Grundlagen der Ökonometrie

Prof. Dr. Susanne Rässler

Otto-Friedrich-Universität Bamberg
Lehrstuhl für Statistik und Ökonometrie

Wintersemester 2016/2017

Gliederung zu Grundlagen der Ökonometrie

Empfohlene Literatur

- Verbeek, M. (2012): A Guide to Modern Econometrics. 4th ed. Wiley, Chichester. Ca. 52 Euro.
- Verbeek, M. (2015): Moderne Ökonometrie. Wiley, Weinheim. Ca. 69 Euro.
- Greene, W.H. (2012): Econometric Analysis. 7th ed. Pearson, New Jersey. Ca. 75 Euro.
- von Auer, L. (2016): Ökonometrie. 7. Auflage. Springer, Berlin. Ca. 35 Euro.

Ökonometrie heute

- In den letzten drei Jahrzehnten hat sich das Gebiet der Ökonometrie rasant entwickelt.
- "Up-to-date" - Ökonometrie ist mittlerweile Standard in der empirischen ökonomischen Arbeit.
- ***Beispiele***: Einheitswurzeltest, Cointegration, GMM-Schätzung, Heteroskedastie- und Autokorrelationskonsistente Standardfehler, Modelle der bedingten Heteroskedastie, Modelle für Paneldaten, Modelle mit beschränkt abhängigen Variablen, endogene Regressoren, Sample Selection.
- Parallel dazu wurde die ökonometrische Anwendungssoftware immer benutzerfreundlicher und aktueller.

Problem

- Benutzer können hochentwickelte Methoden in der empirischen Arbeit einsetzen, ohne die zugrunde liegenden Konzepte und die potentiellen Gefahren eines missbräuchlichen Einsatzes verstanden zu haben.
- Im Gegensatz zu den Belangen der Anwendung wird in vielen einführenden Lehrbüchern großer Wert auf das Standardmodell der linearen Regression mit sehr rigiden Annahmen gelegt, die in der Praxis nur selten erfüllt sind.
- Die fortgeschrittenen Lehrbücher sind häufig für den durchschnittlichen Leser zu "technisch" und zu detailliert, um fokussiert die wichtigsten Ideen vermitteln zu können.

"A Guide to Modern Econometrics"

- Deshalb: Ziel des Lehrbuches von Verbeek ist es, den Leser mit einem breiten Spektrum der Aufgabenstellung der modernen Ökonometrie vertraut zu machen.
- Dabei erfolgt Konzentration auf Inhalte, die wichtig für das **Verständnis** und die **selbständige Durchführung der empirischen Arbeit** sind.
- Das Buch ist mehr ein Führer durch als ein Überblick über alternative ökonometrische Techniken.

Über Ökonometrie

- Ökonomen interessieren sich häufig für die **Beziehungen zwischen** verschiedenen ökonomischen Größen, wie z.B. zwischen dem Individuallohn und dem Ausbildungsniveau.
- Wichtigste Aufgabe der Ökonometrie ist die **Quantifizierung** dieser Beziehungen auf der Basis von verfügbaren Daten und statistischen Methoden und die zulässige **Interpretation** der Ergebnisse.
- Fazit: **Ökonometrie ist das Zusammenspiel von ökonomischer Theorie, beobachteten Daten und statistischen Methoden**.
- Provokation: "Econometrics is much easier without data."
- Traditionell konzentriert sich Ökonometrie auf aggregierte ökonomische Beziehungen. Makroökonomische Modelle bestanden aus wenigen bis vielen hundert Gleichungen, die spezifiziert, geschätzt und für Prognose bzw. Politikevaluation verwendet wurden.

Anwendungsgebiete der Ökonometrie

- Seit den 70er Jahren werden ökonometrische Methoden vermehrt in sog. mikroökonometrischen Modellen des Haushalts und der Unternehmung eingesetzt. Dazu wurden eigens geeignete ökonometrische Modelle und Schätzer entwickelt, die spezifische Probleme wie z.B. diskrete abhängige Variablen und Sample Selection mittels großer Umfragedatensätze und zunehmender Rechnerkapazität lösen.
- Ganz aktuell hat die empirische Analyse von Finanzmärkten viele theoretische Entwicklungen angestoßen und stimuliert.
- Ökonometrie durchdringt die empirische Arbeit in fast allen ökonomischen Bereichen, so dass es nicht mehr ausreicht, einige wenige Regressionen zu rechnen und die Ergebnisse zu interpretieren.

Vorgehen

- Die Ökonomen interessierenden Beziehungen werden zumeist mathematisch formuliert und führen zu ökonometrischen oder statistischen Modellen. In solchen Modellen ist Platz für Abweichungen von den strikten theoretischen Beziehungen, die z.B. auf Messfehlern, unvorhersehbarem Verhalten, Optimierungsfehlern oder unerwarteten Ereignissen beruhen.

Klassifizierung ökonometrischer Modelle

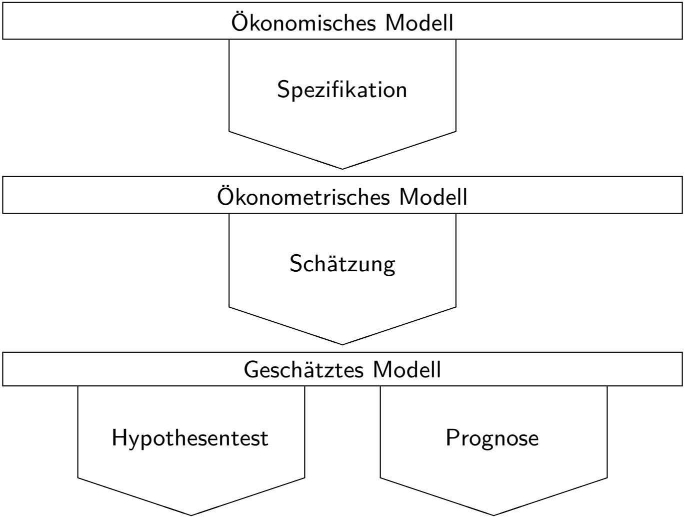
1. Univariate Zeitreihenmodelle
   Beispiel: Wie hängt der Kurzfristzins von der eigenen Vergangenheit ab?
2. Multivariate (Regressions-) Modelle für Zeitreihendaten
   Beispiel: Wie verändert sich der Langfristzins, wenn die Geldpolitik den Kurzfristzins verändert?
3. Multivariate (Regressions-) Modelle für Querschnittsdaten
   Beispiel: Inwieweit lassen sich Differenzen in der Einkommensfunktion durch Geschlechtsunterschiede erklären?
4. Modelle für Paneldaten
   Beispiel: Warum hat Person 1 im Jahre 2001 mehr Einkommen als im Jahr 2002 und tendenziell weniger als Person 2?

Aufgaben der Ökonometrie

- **Spezifikation und Quantifizierung ökonomischer Beziehungen.**
  Ökonometriker formulieren ein statistisches Modell (im allgemeinen basierend auf einer ökonomischen Theorie) und konfrontieren dieses Modell mit den Daten. Die unbekannten Größen in der Spezifikation werden Parameter genannt und sind aus den Daten zu schätzen.
- **Überprüfung der Brauchbarkeit des spezifizierten Modells.**
  - Überprüfung der Annahmen des Modells.
  - Überprüfung, ob das Modell für die gewünschten Zwecke eingesetzt werden kann. Zwecke sind Erklärung, Prognose und Beurteilung der Auswirkungen von Aktionen.
- **Überprüfung ökonomischer Theorien.**
  - Falsifikationsprinzip des kritischen Rationalismus von Popper.

- ***Beispiel***:
  Die Hypothese effizienter Kapitalmärkte impliziert, dass Renditen von Titeln nicht aus der eigenen Vergangenheit vorhergesagt werden können. Die Aufgabe der Ökonometrie ist, diese Hypothese als quantifiziertes Modell zu formulieren und ihre Gültigkeit an den Daten zu testen.

Schema ökonometrischer Analyse



Beispiele und Übungen

- Beispiele sollen selbst ökonomisch interessant sein und stammen aus den wichtigsten Feldern der Anwendung ökonometrischer Methoden: Marketing, Finanzmärkte, Arbeitsmarkt.
- D.h. es werden große, relativ aktuelle Datensätze betrachtet, die in der empirischen Forschung tatsächlich diskutiert worden sind.
- Zielsetzung: Das Erlernen von Ökonometrie beschränkt sich nicht auf das Studieren eines Lehrbuches. "Hands-on experience is crucial in the process of understanding the different methods and how and when to implement them".
- Nur wer sich selbst die Hände mit Daten schmutzig gemacht hat, kann sich davor schützen, den scheinbar vollkommen korrekten Ergebnissen der Ökonometriesoftware zu erliegen, wenn diese für falsche Modelle und falsche Methoden eingesetzt werden.

Ökonometriepakete

- Allgemeine Statistik-Pakete: SPSS, STATISTICA, SAS, BMDP, S-PLUS
- Spezialisierte Software: EVIEWS, STATA, LIMDEP, RATS, TSP, Microfit, ET, PcGive
- Programmiersprachen mit Tools: MATLAB, Gauss
- Freeware: R, ViSta, SciLab

Gliederung Kapitel 2

Ökonomisches Modell

- Regressionsmodelle führen die Ausprägung eines Merkmals auf die Ausprägung eines oder mehrerer anderer Merkmale zurück.
- Gesucht ist

$$y = f(x)$$

  Dabei sind x und y stetige metrische Variablen
  x ist **unabhängige Variable** (Regressor, erklärende oder exogene Variable)
  y ist **abhängige Variable** (Regressand, erklärte oder endogene Variable)

  Es wird nicht nur angegeben, wie groß der Zusammenhang zwischen zwei oder mehreren Variablen ist, sondern auch, um wieviel die abhängige Variable sich verändert, wenn die unabhängige Variable um eine bestimmte Menge zu- oder abnimmt.

Ökonometrisches Modell

- Speziell interessiert ein lineares (stochastisches) Modell

$$Y = X\beta + \varepsilon$$

- Ausgangspunkt: Es liege eine Stichprobe von n individuellen Löhnen (y) und $K-1$ erklärenden Variablen $x_2, \ldots, x_K$ vor.
- Welche Linearkombination von $x_2, \ldots, x_K$ und einer **Konstanten** (Absolutglied) ergibt eine "gute" Approximation von y?
- Linearkombination:

$$\beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K$$

mit den **Parametern** (Koeffizienten) β_k, $k = 1, 2, \ldots, K$.

- Was unterscheidet Parameter $\beta_1, \ldots, \beta_K$ und Variablen y, $x_2, \ldots, x_K$?
  - ▷ y und $x_2, \ldots, x_K$ variieren von Person zu Person in der Stichprobe.
  - ▷ Parameter (Koeffizienten) sind konstant.
- Für die i-te Person werden

  $$y_i,\ x_{i2}, \ldots, x_{iK}$$

  beobachtet für $i = 1, 2, \ldots, n$.
- Nicht alle Zusammenhänge folgen exakt der spezifizierten stochastischen Funktion. Der **Approximationsfehler** ε (Fehlerterm, Störgröße oder Residuum) bezeichnet die Abweichung von geschätzten und tatsächlich beobachteten Werten.

OLS - Ansatz

- Was ist eine "gute" Approximation?
- Vorschlag: Eine Approximation ist umso besser, je kleiner die Approximationsfehler sind.
- Zusammenfassung der Approximationsfehler in der "Residual Sum of Squares" (RSS):

$$S(\beta) = \sum_{i=1}^{n} (y_i - x_i'\beta)^2$$

- "Beste" Approximation ist durch denjenigen Wert von β gegeben, der $S(\beta)$ minimiert (Ordinary Least Squares (OLS) oder Kleinst-Quadrate (KQ) Ansatz):

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - x_i'\beta)^2 \rightarrow min$$

Herleitung

- Lösung durch die Bedingungen 1. Ordnung, d.h. es werden Nullstellen der ersten Ableitung von $S(\beta)$ nach β betrachtet:

$$\frac{\partial S(\beta)}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (y_i - x_i' \beta)$$

  Beachte: System von K Gleichungen in den K unbekannten Parametern $\beta_1, \ldots, \beta_K$ liefert die Schätzer $\widehat{\beta}_1, \ldots, \widehat{\beta}_K$.

- Umformung zum System der K Normalgleichungen:

$$\left(\sum_{i=1}^{n} x_i x_i' \right) \widehat{\beta} = \sum_{i=1}^{n} x_i y_i$$

Lösung

- Auflösung nach dem "besten" $\widehat{\beta}$, sofern $K \times K$-Matrix $\left(\sum_{i=1}^{n} x_i x_i'\right)$ der Summen von Quadraten und Kreuzprodukten invertierbar ist:

$$\widehat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i$$

- Die einzige Annahme (sog. Annahme fehlender Multikollinearität) betrifft die Invertierbarkeit von $\left(\sum_{i=1}^{n} x_i x_i'\right)$ und besagt, dass es keine exakte lineare Abhängigkeit zwischen den Werten der erklärenden Variablen gibt.
- Ergänzung: Durch Untersuchung der Matrix der zweiten Ableitungen zeigt sich, dass $\widehat{\beta}$ ein Minimum von $S(\widehat{\beta})$ produziert.

Konsequenzen für die Residuen

- Aus der 1. Normalgleichung folgt:

$$\sum_{i=1}^{n} x_i(y_i - x_i'\widehat{\beta}) = \sum_{i=1}^{n} x_i\widehat{\varepsilon}_i = 0$$

  D.h. $\widehat{\varepsilon}$ ist orthogonal zu jedem Variablenvektor in X.

- Wegen Achsenabschnitt ergibt sich

$$0 = \sum_{i=1}^{n} \widehat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i'\widehat{\beta}$$

  und damit

$$\overline{y} = \widehat{\beta}'\overline{x} \text{ und } \overline{y} = \overline{\widehat{y}}$$

Nachtrag Matrixschreibweise

- Datenvektoren und -matrizen:

$$X = \begin{pmatrix} 1 & x_{12} & \dots & x_{1K} \\ 1 & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \dots & x_{nK} \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- *Beachte*: X ist eine $n \times K$-Matrix, d.h. n Zeilen und K Spalten.
  - ▷ i-te Zeile gehört zur i-ten Beobachtung
  - ▷ k-te Spalte gehört zur k-ten erklärenden Variablen

Beispiel Einfache lineare Regression

- Spezialfall: Nur eine erklärende Variable, d.h. $K = 2$.
- Beste Approximation durch Minimierung von

$$S(\beta_1, \beta_2) = \sum_{i=1}^{n} (y_i - (\beta_1 + \beta_2 x_i))^2$$

mit den partiellen Ableitungen

$$\frac{\partial S(\beta_1, \beta_2)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - (\beta_1 + \beta_2 x_i))$$

und

$$\frac{\partial S(\beta_1, \beta_2)}{\partial \beta_2} = -2 \sum_{i=1}^{n} x_i (y_i - (\beta_1 + \beta_2 x_i))$$

Nullsetzen und Auflösen (d.h. der durchschnittliche Approximationsfehler ist 0):

$$-2 \sum_{i=1}^{n} (y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 x_i)) \stackrel{!}{=} 0 \text{ und } -2 \sum_{i=1}^{n} x_i (y_i - (\widehat{\beta}_1 + \widehat{\beta}_2 x_i)) \stackrel{!}{=} 0$$

- Lösung:

$$\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2 \overline{x}$$

  und

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

  Die optimale Steigung $\widehat{\beta}_2$ ist Quotient aus Kovarianz und Varianz.

