - 0) / SE_{b_1} = b_1 / SE_{b_1}$
  - There are scenarios, where a different base-line level, such as $H_0 : \beta_1 = 1$ against $H_1 : \beta_1 \neq 1$, becomes relevant. See the later discussion of ***elasticity***.

F-statistic

- An alternative test statistic in bivariate regression is based on the *F*-statistics.

F test better for multiple regression model

- The test statistic is $F = \frac{ESS/(K-1)}{RSS/(n-K)}$. It follows under the null hypothesis $H_0: \beta_1 = 0$ a *F*-distribution with $df_1 = K-1$ and $df_2 = n-K$
- Note: The *F*-statistic will become important in multiple regression analysis. In bivariate analysis it gives significance levels identical to those of the *t*-test.

One-sided test

- If a theory suggests that the regression parameter is expected to be positive (or negative, respectively), then a ***one-sided test*** becomes appropriate.
  That is, $H_0: \beta_1 \leq 0$ against $H_1: \beta_1 > 0$ or $H_0: \beta_1 \geq 0$ against $H_1: \beta_1 < 0$.
- In those cases the reported ***prob***-value of the software, which is based a two-sided test scenario, is only ***half as large*** because just one tail probability is used.
  Therefore, the more specific one-sided hypotheses have a larger chance of being rejected ***as long as*** the estimated regression coefficient point towards the direction of the alternative hypothesis.

CONFIDENCE INTERVALS

- A $1-\alpha$ ***confidence interval***, e.g., 95% interval, around the hypothetical ***population parameter*** β_1 is given by

$$\Pr(t_{\alpha/2,df} < \frac{b_1 - \beta_1}{SE_{b_1}} < t_{1-\alpha/2,df}) = 1-\alpha$$

$$\Rightarrow \Pr(b_1 - SE_{b_1} \cdot t_{1-\alpha/2,df} < \beta_1 < b_1 - SE_{b_1} \cdot t_{\alpha/2,df}) = 1-\alpha$$

If 0 in the confidence interval, we could not reject None hypothesis.

If we assume $H_0 : \beta_1 = 0$ and the confidence interval covers the value zero, that is, $0 \in [b_1 - SE_{b_1} \cdot t_{1-\alpha/2,df}, b_1 + SE_{b_1} \cdot t_{1-\alpha/2,df}]$, then the ***null hypothesis cannot be rejected*** at the significance level α.

- There are two ***confidence intervals*** associated with the ***predictions*** of the dependent variable $\hat{Y}_i$: [a] for the ***estimated regression line*** and [b] for an ***individual point prediction*** $\hat{Y}_i$
  - For the regression line (book calls it ***mean expected value***) we need to account simultaneously for the ***sampling variation*** in the intercept b_0 and slope b_1.
    In this case the variation of the predicted line $\hat{Y}_{i,\text{line}}$ around the true population line at a given value X_i has the standard error of $SE_{\hat{Y}_{i,\text{line}}} = s_e \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{TSS_X}}$

Closer to the mean of x, higher accuracy of prediction.

Regression line using all the information from sample.
If we do point prediction, only use the information around that point.

- [b] For an ***individual point prediction*** of $Y_{i,\text{point}}$ (book calls it an ***individual case's Y value***) we need to account ***not only*** for the sampling variation of the estimated regression parameters but also for the variation of the individual observation at a given value X_i, that is, the ***potential disturbance*** ε_i.

  Therefore, the standard error increases by one unit of the residual standard deviation:

$$SE_{\hat{Y}_{i,\text{point}}} = s_e\sqrt{1+\frac{1}{n}+\frac{\left(X_i-\bar{X}\right)^2}{TSS_X}}$$

- The term $(X_i - \bar{X})^2$ in both equations signifies that the prediction standard errors will increase the further we move away from the mean $\bar{X}$.
- Furthermore, the larger the sample size *n* gets the narrower the confidence intervals will become, because the TSS_X is increasing and $1/n$ shrinks.

