2. OLS-Schätzung linearer Regressionsmodelle

2.1 Formulierung linearer Regressionsmodelle

Einfaches lineares Regressionsmodell:

Das einfache lineare Regressionsmodell ist die simpelste Form eines ökonometrischen Modells und untersucht lediglich den Zusammenhang zwischen zwei Variablen, d.h. den Effekt einer Variablen auf eine andere Variable

Regressionsgleichung:

$$(2.1) \quad y = \beta_0 + \beta_1 x + u$$

Zur Terminologie:

y	x	u	β
Abhängige Variable	Erklärende Variable	Störterm	Regressionsparameter
Zu erklärende Variable	Unabhängige Variable	Fehlerterm	β_0: Konstante (Achsenabschnitt)
Endogene Variable	Exogene Variable		β_1: Steigungsparameter
Regressand	Regressor		

Der Steigungsparameter β_1 beschreibt den linearen Zusammenhang zwischen x und y, falls alle Faktoren in u konstant gehalten werden:

$$(2.2) \quad \Delta y = \beta_1 \Delta x \quad \text{falls } \Delta u = 0$$

Beispiel: Erklärung von Löhnen

Ein einfaches lineares Regressionsmodell, das den Lohn (wage) ausschließlich durch die Ausbildungszeit (educ) sowie andere nicht beobachtbare Einflussfaktoren erklärt, lautet (Terminologie: „wage wird auf educ regressiert"):

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + u$$

Falls der Lohn in Euro pro Stunde und die Ausbildungszeit in Jahren gemessen werden, stellt β_1 die Veränderung des Stundenlohns bei einer Erhöhung der Ausbildungszeit um ein Jahr dar. Die wesentliche Annahme ist dabei, dass alle anderen unbeobachtbaren Einflussfaktoren konstant gehalten werden:

$$\Delta \text{wage} = \beta_1 \Delta \text{educ} \quad \text{falls } \Delta u = 0$$

Wesentliche Voraussetzung für diese Interpretation:

Der Störterm u und die erklärende Variable x dürfen nicht miteinander in Beziehung stehen, wobei hier die Unkorreliertheit bei der Betrachtung des Korrelationskoeffizienten (d.h. der linearen Korrelation) nicht ausreichend ist

Wesentliche Annahme für den Störterm u:

$$(2.3) \quad E(u|x) = E(u)$$

Diese Annahme besagt, dass der Störterm u im Erwartungswert von der erklärenden Variablen x unabhängig (mean independent) ist, d.h. der bedingte Erwartungswert von u, gegeben ein beliebiger Wert von x, ist gleich dem unbedingten Erwartungswert von u und damit konstant.

In der Regel wird zudem angenommen, dass der Erwartungswert von u (also bezogen auf die Grundgesamtheit) gleich null ist:

$$(2.4) \quad E(u) = 0$$

Diese Annahme ist völlig unschädlich und nicht einschränkend, falls die Konstante β_0 in das einfache lineare Regressionsmodell einbezogen wird. Aus der Kombination von (2.3) und (2.4) ergibt sich dann:

$$(2.5) \quad E(u|x) = 0$$

Beispiel: Erklärung von Löhnen

Wie zuvor wird wieder das einfache lineare Regressionsmodell betrachtet, das den Lohn (wage) ausschließlich durch die Ausbildungszeit (educ) sowie weitere nicht beobachtbare Einflussfaktoren erklärt:

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + u$$

Angenommen wird nun, dass der Störterm u nur den unbeobachtbaren Faktor Begabung (abil) beinhaltet. Daraus folgt nach Annahme (2.3), dass die durchschnittliche Begabung bei allen Ausbildungszeiten gleich ist. Das heißt z.B. für den Fall dreier Ausbildungszeiten von 8, 10 und 16 Jahren:

$$E(\text{abil}|x=8) = E(\text{abil}|x=10) = E(\text{abil}|x=16) = E(\text{abil})$$

Die obige Annahme (2.3) wäre dagegen falsch, falls z.B. in der Realität die durchschnittliche Begabung positiv mit der Ausbildungszeit korreliert ist.

Generell ergibt sich aufgrund von Gleichung (2.3) für die Grundgesamtheit:

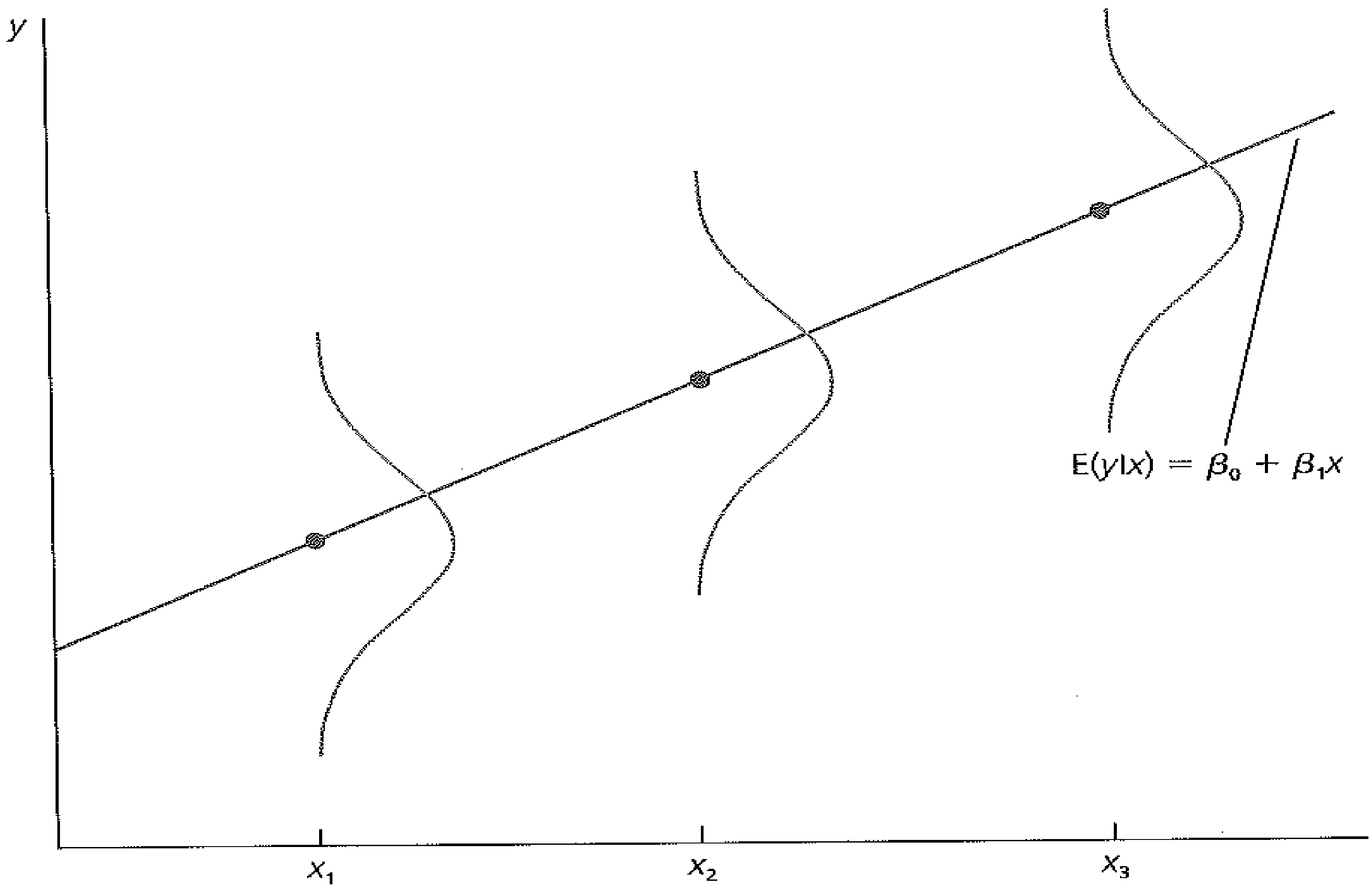
$$y = \beta_0 + \beta_1 x + u$$

$$E(y|x) = \beta_0 + \beta_1 x + E(u|x)$$

$$(2.6) \quad E(y|x) = \beta_0 + \beta_1 x$$

Graphische Darstellung:

E(y|x) as a linear function of x.