- Allgemein: Minimiere

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - x_i'\beta)^2$$

  bezüglich β bzw.

$$S(\beta) = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

- Notwendige Bedingung für ein Minimum:

$$\text{Ableiten: } \frac{\partial S(\beta)}{\partial \beta} = -2X'(y - X\beta)$$

$$\text{Nullsetzen: } -2(X'y - X'X\widehat{\beta}) \stackrel{!}{=} 0$$

- Lösung:

$$\widehat{\beta} = (X'X)^{-1}X'y = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i$$

- Fazit: Beste lineare Approximation (i.S. des OLS-Kriteriums) von y durch $x_1, \ldots, x_K$:

$$\widehat{y_i} = x_i'\widehat{\beta} \text{ für } i = 1, 2, \ldots, n.$$

OLS-Beispiel in Matrixschreibweise

- Beispiel: $n = 4$ mit (x, y) gleich $(1,1), (2,3), (2,4), (3,4)$.
- Vektor y und Matrix X lauten

$$\underset{(4\times 1)}{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix}$$

und

$$\underset{(4\times 2)}{X} = \begin{pmatrix} x_1' \\ x_2' \\ x_3' \\ x_4' \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ x_{13} & x_{23} \\ x_{14} & x_{24} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

- Berechne $\widehat{\beta}_{OLS} = (X'X)^{-1}X'y$:

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 \end{pmatrix} \times \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 8 \\ 8 & 18 \end{pmatrix}.$$

$$(X'X)^{-1} = \begin{pmatrix} 4 & 8 \\ 8 & 18 \end{pmatrix}^{-1} = \frac{1}{72-64} \begin{pmatrix} 18 & -8 \\ -8 & 4 \end{pmatrix} = \begin{pmatrix} 2.25 & -1 \\ -1 & 0.5 \end{pmatrix}.$$

$$X'y = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \\ 4 \\ 4 \end{pmatrix} = \begin{pmatrix} 12 \\ 27 \end{pmatrix}.$$

$$(X'X)^{-1}X'y = \begin{pmatrix} 2.25 & -1 \\ -1 & 0.5 \end{pmatrix} \begin{pmatrix} 12 \\ 27 \end{pmatrix} = \begin{pmatrix} 27-27 \\ -12+13.5 \end{pmatrix} = \begin{pmatrix} 0 \\ 1,5 \end{pmatrix}.$$

- OLS Schätzer:

Achsenabschnitt $\widehat{\beta}_1 = 0$ und Steigungsfaktor $\widehat{\beta}_2 = 1,5$.

- Beispiel für einfaches Regressionsmodell: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$.

$$X'X = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum\limits_{i=1}^n x_i \\ \sum\limits_{i=1}^n x_i & \sum\limits_{i=1}^n x_i^2 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{n\sum\limits_{i=1}^n x_i^2 - (\sum\limits_{i=1}^n x_i)^2} \begin{pmatrix} \sum\limits_{i=1}^n x_i^2 & -\sum\limits_{i=1}^n x_i \\ -\sum\limits_{i=1}^n x_i & n \end{pmatrix}$$

$$= \frac{1}{\sum\limits_{i=1}^n x_i^2 - n\overline{x}^2} \begin{pmatrix} \frac{1}{n}\sum\limits_{i=1}^n x_i^2 & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

$$X'y = \begin{pmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=1}^n y_i \\ \sum\limits_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} n\overline{y} \\ \sum\limits_{i=1}^n x_i y_i \end{pmatrix}$$

$$(X'X)^{-1}X'y = \frac{1}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \begin{pmatrix} \overline{y}\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i y_i \\ -\overline{x}n\overline{y} + \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

$$= \frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \begin{pmatrix} \overline{y}\sum_{i=1}^{n} x_i^2 - \overline{x}\sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \end{pmatrix} = \begin{pmatrix} \overline{y} - \widehat{\beta}_2\overline{x} \\ \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \end{pmatrix}$$

- Also sind $\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2\overline{x}$ und $\widehat{\beta}_2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$.

Ausgangspunkt

- Ökonomen wollen nicht nur optimal approximieren, sondern fundamentale ("naturgesetzliche") Beziehungen entdecken, die auch Aussagen über Variablenwerte erlauben, die nicht beobachtet werden können.
- Annahme: Es gibt in einer wohldefinierten Grundgesamtheit (aller Haushalte, aller Unternehmen, etc.) eine solche fundamentale Beziehung.
- Statistisches Modell für die Grundgesamtheit:

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i, \qquad i = 1, ..., n$$

oder

$$y_i = x_i' \beta + \varepsilon_i, \qquad i = 1, ..., n$$

oder

$$y = X\beta + \varepsilon$$

Kennzeichen der Variablen

- x_i und y_i sind prinzipiell beobachtbar und werden in einer konkreten Stichprobe tatsächlich beobachtet;
- der Approximationsfehler ε_i ist prinzipiell unbeobachtbar.

  Da andere Stichproben auch andere Werte von y_i, x_i und ε_i liefern, sind diese Zufallsvariablen.

  $\widehat{\beta}$ als Vektor von Zufallsvariablen mit einer eigenen Zufallsverteilung.
- **Situation 1:** Zusätzliche Annahme, dass x_i deterministisch ist, d.h. jede Stichprobe liefert denselben Wert von x_i.
  - ▷ Realistisch für Laborexperimente, da Versuchsbedingungen kontrolliert werden können.
  - ▷ Unrealistisch für nicht-experimentelle Daten!
  - ▷ Konvention: Es ist trotzdem üblich, x_i als deterministisch aufzufassen.

- **Situation 2:** Wenn x_i eine Zufallsvariable ist, werden Annahmen über die Beziehung von x_i und ε_i benötigt.
  - ▷ Wenn (y_i, x_i), $i = 1, 2, \ldots, n$ eine Zufallsstichprobe bilden, dann sind z.B. $\varepsilon_1, \ldots, \varepsilon_n$ stochastisch unabhängig (realistisch für Querschnittsdaten).
  - ▷ Wenn (y_i, x_i), $i = 1, 2, \ldots, n$ Zeitreihendaten sind, ist die Annahme der Unabhängigkeit unrealistisch und auch nicht durch einen Prozess der Stichprobenziehung erklärbar (Datengenerierung statt Stichproben).
- *Beachte*:

$$y_i = x_i'\beta + \varepsilon_i$$

ist eine Tautologie, wenn keine zusätzlichen Annahmen getroffen werden!

Annahme der schwachen Exogenität

- Statistische Annahme der **schwachen Exogenität** von x_i:

$$E(\varepsilon_i | x_i) = 0,$$

so dass

$$E(y_i | x_i) = x_i' \beta$$

D.h. β_k misst wie sich im Mittel y_i verändert, wenn sich "ceteris paribus" x_{ik} um eine Einheit verändert.

- *Beachte*: Oft wird β_i als Stärke der Wirkung eines ökonomischen Kausaleffekts interpretiert. Dann besitzt aber ε_i auch eine ökonomische Interpretation. Wenn x_i und ε_i korreliert sind, ist große Vorsicht bei der Interpretation von Regressionskoeffizienten als Maß für die Stärke von Kausaleffekten gegeben.

Gauß-Markov-Annahmen

- Einfache, aber nicht notwendig realistische Annahmen, unter denen sich die Eigenschaften des OLS-Schätzers besonders einfach diskutieren lassen.
- Gauß-Markov-Annahmen:

(**A1**) $E(\varepsilon_i) = 0$ für $i = 1, 2, \ldots, n$ (d.h. die Regressionsebene ist im Mittel korrekt);

(**A2**) Strenge Exogenität: $(x_1, \ldots, x_n)$ und $(\varepsilon_1, \ldots, \varepsilon_n)$ sind stochastisch unabhängig;

(**A3**) Homoskedastie: Alle Störgrößen besitzen dieselbe Varianz, d.h. $V(\varepsilon_i) = \sigma^2$ für $i = 1, 2, \ldots, n$;

(**A4**) Fehlen von Autokorrelation: Verschiedene Störgrößen sind nicht korreliert, d.h. $cov(\varepsilon_i, \varepsilon_j) = 0$ für $i, j = 1, 2, \ldots, n$ und $i \neq j$.

- (**A1**), (**A3**) und (**A4**) in Matrixschreibweise:

$$E(\varepsilon) = 0 \text{ und } V(\varepsilon) = \sigma^2 I_n,$$

  wobei I_n die $n \times n$ Einheitsmatrix ist.

  D.h. die Fehlerterme sind unkorrelierte Ziehungen aus einer Zufallsverteilung mit Erwartungswert Null und konstanter Varianz σ^2.

- Abschwächung von (**A2**):

$$E(\varepsilon|X) = 0 \text{ und } V(\varepsilon|X) = E(\varepsilon\varepsilon'|X) = \sigma^2 I_n$$

  D.h. X liefert keine Information über den Erwartungswert und die Varianz-Kovarianzmatrix von ε.

  ▷ Diese Bedingungen folgen aus (A2) und genügen für

$$E(y_i|x_i) = x_i'\beta$$

Eigenschaften des OLS-Schätzers

- Aus (**A1**) + (**A2**) folgt die **Unverzerrtheit** von $\widehat{\beta}$. D.h. bei wiederholten Stichprobenziehen und Berechnung von $\widehat{\beta}$, wird das arithmetische Mittel der $\widehat{\beta}$'s ungefähr das unbekannte, wahre β liefern. Wegen

$$\widehat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$$

ist

$$E(\widehat{\beta}) = \beta + E((X'X)^{-1}X')E(\varepsilon) = \beta$$

- Wie weit ist im Durchschnitt der unverzerrte OLS-Schätzer vom wahren Wert entfernt?

$$V(\widehat{\beta}) = E[(\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'] = (X'X)^{-1}X'V(\varepsilon)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

Gauß-Markov-Theorem

- Unter den Annahmen (**A1**) bis (**A4**) ist der OLS-Schätzer unter allen linearen unverzerrten Schätzern, derjenige mit der "kleinsten" Varianz-Kovarianz-Matrix.

  Kurz: BLUE = **B**est **L**inear **U**nbiased **E**stimator.

- Doch welche exakte Verteilung besitzt der OLS-Schätzer?

  Das Ergebnis ist abhängig von der für die Störgröße unterstellten Verteilung!

(**A5**) Normalverteiltheit der Störgrößen: $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

Verteilung des OLS-Schätzers

- Wenn die Annahmen (**A1**) bis (**A5**) erfüllt sind, dann ist der OLS-Schätzer K-dimensional normalverteilt:

$$\widehat{\beta} \sim \mathcal{N}_K(\beta, \sigma^2 (X'X)^{-1})$$

- Da die Randverteilungen von multivariaten Normalverteilungen wieder Normalverteilungen sind, gilt:

$$\widehat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2 c_{kk}),$$

  wobei c_{kk} das k-te Hauptdiagonalelement von $(X'X)^{-1}$ bezeichnet.
- *Beachte*: Obwohl die Störgröße unbeobachtbar ist, kann ihre Verteilung nicht beliebig gesetzt werden. Denn: Aus der Normalverteiltheit von ε_i folgt die Normalverteiltheit von y_i und die ist "überprüfbar".

Varianz des OLS-Schätzers

- Wie wird die unbekannte Varianz der Störgrößen geschätzt?

$$\text{Vorschlag:} \quad s^2 = \frac{1}{(n-K)} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 = \frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{(n-K)}$$

  Dann ist unter den Annahmen (**A1**) bis (**A4**) s^2 unverzerrt für σ^2. D.h. $E(s^2) = \sigma^2$.

- Schätzung von $V(\widehat{\beta})$?

$$\hat{V}(\widehat{\beta}) = s^2 (X'X)^{-1} = s^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1}$$

  Zieht man die Wurzel aus dem k-ten Hauptdiagonalelement von $\hat{V}(\widehat{\beta})$, so erhält man den Standardfehler von $\widehat{\beta}_k$:

$$\sqrt{\left[\hat{V}(\widehat{\beta}_k) \right]_{kk}} = s\sqrt{c_{kk}}$$

Beispiel: Individuelle Löhne

- Daten: Stichprobe von 3296 jungen abhängig Beschäftigten
- Datenquelle: US National Longitudinal Survey (NLS) aus dem Jahre 1987
- Variablen: Individuelle Löhne, Geschlecht, Rasse, Schuljahre
- Variablenwerte: Geschlecht erhält Dummy-Codierung (männlich=1)
- Statistisches Modell:

$$wage_i = \beta_1 + \beta_2 \cdot male_i + \varepsilon_i$$

Ergebnis der OLS-Schätzung

OLS-Schätzergebnisse für die Lohngleichung

Abhängige Variable: $wage$		
Variable	Schätzwert	Standardfehler
Konstante	5.1469	0.1011
$male$	1.2777	0.1397
$s = 4.0048,\ R^2 = 0.0248,\ F = 83.68$		

- *Beachte*: Mit $E(\varepsilon_i) = 0$ ist

$$\beta_1 = E(wage_i | male_i = 0)$$

und

$$\beta_2 = E(wage_i | male_i = 1) - \beta_1 = E(wage_i | male_i = 1) - E(wage_i | male_i = 0)$$

Goodness of Fit

- **Bestimmtheitsmaß**

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

- *Beachte*: Falls ein Achsenabschnitt im Modell vorhanden ist, gilt:

$$\hat{V}(y_i) = \hat{V}(\widehat{y}_i) + \hat{V}(\widehat{\varepsilon}_i)$$

  (vgl. Normalgleichungen).

- Woraus das Bestimmtheitsmaß sich umschreiben lässt zu:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Interpretation des Bestimmtheitsmaßes

- **Interpretation 1**: Teilt man Zähler und Nenner durch $n-1$ oder n, so misst das Bestimmtheitsmaß den Anteil der Streuung von y, der sich durch $x_2, \ldots, x_K$ (d.h. das Modell) erklären lässt.
- **Interpretation 2**: Betrachtet man als einfachstes Modell ein Modell nur mit Achsenabschnitt $y_i = \beta_1 + \varepsilon_i$, dann ist offensichtlich $\overline{y}$ der OLS-Schätzer für β_1 und die Residuenquadratsumme lautet:

$$\sum_{i=1}^{n} (y_i - \overline{y})^2$$

Somit misst

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

um wieviel Prozent sich die Residuenquadratsumme verringert, wenn neben dem Achsenabschnitt weitere erklärende Variablen im Modell aufgenommen werden.

Eigenschaften des Bestimmtheitsmaßes

1. $0 \leq R^2 \leq 1$.
2. $\widehat{\varepsilon}_i = 0$ für $i = 1, 2, \ldots, n \Leftrightarrow R^2 = 1$, d.h. alle Punkte liegen auf einer "Gerade" (allgemein: Hyperebene).
3. $\widehat{y}_i = \overline{y}$ für $i = 1, 2, \ldots, n \Leftrightarrow R^2 = 0$, d.h. die Regressions"gerade" verläuft (horizontal) parallel zur x-Achse.