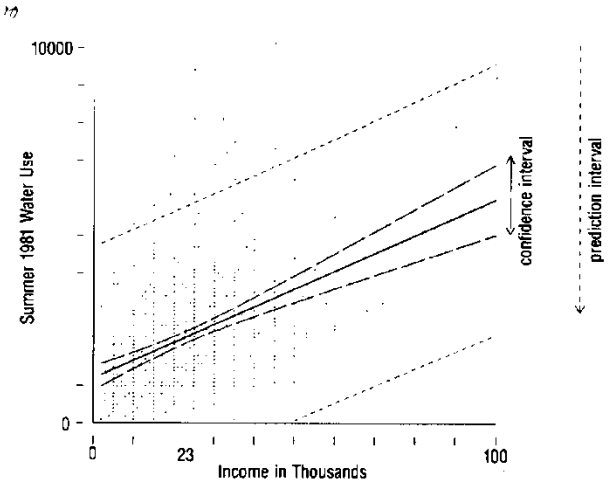


Figure 2.7 Confidence and prediction intervals around regression line.

REGRESSION THROUGH THE ORIGIN

- Regression through the origin should only be considered if there are ***strong theoretical reasons*** to assume that the intercept $\beta_0 = 0$.

- A test outcome, that indicates that the intercept is zero, ***does <u>not</u> justify*** dropping the intercept.
- In R we can suppress the intercept with the statement **`model.lm <- lm(y~-1+x.1)`**, where the negative term **`-1`** denotes to drop the intercept
- Most of the standard test statistics of regression analysis ***become invalid*** if we suppress the intercept. For instance:
    - ***Without an intercept*** the sum of the regression residuals will not necessarily remain zero, that is, $0 \neq \sum_{i=1}^{n} e_i$
    - The goodness of fit measure R^2 is no longer defined.

PROBLEMS ASSOCIATED WITH BIVARIATE REGRESSION ANALYSIS

- ***Omitted additional relevant variable***. The estimate regression parameters become ***biased*** (on average, over many samples from the population, not identical to the population parameter).
- ***Non-linear relationship***. Mis-specified model. Perhaps the residuals indicate autocorrelation.
- Non-constant disturbance variance (***heteroscedasticity***). The variance of the regression residuals changes systematically with either [a] an independent variable or [b] some other external factors currently not considered in the regression model.

- ***Autocorrelation***. The disturbances are no longer independent.
- ***Non-normal disturbances***. Test statistics, that are based on the normal assumptions, become unreliable in small samples.
- ***Influential cases***. Regression analysis is not resistant to outliers. Large squared residuals $\hat{e}_i^2 = \left(y_i - \hat{y}_i\right)^2$ have a strong impact on the ordinary least squares estimation (a squared large distance becomes even larger).
- Some of these problems can be identified by an inspection of the regression residuals.
  Recall: the residuals e_i are ***uncorrelated*** with the predicted values $\hat{Y}_i = b_0 + b_1 \cdot X_i$.
  The same holds for the exogenous variable X_i and the regression residuals e_i.
- Therefore, a scatterplot of the residuals against either the ***predicted value*** or an ***independent variable***, which already is included in the model, ***should not show a systematic pattern***.

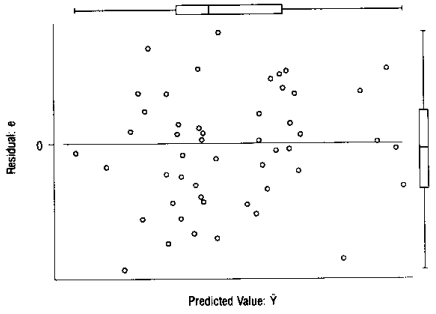


Figure 2.10 "All clear" e-versus-$\hat{Y}$ plot (artificial data).