Generell ist die Annahme der Unkorreliertheit zwischen der erklärenden Variablen x und den unbeobachtbaren Einflussfaktoren im Störterm u unrealistisch, so dass Gleichung (2.3) in der Regel verletzt ist. Damit ist aber auch die ceteris paribus Betrachtung problematisch. Aus diesem Grund sollten weitere (d.h. sämtliche relevante) erklärenden Variablen in das lineare Regressionsmodell aufgenommen werden.

Einfachste Form eines multiplen linearen Regressionsmodells (mit zwei erklärenden Variablen):

$$(2.7) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Dabei gilt:

x_1, x_2: Erklärende Variablen
β_0: Konstante
β_1: Dieser Parameter misst den Effekt einer Erhöhung von x_1 auf y, falls alle anderen beobachteten und unbeobachteten Faktoren konstant sind
β_2: Dieser Parameter misst den Effekt einer Erhöhung von x_2 auf y, falls alle anderen beobachteten und unbeobachteten Faktoren konstant sind
u: Störterm

Wesentliche Annahme für den Störterm u:

$$(2.8) \quad E(u|x_1, x_2) = 0$$

Beispiel: Erklärung von Löhnen

Nun wird ein multiples lineares Regressionsmodell betrachtet, das den Lohn (wage) durch die Ausbildungszeit (educ), die Berufserfahrung (exper) sowie weitere nicht beobachtbare Einflussfaktoren erklärt (Terminologie: „wage wird auf educ und exper regressiert"):

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u$$

Im Vergleich zum vorherigen einfachen linearen Regressionsmodell wird hier der Faktor Berufserfahrung aus dem Störterm herausgenommen und explizit in das Regressionsmodell einbezogen. Damit lässt sich nun der Effekt der Ausbildungszeit auf den Lohn unter der Annahme bestimmen, dass die Berufserfahrung konstant bleibt, wohingegen man im einfachen linearen Regressionsmodell unterstellen musste, dass die Berufserfahrung nicht mit der Ausbildungszeit korreliert.

Dabei muss jetzt folgende Annahme getroffen werden:

$$E(u|\text{educ}, \text{exper}) = 0$$

Allgemeine Form eines multiplen linearen Regressionsmodells (mit mehr als zwei erklärenden Variablen):

$$(2.9) \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{k-1} x_{k-1} + \beta_k x_k + u$$

Dabei gilt:

$x_1, x_2, x_3, \ldots, x_{k-1}, x_k$: Erklärende Variablen
β_0: Konstante
β_1: Dieser Parameter misst den Effekt einer Erhöhung von x_1 auf y, falls alle anderen beobachteten und unbeobachteten Faktoren konstant sind
β_2: Dieser Parameter misst den Effekt einer Erhöhung von x_2 auf y, falls alle anderen beobachteten und unbeobachteten Faktoren konstant sind
:
β_k: Dieser Parameter misst den Effekt einer Erhöhung von x_k auf y, falls alle anderen beobachteten und unbeobachteten Faktoren konstant sind
u: Störterm

Wesentliche Annahme für den Störterm u:

$$(2.10) \quad E(u|x_1, x_2, \ldots, x_k) = 0$$

Beispiel: Erklärung von Löhnen

Jetzt wird ein multiples lineares Regressionsmodell betrachtet, das den Lohn (wage) durch die Ausbildungszeit (educ), die Berufserfahrung (exper), die betriebliche Weiterbildung (training), die Ausbildung der Mutter (motheduc) sowie weitere nicht beobachtbare Einflussfaktoren erklärt (Terminolgie: „wage wird auf educ, exper, training und motheduc regressiert"):

$$\text{wage} = \beta_0 + \beta_1\text{educ} + \beta_2\text{exper} + \beta_3\text{training} + \beta_4\text{motheduc} + u$$

Im Vergleich zum vorherigen multiplen linearen Regressionsmodell mit zwei erklärenden Variablen werden jetzt die Faktoren betriebliche Weiterbildung und Ausbildung der Mutter aus dem Störterm herausgenommen und explizit in das Regressionsmodell einbezogen.

Dabei muss jetzt folgende Annahme getroffen werden:

$$E(u|\text{educ, exper, training, motheduc}) = 0$$

Zur Linearität von Regressionsmodellen:

Die Linearität der entsprechenden Gleichungen bezieht sich auf die unbekannten Regressionsparameter β_0, β_1, β_2,…, nicht jedoch auf die Struktur der erklärenden und abhängigen Variablen. Somit können auch Nichtlinearitäten in lineare Regressionsmodelle einbezogen werden.

Populäre Ansätze zur Einbeziehung nichtlinearer Zusammenhänge in lineare Regressionsmodelle sind die Betrachtung von logarithmierten oder quadrierten Variablen wie z.B.:

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{k-1} x_{k-1} + \beta_k x_k + u$$

$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_{k-1} x_{k-1} + \beta_k x_k + u$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \cdots + \beta_{k-1} x_{k-2} + \beta_k x_{k-1} + u$$

$$\log y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log x_2 + \beta_4 x_3 + \cdots + \beta_{k-1} x_{k-2} + \beta_k x_{k-1} + u$$

Das untere Beispiel kann mit y = logy, $x_2 = x_1^2$, $x_3 = \log x_2$, $x_4 = x_3$,…, $x_k = x_{k-1}$ in das in (2.9) formulierte multiple lineare Regressionsmodell überführt werden.

Der Umgang mit derart spezifizierten linearen Regressionsmodellen wird später behandelt. Nichtlineare Regressionsmodelle (in Bezug auf die Regressionsparameter) werden dagegen in dieser Veranstaltung nicht untersucht.

2.2 Schätzung der unbekannten Parameter mit der OLS-Methode

Für die weitere Analyse linearer Regressionsmodelle benötigt man eine Stichprobe vom Umfang n aus der Grundgesamtheit.

Einfaches lineares Regressionsmodell:

$\{(x_i, y_i), i = 1,\ldots, n\}$

Multiples lineares Regressionsmodell mit k erklärenden Variablen:

$\{(x_{i1}, x_{i2},\ldots, x_{ik}, y_i), i = 1,\ldots, n\}$

Unter Einbeziehung der Beobachtungen i = 1,…, n ergeben sich folgende lineare Regressionsmodelle:

$$(2.11) \quad y_i = \beta_0 + \beta_1 x_i + u_i$$

$$(2.12) \quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

Dabei ist z.B. x_{ik} der Wert der erklärenden Variablen k bei Beobachtung i.