Ohne Achsenabschnitt stimmen die verschiedenen Darstellungen für das Bestimmtheitsmaß nicht überein; die letztgenannte Formel kann sogar negativ werden. Dann lässt sich das **unzentrierte Bestimmtheitsmaß**

$$\text{uncentered } R^2 = \frac{\sum_{i=1}^n \widehat{y}_i^2}{\sum_{i=1}^n y_i^2} = 1 - \frac{\sum_{i=1}^n \widehat{\varepsilon}_i^2}{\sum_{i=1}^n y_i^2}$$

betrachten, das wieder die gewünschten Eigenschaften besitzt.

Nachteile des Bestimmtheitsmaßes

- **Problem 1**: Abhängigkeit der Interpretation des Bestimmtheitsmaßes von der Variation der zu erklärenden Variable y.

  Ist $R^2 = 0.02$ groß oder klein?

- **Problem 2**: Bestimmtheitsmaß ist kein Maß für die Modellgüte, sondern ein Maß für die Anpassungsgüte der OLS-Methode.
  D.h. jede andere Schätzmethode muss zwangsläufig ein kleineres Bestimmtheitsmaß liefern.
  Für andere Schätzmethoden gilt die Äquivalenz zwischen den beiden Darstellungen des Bestimmtheitsmaßes nicht.

- **Lösung**: Definition eines Gütemaßes, das für OLS in das Bestimmtheitsmaß übergeht und für jede Schätzmethode Werte zwischen 0 und 1 liefert:

$$corr^2\{y_i, \widehat{y_i}\} = \frac{\left(\sum_{i=1}^n (y_i - \overline{y})(\widehat{y_i} - \overline{\widehat{y}})\right)^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \sum_{i=1}^n (\widehat{y_i} - \overline{\widehat{y}})^2}$$

  Wie gut wird die Variation von y_i durch die Variation der angepassten Werte $\widehat{y_i}$ erklärt?

- *Beachte*: Weiterhin wird aber lediglich die Anpassungsgüte und nicht die Modellgüte gemessen.
  Lösung: Es werden noch sog. Informationskriterien verwendet.

- **Problem 3**: Bestimmheitsmaß wächst monoton mit der Anzahl der erklärenden Variablen, d.h. Modelle werden umso besser beurteilt, je größer sie sind. D.h. Verletzung des Prinzips der Sparsamkeit. Lösungsvorschlag: **Angepasstes (adjusted) Bestimmtheitsmaß** (bei Achsenabschnitt)

$$\text{adjusted } R^2 = \overline{R}^2 = 1 - \frac{1/(n-K)\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{1/(n-1)\sum_{i=1}^{n}(y_i-\overline{y})^2}$$

  ist monoton abnehmend in K.
- *Beachte*: Das angepasste Bestimmtheitsmaß kann negativ werden.

t-Verteilung

- Die Verteilung der Regressionsparameter ist bekannt für $k = 1, 2, ...K$:

$$\widehat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2 c_{kk}) \quad \Rightarrow \quad \frac{\widehat{\beta}_k - \beta_k}{\sigma\sqrt{c_{kk}}} \sim \mathcal{N}(0,1)$$

- Problem: σ^2 unbekannt $\Rightarrow s^2$

$$\frac{(n-K)s^2}{\sigma^2} = \frac{\sum \widehat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2_{n-K},$$

so dass (wegen der stoch. Unabhängigkeit von $\widehat{\beta}$ und s^2):

$$t_k = \frac{\widehat{\beta}_k - \beta_k}{s\sqrt{c_{kk}}} \sim t_{n-K}$$

Einfacher t-Test (zweiseitig)

- $H_0 : \beta_k = \beta_k^0$ gegen $H_1 : \beta_k \neq \beta_k^0$
- Signifikanzniveau $\alpha \Rightarrow t_{n-K;\alpha/2}$ mit

$$P(|t_k| > t_{n-K;1-\alpha/2}) = \alpha$$

  mit kritischer Schranke $t_{n-K;1-\alpha/2}$ ($(1-\alpha/2)$-Quantil der t-Verteilung mit $n-K$ Freiheitsgraden).
- $n-K$ nicht zu klein, $\alpha = 0.05 \Rightarrow t_{n-K;1-\alpha/2} \approx 1.96$.
- *Beachte*: Test auf i.a. $\beta_k^0 = 0$ in Softwarepaketen.
  Bei Ablehnung der Nullhypothese wird davon gesprochen, dass $\widehat{\beta}_k$ signifikant von 0 verschieden (oder kurz signifikant) ist.

Einfacher t-Test (einseitig)

- $H_0 : \beta_k \leq \beta_k^0$ gegen $H_1 : \beta_k > \beta_k^0$
- Signifikanzniveau $\alpha \Rightarrow t_{n-K;\alpha}$ mit

$$P(t_k > t_{n-K;1-\alpha}) = \alpha$$

  mit kritischer Schranke $t_{n-K;1-\alpha}$ ($(1-\alpha)$-Quantil der t-Verteilung mit $n-K$ Freiheitsgraden).
- $n - K$ nicht zu klein, $\alpha = 0.05 \Rightarrow t_{n-K;1-\alpha} \approx 1.64.$

Konfidenzintervall zum Niveau $1 - \alpha$

- Entspricht dem Nicht-Ablehnungsbereich eines zweiseitigen t-Testes:

$$P(\widehat{\beta}_k - t_{n-K;1-\alpha/2} s\sqrt{c_{kk}} < \beta_k < \widehat{\beta}_k + t_{n-K;1-\alpha/2} s\sqrt{c_{kk}}) = 1 - \alpha$$

bzw. kurz

$$\left[\widehat{\beta}_k - t_{n-K;\alpha/2} s\sqrt{c_{kk}}\ ,\ \widehat{\beta}_k + t_{n-K;\alpha/2} s\sqrt{c_{kk}}\right]$$

Beispiel: Individuelle Löhne

OLS-Schätzergebnisse für die Lohngleichung

Abhängige Variable: $wage$		
Variable	Schätzwert	Standardfehler
Konstante	5.1469	0.1011
$male$	1.2777	0.1397
$s = 4.0048,\ R^2 = 0.0248,\ \overline{R}^2 = 0.0245,\ F = 83.68$		

D.h.

1. Nullhypothese $H_0 : \beta_2 = 0$ bedeutet, dass das erwartete Lohndifferential zwischen Männern und Frauen gleich 0 ist.
2. Für β_2 ist $t = 1.2777/0.1397 = 9.15 > 1.96 \approx 2$, so dass der Einfluss des Geschlechts signifikant auf dem 5%-Niveau ist.
3. Realisiertes Konfidenzintervall zum 95%-Niveau für β_2: $1.28 \pm 1.96 \times 0.14$.

Testen einer linearen Restriktion

- $H_0 : r_1\beta_1 + ... + r_K\beta_K = r'\beta = q$
- Verteilung von $r'\widehat{\beta}$ für normalverteilte Residuen:

$$\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}) \quad \Rightarrow \quad r'\widehat{\beta} \sim \mathcal{N}(r'\beta, \sigma^2 r'(X'X)^{-1}r))$$

- Verteilung von $r'\widehat{\beta}$ unter H_0:

$$t = \frac{r'\widehat{\beta} - q}{s\sqrt{r'(X'X)^{-1}r}} \sim t_{n-K}$$

Gemeinsamer Test auf Signifikanz

- Ausgangspunkt: J der K Regressoren haben gemeinsam keinen signifikanten Einfluss.
  Beachte: Achsenabschnitt zählt nicht zu diesen Regressoren.
- Nullhypothese:
$$H_0 : \beta_{K-J+1} = \ldots = \beta_K = 0$$
- Residuenquadratsummen:
  - S_0 bezeichne Fehlerquadratsumme unter H_0.
  - S_1 bezeichne Fehlerquadratsumme bei freier Schätzung.
- Testidee: Wenn die Nullhypothese falsch ist, muss die Residuenquadratsumme bei freier Schätzung deutlich kleiner sein als bei Schätzung unter H_0.

- Wegen der folgenden Eigenschaften
  - ▷ $\frac{S_0 - S_1}{\sigma^2} \sim \chi^2(J)$ unter H_0,
  - ▷ $\frac{(n-K)s^2}{\sigma^2} = \frac{S_1}{\sigma^2} \sim \chi^2(n-K)$,
  - ▷ $S_0 - S_1$ und S_1 stochastisch unabhängig

  folgt unter den Annahmen (**A5**) und (**A1**) bis (**A4**):

$$f = \frac{(S_0 - S_1)/J}{S_1/(n-K)} = \frac{(S_0 - S_1)/J}{s^2} \quad \overset{H_0}{\sim} \quad F(J; n-K)$$

- Alternative Darstellung mittels Bestimmtheitsmaß: Mit

$$R_0^2 = 1 - \frac{S_0}{\sum_{i=1}^n (y_i - \overline{y})^2} \qquad \text{und} \qquad R_1^2 = 1 - \frac{S_1}{\sum_{i=1}^n (y_i - \overline{y})^2}$$

  folgt

$$f = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(n-K)} = \frac{(S_0 - S_1)/J}{S_1/(n-K)}$$

- *Beachte*: Trotz der zweiseitigen Alternativhypothese ist der Ablehnungsbereich einseitig:
$$P(f > F^J_{n-K;\alpha}) = \alpha$$
- Typische Werte: Für $\alpha = 0.05$, $n - K = 60$ und $J = 3$ ist $F^3_{60,0.05} = 2.76$
- *Beachte*: $\widehat{\beta}_1, \ldots, \widehat{\beta}_K$ sind korreliert!
D.h. es können u.U. kleine t-Werte für $\beta_2, \ldots, \beta_K$ mit einem relativ großen F-Wert für $(\beta_2, \ldots, \beta_K)$ einhergehen.
D.h. es können sämtliche Regressoren einzeln insignifikant und gemeinsam signifikant sein (Multikollinearität).

Spezialfall: Test auf die Gültigkeit des gesamten Modells

- Nullhypothese:

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_K = 0$$

- Prüfmaß:

$$f = \frac{(S_0 - S_1)/(K-1)}{S_1/(n-K)}$$

- Mit $S_0 = \sum (y_i - \bar{y})^2$ (und damit $R_0^2 = 0$) gilt unter H_0:

$$f = \frac{R^2/(K-1)}{(1-R^2)/(n-K)} \quad \overset{H_0}{\sim} \quad F(K-1; n-K)$$

Interpretation des Tests

- Wenn dieser Test die Nullhypothese nicht ablehnt, muss das Modell sehr schlecht sein. Eine Modell nur mit Achsenabschnitt hat eine “ähnlich” große Residuenquadratsumme wie das Modell mit Regressoren.
- *Beachte*: Wenn die Nullhypothese abgelehnt wird, kann aber nicht geschlossen werden, dass das Modell gut, perfekt, gültig oder das beste ist.

Beispiel: Individuelle Löhne

- Als Regressoren werden neben Geschlecht ($male$), Anzahl der Schuljahre ($school$) und Berufserfahrung in Jahren ($exper$) betrachtet.
- Modell beschreibt den bedingten erwarteten Lohn eines Individuums gegeben Geschlecht, Anzahl der Schuljahre und Berufserfahrung.
- Der Regressionskoeffizient β_2 der Variablen Geschlecht ($male$) misst die Differenz des erwarteten Lohns zwischen Männern und Frauen mit derselben Schulausbildung und Berufserfahrung.
- Der Regressionskoeffizient β_3 der Variablen Anzahl der Schuljahre misst die erwartete Lohndifferenz zweier Individuen desselben Geschlechts und mit derselben Berufserfahrung, wenn die Schulausbildung des einen ein Jahr länger gedauert hat.

Ergebnis der OLS-Schätzung

OLS-Schätzergebnisse für die Lohngleichung

Abhängige Variable: $wage$

Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-2.8901	0.5916	-4.884
$male$	1.4702	0.1370	10.729
$school$	0.6204	0.0417	14.861
$exper$	0.0896	0.0302	2.964

$s = 3.87797,\ R^2 = 0.0861,\ \overline{R}^2 = 0.0853,\ F = 103.387$

Interpretation:

1. Das erwartete Lohndifferential zwischen Männern und Frauen ist von \$1.28 auf \$1.47 gestiegen. Der Standardfehler ist \$0.14, so dass dieser Unterschied signifikant ist.
2. Die t-Statistik zeigt, dass die Nullhypothese "Anzahl der Schuljahre hat keine Einfluss auf den Lohn gegeben Geschlecht und Berufserfahrung" abgelehnt werden kann $(14.861 > 1.96)$.
   Der geschätzte Lohnzuwachs eines zusätzlichen Schuljahres gegeben Geschlecht und Berufserfahrung beträgt \$0.62.

3. Der F-Wert zeigt, dass die drei Regressoren gemeinsam einen signifikanten Einfluss haben $(103.4 > 2.61)$.
4. Testet man die Nullhypothese, dass Anzahl der Schuljahre und Berufserfahrung gemeinsam keinen Einfluss auf den erwarteten Lohn besitzen gegeben das Geschlecht, so sieht man, dass der Wert des Bestimmtheitsmaßes von 0.0248 auf 0.0861 gestiegen ist. Damit ist
$$f = \frac{(0.0861 - 0.0248)/2}{(1 - 0.0861)/(3296 - 4)} = 110.41 > 3.00,$$
so dass die Nullhypothese abgelehnt werden kann (mit einer Irrtumswahrscheinlichkeit von 5%).

Fazit: Das Modell mit den Variablen Geschlecht, Anzahl der Schuljahre und Berufserfahrung ist signifikant besser als ein Modell lediglich mit der Variable Geschlecht.

Allgemeiner Fall

- Nullhypothese: J lineare Restriktionen zwischen den K Regressionskoeffizienten:

$$H_0 : R\beta = q$$

  mit R als $J \times K$-Matrix mit Rang J.
  Beachte: $J \leq K$.

- ***Beispiel***: Seien

$$\begin{aligned} 1 &= \beta_2 + \beta_3 + \ldots + \beta_k \\ \beta_2 &= \beta_3, \end{aligned}$$

  dann sind

$$R = \begin{pmatrix} 0 & 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & -1 & 0 & \ldots & 0 \end{pmatrix}, \qquad q = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

- Verteilung von $R\widehat{\beta}$ für normalverteilte Residuen:

$$R\widehat{\beta} \sim \mathcal{N}(R\beta, \sigma^2 R(X'X)^{-1}R')$$

- Unter H_0 gilt:

$$\xi = (R\widehat{\beta} - q)'(R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - q)/\sigma^2 \quad \sim \quad \chi^2(J)$$

- Problem: σ^2 ist unbekannt.

Mögliche Lösungen

1. **WALD-Test**: Ersetze σ^2 durch die Residuenvarianz einer OLS-Schätzung s^2. Dann ist für große Stichprobenumfänge

$$W = (R\widehat{\beta} - q)'(R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - q)/s^2 \quad \overset{approx}{\sim} \quad \chi^2(J)$$

2. F-**Test mit Prüfmaß**:

$$\begin{aligned} f &= \frac{(R\widehat{\beta} - q)'(R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - q)/(\sigma^2 J)}{((n-K)s^2/\sigma^2)/(n-K)} \\ &= \frac{(R\widehat{\beta} - q)'(R(X'X)^{-1}R')^{-1}(R\widehat{\beta} - q)/J}{s^2} \quad \overset{H_0}{\sim} \quad F(J; n-K) \end{aligned}$$

Größe, Macht (Power) und p-Werte

- Fehler beim Testen statistischer Hypothesen:
  - Fehler 1. Art: Ablehnung der Nullhypothese, obwohl sie wahr ist,
  - Fehler 2. Art: Nicht-Ablehnung der Nullhypothese, obwohl sie falsch ist.
- Die Wahrscheinlichkeit des Fehlers 1. Art wird durch die Festlegung des Signifikanzniveaus (auch Größe des Testes genannt) gesteuert.
- Die Gegenwahrscheinlichkeit des Fehlers 2. Art wird groß (u.U. nahe 1) sein, wenn der wahre Parameterwert weit von dem der Nullhypothese entfernt liegt.
- Die Macht (oder Güte) eines Testes ist die Gegenwahrscheinlichkeit der Wahrscheinlichkeit des Fehlers 2. Art: Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie falsch ist.
- *Beachte*: Reduzierung des Signifikanzniveaus verringert i.a. die Güte eines Testes.