- Some violations are indicated in the plots

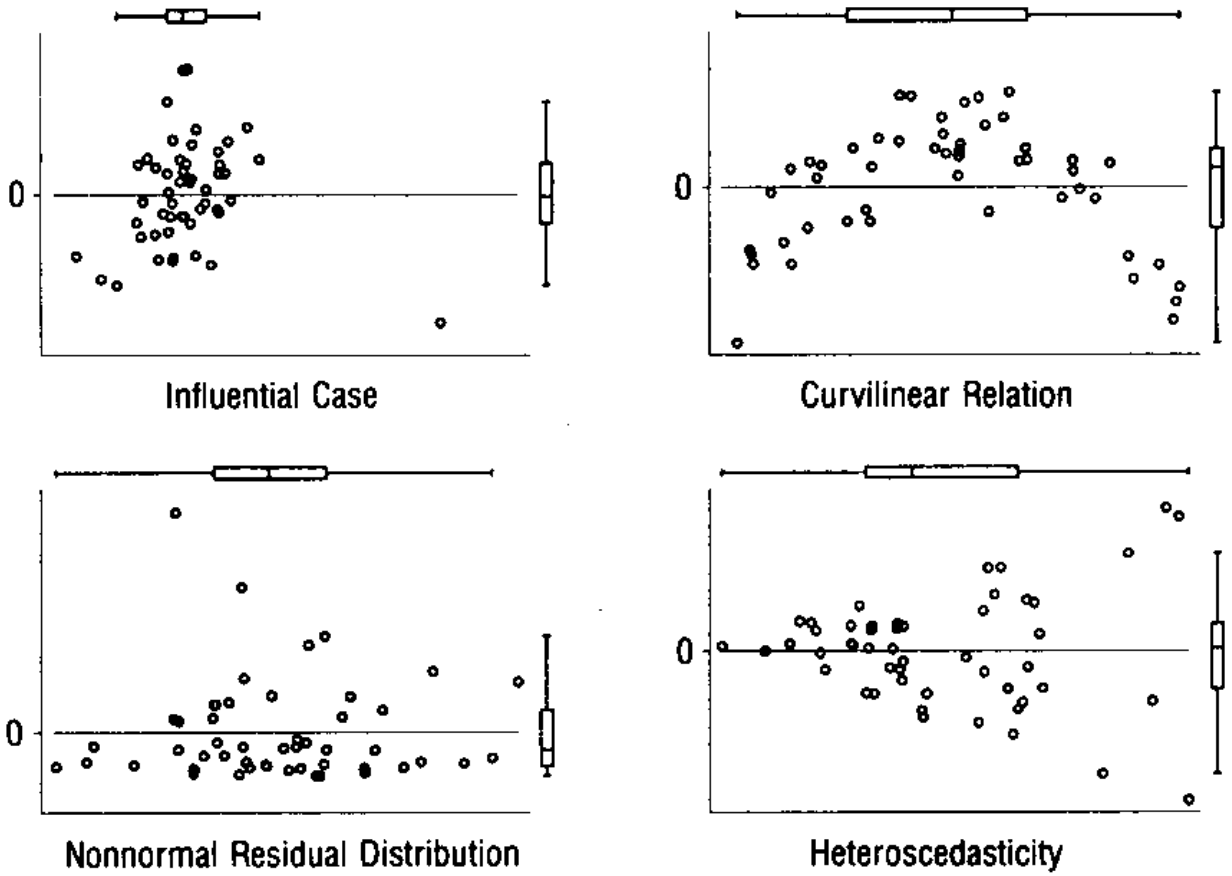


Figure 2.11 Examples of trouble seen in e-versus-$\hat{Y}$ plots (artificial data).

TRANSFORMATIONS OF THE ENDOGENOUS AND EXOGENOUS VARIABLES (FOLLOWING HAMILTON'S NAÏVE APPROACH)

- A scatterplot of $Y \sim X$ or the analysis of the regression residuals may indicate that a transformation of the dependent and/or the independent variables is required.
    - to fix influential cases by pulling them into the population
    - to fix a curvilinear relationship and making the relationship linear
    - to bring the distribution of the regression residuals closer to a symmetric distribution and/or even better the normal distribution
    - to stabilize the variability of the regression residuals by a transformation of the dependent variable.
- This requires experimentation. Both variables could be simultaneously transformed, that is, $Y_i^* = f(Y_i)$ and $X_i^* = g(X_i)$ and the regression residuals must be evaluated until a set of transformations is found, for which no violations of the underlying regression assumptions are detected.
- The functions for the dependent and independent variable $f(\cdot)$ and $g(\cdot)$, respectively, can be different.