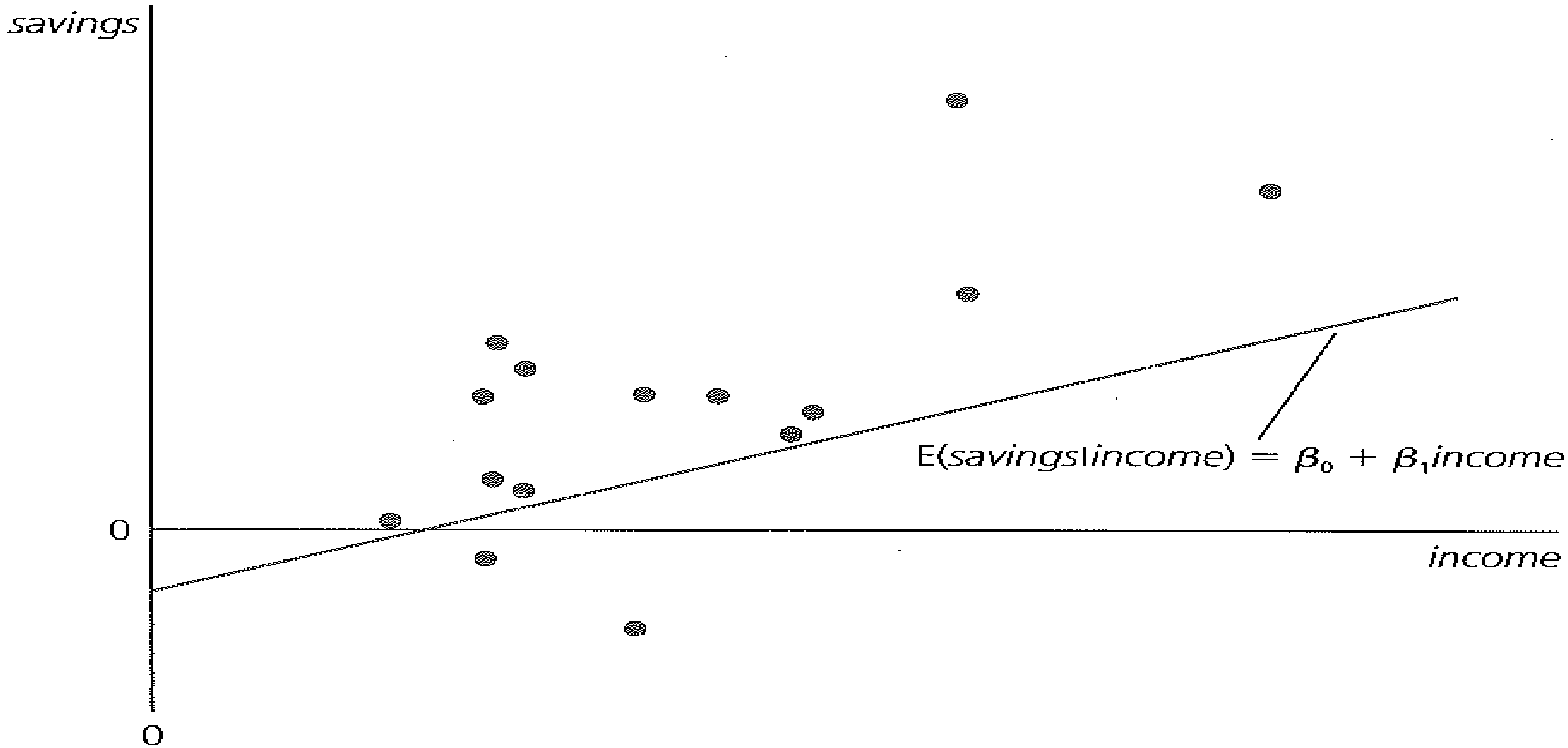
Wesentliche Aufgabe der Regressionsanalyse:

Schätzung der unbekannten Regressionsparameter $\beta_0, \beta_1, \beta_2,\ldots$

Beispiel: Effekt von Einkommen auf Ersparnisse

Für das einfache lineare Regressionsmodell werden Daten von 15 Familien untersucht, d.h. $n = 15$.

Scatterplot of savings and income for 15 families, and the population regression $E(savings|income) = \beta_0 + \beta_1 income$.



Optimierungsproblem bei der Methode der kleinsten Quadrate („OLS, ordinary least squares“) im einfachen linearen Regressionsmodell:

$$\min_{b_0,\, b_1} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Daraus folgen die Bedingungen erster Ordnung für die beiden geschätzten Regressionsparameter:

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

Daraus folgen die OLS-Schätzer für die beiden unbekannten Parameter:

$$\text{(2.13)} \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = \frac{1}{n} \sum_{i=1}^{n} y_i - \hat{\beta}_1 \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)$$

$$\text{(2.14)} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Optimierungsproblem bei der Methode der kleinsten Quadrate im multiplen linearen Regressionsmodell:

$$\min_{b_0, b_1, \ldots} \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik})^2$$

Daraus folgen die Bedingungen erster Ordnung für die k+1 geschätzten Regressionsparameter:

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n} x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n} x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{ik} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

OLS-Regressionswerte („fitted values“) sind geschätzte Werte der abhängigen Variablen:

$$(2.15)\quad \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} \quad \text{für } i = 1, \ldots, n$$

OLS-Regressionsfunktion:

$$(2.16)\quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

Interpretation des geschätzten Steigungsparameters im einfachen linearen Regressionsmodell entsprechend (2.2):

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$(2.17)\quad \Delta\hat{y} = \hat{\beta}_1 \Delta x$$

Der geschätzte Steigungsparameter gibt somit die Veränderung des Regressionswertes an, falls die erklärende Variable x um eine Einheit steigt. Damit kann für jegliche positive oder negative Veränderung von x die Veränderung der abhängigen Variablen y geschätzt werden.

Der geschätzte Achsenabschnitt gibt dagegen den Regressionswert an, wenn x null ist.

Beispiel 1: Erklärung von Löhnen

Die OLS-Schätzung eines lineares Regressionsmodells, das den Stundenlohn in Dollar (wage) durch die Ausbildungszeit in Jahren (educ) erklären soll, hat mit Hilfe einer Stichprobe von n = 526 Personen folgende Regressionsfunktion ergeben:

$$\widehat{wage} = -0{,}90 + 0{,}54 educ$$

Interpretation:

- educ = 0: In diesem Fall beträgt der geschätzte Lohn -90 Cent. Dieser Wert ist aber nicht sehr sinnvoll zu interpretieren, zumal auch keine Beobachtungen in der Stichprobe in diesen Bereich fallen.
- Geschätzter Steigungsparameter beträgt 0,54: Ein zusätzliches Jahr an Ausbildung führt somit zu einem geschätzten Anstieg des Stundenlohns um 54 Cent
- Bei vier Jahren mehr Ausbildung ergibt sich damit eine geschätzte Erhöhung des Stundenlohns um 4·0,54 Dollar = 2,16 Dollar

→ Zu beachten ist, dass eine andere Stichprobe in der Regel zu einer anderen Regressionsfunktion führen würde

Beispiel 1: Erklärung von Löhnen (STATA-Output)

Mit STATA haben sich folgende OLS-Schätzergebnisse gezeigt:

```
reg wage educ

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =  103.36
       Model |  1179.73205     1  1179.73205           Prob > F      =  0.0000
    Residual |  5980.68226   524  11.4135158           R-squared     =  0.1648
-------------+------------------------------           Adj R-squared =  0.1632
       Total |  7160.41431   525  13.6388844           Root MSE      =  3.3784

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5413593    .053248    10.17   0.000     .4367534    .6459651
       _cons |  -.9048517   .6849678    -1.32   0.187    -2.250472    .4407687
------------------------------------------------------------------------------
```

Beispiel 2: Erklärung von Vorstandsgehältern

Mit Hilfe eines linearen Regressionsmodells soll der Effekt der Eigenkapitalrendite von Unternehmen in % (roe) auf das Gehalt von Vorständen dieser Unternehmen in 1000 Dollar (salary) untersucht werden:

$$\text{salary} = \beta_0 + \beta_1 \text{roe} + u$$

Die wesentliche zu untersuchende Hypothese ist, ob eine höhere Eigenkapitalrendite zu einem höheren Gehalt der Vorstände führt, d.h. ob $\beta_1 > 0$. Mit Hilfe einer Stichprobe von n = 209 Vorstandsmitgliedern wurde folgende OLS-Regressionsfunktion geschätzt:

$$\widehat{\text{salary}} = 963{,}191 + 18{,}501\text{roe}$$

Interpretation:

- roe = 0: In diesem Fall ergibt sich ein geschätztes Gehalt von 963191 Dollar
- roe = 30: Hier ergibt sich ein geschätztes Gehalt von 963,191 + 18,501·30 = 1518,221 = 1518221 Dollar
- Geschätzter Steigungsparameter beträgt 18,501: Eine Erhöhung der Eigenkapitalrendite um einen Prozentpunkt führt zu einem geschätzten Anstieg des Gehalts um 18501 Dollar

Beispiel 2: Erklärung von Vorstandsgehältern (STATA-Output)

Mit STATA haben sich folgende OLS-Schätzergebnisse gezeigt:

```
reg salary roe

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  1,   207) =    2.77
       Model |  5166419.33     1  5166419.33           Prob > F      =  0.0978
    Residual |   386566563   207  1867471.32           R-squared     =  0.0132
-------------+------------------------------           Adj R-squared =  0.0084
       Total |   391732982   208  1883331.64           Root MSE      =  1366.6

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         roe |   18.50119   11.12325     1.66   0.098    -3.428195    40.43057
       _cons |   963.1913   213.2403     4.52   0.000     542.7902    1383.592
------------------------------------------------------------------------------
```

Interpretation der geschätzten Parameter in multiplen linearen Regressionsmodellen:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

$$(2.18) \quad \Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \cdots + \hat{\beta}_k \Delta x_k$$

Falls die erklärenden Variablen x_2, x_3, x_4,…, x_k konstant gehalten werden, folgt:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1$$

Der geschätzte Parameter für die erklärende Variable x_1 gibt somit in diesem Fall die Veränderung des Regressionswertes an, falls x_1 um eine Einheit steigt.