- Wenn z.B. der wahre Wert für β_2 gleich 0.1 ist und $H_0 : \beta_2 = 0$ getestet wird, dann hängt die Wahrscheinlichkeit, die Nullhypothese abzulehnen, von dem Standardfehler von $\widehat{\beta_2}$ ab, und dieser wird vom Stichprobenumfang beeinflusst.
  Je größer der Stichprobenumfang ist, desto kleiner ist der Standardfehler und desto größer ist die Wahrscheinlichkeit, die Nullhypothese abzulehnen.
- Problem: In großen Stichproben wird die Nullhypothese fast immer abgelehnt und der Fehler 2. Art besitzt eine sehr kleine Wahrscheinlichkeit.
  Lösung des Praktikers:
  - In großen Stichproben wird statt des 5%-Signifikanzniveaus ein 1%-Niveau betrachtet.
  - In kleinen Stichproben umgekehrt das 10%-Niveau.

- Gemeinhin wird die Nullhypothese solange als wahr angesehen, solange sie nicht abgelehnt werden konnte.
- Problem: Es können aber u.U. verschiedene, sich widersprechende Nullhypothesen betrachtet und nicht abgelehnt werden.
  Lösung: Die einzig "erlaubte" Aussage ist, dass keiner der alternativen Nullhypothesenwerte abgelehnt werden konnte.
- *Beachte*: Häufig sind ökonometrische Tests nicht sehr trennscharf (powerful), so dass sehr große Stichproben benötigt werden.
- Der p-Wert eines Testes ist die kleinste Wahrscheinlichkeit, für die bei gegebenem Stichprobenbefund die Nullhypothese gerade noch abgelehnt werden kann.
  Entscheidungsregel: Ist der p-Wert kleiner als das Signifikanzniveau wird die Nullhypothese abgelehnt.
  Vorteil: Es werden keine "Tabellen" der kritischen Werte benötigt.

OLS Konsistenz

- Konsistenz
  - bedeutet, dass der Wahrscheinlichkeitslimes (plim) von $\widehat{\beta}$ gleich β ist.
  - D.h.: $\lim\limits_{n\to\infty} P[|\widehat{\beta} - \beta| < \epsilon] = 1$ für alle $\epsilon > 0$.
- Wir erhalten also

$$
\begin{aligned}
\text{plim } \widehat{\beta} &= \text{plim } \{\beta + (X'X)^{-1}X'\varepsilon\} \\
&= \text{plim } \beta + \text{plim } \{(\sum_{i=1}^{n} x_i x_i')^{-1} \sum_{i=1}^{n} x_i \varepsilon_i\} \\
&= \text{plim } \beta + \text{plim}(\frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1} \times \text{plim } \frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i \\
&= \beta + (\text{plim } \frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1} \times 0 \\
&= \beta
\end{aligned}
$$

  - $\text{plim } \{A_n \times b_n\} = \text{plim } A_n \times \text{plim } b_n$, wenn die plims Konstanten sind.
  - Die Wahrscheinlichkeitslimes existieren aufgrund des Gesetzes der Großen Zahlen.
  - Für $\text{plim } \frac{1}{n}\sum\limits_{i=1}^{n} x_i \varepsilon_i = 0$ wird $E(\varepsilon_i | x_i) = 0$ angenommen.

Asymptotische Normalverteiltheit

- **Zentraler Grenzwertsatz**: $X_1, \ldots, X_n$ i.i.d. mit μ, σ^2. Dann gilt

$$\frac{\frac{1}{n}\sum X_i - \mu}{\sigma/\sqrt{n}} \longrightarrow \mathcal{N}(0,1) \text{ mit } n \to \infty$$

  Für $\mu = 0$ folgt

$$\frac{\sqrt{n}}{n}\sum X_i = \sqrt{n}\bar{X}_n \longrightarrow \mathcal{N}(0,\sigma^2) \quad \text{mit} \quad n \to \infty$$

- Übertragung des zentralen Grenzwertsatzes: Da $x_1\varepsilon_1, \ldots, x_n\varepsilon_n$ i.i.d. unter (**A1**) - (**A4**) folgt

$$\sqrt{n}(\widehat{\beta} - \beta) \to \mathcal{N}(0, \sigma^2\Sigma_{xx}^{-1})$$

- n endlich, aber groß führt zur folgenden approximativen Verteilung:

$$\widehat{\beta} \overset{approx}{\sim} \mathcal{N}(\beta, \sigma^2 \Sigma_{xx}^{-1}/n)$$

  Problem: Diese Approximation ist nicht hilfreich, da σ^2 und Σ_{xx} unbekannt sind.
- Lösung: Für n noch etwas größer ist

$$\widehat{\beta} \overset{approx}{\sim} \mathcal{N}\left(\beta, s^2 \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\right)$$

- **Wichtige Folgerung**: Verteilungseigenschaften des OLS-Schätzers und der t- bzw. F-Statistiken bleiben für große n approximativ gültig, auch wenn $\varepsilon_1, \ldots, \varepsilon_n$ *nicht normalverteilt* sind.

Ergebnisinterpretation

- Dabei wird derjenige Wert y_0 gesucht, der sich für Werte x_0 ergibt, wenn das Modell für alle potentiellen Beobachtungen (und damit auch für y_0 und x_0) als gültig angenommen wird:
$$y_0 = x_0'\beta + \varepsilon_0$$
- Prädiktor (Zufallsvariable) für y_0 (ebenfalls Zufallsvariable) ergibt sich durch Extrapolation der OLS-Schätzgleichung:
$$\hat{y}_0 = x_0'\widehat{\beta}$$
- Unverzerrtheit des OLS-Prädiktors (wenn der OLS-Schätzer unverzerrt ist):
$$E(\hat{y}_0 - y_0) = x_0' E(\widehat{\beta}) - E(y_0) = x_0'\beta - x_0'\beta = 0$$

Verlässlichkeit der Punktprognose

- Varianz der OLS-Prädiktors (unter den Annahmen (**A1**) bis (**A4**)):

$$V(\hat{y}_0) = V(x_0'\widehat{\beta}) = x_0'V(\widehat{\beta})x_0 = \sigma^2 x_0'(X'X)^{-1}x_0$$

- Prognosefehler:

$$y_0 - \hat{y}_0 = x_0'\beta + \varepsilon_0 - x_0'\widehat{\beta} = \varepsilon_0 + x_0'(\beta - \widehat{\beta})$$

- Varianz des Prognosefehlers (sofern ε_0 und $\widehat{\beta}$ unkorreliert sind):

$$V(y_0 - \hat{y}_0) = \sigma^2(1 + x_0'(X'X)^{-1}x_0)$$

- Spezialfall: Einfachregression (mit Achsenabschnitt)

$$V(y_0 - \hat{y}_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)$$

Prognoseintervalle

- *Beachte*: Varianz des Prognosefehlers ist umso größer, je weiter x_0 vom Stichprobenmittel $\overline{x}$ entfernt liegt.
  Mithin können für Werte von x_0, die weit von der Punktwolke entfernt liegen, keine genauen Prognosen von y erwartet werden.
- Prognoseintervall zum Niveau $1-\alpha$ (exakt für normalverteilte Residuen):

$$\left[x_0'\widehat{\beta} - t_{n-K;1-\alpha/2} s\sqrt{1 + x_0'(X'X)^{-1}x_0}\ ,\ x_0'\widehat{\beta} + t_{n-K;1-\alpha/2} s\sqrt{1 + x_0'(X'X)^{-1}x_0}\right]$$

- Approximatives Prognoseintervall zum Niveau 95 % (für großes n):

$$\left[x_0'\widehat{\beta} - 1.96 \cdot s\sqrt{1 + x_0'(X'X)^{-1}x_0}\ ,\ x_0'\widehat{\beta} + 1.96 \cdot s\sqrt{1 + x_0'(X'X)^{-1}x_0}\right]$$

Gliederung Kapitel 3

Multikollinearität: Kennzeichen

- *Beachte*: Im allgemeinen ist es kein Problem, wenn korrelierte Regressoren (z.B. Alter und Erfahrung) aufgenommen werden.
- Wenn die Korrelation zu groß wird, kann es zu Problemen kommen.
- Technisches Problem: $(X'X)$ ist "kaum" noch invertierbar. Dies führt zu großen Standardfehlern und insignifikanten Regressoren.
- Inhaltliches Problem: Es ist keine Zurechnung (Identifikation) von Partialeffekten mehr möglich.
- **Lösung 1**: Erst durch eine Erhöhung des Stichprobenumfangs ist u.U. eine Identifikation möglich.
- **Lösung 2**: Setzen von Restriktionen.

Multikollinearität

- Multikollinearität liegt vor, wenn eine "sehr starke" lineare Beziehung zwischen den Regressoren besteht.
- Exakte Multikollinearität liegt vor, wenn eine exakte lineare Beziehung zwischen zwei oder mehr Regressoren besteht. In diesem Falle ist der OLS-Schätzer nicht mehr durch die Normalgleichungen definiert.
- Häufige Ursache für exakte Multikollinearität: Übermäßiger Gebrauch von Dummy-Variablen.
  Beispiel: Dummy für "female", Dummy für "male" und Achsenabschnitt, da

$$female_i + male_i = 1$$

Lösung: Weglassen des Achsenabschnitts oder eines Dummies.

- Neues Problem, falls Weglassen des Achsenabschnitts (z.B. mit Bestimmtheitsmaß).
- Weitere Ursache für exakte Multikollinearität: Explizite Berücksichtigung von Regressoren, die Linearkombination bereits berücksichtigter Regressoren sind.

Beispiel: In der Lohngleichung kommen Alter, Länge der Schulausbildung und potentielle Berufserfahrung (Alter minus Länge der Schulausbildung minus 6) vor.

Wirkung der Multikollinearität auf den OLS-Schätzer

- Modell:

$$y = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

- Es seien o.B.d.A.
  - ▷ $\overline{y} = \overline{x}_1 = \overline{x}_2 = 0$
  - ▷ $1/n \sum_{i=1}^{n} (x_{i1} - \overline{x}_1)^2 = 1/n \sum_{i=1}^{n} (x_{i2} - \overline{x}_2)^2 = 1$
  - ▷ $r_{12} = 1/n \sum_{i=1}^{n} (x_{i1} - \overline{x}_1)(x_{i2} - \overline{x}_2)$

  Dann ist

$$\begin{aligned} V(\widehat{\beta}) &= \sigma^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \\ &= \sigma^2 \frac{1}{n} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}^{-1} \\ &= \frac{\sigma^2/n}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix} \end{aligned}$$

- Wächst r_{12}^2, so werden die Varianzen von $\widehat{\beta}_1$ und $\widehat{\beta}_2$ größer und die t-Werte kleiner.
- Ist $r_{12} > 0$, sind $\widehat{\beta}_1$ und $\widehat{\beta}_2$ negativ korreliert.

  D.h. wenn r_{12} in der Nähe von 1 liegt, ist die Varianz der Summe klein und die Varianz der Differenz groß, so dass die Summe (gesamter Effekt) relativ genau geschätzt (identifiziert) werden kann.

  Fazit: Bestimmtheitsmaß (globaler Fit) und die Prognoseeignung werden durch Multikollinearität nicht tangiert.

Multikollinearität: Beispiel der Individuallöhne

Abhängige Variable: $wage$

Spezifikation	A	B	C
Konstante	5.147	6.425	—
	(0.101)	(0.096)	
$male$	1.278	—	6.425
	(0.140)		(0.096)
$female$	—	-1.278	5.147
		(0.140)	(0.101)
R^2, unzentr. R^2	0.0248	0.0248	0.6811

Interpretation des linearen Modells

- Wiederholung: Das lineare Modell

$$y_i = x_i'\beta + \varepsilon_i$$

besitzt nur eine geringe Bedeutung, wenn keine Annahmen über ε_i gemacht werden.

- Eine Annahme ist

$$E(\varepsilon_i|X) = E(\varepsilon|x_i) = 0, \qquad \text{(*)}$$

mit der das Regressionsmodell als bedingter Erwartungswert von y_i gegeben x_i interpretiert werden kann.

Beispiele

1. Erwarteter Lohn einer beliebigen Frau mit 40 Jahren, Universitätsausbildung und 14 Jahren Berufserfahrung.
2. Erwartete Arbeitslosenrate für gegebenen Lohn, Inflationsrate, und Bruttoinlandsprodukt einer Volkswirtschaft.

Absolute Marginaleffekte

- Konsequenz der Annahme (*): Interpretation der einzelnen Regressionskoeffizienten β_k als partielle Ableitung des Erwartungswertes

$$\beta_k = \frac{\partial E(y_i|x_i)}{\partial x_{ik}}$$

D.h.: Um wieviel verändert sich y_i, wenn x_{ik} um eine Einheit steigt und alle anderen Variablen konstant gehalten werden (**ceteris paribus-Bedingung**).
Beispiel: Um wieviele Einheiten steigt der erwartete Lohn einer Frau bei einem zusätzlichen Jahr Lebensalter, wenn Ausbildungsniveau und Anzahl der Jahre an Berufserfahrung konstant bleiben.

Konsequenz der ceteris paribus-Bedingung: Einzelne Koeffizienten sind nicht interpretierbar, wenn nicht bekannt ist, welche Variablen das Regressionsmodell insgesamt umfasst.

Probleme der ceteris paribus-Bedingung

1. Kollinearität von Variablen macht es in einer Stichprobe unmöglich, eine Variable zu variieren (z.B. Lebensalter) und andere zu fixieren (z.B. Berufserfahrung).
2. Gelegentlich wird eine Variable auch in transformierter Form in einem Regressionsmodell berücksichtigt. Dann kann bei einer Variation der Variablen die transformierte Variable nicht konstant gehalten werden.
   Beispiel: Lebensalter (age) und Lebensalter^2:

$$\frac{\partial E(y_i|x_i)}{\partial age} = \beta_2 + 2age_i\beta_3$$

   misst den Marginaleffekt einer Variation des Lebensalters, wenn alle anderen Variablen (die nicht Transformation des Lebensalters sind) fixiert werden.

3. Es lassen sich auch Interaktionen (multiplikative Terme) in einem Regressionsmodell berücksichtigen (wie z.B. zwischen Geschlecht und Alter).
Beispiel:

$$y_i = \beta_1 + \beta_2 age_i + \beta_3 male_i \cdot age_i + \beta_4 male_i + \varepsilon_i,$$

dann ist

$$\frac{\partial E(y_i|x_i)}{\partial age} = \beta_2 + male_i \beta_3$$

D.h der Marginaleffekt der Variation des Alters ist β_2 für Frauen und $\beta_2 + \beta_3$ für Männer.