Do the transformation because we want the relationship between variables become linear

- In the transformed model predictions can be made $\hat{Y}_i^* = b_0 + b_1 \cdot X_i^*$
- The inverse function is applied on the predicted values $\hat{Y}_i = f^{-1}(\hat{Y}_i^*)$ and the curvilinear relationship between $\hat{Y}_i$ and X_i can be graphed in the original measurement units.
- For the Box-Cox transformation $Z^* = \frac{Z^\lambda - 1}{\lambda}$ its inverse transformation is $Z = \left(Z^* \cdot \lambda + 1\right)^{1/\lambda}$.

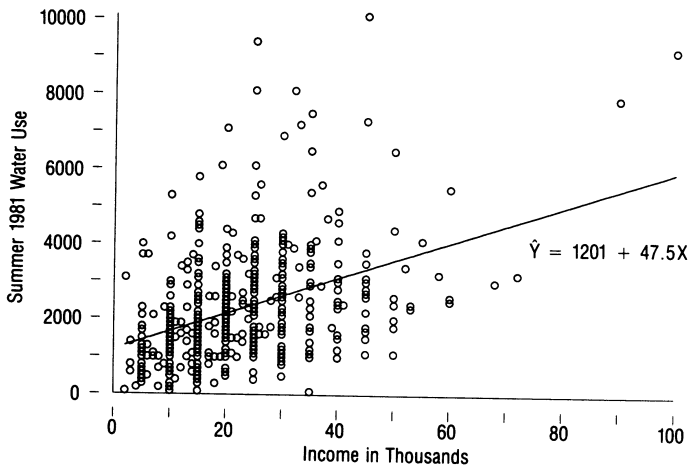


Figure 2.4 Scatterplot with regression line.

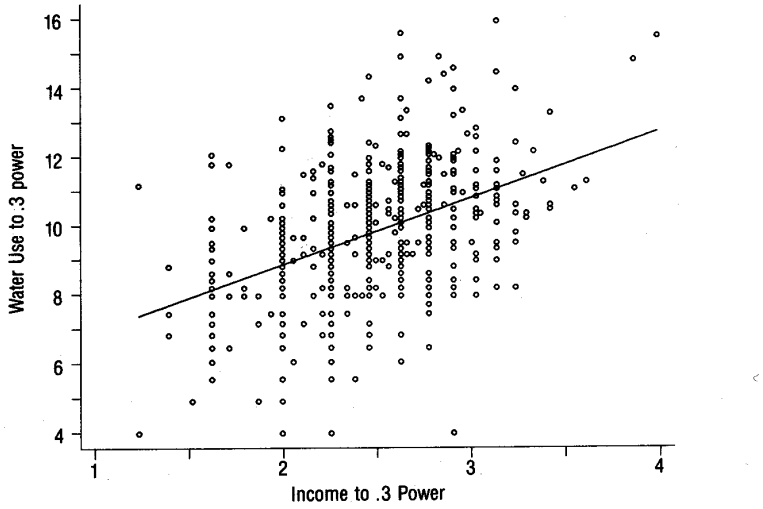


Figure 2.15 Transformed water use versus transformed household income.

Problems in this approach:
1. Linear regression assume residuals are symetric, not indecent and dependent variable are normal distributed