Falls x_1, x_2, x_3,…, x_{k-1} konstant gehalten werden, ergibt sich:

$$\Delta\hat{y} = \hat{\beta}_k \Delta x_k$$

Der geschätzte Parameter für die erklärende Variable x_k gibt somit in diesem Fall die Veränderung des Regressionswertes an, falls x_k um eine Einheit steigt.

Die geschätzten Parameter können somit als geschätzte partielle Effekte interpretiert werden, d.h. bei der Schätzung des Effektes einer Variablen wird für die anderen erklärenden Variablen kontrolliert. Dies ist der große Vorteil der Regressionsanalyse (bzw. allgemein ökonometrischer Analysen), d.h. es kann eine ceteris paribus Betrachtung vorgenommen werden, ohne dass ein entsprechendes kontrolliertes Experiment durchgeführt werden muss.

Beispiel 1: Erklärung von College-Noten

Mit Hilfe eines linearen Regressionsmodells soll der Effekt der durchschnittlichen Punktzahl in der Highschool (hsGPA) und der Punktzahl im College-Aufnahmetest (ACT) auf die durchschnittliche College-Punktzahl (colGPA) untersucht werden:

$$\text{colGPA} = \beta_0 + \beta_1 \text{hsGPA} + \beta_2 \text{ACT} + u$$

Dabei wurde mit Hilfe einer Stichprobe von n = 141 Studierenden folgende OLS-Regressionsfunktion geschätzt:

$$\text{col}\hat{\text{G}}\text{PA} = 1{,}29 + 0{,}453\text{hsGPA} + 0{,}0094\text{ACT}$$

Interpretation:

- Die geschätzte Konstante kann nicht sinnvoll interpretiert werden, da hier ACT und hsGPA realitätsfern mit null gleichgesetzt werden
- Geschätzter positiver Effekt der Highschool-Note auf die College-Note: Falls ACT konstant gehalten wird, führt eine um einen Punkt höhere Highschool-Note zu einer geschätzten Erhöhung der College-Punktzahl um 0,453 Punkte
- Dagegen hat die Punktzahl im College-Aufnahmetest einen deutlich geringeren geschätzten Effekt

Beispiel 1: Erklärung von College-Noten (STATA-Output)

Mit STATA haben sich folgende OLS-Schätzergebnisse gezeigt:

```
reg colGPA hsGPA ACT

      Source |       SS       df       MS              Number of obs =     141
-------------+------------------------------           F(  2,   138) =   14.78
       Model |  3.42365461     2  1.7118273            Prob > F      =  0.0000
    Residual |  15.9824447   138  .115814817           R-squared     =  0.1764
-------------+------------------------------           Adj R-squared =  0.1645
       Total |  19.4060993   140  .138614995           Root MSE      =  .34032

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .4534558   .0958129     4.73   0.000     .2640046    .6429071
         ACT |    .009426   .0107772     0.87   0.383    -.0118838    .0307358
       _cons |   1.286328   .3408221     3.77   0.000     .6124191    1.960237
------------------------------------------------------------------------------
```

Beispiel 2: Erklärung von Löhnen

Mit Hilfe eines linearen Regressionsmodells soll der Effekt der Ausbildungszeit in Jahren (educ), der Berufserfahrung in Jahren (exper) und der Betriebszugehörigkeit in Jahren (tenure) auf den Logarithmus des Stundenlohns (logwage) untersucht werden:

$$\text{logwage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + u$$

Dabei wurde folgende OLS-Regressionsfunktion geschätzt:

$$\text{log}\hat{\text{w}}\text{age} = 0{,}284 + 0{,}092\text{educ} + 0{,}0041\text{exper} + 0{,}022\text{tenure}$$

Interpretation:

- Geschätzter positiver Effekt der Ausbildungszeit: Falls exper und tenure konstant gehalten werden, führt eine um ein Jahr höhere Ausbildungszeit zu einer geschätzten Erhöhung des Logarithmus des Lohnes um 0,092 (genaue Interpretation von logarithmierten Variablen siehe später)
- Entsprechend liegen geschätzte positive Effekte von exper und tenure vor, wenn jeweils die anderen erklärenden Variablen konstant gehalten werden

---

Beispiel 2: Erklärung von Löhnen (STATA-Output)

Mit STATA haben sich folgende OLS-Schätzergebnisse gezeigt:

```
reg logwage educ exper tenure
      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  3,   522) =   80.39
       Model |  46.8741806     3  15.6247269           Prob > F      =  0.0000
    Residual |  101.455582   522  .194359353           R-squared     =  0.3160
-------------+------------------------------           Adj R-squared =  0.3121
       Total |  148.329763   525  .282532881           Root MSE      =  .44086

------------------------------------------------------------------------------
     logwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    .092029   .0073299    12.56   0.000     .0776292    .1064288
       exper |   .0041211   .0017233     2.39   0.017     .0007357    .0075065
      tenure |   .0220672   .0030936     7.13   0.000     .0159897    .0281448
       _cons |   .2843595   .1041904     2.73   0.007     .0796755    .4890435
------------------------------------------------------------------------------
```

Residuen (geschätzte Störterme): Differenz zwischen den tatsächlich beobachteten Werten der abhängigen Variablen und den OLS-Regressionswerten

$$(2.19) \quad \hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik} \quad \text{für } i = 1, \ldots, n$$

Falls Residuum positiv:

Tatsächlich beobachtete abhängige Variable ist größer als der entsprechende Regressionswert und wird somit unterschätzt

Falls Residuum negativ:

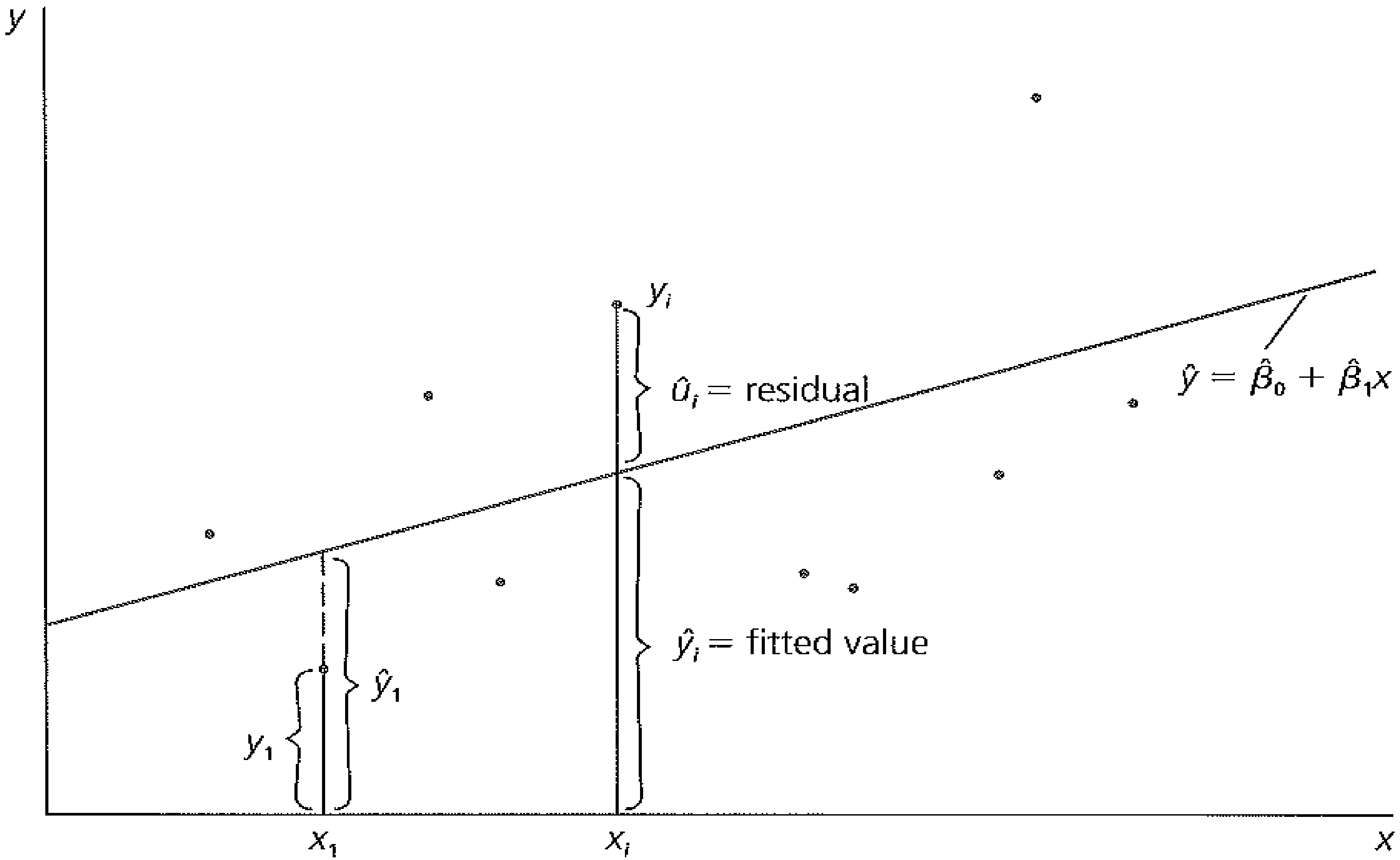
Tatsächlich beobachtete abhängige Variable ist kleiner als der entsprechende Regressionswert und wird somit überschätzt

Damit lässt sich die OLS-Methode folgendermaßen alternativ charakterisieren: Die OLS-Schätzer werden dadurch ermittelt, dass die Summe der quadrierten Residuen minimal wird.

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots} \sum_{i=1}^{n} \hat{u}_i^2 = \min_{\hat{\beta}_0, \hat{\beta}_1, \ldots} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik})^2$$

Darstellung im einfachen linearen Regressionsmodell:

Fitted values and residuals.



Beispiel: Erklärung von Vorstandsgehältern

Betrachtet wird wieder die zuvor untersuchte OLS-Regressionsfunktion zur Analyse der Determinanten des Gehalts von Vorstandsmitgliedern:

$$\widehat{salary} = 963{,}191 + 18{,}501 roe$$

Dabei ergibt sich für die ersten fünf untersuchten Vorstandsmitglieder:

Beobachtung	roe	salary	$\widehat{salary}$	$\hat{u}$
1	14,1	1095	1224,058	-129,058
2	10,9	1001	1164,854	-163,854
3	23,5	1122	1397,969	-275,969
4	5,9	578	1072,348	-494,348
5	13,8	1368	1218,508	149,492

Für das erste Vorstandsmitglied gilt z.B. für das Residuum:

$$1095 - (963{,}191 + 18{,}501 \cdot 14{,}1) = 1095 - 1224{,}058 = -129{,}058$$

Die ersten vier Vorstandsmitglieder erhalten also ein geringeres Gehalt als durch die entsprechenden Regressionswerte impliziert. Das Gehalt des fünften Vorstandsmitglieds wird dagegen unterschätzt.

Wichtige Eigenschaften von Regressionswerten und Residuen:

- (2.20) $\sum_{i=1}^{n} \hat{u}_i = 0$

  Diese Eigenschaft folgt aus der ersten Bedingung erster Ordnung bei der OLS-Schätzung der Regressionsparameter.

  Daraus folgt:

  $$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0$$

  $$\overline{y} = \overline{\hat{y}}$$

- (2.21) $\sum_{i=1}^{n} x_{ij}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}) = \sum_{i=1}^{n} x_{ij}\hat{u}_i = 0 \quad \text{für } j = 1,\ldots,k$

  Das heißt, die Kovarianz zwischen jeder erklärenden Variablen und den Residuen ist null. Auch diese Eigenschaft folgt aus den Bedingungen erster Ordnung bei der OLS-Schätzung der Regressionsparameter.

- Der folgende Punkt liegt immer auf der OLS-Regressionsfunktion:

  $(\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k, \overline{y})$

Alternative Darstellung linearer Regressionsmodelle:

$$(2.22)\quad y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i \quad \text{für } i = 1,\ldots, n$$

Gesamte Abweichungsquadratsumme („total sum of squares“) :

$$(2.23)\quad SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Erklärte Abweichungsquadratsumme („explained sum of squares“):

$$(2.24)\quad SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Residualabweichungsquadratsumme („residual sum of squares“):

$$(2.25)\quad SSR = \sum_{i=1}^{n} (\hat{u}_i - \bar{\hat{u}})^2 = \sum_{i=1}^{n} \hat{u}_i^{\,2}$$

Es gilt:

$$(2.26)\quad SST = SSE + SSR$$

$$\frac{SSR}{SST} + \frac{SSE}{SST} = 1$$

Bestimmtheitsmaß (Determinationskoeffizient): Anteil der Variation der abhängigen Variablen y_i, der durch die OLS-Regressionsfunktion erklärt wird

$$(2.27) \quad R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Das Bestimmtheitsmaß entspricht auch dem quadrierten Korrelationskoeffizienten zwischen den abhängigen Variablen und den OLS-Regressionswerten:

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right)} = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})\right)^2}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2\right)}$$

Eigenschaften des Bestimmtheitsmaßes:

- $0 \leq R^2 \leq 1$
- R^2 sinkt niemals, wenn eine weitere (möglicherweise irrelevante) erklärende Variable hinzugefügt wird (da SSR in diesem Fall niemals ansteigt)
- Aus diesem Grund ist R^2 ein schlechtes Maß zur Beurteilung der Güte eines linearen Regressionsmodells

Beispiel 1: Erklärung von Vorstandsgehältern

Betrachtet wird wieder die zuvor untersuchte OLS-Regressionsfunktion zur Analyse der Determinanten des Gehalts von Vorstandsmitgliedern:

$$\widehat{salary} = 963{,}191 + 18{,}501 roe$$

Das R^2 beträgt 0,0132, d.h. die Variation der Eigenkapitalrendite roe erklärt zu 1,32% die Variation der Vorstandsgehälter.

Mit STATA hatten sich folgende OLS-Schätzergebnisse gezeigt:

```
reg salary roe

      Source |       SS       df       MS              Number of obs =     209
-------------+------------------------------           F(  1,   207) =    2.77
       Model |  5166419.33     1  5166419.33           Prob > F      =  0.0978
    Residual |   386566563   207  1867471.32           R-squared     =  0.0132
-------------+------------------------------           Adj R-squared =  0.0084
       Total |   391732982   208  1883331.64           Root MSE      =  1366.6

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         roe |   18.50119   11.12325     1.66   0.098    -3.428195    40.43057
       _cons |   963.1913   213.2403     4.52   0.000     542.7902    1383.592
------------------------------------------------------------------------------
```

Beispiel 2: Erklärung von College-Noten

Betrachtet wird wieder die zuvor untersuchte OLS-Regressionsfunktion zur Analyse der Determinanten der durchschnittlichen College-Punktzahl:

$$\text{col}\hat{\text{G}}\text{PA} = 1{,}29 + 0{,}453\text{hsGPA} + 0{,}0094\text{ACT}$$

Das R^2 beträgt 0,176, d.h. die Variation der Highschool-Note hsGPA und der Punktzahl im College-Aufnahmetest ACT erklärt zu 17,6% die Variation der College-Note colGPA.