Elastizitäten

- Ökonomen interessieren sich häufig für dimensionslose Elastizitäten statt für dimensionsgebundene Marginaleffekte.
  Elastizitäten messen die relative (zumeist prozentuale) Änderung einer Variablen bei einer relativen (zumeist prozentualen) Änderung einer anderen Variablen.
  Elastizitäten lassen sich direkt in sog. loglinearen Regressionsmodellen

  $$\log y_i = (\log x_i)'\gamma + v_i$$

  schätzen, wobei

  $$\log x_i = (1, \log x_{i2}, \log x_{i3}, \dots, \log x_{iK})'$$

  ist. Dieses Modell setzt zunächst voraus, dass alle involvierten Variablen nur positive Werte annehmen können (d.h. verhältnisskaliert sind).

- Modellannahme: $E(v_i|\log x_i) = 0$
  Beachte:
  $$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} \cdot \frac{x_{ik}}{E(y_i|x_i)} \approx \frac{\partial E(\log y_i|\log x_i)}{\partial \log x_{ik}} = \gamma_k$$
  D.h. die Elastizitäten sind im loglinearen Modell (approximativ) konstant. Diese Beziehung gilt umso exakter, je kleiner die Elastizitäten sind.
- Elastizitäten im linearen Modell sind nicht-konstant und variieren mit x_i:
  $$\frac{\partial E(y_i|x_i)}{\partial x_{ik}} \cdot \frac{x_{ik}}{E(y_i|x_i)} = \frac{x_{ik}}{x_i'\beta}\beta_k$$

Semi-Elastizitäten

- Ökonomische Theorie gibt einem nur selten Anhaltspunkte für die Funktionsform. Da aber die logarithmische Transformation varianzstabilisierend wirkt, kann u.U. mit ihr das Heteroskedastieproblem reduziert werden (siehe Nachweis in den Veranstaltungen zur Zeitreihenanalyse).
- Wenn z.B. die erklärenden Variablen negative Werte annehmen können oder Dummy-Variablen sind, dann ist auch ein halblogarithmisches Modell

  $$\log y_i = x_i'\beta + \varepsilon_i$$

  angesagt. Dann misst β_k den relativen (prozentualen) Marginaleffekt einer absoluten Veränderung von x_{ik} auf y_i.

  Beispiel: Wenn x_{ik} eine Dummy-Variable für Geschlecht "männlich" ist, dann misst β_k das relative Lohndifferential zwischen Männern und Frauen.

- *Beachte*: Die Ungleichheit

$$E(\log y_i|x_i) \neq \log E(y_i|x_i)$$

hat Konsequenzen für die Prognose.
Problem: Unter der Annahme $E(v_i|\log x_i) = 0$ liefert im loglinearen Modell $(\log x_i)'\gamma$ eine Prognose für $\log y_i$. Da man an der Prognose von y_i interessiert ist, könnte $\exp((\log x_i)'\gamma)$ als Prädiktor für y_i verwendet werden. Es gilt aber

$$E(y_i|x_i) \neq \exp((\log x_i')\gamma).$$

Lösung mittels spezifischer Verteilungsannahmen: Wenn v_i normalverteilt mit Mittelwert 0 und Varianz σ_v^2 sind, ist die bedingte Verteilung von y_i lognormal mit Mittelwert

$$E(y_i|x_i) = \exp(E(\log y_i|x_i) + 1/2\sigma_v^2) = \exp((\log(x_i)'\gamma + 1/2\sigma^2).$$

D.h. der "naive" Prädiktor wird um $1/2\sigma_v^2$ korrigiert.
Beachte: Gelegentlich wird diese Korrektur auch dann vorgenommen, wenn keine Lognormalverteilung unterstellt wird.

- *Beachte*: Die für die Interpretation wichtige Bedingung $E(\varepsilon_i|x_i) = 0$ ist keineswegs trivial und in vielen ökonomischen Modellen nicht erfüllt.
  Z.B.: Problem der Identifizierbarkeit in Kapitel 5.
- Wenn zwei Modelle

$$\begin{aligned} E(y_i|x_i) &= x_i'\beta \\ E(y_i|z_i) &= z_i'\gamma \end{aligned}$$

  betrachtet werden, dann können beide "wahr" sein.

  Aber: Sollte eine Variable in x_i und z_i enthalten sein, dann ist ihre Interpretation (ceteris paribus) in beiden Modellen unterschiedlich, da verschiedene Variablen konstant gehalten werden.

- Problem: Betrachtet man aber die Modelle

$$\begin{aligned} E(y_i|x_i, z_i) &= x_i'\beta \\ E(y_i|x_i, z_i) &= z_i'\gamma, \end{aligned}$$

so wird jeweils unterstellt, dass ein Variablensatz explizit keinen Einfluss besitzt. Damit kann auch nur **eines** der beiden Modelle "wahr" sein.

Übersicht Marginaleffekte

	Linear	**Halblogarithmisch**	**Log-linear**
Marginaleffekt	β_k	$x'\beta \cdot \beta_k$	$\beta_k \cdot \frac{x'\beta}{x_k}$
Semi-Elastizität	$\beta_k \cdot \frac{1}{x'\beta}$	β_k	$\beta_k \cdot x_k$
Elastizität	$\beta_k \cdot \frac{x_k}{x'\beta}$	$\beta_k \cdot x_k$	β_k

Interpretation:

- (Absolute) Marginaleffekt gibt an, um wieviel sich y änderrt, wenn sich x um eine Einheit ändert.
- Semi-Elastizität gibt an, um wieviel Prozent sich y ändert, wenn sich x um eine Einheit ändert.
- Elastizität gibt an, um wieviel Prozent sich y ändert, wenn sich x um 1% ändert.

Fehlspezifikation der Menge der Regressoren

- Fälle:
  1. Weglassen relevanter Regressoren
  2. Aufnehmen irrelevanter Regressoren
- Betrachte die beiden Modelle

$$\begin{array}{lrcl} \text{Modell 1} & y_i & = & x_i'\beta + z_i'\gamma + \varepsilon_i \\ \text{Modell 2} & y_i & = & x_i'\beta + v_i, \end{array}$$

  wobei $E(\varepsilon_i|x_i, z_i) = E(v_i|x_i) = 0$ ist.
- *Beachte*: Modell 2 ist im Modell 1 genestet, da sich Modell 2 für $\gamma = 0$ als Spezialfall ergibt.

Fall 1: Weglassen relevanter Variablen

Was passiert, wenn Modell 1 korrekt ist, aber Modell 2 geschätzt wird?

- Der OLS-Schätzer $\widehat{\beta}_2$ für β im Modell 2 lautet:

$$\widehat{\beta}_2 = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i y_i$$

Das wahre (größere) Modell für y_i eingesetzt ergibt:

$$\widehat{\beta}_2 = \beta + \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} x_i z_i' \gamma + \sum_{i=1}^{n} x_i \varepsilon_i \right)$$

Problem: Selbst wenn $E(\varepsilon_i|x_i) = 0$ ist, kann $\widehat{\beta}_2$ im allg. weder unverzerrt noch konsistent sein wegen

$$E(\widehat{\beta}_2) = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i z_i' \gamma$$

Die Verzerrung $E(\widehat{\beta}_2) - \beta$ wird "**omitted variable bias**" genannt.
Keine Verzerrung wird auftreten, wenn

1. $\gamma = 0$ ist (d.h. beide Modelle sind identisch),
2. $\sum_{i=1}^{n} x_i z_i' = 0$ (bzw. asymptotisch $E(x_i z_i') = 0$)

sind. Im zweiten Fall sind die Beobachtungsvektoren x_i und z_i orthogonal, was für ökonomische Daten sehr selten vorliegt.

Beachte: Wenn x_i eine 1 für den Achsenabschnitt enthält, muss insbesondere $\sum_{i=1}^{n} z_i = 0$ (asymptotisch $E(z_i) = 0$) sein, damit der Verzerrung wegfällt. Alle weggelassenen relevanten Variablen müssen mithin Mittelwert 0 besitzen.

Fall 2: Aufnehmen irrelevanter Variablen

Modell 2 ist korrekt und Modell 1 wird geschätzt.
Dieser Fall hat keinen Einfluss auf Unverzerrtheit und Konsistenz. Der OLS-Schätzer kann aber nicht mehr BLUE (d.h. insbesondere effizient) sein. Die Standardfehler der Schätzer sind größer (und die t-Werte kleiner) als sie sein könnten.
Fazit:

1. Damit ist es nicht sinnvoll, zunächst möglichst viele Regressoren in das Modell aufzunehmen.
2. Schwer wiegende Probleme entstehen aber, wenn zu wenige Regressoren aufgenommen werden.
3. Somit ist die Bestimmung der "richtigen" Regressoren eine zentrale Aufgabe.

Auswahl der Regressoren

- Diskussion: Wahl der Regressoren durch substanzwissenschaftliche (hier: ökonomische) oder durch statistische (datenbezogene) Kriterien.
- Hilfestellung durch die ökonomische Theorie.
  Beispiel: Lohngleichung bestimmt sich aus der Humankapitaltheorie.
- Angeklagt zu werden, "Data Snooping" oder "Data Mining" zu betreiben, ist wenig ehrenhaft.

Data Snooping

- **Data Snooping** bezeichnet die Verwendung eines Datensatzes, um sowohl Modellspezifikation als auch Hypothesentesten zu betreiben.
- Obwohl statistische Programmpakete Techniken der automatischen Regressorenauswahl anbieten, sind diese für die empirische ökonomische Arbeit nicht zu empfehlen.
- Wenn 20 Regressoren auf Signifikanz getestet werden, besteht die Gefahr, dass auch dann Signifikanz festgestellt wird, wenn die zugehörige Variable ökonomisch keinen Sinn macht.
- Es besteht die Gefahr, aus der Stichprobe Zusammenhänge zu identifizieren, die außerhalb der Stichprobe keine Gültigkeit besitzen.
- In der praktischen Arbeit ist kaum zu vermeiden, dass partiell Data Snooping betrieben wird.

- So ist es denkbar, dass das Wissen um die vermeintlich richtige Modellspezifikation Erfahrungen entstammt, die in früheren Untersuchungen gemacht wurden.
- Gerade bei der empirischen Analyse von Finanzmarktrenditen sind in jüngster Zeit Verzerrungen aufgrund des Data Snoopings aufgetreten. Lo & MacKinley (1990) haben Asset Pricing Modelle untersucht und Sullivan et al. (1998) Kalendereffekte (wie z.B. den Januareffekt).
- Strategie: "From specific to general":
  Die Gefährlichkeit des data mining ist besonders hoch, wenn eine Suchstrategie "bottom up" oder "from specific to general" gewählt wird, da dann eine Vielzahl einzelner, nicht unabhängiger Tests durchzuführen ist, so dass das Signifikanzniveau nicht mehr kontrolliert werden kann.

- Alternative: LSE- (=London School of Economics) Strategie "from general to specific".
  Ausgangspunkt ist ein möglichst umfassendes Modell, in dem Restriktionen getestet werden, die zum Ausschluss von Variablen führen.
- Praktische Arbeit beginnt in der Mitte. D.h. es werden Tests
  1. auf die Korrektheit von unterstellten Modellrestriktionen (Fehlspezifikationstest für weggelassene Variablen, Test auf Heteroskedastie und Autokorrelation)
  2. auf zusätzliche Restriktionen, die das Modell weiter vereinfachen können,

  durchgeführt.
- Wichtig: Auch Ausweis insignifikanter Ergebnisse, da diese Information für den Nutzer der Schätzergebnisse wichtig ist.
- Dabei das Problem der Multikollinearität beachten, das bereits insignifikante Schätzergebnisse erzeugen kann.

Informationskriterien

- Wahl des besten Modells mittels des Bestimmtheitsmaßes:
  Problem: Bestimmtheitsmaß wächst mit der Zahl der Regressoren.
  Deshalb: Wahl des besten Modells mittels des korrigierten Bestimmtheitsmaßes

$$\overline{R}^2 = 1 - \frac{1/(n-K)\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{1/(n-1)\sum_{i=1}^{n}(y_i - \overline{y})^2},$$

  das einen Trade-off zwischen "Goodness of Fit" and "Law of Parsimony" anzeigt.
- Diesen Trade-off messen auch die sog. **Informationskriterien**, die die Residuenquadratsumme der OLS-Schätzung um einen sog. Strafterm erweitern, der mit der Anzahl der Regressoren K steigt.

1. **Akaikes Informationskriterium (AIC)**:

$$AIC = \log \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 + \frac{2K}{n}$$

2. **Schwarzsche oder Bayesianisches Informationskriterium (BIC)**:

$$BIC = \log \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 + \frac{K}{n} \log n$$

Beachte: Der Strafterm von BIC wächst schneller mit K als der von AIC, so dass mit BIC "sparsameren" Modellen der Vorzug gegeben wird.

- Modellwahl mittels F-Test auf den Zuwachs des Bestimmtheitsmaßes: Seien R_0 bzw. R_1 die Bestimmtheitsmaße eines Modells ohne bzw. mit J zusätzlichen Variablen z_i, dann ist

$$f = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(n-K)}$$

unter der Nullhypothese, dass die J Variablen in z_i gemeinsam keinen Einfluss haben, F-verteilt mit J Zähler- und $n-K$ Nennerfreiheitsgraden.
Beachte: Verwendet man die korrigierten Bestimmtheitsmaße, so geht die Verteilungseigenschaft bei Gültigkeit der Nullhypothese verloren. Es lässt sich aber zeigen, dass

$$\overline{R}_1 > \overline{R}_0 \iff f > 1$$

- Alternative zum F-Test: WALD-Test, für den nur eine Regression unter Einschluss von z_i nötig ist.
  Sei γ der zu z_i gehörende Vektor der Regressionskoeffizienten, $\hat{\gamma}$ der OLS-Schätzer und $\hat{V}(\hat{\gamma})$ die geschätzte Varianz-Kovarianzmatrix, dann ist unter $H_0: \ \gamma = 0$

  $$\xi = \hat{\gamma}'\hat{V}(\hat{\gamma})\hat{\gamma} \sim \chi^2(J)$$

  für große Stichprobenumfänge n.
  Beachte: Unter den Gauß-Markov-Annahmen ist $\xi = Jf$.
- Warnung: **Zwei** einzelne Tests sind im Allg. nicht äquivalent zu **einem** gemeinsamen Test.

Vergleich nicht-genesteter Modelle

- Betrachte die beiden Modelle

$$\begin{array}{lrcl} \text{Modell A} & y_i & = & x_i'\beta + \varepsilon_i \\ \text{Modell B} & y_i & = & z_i'\gamma + v_i, \end{array}$$

  wobei $E(\varepsilon_i|x_i) = E(v_i|z_i) = 0$ ist.
  Diese Modelle sind nicht-genestet, wenn z_i eine Variable enthält, die nicht in x_i ist und umgekehrt.
  Da jeweils dieselbe Variable y_i erklärt wird, können zum Modellvergleich das korrigierte Bestimmtheitsmaß oder die Informationskriterien verwendet werden.