Relationship becomes linear

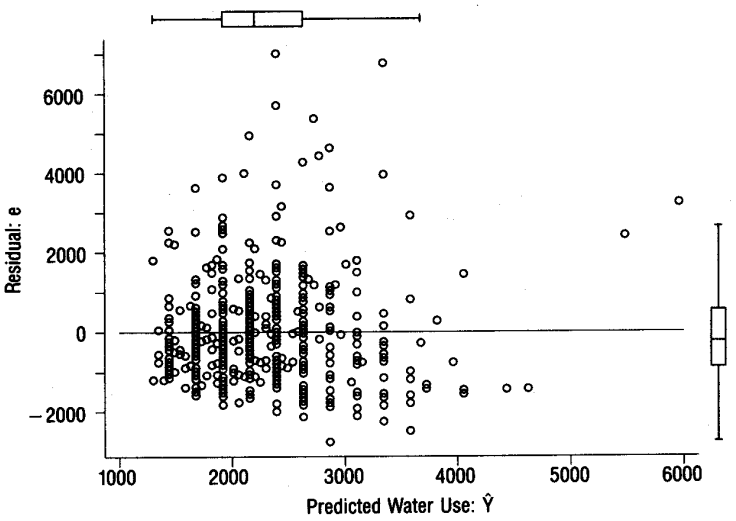


Figure 2.9 e-versus-$\hat{Y}$ plot for water-use regression.

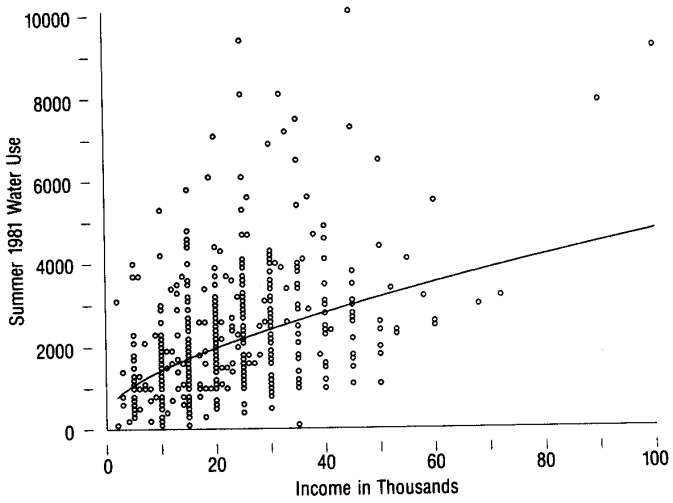


Figure 2.17 Curvilinear relation of water use to income.

STATISTICALLY RIGOROUS TRANSFORMATION APPROACH

- The naïve approach exhibits two problems:
  1. The objective is not that the dependent variable is symmetrically or normal distributed, but rather that the ***regression residuals*** are ***symmetrically or normal distributed***.
  2. The predicted values after the inverse transformation back into the original measurement units become ***conditional medians*** rather than conditional expectations $\mu_{y_i|x_{i1}} \equiv E(y_i \mid x_{i1})$

- To overcome these problems the best approach is

1. Transform all independent variable to become approximately symmetry. This enhances the linearity of the model $\Rightarrow X^* = g(X)$.
2. Find a transformation $f(\,)$ for the dependent variable Y so that the regression residuals become approximately normal distributed $\Rightarrow e^* = f(Y) - b_0^* + b_1^* \cdot X^*$ with $e \sim N(0, \sigma)$.
3. Predict $\widehat{f(Y)} = b_0^* + b_1^* \cdot X^*$ in the transformed system. Note: all model assumptions should be satisfied in the transformed system.
4. Map $\widehat{f(Y)}$ back into the original measurement units either by the inverse transformation
   a. $f^{-1}\left(\widehat{f(Y)}\right)$ for the predicted median $\hat{Y}_{median}$, or
   b. $f_{corrected}^{-1}\left(\widehat{f(Y)}\right)$ for the predicted expected value $\mu_{y_i|x_{i1}} \equiv E(y_i \mid x_{i1})$
5. The script **BOXCOXBIVARIATEREGRESSION.R** outlines the procedure.

- The same logic can be applied in multiple regression models with more than one independent variable.