Mit STATA hatten sich folgende OLS-Schätzergebnisse gezeigt:

```
reg colGPA hsGPA ACT

      Source |       SS       df       MS              Number of obs =     141
-------------+------------------------------           F(  2,   138) =   14.78
       Model |  3.42365461     2   1.7118273           Prob > F      =  0.0000
    Residual |  15.9824447   138  .115814817           R-squared     =  0.1764
-------------+------------------------------           Adj R-squared =  0.1645
       Total |  19.4060993   140  .138614995           Root MSE      =  .34032

------------------------------------------------------------------------------
      colGPA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hsGPA |   .4534558   .0958129     4.73   0.000     .2640046    .6429071
         ACT |    .009426   .0107772     0.87   0.383    -.0118838    .0307358
       _cons |   1.286328   .3408221     3.77   0.000     .6124191    1.960237
------------------------------------------------------------------------------
```

Korrigiertes bzw. angepasstes („adjusted“) Bestimmtheitsmaß:

$$(2.28) \quad \bar{R}^2 = 1 - \frac{\frac{SSR}{n-k-1}}{\frac{SST}{n-1}} = 1 - \frac{n-1}{n-k-1}(1-R^2)$$

Das angepasste Bestimmtheitsmaß berücksichtigt im Gegensatz zum herkömmlichen Bestimmtheitsmaß die Anzahl der erklärenden Variablen. Durch diese Einbeziehung kann das angepasste Bestimmtheitsmaß in einzelnen Fällen negativ werden. Auch das angepasste Bestimmtheitsmaß ist kein generell geeignetes Maß zur Beurteilung der Güte eines linearen Regressionsmodells.

Beispiel: Erklärung von Vorstandsgehältern

Betrachtet wird wieder die zuvor untersuchte OLS-Regressionsfunktion zur Analyse der Determinanten des Gehalts von Vorstandsmitgliedern (n = 209):

$$\widehat{salary} = 963{,}191 + 18{,}501 roe$$

Mit R^2 = 0,0132 ergibt sich das angepasste Bestimmtheitsmaß:
1 - (208/207)·0,9868 = 0,0084

Regression durch den Ursprung:

Manchmal können eine ökonomische Theorie oder aber ökonomische Plausibilitäten auf eine Konstante von null in einem linearen Regressionsmodell hinweisen

Damit ergibt sich:

(2.29) $y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ bzw.

$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$

Falls alle erklärenden Variablen gleich null sind, ist auch der OLS-Regressionswert null:

- In diesem Fall ist das arithmetische Mittel der Residuen ungleich null
- Dabei kann das (herkömmliche) Bestimmtheitsmaß negative Werte annehmen
- Regressionen durch den Ursprung werden in der Praxis nicht häufig angewendet, da schwerwiegende Probleme auftauchen können, falls die Konstante tatsächlich ungleich null ist

2.3 Erwartungswert von OLS-Schätzern

→ Zu beachten ist, dass sich statistische Eigenschaften (d.h. Erwartungswert und Varianz) von OLS-Schätzern nicht auf eine bestimmte Stichprobe, sondern auf wiederholte Zufallsstichproben aus der Grundgesamtheit beziehen

Annahmen zur Betrachtung des Erwartungswerts von OLS-Schätzern:

- Annahme A1: Linearität in den Parametern
  Der Zusammenhang zwischen der abhängigen Variablen y und den erklärenden Variablen $x_1, x_2, \ldots, x_k$ ist linear in den Parametern, d.h. es gilt (2.9) mit $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
- Annahme A2: Zufallsstichprobe
  Es liegt eine zufällige Stichprobe vom Umfang n aus der Grundgesamtheit vor mit $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i), i = 1, \ldots, n\}$, so dass für eine zufällig ausgewählte Beobachtung i (2.12) gilt mit $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$
- Annahme A3: Keine perfekte Kollinearität
  In der Stichprobe (und daher auch in der Grundgesamtheit) ist keine der erklärenden Variablen konstant und es besteht keine exakte lineare Beziehung zwischen den erklärenden Variablen
- Annahme A4: Bedingter Erwartungswert von u ist null
  Es gilt also (2.10) mit $E(u|x_1, x_2, \ldots, x_k) = 0$

Unter diesen vier Annahmen sind alle mit der OLS-Methode geschätzten Parameter erwartungstreu, d.h. der Erwartungswert der Schätzer ist gleich den unbekannten Regressionsparametern:

$$(2.30) \quad E(\hat{\beta}_j) = \beta_j \quad \text{für } j = 0, 1, \ldots, k$$

Zu Annahme A3:

Diese Annahme schließt nicht generell eine Korrelation zwischen den erklärenden Variablen aus, sondern besagt, dass diese nicht exakt linear sein darf. Falls A3 verletzt ist, kann keine OLS-Schätzung durchgeführt werden.

- Die einfachste Verletzung von A3 liegt für zwei erklärende Variablen vor, falls die eine Variable ein Vielfaches der anderen Variablen ist, z.B. wenn sowohl Einkommen in Euro als auch Einkommen in Dollar (bzw. Einkommen in 1000 Euro) als erklärende Variablen einbezogen werden
- A3 wird auch verletzt, wenn eine erklärende Variable als lineare Funktion zweier oder mehrerer anderer erklärender Variablen formuliert werden kann, z.B. wenn im obigen Beispiel zur Erklärung von College-Noten neben der durchschnittlichen Punktzahl in der Highschool und der Punktzahl im College-Aufnahmetest auch die Summe dieser Punktzahlen als erklärende Variable einbezogen wird
- A3 wird auch bei einem zu kleinen Stichprobenumfang verletzt, d.h. wenn $n < k+1$

→ Dagegen wird A3 nicht verletzt, wenn verschiedene nichtlineare Funktionen derselben erklärenden Variablen einbezogen werden, z.B. das Einkommen und das quadrierte Einkommen. Dagegen können z.B. der natürliche Logarithmus des Einkommens sowie der natürliche Logarithmus des quadrierten Einkommens nicht gleichzeitig als erklärende Variablen einbezogen werden, da der natürliche Logarithmus des quadrierten Einkommens doppelt so groß ist wie der natürliche Logarithmus des Einkommens und somit eine lineare Beziehung zwischen den erklärenden Variablen vorliegt.

Zu Annahme A4:

Falls diese Annahme zutrifft, liegen exogene erklärende Variablen vor. Falls dagegen A4 verletzt wird, liegen endogene erklärende Variablen bzw. Endogenität vor.

- Eine Verletzung von A4 liegt z.B. vor, falls Messfehler in den erklärenden Variablen existieren oder der funktionale Zusammenhang zwischen den abhängigen und erklärenden Variablen fehlspezifiziert ist (siehe später)
- Eine der wichtigsten Verletzungen von A4 liegt vor, wenn eine relevante erklärende Variable, die mit den anderen erklärenden Variablen korreliert ist, vernachlässigt wird

Mögliche Verzerrungen bei der Vernachlässigung relevanter erklärender Variablen („omitted variable bias“)

Es wird das folgende korrekte lineare Regressionsmodell betrachtet (wobei obige Annahmen A1 bis A4 erfüllt sind):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + \beta_k x_k + u$$

Geschätzt wird dagegen folgendes fehlspezifizierte lineare Regressionsmodell, das x_k vernachlässigt (z.B. aufgrund von Unkenntnis oder fehlender Daten):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + u$$

Damit ergeben sich folgende OLS-Regressionsfunktionen in den korrekten und fehlspezifizierten linearen Regressionsmodellen:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_{k-1} x_{k-1} + \hat{\beta}_k x_k$$

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \cdots + \tilde{\beta}_{k-1} x_{k-1}$$

Dabei gilt folgender Zusammenhang (j = 1,…, k-1):