Alternative: Encompassing-Test

- Idee: Wenn Modell A das wahre Modell ist, muss es in der Lage sein, Ergebnisse von Modell B zu "erklären". Wenn Modell A dazu nicht in der Lage ist, muss es abgelehnt werden.
  Beachte: Es können beide Modelle abgelehnt werden, ohne dass dies auf den Fehler 1. Art zurückzuführen ist.
  Wird Modell 1 nicht abgelehnt, wird es gegen weitere Modelle getestet und solange als akzeptabel bewertet, wie es nicht abgelehnt werden kann.
  Beachte: Das Encompassing-Prinzip gilt auch für genestete Modelle, da das allgemeinere Modell immer in der Lage ist, Ergebnisse des spezielleren Modells zu erklären.

Non-nested F-Test

- Es soll die Nullhypothese getestet werden, dass Modell B gilt und somit Modell A "encompassed" (umschließt).
- Ausgangspunkt: $x_i' = (x_{1i}', x_{2i}')$, wobei x_{1i} Variablen umfasst, die in beiden Modellen enthalten sind. x_{2i} fasst diejenigen Variablen zusammen, die nicht Bestandteil von Modell B sind.
  Künstlich genestetes Modell
  $$y_i = z_i'\gamma + x_{2i}'\delta_A + v_i.$$
  Wenn Modell B das Modell A nicht umschließt, so dass Modell B zu verwerfen ist, müssen diejenigen Variablen aus Modell A, die zusätzlich in Modell B aufgenommen wurden, einen signifikanten Erklärungsbeitrag leisten.
- Zu testen ist somit die Nullhypothese: $H_0 : \delta_A = 0$.
  Unter den üblichen Annahmen des linearen Normalverteilungsmodells geschieht dies mit einem F-Test.

- *Beachte*: Es muss auch umgekehrt im genesteten Modell

$$y_i = x_i'\beta + z_{2i}'\delta_B + \varepsilon_i$$

getestet werden, dass $\delta_B = 0$ ist und Modell A Modell B umschließt, wobei z_{2i} die Variablen umfasst, die nicht Bestandteil von Modell A sind.

- 4 mögliche Ergebnisse:
    1. Modell A wird verworfen.
    2. Modell B wird verworfen.
    3. Modell A und Modell B werden verworfen.
    4. Weder Modell A noch Modell B werden verworfen.

J-Test

- Ausgangspunkt: Künstlich genestetes Modell

$$y_i = (1 - \delta) x_i' \beta + \delta z_i' \gamma + u_i,$$

das sich für $\delta = 0$ ($\delta = 1$) auf Modell A (Modell B) reduziert.

- Problem: Parameter δ, β und γ sind nicht gemeinsam identifiziert und können nicht geschätzt werden.

- Lösungsvorschlag von Davidson & MacKinnon (1981):
  Berechne den OLS-Schätzwert $\hat{\gamma}$ für γ aus Modell B und setze diesen für γ im genesteten Modell ein:

$$\begin{aligned} y_i &= (1-\delta)x_i'\beta + \delta z_i'\hat{\gamma} + u_i \\ &= x_i'\beta^* + \delta \hat{y}_{iB} + u_i, \end{aligned}$$

  mit $\hat{y}_{iB}$ als den ausgeglichenen Werten der OLS-Regression des Modells B und $\beta^* = (1-\delta)\beta$.
  Mittels eines t-Testes kann $H_0 : \delta = 0$ getestet werden. Wird diese Nullhypothese verworfen, wird auch Modell A verworfen, da das Modell B einen signifikanten Erklärungsbeitrag liefert.
  Vorteil: Es muss nur auf einen Koeffizienten getestet werden.
- Wenn nur ein Regressor als zusätzlicher Regressor im genesteten Modell aufgenommen wird, stimmen nicht-genesteter F-Test und J-Test überein.

Nicht-Nestung wegen Variablentransformation

- Spezialfall: Lineare und loglineare Modelle sind z.B. nicht genestet.
- Problem: Da sich die erklärte Variable unterscheidet (einmal y_i und einmal $\log y_i$) können Bestimmtheitsmaße und Informationskriterien nicht zum Modellvergleich herangezogen werden.
- **Lösung 1**: Suche eines größeren Modells, in dem lineares und loglineares Modell als Spezialfall enthalten sind.
  Box-Cox-Transformation ist
  $$\frac{y_i^\delta - 1}{\delta}, \quad \delta > 0$$
  und liefert für $\delta = 1$ ein lineares bzw. für $\delta \to 0$ ein loglineares Modell.

PE-Test

- **Lösung 2**: **PE-Test** von Davidson et al. (1993), der analog zum J-Test funktioniert. Schätze zunächst das lineare und das loglineare Modell separat mit OLS. Berechne die ausgeglichenen Werte $\hat{y}_i$ und $\widehat{\log y_i}$.
  Ein t-Test der Nullhypothese des linearen gegen die Alternative des loglinearen Modells testet $H_0 : \delta_{LIN} = 0$ in

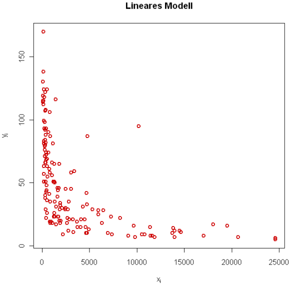
$$y_i = x_i'\beta + \delta_{LIN}(\log \hat{y}_i - \widehat{\log y_i}) + u_i.$$

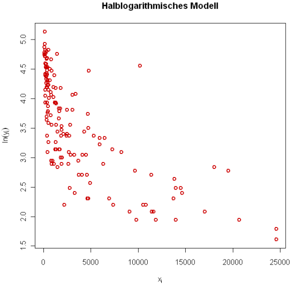
  Analog kann die Nullhypothese eines loglinearen Modells gegen die lineare Alternative getestet werden.

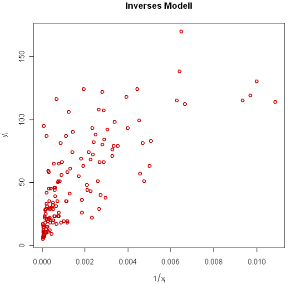
Alternative Funktionsformen

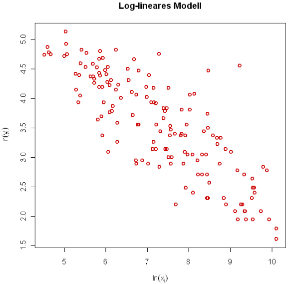
- Viele in der Wirklichkeit auftretende Zusammenhänge liegen in nicht-linearer Form vor. Würde man dennoch ein lineares Modell schätzen, läge eine Fehlspezifikation der funktionalen Form vor.
  Häufig können Zusammenhänge aber durch geeignete Tranformationen in linearer Form dargestellt und dann mittels OLS geschätzt werden.
- Die prominentesten der linear formalisierten Modelle sind:
  - Halblogarithmisches Modell
  - Inverses Modell
  - Log-lineares Modell
- ***Beispiel***: Zusammenhang zwischen Bruttosozialprodukt (x_i) und Kindersterblichkeitsrate (y_i) in den UNO-Mitgliedsstaaten 1998.

Datensituation









Beschreibung der Plots

1. Das **Lineares Modell** $\boxed{y_i = \beta_1 + \beta_2 x_i + \varepsilon}$ scheint nicht geeignet, den wahren Zusammenhang zwischen x_i und y_i wiederzugeben. Deswegen Prüfung, ob andere Funktionen von x_i und y_i eine lineare Beziehung aufweisen.
2. Nach Logarithmieren der Werte der abhängigen Variablen erhält man das **halblogarithmische Modell**, welches ebenso linear in den Parametern β_1 und β_2 ist. Das Modell $\boxed{\log(y_i) = \beta_1 + \beta_2 x_i + \varepsilon}$ scheint dem postulierten Zusammenhang zwischen x_i und $\log(y_i)$ zu widersprechen.
3. Durch Abtragen der reziproken Werte der unabhängigen Variablen auf der Ordinate gelangt man zu dem **inversen Modell** $\boxed{y_i = \beta_1 + \beta_2 \left(\frac{1}{x_i}\right) + \varepsilon}$ Es legt einen nicht linearen Zusammnehang zwischen y_i und $(\frac{1}{x_i})$ nahe.
4. Das **Log-linearer Modell** $\boxed{\log(y_i) = \beta_1 + \beta_2 \log(x_i) + \varepsilon}$ liefert eine plausible Beschreibung des formalisierten linearen Zusammenhangs zwischen x_i und y_i.

Vergleich der Modelle

OLS-Schätzergebnisse für das lineare Modell

Abhängige Variable: $infmort$		
Variable	Schätzwert	Standardfehler
Konstante	61.6606	3.0798
gdp	-0.0038	0.0005

$s = 30.6153,\ R^2 = 0.2861,\ F = 59.31$

OLS-Schätzergebnisse für das Log-lineare Modell

Abhängige Variable: $log(infmort)$		
Variable	Schätzwert	Standardfehler
Konstante	7.1448	0.2122
$log(gdp)$	-0.4933	0.0286

$s = 0.4937,\ R^2 = 0.6687,\ F = 298.7$

Test auf die funktionale Form

- Einfache Idee: Um
$$H_0 : E(y_i|x_i) = x_i'\beta$$
zu testen, müssen nur weitere nicht-lineare Regressoren aufgenommen und entweder einzeln (t-Test) oder gemeinsam (F-Test) auf Signifikanz geprüft werden.
Problem: Die spezifische Form der Alternativhypothese muss bekannt sein.
- **Lösung**: **RESET**- (**R**egression **E**quation **S**pecification **E**rror **T**est) Verfahren von Ramsey (1969):
Berechne zunächst die ausgeglichenen Werte $\hat{y}_i = x_i'\beta$ einer OLS-Regression. Dann werden in einem zweiten Schritt Potenzen von $\hat{y}_i$ zusätzlich zum linearen Teil als Regressoren in das Modell eingesetzt:
$$y_i = x_i'\beta + \alpha_2\hat{y}_i^2 + \alpha_3\hat{y}_i^3 + \ldots + \alpha_Q\hat{y}_i^Q + v_i$$

- Die Nullhypothese der Linearität ist zu verwerfen, wenn im Normalverteilungsfall die $Q-1$ Restriktionen in

$$H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_Q = 0$$

mittels eines F-Testes verworfen werden können.
Beachte: Der Test wird zumeist für $Q = 2$ durchgeführt. Eine Ablehnung der Nullhypothese kann aber auch auf vernachlässigte relevante Variablen zurückzuführen sein.

Fallstudie: Individuelle Löhne

- Daten entstammen einer Erhebung durch das Nationale Bildungspanel (NEPS) im Jahre 2011,[1] in der 8840 Personen in Deutschland zufällig ausgewählt wurden.[2] Verfügbare Variablen sind:
  - Bruttomonatslöhne in Euro (EUR) ($wage$)
  - Geschlecht mit 1=männlich ($male$)
  - Ausbildungsniveau ($educ$) mit
    1=kein Schulabschluss
    2=Hauptschule
    3=Realschule
    4=Abitur
    5=Hochschule
  - Berufserfahrung in Jahren ($exper$)

[1] NEPS Scientific Use File of Starting Cohort 6 – Adults (http://dx.doi.org/10.5157/NEPS:SC6:1.0.0).

[2] Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special issue]. Zeitschrift für Erziehungswissenschaft,14. Wiesbaden: VS Verlag für Sozialwissenschaften.

Deskriptive Statistiken

Zusammenfassende Statistik, 8840 Beobachtungen

	Männer		Frauen	
Variable	Mittelwert	Standardfehler	Mittelwert	Standardfehler
wage	3728	3209.102	2024	1708.368
educ	3.557	1.041	3.549	1.160
exper	24.077	12.729	22.247	11.707

Schätzung eines linearen Modells

OLS-Schätzergebnisse für Spezifikation 1

Abhängige Variable: *wage*			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-1079.748	122.612	-8.806
male	1630.582	57.567	28.325
educ	666.932	26.441	25.223
exper	33.818	2.381	14.204

$s = 2477, R^2 = 0.178, \overline{R}^2 = 0.178, F = 540.1$

Interpretation der Schätzergebnisse

1. C.p.-Einflüsse von Geschlecht, Ausbildung und Erfahrung auf den erwarteten Lohn sind hochsignifikant.
2. Die Steigerung der Erfahrung um ein Jahr hat c.p. mit ca. 34 EUR den erwarteten positiven Einfluss auf den erwarteten Lohn.
3. Zwei Personen mit benachbartem Ausbildungsniveau besitzen bei gleichem Geschlecht und gleicher Erfahrung eine Differenz der erwarteten Löhne von ca. 667 EUR.
4. $R^2 = 0.178$ besagt, dass ca. 17.8% der Variation des Individuallohns linear durch Unterschiede im Geschlecht, Erfahrung und Ausbildungsniveau erklärt werden kann.

Berücksichtigung eines quadratischen Terms

- Berücksichtigung von $(exper^2)$.
  Grund ist die Hypothese: Mit zunehmender Berufserfahrung c.p. wird der Effekt eines zusätzlichen Jahres auf den Lohn zunehmend geringer.
  D.h. $exper^2$ müsste ein negatives Vorzeichen haben.

OLS-Schätzergebnisse für Spezifikation 2

Abhängige Variable: *wage*			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-1575.602	136.059	-11.58
male	1644.544	57.335	28.68
educ	669.920	26.326	25.45
exper	82.674	6.379	12.96
exper^2	-0.949	0.115	-8.25

$s = 2466,\ R^2 = 0.186,\ \overline{R}^2 = 0.185,\ F = 425.7$

Interpretation der Schätzergebnisse

1. Das Vorzeichen ist wie erwartet negativ und der Schätzwert des Koeffizienten von $\textit{exper}^2$ hochsignifikant.
2. Sogar das korrigierte Bestimmtheitsmaß hat sich von 0.178 auf 0.185 erhöht.
3. Beachte: *exper* und $\textit{exper}^2$ können nicht isoliert verändert werden (d.h. c.p.-Interpretation gilt nicht).
   Der c.p.-Effekt der Steigerung der Erfahrung um ein Jahr auf den erwarteten Lohn ist (partielle Ableitung nach *exper*):

   $$82.674 - 2 \cdot 0.949 \cdot \textit{exper}$$

   Dieser Effekt hängt von dem Niveau von *exper* ab, und reduziert sich von 82.674 EUR (keine Erfahrung) auf 25.734 EUR (bei 30-jähriger Berufserfahrung).

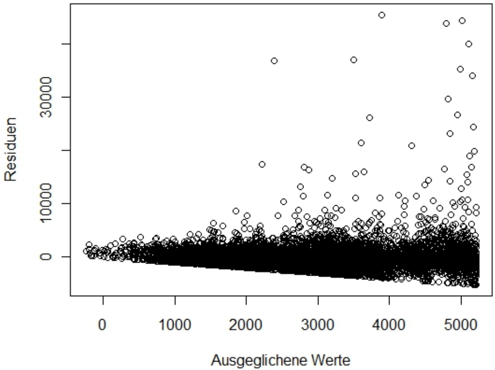
4. *Beachte*: Prognostiziert man die erwarteteten Löhne zweier Personen mit 31 bzw. 30 Jahren Berufserfahrung bei gleichem Geschlecht und gleichem Ausbildungsniveau, dann ergibt sich eine Differenz der erwarteten Löhne von

$$82.674 \cdot (31 - 30) - 0.949 \cdot (31^2 - 30^2) = 24.785 \neq 25.734$$

Der Unterschied in den Werten erklärt sich durch die Punktbetrachtung (Steigung der Tangente) der partiellen Ableitung im Gegensatz zur Betrachtung der Sehne, wenn tatsächlich die Erfahrung um ein Jahr erhöht wird.