ELASTICITY

- To estimate a non-linear exponential model

$$y = \exp(b_0) \cdot x^{b_1} \cdot \exp(\varepsilon)$$

by linear regression the model can be transformed into the *log-log* form. This gives the transformed model of the form

$$\ln(y) = b_0 + b_1 \cdot \ln(x) + \varepsilon.$$

- In this model the estimated regression coefficient b_1 is interpreted as a ***relative rate of change*** (i.e., percentage change) at a given value y_0 and x_0

$$b_1 = \frac{\%\Delta y}{\%\Delta x} = \frac{\frac{\Delta y}{y_0}}{\frac{\Delta x}{x_0}} = \frac{\frac{y_0 - y}{y_0}}{\frac{x_0 - x}{x_0}}.$$

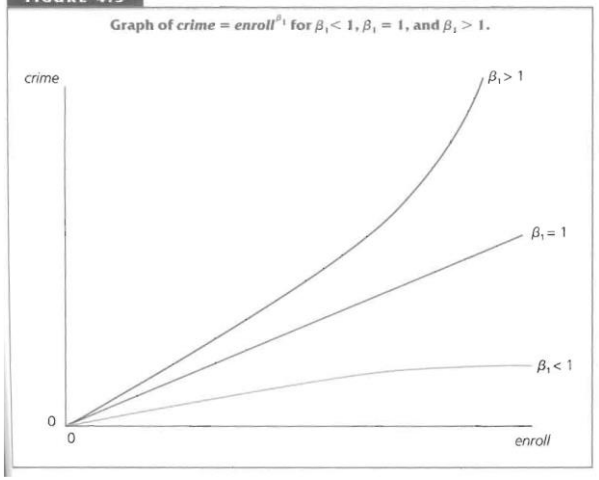
Which may be evaluated at any feasible value x_0 and in particular for $\bar{x}$. The estimate b_1 in the log-log model is called in economics the "***elasticity***".

- To learn more about the economic concept of elasticity start with http://en.wikipedia.org/wiki/Elasticity_(economics)

- The value $b_1 = 1$ is the neutral value where any relative change of x is equal to a relative change in y.
Therefore, the meaningful null hypothesis becomes $H_0: \beta_1 = 1$ against $H_1: \beta_1 \neq 0$ Should be 1

FIGURE 4.5

Graph of *crime* = *enroll*$^{\beta_1}$ for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.



- Interpretation:
  - For $b_1 > 1$ one observes increasing rates of return, that is, y changes relatively faster than x.
  - Whereas for $0 < b_1 < 1$ one observes decreasing rates of return.
- See for example graphs of the exponential model $crime = \exp(b_0) \cdot enroll^{b_1}$.
- The interpretation of models with mixtures of log-transformed variables is given in the table below:

Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

interpreting for linear regression

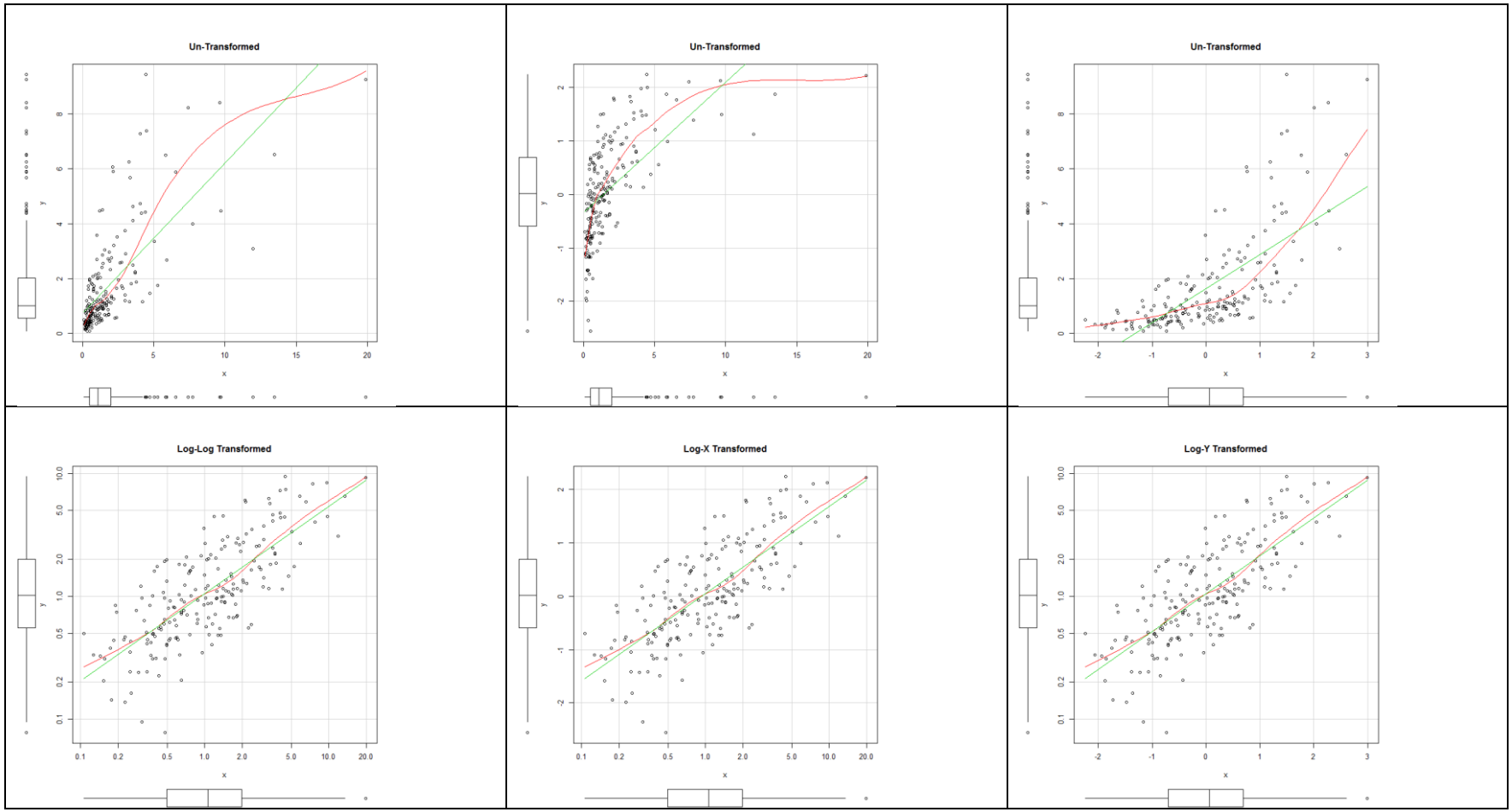
Example 2.11, in the log-log model, β_1 is the **elasticity** of y with respect to x. Table 2.3 warrants careful study, as we will refer to it often in the remainder of the text.

EXCURSION: MATHEMATICAL BACKGROUND RULES (SKIPPED)

- The derivative of the log-function is $\frac{\partial \ln(x)}{\partial x} = \frac{1}{x}$
- The chain rule of derivatives gives $\frac{\partial f(y)}{\partial x} = \frac{\partial f(y)}{\partial y} \cdot \frac{\partial g(x)}{\partial x}$ with $y = g(x)$.
- Let us explore the exponential model $\ln(y) = b_0 + b_1 \cdot \ln(x)$.
  Then for a change ∂x in X we get the change in $\frac{\partial \ln(y)}{\partial x} = b_1 \cdot \frac{\partial \ln(x)}{\partial x} \Leftrightarrow \frac{1}{y} \cdot \frac{\partial y}{\partial x} = b_1 \cdot \frac{1}{x}$.
  Therefore, solving the expression for b_1 gets the x elasticity of y: $b_1 = \left(\frac{x}{y}\right) \cdot \frac{\partial y}{\partial x} = \frac{\frac{\partial y}{y}}{\frac{\partial x}{x}}$
  which is approximately $b_1 \approx \frac{\%\Delta y}{\%\Delta x} = \frac{\frac{y_0 - y}{y_0}}{\frac{x_0 - x}{x_0}}$.

Other Nonlinear Bivariate Relationships (skipped)

- Positive Relationship ($b_1 > 0$) between $Y \sim X$



Red: Loess line, Green: Trend line

- Negative Relationship ($b_1 < 0$) between Y~X