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j$$

$\tilde{\delta}_j$ (j = 1,…, k-1) ist der mit OLS geschätzte Steigungsparameter für x_j bei einer Regression von x_k auf alle anderen erklärenden Variablen (einschließlich einer Konstante). Es ergibt sich:

$$(2.31) \quad E(\tilde{\beta}_j) = E(\hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j) = \beta_j + \beta_k \tilde{\delta}_j$$

Damit ergibt sich folgende Verzerrung (bias) für die OLS-Schätzung des Steigungsparameters β_j im fehlspezifizierten linearen Regressionsmodell:

$$\text{(2.32)} \quad \text{Bias}(\tilde{\beta}_j) = E(\tilde{\beta}_j) - \beta_j = \beta_k \tilde{\delta}_j$$

Es liegt nur dann keine Verzerrung vor, wenn β_k oder $\tilde{\delta}_j$ null ist:

- Im ersten Fall ist das „fehlspezifizierte“ lineare Regressionsmodell korrekt spezifiziert
- Im zweiten Fall sind x_j und x_k in der Stichprobe unkorreliert

Im Regelfall sind aber x_j und x_k in der Stichprobe korreliert, wobei die Richtung der Verzerrung folgendes Aussehen hat:

- Positiver bias: $\beta_k > 0$ und corr $(x_j, x_k) > 0$ oder $\beta_k < 0$ und corr $(x_j, x_k) < 0$
- Negativer bias: $\beta_k < 0$ und corr $(x_j, x_k) > 0$ oder $\beta_k > 0$ und corr $(x_j, x_k) < 0$

Statt der Richtung des bias ist aber vor allem die Höhe wichtig. Allerdings kann die Verzerrung in der Regel nicht berechnet oder geschätzt werden, da β_k unbekannt ist.

Dagegen:

Die Einbeziehung irrelevanter erklärender Variablen (d.h. einer oder mehrerer erklärender Variablen, die keinen (partiellen) Effekt auf die abhängige Variable haben) hat keine Auswirkung auf die Erwartungstreue der mit OLS geschätzten Parameter, führt also nicht zu Verzerrungen

Betrachtet wird das folgende korrekte lineare Regressionsmodell:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + u$$

Geschätzt wird dagegen folgendes fehlspezifizierte lineare Regressionsmodell, das x_k zusätzlich als erklärende Variable einbezieht, wobei der entsprechende Steigungsparameter β_k null ist:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + \beta_k x_k + u$$

Es gilt dennoch:

$$E(\hat{\beta}_0) = \beta_0,\ E(\hat{\beta}_1) = \beta_1,\ E(\hat{\beta}_2) = \beta_2, \ldots,\ E(\hat{\beta}_{k-1}) = \beta_{k-1},\ E(\hat{\beta}_k) = \beta_k = 0$$

→ Allerdings hat die Einbeziehung irrelevanter erklärender Variablen einen Einfluss auf die Varianz der OLS-Schätzer

2.4 Varianz von OLS-Schätzern

Annahmen zur Betrachtung der Varianz von OLS-Schätzern:

- Annahmen A1 bis A4 bei der Betrachtung des Erwartungswertes von OLS-Schätzern
- Annahme A5: Homoskedastizität
  Die bedingte Varianz des Fehlerterms u ist konstant, d.h. es gilt $\mathrm{Var}(u|x_1, x_2,\ldots, x_k) = \sigma^2$. Falls dies nicht zutrifft, d.h. wenn die Varianz von den erklärenden Variablen abhängt, liegt Heteroskedastizität vor.

Beispiel: Erklärung von Löhnen

Mit Hilfe eines multiplen linearen Regressionsmodells soll wiederum der Effekt der Ausbildungszeit in Jahren (educ), der Berufserfahrung in Jahren (exper) und der Betriebszugehörigkeit in Jahren (tenure) auf den Stundenlohn (wage) untersucht werden:

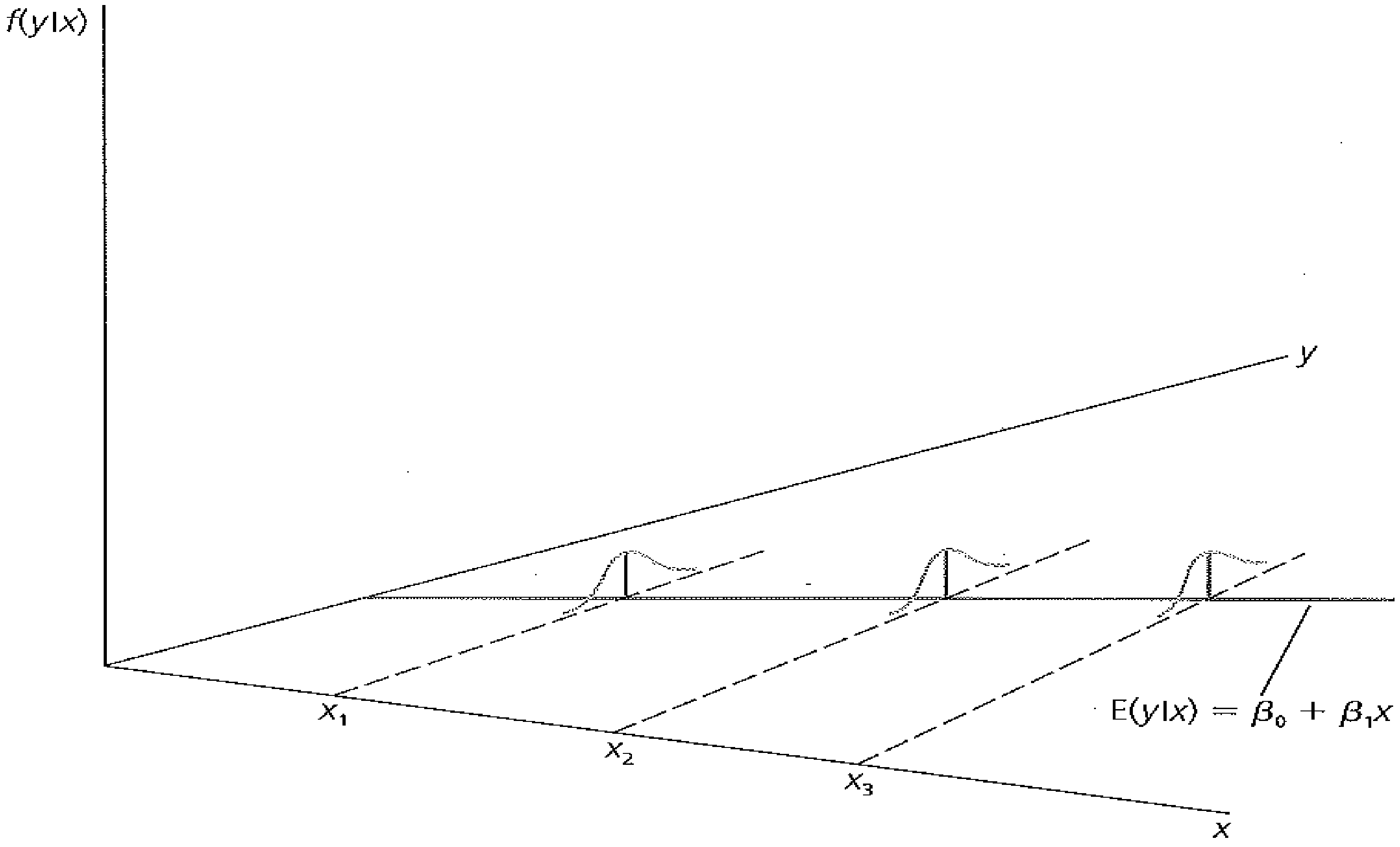
$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

Unter den Annahmen A1 bis A5 gilt dann insbesondere:

$\mathrm{Var}(u|educ, exper, tenure) = \sigma^2$

Darstellung von Homoskedastizität im einfachen linearen Regressionsmodell:

The simple regression model under homoskedasticity.