Überprüfung der Annahmen des klassischen linearen Modells

1. Autokorrelation ist unwahrscheinlich, da die Personen zufällig (ohne irgendeine systematische Reihenfolge) ausgewählt wurden.
2. Heteroskedastie erkennt man häufig am Scatter- (Streuungs-) Plot der angepassten Werte gegen die Residuen:



Bei Homoskedastie dürfte die Variation der OLS-Residuen nicht von den gefitteten Werten abhängen.
Hier: Die Variation wächst mit dem Niveau der gefitteten Werte. D.h. es ist Heteroskedastie zu vermuten, was die standardmäßig berechneten Standardfehler und die t-Werte unbrauchbar macht.

- Behandlung des Heteroskedastieproblems durch logarithmische Transformation der Löhne:
  Ausgangspunkt: Transformiertes Modell

$$w_i = g(x_i) + \varepsilon_i$$

mit $g(x_i) = x_i'\beta$.
Alternatives multiplikatives Modell:

$$w_i = g(x_i) \exp(\eta_i)$$

Beachte: η_i kann homoskedastisch sein, wenn ε_i heteroskedastisch ist.
Logarithmierung des multiplikativen Modells führt zum loglinearen Modell:

$$\log w_i = \log g(x_i) + \eta_i$$

Beachte: Dummy-Variablen werden nicht logarithmiert.

Schätzung eines loglinearen Modells

OLS-Schätzergebnisse für Spezifikation 3

Abhängige Variable: log(*wage*)			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	5.982	0.064	94.132
male	0.636	0.020	31.850
log(*educ*)	0.731	0.031	23.902
log(*exper*)	0.073	0.036	2.038
$log^2(exper)$	0.029	0.008	3.825

$s = 0.8358,\ R^2 = 0.1999,\ \overline{R}^2 = 0.1994,\ F = 441.1$

Interpretation der Schätzergebnisse

1. *Beachte*: Regressor ist $\log^2(exper)$ und nicht $\log exper^2 = 2\log exper$ (perfekte Multikollinearität).
2. *Beachte*: Das Bestimmtheitsmaß muss trotz ähnlicher Werte anders interpretiert werden und kann nicht mit dem Wert aus dem linearen Modell verglichen werden.
3. Der Schätzwert für die Variable *male* misst jetzt die relative (prozentuale) Lohndifferenz von Männern und Frauen bei gleicher Erfahrung und gleichem Ausbildungsniveau.
   Diese relative Differenz beträgt 0.636 oder 63.6$\%$.
   Sei w_i der Lohn einer Frau und w_j der Lohn eines Mannes, dann ist

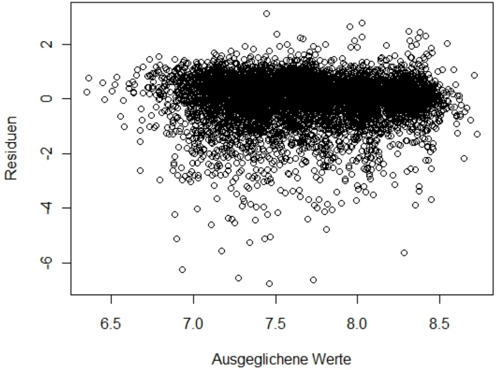
$$w_j = exp(log w_i + 0.636) = w_i exp(0.636) = w_i \cdot 1.889$$

   Damit ist die prozentuale Lohndifferenz exakt 88.8% ≈ 63.6%.
   Eine ungefähre Übereinstimmung gilt wegen

$$exp(a) \approx 1 + a$$

   nur für kleine a.

Ein Plot der gefitteten Werte im logarithmierten Modell ($\widehat{\log y_i}$) gegen die OLS-Residuen zeigt ein deutlich weniger ausgeprägtes Heteroskedastieverhalten:



Folgerung: Die Standardfehler und t-Werte im loglinearen Modell erlauben die übliche Interpretation.

Ergebnisse für die Variable Berufserfahrung $exper$

1. Unter Verzicht auf den quadratischen Term repräsentiert der Wert von 0.073 eine Elastizität, d.h. wenn die Berufserfahrung c.p. um ein Prozent steigt, erhöht sich der erwartete Lohn um 0.073$\%$.
2. Wenn der quadratische Term berücksichtigt wird, ist die Elastizität

$$\frac{\partial E(log\ \mathsf{wage}_i)}{\partial log\ \mathsf{exper}_i} = 0.073 + 2 \cdot 0.029 \cdot log(\mathsf{exper})$$

Wegen $0.029 > 0$ wächst die Elastizität mit der Erfahrung.

- *Beachte*: Die Schätzwerte der Koeffizienten für log(*exper)* und log^2(*exper*) sind signifikant. log(*exper)* auf dem 5% Niveau und log^2(*exper*) hochsignifikant. log(*exper)* ist insignifikant auf dem 1% Niveau. Damit sind nicht automatisch beide Schätzwerte gemeinsam insignifikant, wie ein F-Test zeigt, wenn $R^2 = 0.156$ in einem Modell ohne die Erfahrungsvariable ist:

$$F = \frac{(0.1999 - 0.156)/2}{(1 - 0.1999)/(8840 - 5)} = 242.38$$

mit einem p-Wert von 0.000.

- Verzichtet man auf die Variable $\log^2$ exper, so verschlechtert sich die Anpassungsgüte kaum:

OLS-Schätzergebnisse für Spezifikation 4

Abhängige Variable: log(*wage*)			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	5.893	0.059	99.45
male	0.640	0.02	32.08
log(*educ*)	0.719	0.0305	23.61
log(*exper*)	0.201	0.013	15.38

$s = 0.8366,\ R^2 = 0.1982,\ \overline{R}^2 = 0.1979,\ F = 582.1$

Ergebnisse für die Variable Berufserfahrung $educ$

1. Die c.p. Differenz des erwarteten logarithmierten Lohnes zwischen den beiden Ausbildungsniveaus $educ = 1$ und $educ = 2$ ist:

$$0.719 \cdot (log(\textbf{\textit{educ=2}}) - log(\textbf{\textit{educ=1}}))$$

   Setzt man für $educ = 1$ ein und lässt $educ = 2$ von 2 bis 5 laufen, so erhält man die folgenden c.p.-Differenzen:

$$0.498,\ 0.790,\ 0.997,\ 1.157.$$

2. Alternative Vorgehensweise: Einführung von Dummy-Variablen (0-1) für jede Ausbildungsstufe.
   Beachte: Berücksichtigt man 5 Dummy-Variablen (d.h. für jede Stufe eine), so tappt man in die "Dummy variable trap" der perfekten Multikollinearität.
   Da: Variable für Achsenabschnitt entspricht der Summe der Dummy-Variablen.
   Lösung: Nur 4 Dummy-Variablen für die Ausbildungsniveaus 2 bis 5. Dann ist die Ausbildungsstufe 1 die Referenzkategorie, auf die sämtliche Interpretationen bezogen sind.

Schätzergebnisse des Dummy-Variablenmodells

OLS-Schätzergebnisse für Spezifikation 5

Abhängige Variable: log(*wage*)			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	6.452	0.254	25.415
male	0.622	0.020	30.996
educ=2	0.048	0.252	0.189
educ=3	0.226	0.252	0.898
educ=4	0.356	0.252	1.412
educ=5	0.718	0.251	2.852
log(*exper*)	0.191	0.013	14.459

$s = 0.833,\ R^2 = 0.2054,\ \overline{R}^2 = 0.2048,\ F = 304.3$

Interpretation der Schätzergebnisse

1. Die c.p. Differenz der erwarteten logarithmieren Löhne sind alle außer *educ5* insignifikant und betragen
$$0.048,\ 0.226,\ 0.356,\ 0.718$$
D.h. 0.718 gibt an, dass die erwartete logarithmierte Lohndifferenz zwischen dem niedrigsten und höchsten Ausbildungsniveau c.p. 0.718 beträgt.
2. Da das Modell mit logarithmierten Ausbildungsniveau in dem Modell mit Dummy-Variablen genestet ist (Nachweis ist mühsam!), kann ein Vergleich der beiden Modelle mit dem F-Test durchgeführt werden:
$$F = \frac{(0.2054 - 0.1982)/3}{(1 - 0.2054)/(8840 - 7)} = 26.679$$
Dieser Wert ist größer als das 99%-Quantil der F-Verteilung mit 3 Zähler- und 8833 Nennerfreiheitsgraden, so dass das Modell mit Dummy-Variablen signifikant besser ist.

- Einfluss des Geschlechts:
  Bisher: Der Einfluss des Geschlechts ist unabhängig vom Ausbildungsniveau und der Erfahrung für alle Personen gleich.
  Problem: Männer können u.U. besser gefördert werden als Frauen, so dass sie leichter ein höheres Ausbildungsniveau erreichen.
  Lösung: Interaktionsterme zwischen dem Dummy für Geschlecht und den 4 Dummies für die Ausbildungsniveaus.

Schätzergebnisse bei Berücksichtigung von Interaktionstermen für Geschlecht und Ausbildung

Abhängige Variable: log(*wage*)			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	6.160	0.420	14.682
male	1.079	0.526	2.051
educ=2	0.223	0.417	0.535
educ=3	0.553	0.416	1.328
educ=4	0.730	0.417	1.751
educ=5	1.065	0.417	2.558
log(*exper*)	0.182	0.0193	9.439
educ=2 × *male*	-0.277	0.523	-0.529
educ=3 × *male*	-0.521	0.522	-0.998
educ=4 × *male*	-0.622	0.523	-1.190
educ=5 × *male*	-0.555	0.522	-1.062
log(*exper*)× *male*	0.016	0.026	0.597

$s = 0.8312,\ R^2 = 0.2093,\ \overline{R}^2 = 0.2081,\ F = 169.8$

Interpretation der Schätzergebnisse

1. *Beachte:* Statt eine Regression mit einer Dummy-Variablen für Geschlecht, hätten auch zwei separate Regressionen für "männlich" und "weiblich" gerechnet werden können.
   Allerdings: Separate Regressionen über die beiden Teilstichproben lassen für diese unterschiedliche Varianz zu (Heteroskedastie).
   Fazit: Schätzwerte für die Koeffizienten sind zwar identisch, aber die Schätzfehler sind unterschiedlich.
2. Es scheint keine signifikanten Differenzen des Effekts der Variablen $\log(exper) \times male$ für Männer und Frauen zu geben.
3. Die Interaktionen von Ausbildung und Geschlecht geben Hinweise, dass der Ausbildungseffekt auf den Lohn für Männer geringer ist.
4. *Beachte*: Der Koeffizient der Variablen Geschlent kann nicht mehr c.p. interpretiert werden, da andere Variablen eine Funktion von *male* sind.
5. Das geschätzte Lohndifferential zwischen Männern und Frauen bei 20-jähriger Berufserfahrung und Ausbildungsniveau 2 ist

$$1.079 + 0.016 \cdot log(20) - 0.277 = 0.8499$$

- F-Test auf Signifikanz der Interaktionen mittels des Vergleichs von Bestimmtheitsmaßen:

$$F = \frac{(0.2093 - 0.2054)/5}{(1 - 0.2093)/(8840 - 12)} = 8.708$$

  Beachte: Das 99%-Quantil der F-Verteilung mit 5 Zähler- und 8828 Nennerfreiheitsgraden ist 3.02.

- RESET-Test auf funktionale Form ($Q = 2$): $t = 3.989$

Schätzergebnisse bei Berücksichtigung von Interaktionstermen zwischen Ausbildungsniveau und Berufserfahrung

OLS-Schätzergebnisse für Spezifikation 7

Abhängige Variable: log(*wage*)

Variable	Schätzwert	Standardfehler	t-Wert
Konstante	6.535	0.744	8.788
male	0.623	0.020	31.043
educ=2	0.067	0.751	0.089
educ=3	0.328	0.748	0.439
educ=4	0.007	0.747	0.010
educ=5	0.690	0.747	0.924
log(*exper*)	0.160	0.262	0.610
log(*exper*)×*educ=2*	-0.001	0.264	-0.004
log(*exper*)×*educ=3*	-0.030	0.263	-0.112
log(*exper*)×*educ=4*	0.132	0.263	0.502
log(*exper*)×*educ=5*	0.012	0.263	0.046

$s = 0.8319,\ R^2 = 0.208,\ \overline{R}^2 = 0.2069,\ F = 185.4$

Interpretation der Schätzergebnisse

1. Der Koeffizient von log(*exper*) $\times$ *educ*=2 gibt c.p. die Lohndifferenz einer Änderung der Berufserfahrung für Personen der zweiten Ausbildungsstufe gegenüber Personen der ersten Ausbildungsstufe (Referenzkategorie) an.
2. *Beachte*: Sämtliche Interaktionensvariablen sind insignifikant.
3. Mögliche Ursache: Multikollinearität, da das Bestimmtheitsmaß hoch ist, obwohl sämtliche t-Werte (bis auf Achsenabschnitt und die Variable *male*) betragsmäßig klein sind.

- Test auf globalen Fit ergibt $F = 185.4$.

Gliederung Kapitel 4

Konsequenzen für den OLS-Schätzer

- Modell

$$y = X\beta + \varepsilon \text{ mit } V(\varepsilon|X) = \sigma^2\Psi$$

- $\widehat{\beta}$ bleibt unverzerrt und konsistent, da z.B.

$$E(\widehat{\beta}|X) = (X'X)^{-1}X'(X\beta + E(\varepsilon|X)) = \beta$$

- Varianz des OLS-Schätzers $\widehat{\beta}$:

$$V(\widehat{\beta}|X) = V(\beta + (X'X)^{-1}X'\varepsilon|X) = \sigma^2(X'X)^{-1}X'\Psi X(X'X)^{-1}$$

- **Lösungen**:
  1. Neuer Schätzer
  2. Anpassung der Standardfehler
  3. Neues Modell

Der GLS - Schätzer

- Ausgangspunkt: Zerlegung von Ψ:

$$\Psi^{-1} = P'P, \;\; \Psi = (P'P)^{-1} = P^{-1}(P')^{-1}, \;\; P\Psi P' = I$$

- GLS-Schätzer (**G**eneralized **L**east **S**quares):

$$\widehat{\beta}_{GLS} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$$

- Alternative: Betrachte das transformierte Modell

$$Py = PX\beta + P\varepsilon$$

- Dann sind
  1. $E(P\varepsilon|X) = PE(\varepsilon|X) = 0,$
  2. $V(P\varepsilon|X) = PV(\varepsilon|X)P' = \sigma^2 P\Psi P' = \sigma^2 I$

  D.h. im transformierten Modell sind die Annahmen (**A1**) bis (**A4**) erfüllt.
- OLS-Schätzer im transformierten Modell:

$$\widehat{\beta} = [(PX)'(PX)]^{-1}(PX)'Py = [X'(P'P)X]^{-1}X'P'Py$$

- Fazit: GLS-Schätzer und OLS-Schätzer im transformierten Modell sind identisch.