→ Die Annahmen A1 bis A5 werden (im Falle von Regressionsanalysen mit Querschnittsdaten) auch als Gauss-Markov-Annahmen bezeichnet

Damit hängt der (bedingte) Erwartungswert der abhängigen Variablen y von den erklärenden Variablen ab, nicht aber die (bedingte) Varianz, d.h.:

$$E(y|x_1, x_2,\ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

$$Var(y|x_1, x_2,\ldots, x_k) = \sigma^2$$

Damit ergibt sich unter den Annahmen A1 bis A5 für die Varianz der mit OLS geschätzten Steigungsparameter in linearen Regressionsmodellen:

$$(2.33) \quad Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2} = \frac{\sigma^2}{(1-R_j^2)SST_j} \quad \text{für } j = 1,\ldots, k$$

Dabei stellt R_j^2 das Bestimmtheitsmaß bei einer Regression von x_j auf alle anderen erklärenden Variablen (einschließlich einer Konstante) dar.

→ Während die Annahme der Homoskedastizität unwesentlich für die Erwartungstreue der geschätzten Parameter ist, gilt obige Varianz nur unter dieser Annahme, nicht aber bei Heteroskedastizität

Im einfachen linearen Regressionsmodell ergibt sich speziell:

$$(2.34)\quad \mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ sowie } \mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Zu den Komponenten der Varianz (2.33) der mit OLS geschätzten Steigungsparameter in linearen Regressionsmodellen:

- Varianz σ^2 des Fehlerterms u:
  Bei einer größeren Varianz von u ist es schwieriger, den (partiellen) Effekt einer erklärenden Variablen zu schätzen. Daher ergibt sich in diesem Fall auch eine größere Varianz der geschätzten Steigungsparameter. Die Varianz von u (die nicht vom Stichprobenumfang abhängt) kann lediglich durch die Einbeziehung weiterer erklärender Variablen reduziert werden.
- Stichprobenvariation SST_j der erklärenden Variablen x_j:
  Ein höheres SST_j führt zu einer geringeren Varianz des geschätzten entsprechenden Steigungsparameters. Wenn SST_j dagegen gering ist, ergibt sich eine höhere Varianz. SST_j kann durch eine Erhöhung des Stichprobenumfangs erhöht werden. Ein SST_j von null ist durch Annahme A3 ausgeschlossen

Wesentliche Komponente der Varianz (2.33): R_j^2

- R_j^2 ist der Anteil der Variation der erklärenden Variablen x_j, der durch die anderen erklärenden Variablen erklärt wird
- Die geringste Varianz des geschätzten entsprechenden Steigungsparameters (bei gegebenen σ^2 und SST_j) ergibt sich, wenn $R_j^2 = 0$, d.h. wenn x_j in der Stichprobe mit allen anderen erklärenden Variablen unkorreliert ist
- Der andere Extremfall $R_j^2 = 1$, d.h. wenn x_j eine perfekte Linearkombination von einigen anderen erklärenden Variablen ist, wird durch Annahme A3 ausgeschlossen
- In empirischen Anwendungen liegt aber häufig der Fall vor, dass R_j^2 hohe Werte annimmt. In diesem Fall der Multikollinearität kann die Varianz des geschätzten entsprechenden Steigungsparameters sehr hohe Werte annehmen.

Multikollinearität:

Hohe, aber nicht perfekte Korrelation zwischen zwei oder mehreren erklärenden Variablen. Dadurch liegt bei Multikollinearität keine Verletzung von Annahme A3 oder der anderen Annahmen vor.

Zur Multikollinearität:

- Eine durch Multikollinearität ausgelöste hohe Varianz von geschätzten Steigungsparametern kann zu Problemen bei der Durchführung von statistischen Tests führen
- Wenngleich Multikollinearität häufiger problematisiert wird, unterscheidet sich dieses Problem nicht von geringen Stichprobenumfängen (und damit kleinen SST_j) oder hohen σ^2
- Eine Möglichkeit der Eindämmung von Multikollinearität ist die Unterdrückung von erklärenden Variablen. Allerdings kann dieses Vorgehen zu Verzerrungen führen, so dass bei der Abwägung eine solche Möglichkeit in empirischen Untersuchungen häufig nicht besteht.
- Hohe Korrelationen zwischen einzelnen Variablen haben keinen Effekt auf die Varianz von geschätzten Steigungsparametern anderer erklärender Variablen
- Ein häufig verwendetes Maß zur Stärke der Multikollinearität ist der Varianzinflationsfaktor $VIF_j = 1/(1- R_j^2)$ in Bezug auf eine erklärende Variable x_j. Dieser Faktor misst die Erhöhung der Varianz des geschätzten Steigungsparameters, weil x_j nicht mit den anderen erklärenden Variablen unkorreliert ist. Faustregeln, bei denen ein Problem der Multikollinearität diagnostiziert werden (z.B. $VIF_j = 10$), sind willkürlich und deshalb nicht generell hilfreich.

Schätzung der Varianz σ^2 des Fehlerterms u:

Die Schätzung von σ^2 ist die Grundlage für die Schätzung der Varianz der mit OLS geschätzten Regressionsparameter

Da $\sigma^2 = E(u^2)$, wäre folgender Schätzer für σ^2 denkbar:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^{\,2} = \frac{SSR}{n}$$

Allerdings ist dieser Schätzer verzerrt. Ein erwartungstreuer Schätzer ergibt sich dagegen durch den Quotienten zwischen SSR und der Differenz zwischen dem Stichprobenumfang n und der Anzahl k+1 an Regressionsparametern:

$$(2.35)\quad \hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^{\,2} = \frac{SSR}{n-k-1}$$

Der entsprechende (zwar konsistente, aber nicht erwartungstreue) Schätzer für die Standardabweichung σ des Fehlerterms u („standard error of the regression, SER“) lautet dann:

$$(2.36)\quad \hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^{\,2}}$$

Damit kann nun die Varianz (2.33) der mit OLS geschätzten Steigungsparameter in linearen Regressionsmodellen erwartungstreu geschätzt werden:

$$(2.37) \quad \hat{\text{Var}}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{(1-R_j^2)SST_j} \quad \text{für } j = 1, \ldots, k$$

Die Standardabweichung der mit OLS geschätzten Steigungsparameter lautet:

$$(2.38) \quad \sqrt{\text{Var}(\hat{\beta}_j)} = \frac{\sigma}{\sqrt{(1-R_j^2)SST_j}} \quad \text{für } j = 1, \ldots, k$$

Diese Standardabweichung kann dann folgendermaßen geschätzt werden:

$$(2.39) \quad \sqrt{\hat{\text{Var}}(\hat{\beta}_j)} = \frac{\hat{\sigma}}{\sqrt{(1-R_j^2)SST_j}} \quad \text{für } j = 1, \ldots, k$$

Wichtig ist dabei, dass die Verwendung dieser Schätzungen insbesondere auf der Annahme A5 der Homoskedastizität beruht. Falls dagegen Heteroskedastizität vorliegt (siehe später), ist (2.37) eine verzerrte Schätzung für die Varianz (2.33) der mit OLS geschätzten Steigungsparameter (obwohl Heteroskedastizität keinen Einfluss auf die Erwartungstreue der geschätzten Regressionsparameter hat).

2.5 Effizienz von OLS-Schätzern

Falls die Annahmen A1 bis A5 gelten, ergibt sich:

Die OLS-Schätzer sind die besten linearen unverzerrten Schätzer der Regressionsparameter in linearen Regressionsmodellen („BLUE, best linear unbiased estimator“)

Bestandteile von BLUE:

- „Unverzerrt“ bedeutet, dass der Schätzer erwartungstreu ist
- „Linear“ bedeutet, dass der Schätzer eine lineare Funktion der Daten und der abhängigen Variablen darstellt
- „Beste“ bedeutet, dass der Schätzer die geringste Varianz besitzt

Im Einklang mit dem Gauss-Markov-Theorem haben OLS-Schätzer damit in der Klasse aller linearen und unverzerrten Schätzer die geringste Varianz. Voraussetzung für diese Eigenschaft ist aber die Gültigkeit aller Annahmen A1 bis A5.