- Varianz-Kovarianzmatrix des GLS-Schätzers:

$$V(\widehat{\beta}_{GLS}|X) = \sigma^2[(PX)'(PX)]^{-1} = \sigma^2(X'\Psi^{-1}X)^{-1}$$

- Ineffizienz des OLS-Schätzers: GLS-Schätzer ist als OLS-Schätzer im transformierten Modell BLUE.
  Man kann zeigen, dass

$$V(\widehat{\beta}_{GLS}|X) - V(\widehat{\beta}_{OLS}|X) = \sigma^2[(X'\Psi^{-1}X)^{-1} - (X'X)^{-1}X'\Psi X(X'X)^{-1}]$$

  negativ semidefinit ist.
- Unverzerrte Schätzung von σ^2:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-K}(Py - PX\widehat{\beta}_{GLS})'(Py - PX\widehat{\beta}_{GLS}) \\
&= \frac{1}{n-K}(y - X\widehat{\beta}_{GLS})'P'P(y - X\widehat{\beta}_{GLS}) \\
&= \frac{1}{n-K}(y - X\widehat{\beta}_{GLS})'\Psi^{-1}(y - X\widehat{\beta}_{GLS})
\end{aligned}$$

FGLS- bzw. EGLS-Schätzer bei unbekannter Varianz-Kovarianzmatrix

- Problem: Ψ ist unbekannt und muss geschätzt werden.
- **Lösung**: Zweistufiges Verfahren:
  1. OLS-Schätzung und Schätzung $\widehat{\Psi}$ von Ψ aus den OLS-Residuen $\widehat{\varepsilon}$,
  2. Berechnung des FGLS- (**F**easible **G**eneralized **L**east **S**quares) oder alternativ genannt EGLS-Schätzer (**E**stimated **G**eneralized **L**east **S**quares):

$$\widehat{\beta}_{FGLS} = (X'\widehat{\Psi}^{-1}X)^{-1}X'\widehat{\Psi}^{-1}y$$

- *Beachte*: Man unterscheidet nicht-iterierte und voll-iterierte Varianten des FGLS-Schätzers.

Heteroskedastie

- Die Varianz-Kovarianzmatrix des Vektors der Störgrößen ist nicht mehr skalar, aber noch diagonal, d.h.

$$V(\varepsilon_i) = \sigma_i^2 \text{ für i=1,2,...,n}$$

  Beachte: Es wird weiterhin die Unkorreliertheit der Störgrößen unterstellt.
- Anwendungsbeispiel: Heteroskedastie tritt häufig in Querschnittsstudien auf.
- Wenn y_i Ausgaben für Nahrungsmittel sind und x_i umfasst einen Konstante und das verfügbare Einkommen DPI_i, dann steht zu erwarten, dass die Engel-Kurve für Nahrungsmittel monoton steigend verläuft, wobei die Steigung mit dem Einkommen abnimmt. Zusätzlich ist zu erwarten, dass die Variation der Ausgaben für Nahrungsmittel für höhere Einkommen deutlich größer ist als für niedrige Einkommen.
- D.h. ein geeignetes Modell für die Heteroskedastie könnte sein:

$$V(\varepsilon_i|DPI_i) = \sigma_i^2 = \sigma^2 \exp(\alpha_2 DPI_i) = \exp(\alpha_1 + \alpha_2 DPI)$$

  mit $\alpha_1 = \log \sigma^2$.

- Varianz-Kovarianzmatrix von ε bei allgemeiner Form der Heteroskedastie

$$V(\varepsilon_i|X) = V(\varepsilon_i|x_i) = \sigma^2 h_i^2$$

lautet

$$V(\varepsilon|X) = \sigma^2 \text{diag}\{h_i^2\} = \sigma^2 \Psi \qquad \textbf{(A9)}$$

mit

$$\Psi = \text{diag}\{h_i^2\} = \begin{pmatrix} h_1^2 & 0 & \ldots & 0 \\ 0 & h_2^2 & \ldots & \ldots \\ .. & & \ldots & \ldots \\ 0 & \ldots & \ldots & h_n^2 \end{pmatrix}$$

- *Beachte*: (**A9**) ersetzt die Annahmen (**A3**) und (**A4**). Wegen der funktionalen Abhängigkeit der Varianz von ε von X, ist die Annahme (**A2**) der stochastischen Unabhängigkeit von ε und X nicht länger haltbar. Stattdessen wird als Ersatz für (**A1**) und (**A2**)

$$E(\varepsilon|X) = 0 \qquad ; \quad \textbf{(A10)}$$

unterstellt, was stärker als die Unkorreliertheit von ε und X ist.

- Der GLS-Schätzer für β im Modell

$$y = X\beta + \varepsilon \text{ mit } V(\varepsilon|X) = \sigma^2\Psi = \sigma^2 \text{diag}\{h_i^2\}$$

ergibt sich aus der offensichtlichen Zerlegung von

$$\Psi^{-1} = \text{diag}\{1/h_i^2\} = \text{diag}\{1/h_i\}\text{diag}\{1/h_i\},$$

so dass

$$P = P' = \text{diag}\{1/h_i\}$$

gesetzt werden kann.

- Die Transformation Py und PX bewirkt eine Gewichtung

$$y_i/h_i, \;\; x_i/h_i \text{ für } i = 1, 2, \ldots, n$$

D.h. Beobachtungen mit großer Varianz werden stärker “herunter”gewichtet als solche mit kleiner Varianz.

- Das transformierte Modell lautet

$$\frac{y_i}{h_i} = \left(\frac{x_i}{h_i}\right)' \beta + \frac{\varepsilon_i}{h_i}$$

- Der GLS-Schätzer ist dann

$$\widehat{\beta}_{GLS} = \left(\sum_{i=1}^{n} h_i^{-2} x_i x_i'\right)^{-1} \sum_{i=1}^{n} h_i^{-2} x_i y_i$$

- *Beachte*: Der GLS-Schätzer wird auch als "gewichteter" OLS-Schätzer bezeichnet, da jede Beobachtung mit einem Faktor gewichtet wird, der proportional zur Varianz der Störgröße ist.
- *Beachte*: Im transformierten Modell wird auch der Achsenabschnitt gewichtet, so dass das transformierte Modell über keinen Achsenabschnitt verfügt.
- *Beachte*: Sinnvolle Interpretation der geschätzten Parameterwerte ist nur im Kontext des Originalmodells und nicht des transformierten Modells möglich.

Eigenschaften des GLS-Schätzers und Hypothesentests

- Die Varianz-Kovarianzmatrix des GLS-Schätzers ergibt sich durch

$$V(\widehat{\beta}_{GLS}|X) = \sigma^2(X'P'PX)^{-1} = \sigma^2 \left(\sum_{i=1}^{n} h_i^{-2} x_i x_i' \right)^{-1}$$

- Unverzerrter Schätzer für das unbekannte σ^2 (Residuenvarianz im transformierten Modell) ist

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^{n} h_i^{-2} (y_i - x_i' \widehat{\beta}_{GLS})^2$$

- Damit ergibt sich als Schätzer für Varianz-Kovarianzmatrix des GLS-Schätzers

$$\hat{V}(\widehat{\beta}_{GLS}|X) = \frac{1}{n-K} \sum_{i=1}^{n} h_i^{-2} (y_i - x_i' \widehat{\beta}_{GLS})^2 \left(\sum_{i=1}^{n} h_i^{-2} x_i x_i' \right)^{-1}$$

  Die Wurzel der Elemente auf der Hauptdiagonalen dieser Matrix gibt wieder die Schätzfehler der einzelnen GLS-Schätzer an.

- Wenn weiterhin noch die Annahme (A5) der Normalverteiltheit der Residuen unterstellt wird, ist

$$\widehat{\beta}_{GLS} \sim \mathcal{N}\left(\beta, \sigma^2 \left(\sum_{i=1}^{n} h_i^{-2} x_i x_i'\right)^{-1}\right),$$

womit sich wiederum die bekannten t- und F-Tests auf lineare Restriktionen ableiten lassen.

Beispiel 1

- Ein Test
$$H_0 : \beta_2 = 1 \text{ gegen } H_1 : \beta_2 \neq 1$$
kann mittels der t-Statistik
$$t = \frac{\widehat{\beta}_{GLS\ 2} - 1}{\sqrt{\hat{V}(\widehat{\beta}_{GLS\ 2})}}$$
durchgeführt werden. Diese Teststatistik ist bei bekannten Gewichten h_i wieder t-verteilt mit $n - K$ Freiheitsgraden.
- Ohne die Annahme der Normalverteilung der Residuen, gilt die Normalverteilungseigenschaft des GLS-Schätzers nur asymptotisch.
Im Beispiel des t-Testes können die kritischen Schranken dann aus einer Normalverteilung abgelesen werden, da die t-Verteilung mit steigender Zahl von Freiheitsgraden in eine Standardnormalverteilung übergeht.

Beispiel 2

- Ein Test auf die allgemeine lineare Restriktion $H_0 : R\beta = q$, kann für normalverteilte Residuen mittels der unter H_0 F-verteilten Prüfgröße

$$f = \xi/J = (R\widehat{\beta}_{GLS} - q)'(R\hat{V}(\widehat{\beta}_{GLS})R')^{-1}(R\widehat{\beta}_{GLS} - q)/J \sim F(J, n-K)$$

getestet werden.
Alternativ können für nicht-normalverteilte Residuen und große Stichprobenumfänge die Quantile der Teststatistik $\xi = fJ$ bei Gültigkeit von H_0 aus der χ^2-Tabelle mit J Freiheitsgraden entnommen werden (sog. WALD-Test).

Falls die Varianzen unbekannt sind

- Bislang wurde h_i als bekannt vorausgesetzt. Dies ist nur im Fall

$$V(\varepsilon_i|x_i) = \sigma^2 x_{i2}^2$$

  sinnvoll, wenn die Heteroskedastie durch genau eine Variable erklärt wird.
- Problem: Wenn h_i^2 unbekannt sind, kann der GLS-Schätzer nicht mehr berechnet werden (Gewichtung ist unbekannt).
- **Lösung**: Unverzerrte oder konsistente Schätzung für h_i^2 und Berechnung des FGLS-Schätzers.

- Neues Problem: Es gibt n h_i's und n Beobachtungen.
  Schätzt man in einer ersten Stufe mittels OLS (d.h. Heteroskedastie wird ignoriert), dann erhält man n Regressionsfehler $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$, so dass man die Varianz σ_i^2 nur aus einem Datum $\widehat{\varepsilon}_i$ schätzen kann. Dies ist konsistent nicht möglich, wenn nicht zusätzliche Annahmen getroffen werden.
- Diese Annahmen betreffen die Form der Heteroskedastie. Sie stellen insbesondere die Varianz als Funktion der beobachteten Variablen dar.

Beispiele:

$$V(\varepsilon_i | x_{ik}) = \sigma^2 x_{ik}^{\alpha}$$

oder

$$V(\varepsilon_i | x_{ik}, x_{il}) = \sigma^2 \left(x_{ik}^{\alpha_1} + x_{il}^{\alpha_2} \right)$$

- Dabei müssen die Parameter α bzw. α_1, α_2 konsistent geschätzt werden, woraus sich ein konsistente Schätzung $\hat{h}_i$ ergibt.
  Wenn dies möglich ist, lautet der FGLS- (EGLS-) Schätzer

$$\widehat{\beta}_{FGLS} = \left(\sum_{i=1}^{n} \hat{h}_i^{-2} x_i x_i' \right)^{-1} \sum_{i=1}^{n} \hat{h}_i^{-2} x_i y_i$$

  und der Schätzer für die Varianz-Kovarianzmatrix des FGLS-Schätzers ist

$$\hat{V}(\widehat{\beta}_{FGLS} | X) = \hat{\sigma}^2 \left(\sum_{i=1}^{n} \hat{h}_i^{-2} x_i x_i' \right)^{-1},$$

  wobei

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^{n} \hat{h}_i^{-2} (y_i - x_i' \widehat{\beta}_{FGLS})^2$$

  ist.

- *Beachte*: Wenn h_i, $i = 1, 2, \ldots, n$ konsistent geschätzt werden können, sind der FGLS- und GLS-Schätzer asymptotisch äquivalent. D.h. sie besitzen dieselben asymptotischen Eigenschaften. Asymptotisch kann man dann also das Problem unbekannter Varianzen der Störgrößen ignorieren.
  Allerdings sind die Eigenschaften von FGLS- und GLS-Schätzer in kleinen Stichproben keineswegs identisch. So ist insbesondere der FGLS-Schätzer nicht BLUE.
  Man kann lediglich sagen, dass unter den Annahmen (**A9**) und (**A10**) und für gewisse Annahmen über die Form der Heteroskedastie der FGLS-Schätzer konsistent und asymptotisch effizient ist.

Heteroskedastie-konsistente Standardfehler für OLS

- Idee: Statt eines neuen Schätzers (GLS oder FGLS) wird auch bei Heteroskedastie weiterhin OLS geschätzt. Man korrigiert aber die Standardfehler des OLS-Schätzers.
- Der OLS-Schätzer $\widehat{\beta} = (X'X)^{-1}X'y$ besitzt im Falle einer bestimmten Form der Heteroskedastie die Varianz-Kovarianzmatrix

$$V(\widehat{\beta}|X) = (X'X)^{-1} \underbrace{X'\text{Diag}\{\sigma_i^2\}X}_{\sum_{i=1}^{n} \sigma_i^2 x_i x_i'}(X'X)^{-1}$$

- White (1980) hat vorgeschlagen die unbekannte Varianz-Kovarianzmatrix Psi der Störgröße durch

$$S = \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 x_i x_i'$$

zu schätzen. Dieser Schätzer ist konsistent für

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 x_i x_i'.$$

- Mittels

$$\hat{V}(\widehat{\beta}|X) = (X'X)^{-1} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 x_i x_i' (X'X)^{-1}$$

lässt sich die wahre Varianz-Kovarianzmatrix (und damit die wahren Standardfehler) des OLS-Schätzers schätzen.

- **Vorteil:** Diese Heteroskedastie-konsistenten Standardfehler (kurz: White-Standardfehler) erfordern nicht die Spezifikation der Form der Heteroskedastie.

Spezialfall: Modell mit zwei unbekannten Varianzen

- Spezialfall: Die Beobachtungen $i = 1, 2, \ldots, n$ lassen sich in zwei Gruppen A und B zerlegen. Die einfachste Form der Heteroskedastie liegt dann vor, wenn die Varianz der Störgröße nur zwei Werte σ_A^2 und σ_B^2 annehmen kann.
  D.h. in jeder Teilgruppe ist die Varianz homoskedastisch.
- **Beispiel**: Die Varianz der Störgröße in einer Lohngleichung unterscheidet sich für Männer und Frauen.
- Da die Varianz in den Teilgruppen homoskedastisch ist, können für beide Gruppen isoliert OLS-Schätzungen $\widehat{\beta}_A$ bzw. $\widehat{\beta}_B$ durchgeführt werden, die aufgrund der Homoskedastie auch BLUE sind.
  Aus den n_A Regressionsfehlern für die Regression A lässt sich σ_A^2 mittels

  $$\widehat{\sigma}_A^2 = s_A^2 = \frac{1}{n_A - K} \sum_{i \in A} (y_i - x_i' \widehat{\beta}_A)^2$$

  konsistent schätzen. Es gibt jetzt n_A Werte der Regressionsfehler zur Schätzung einer Varianz. Analog erhält man s_B^2 als Schätzer für σ_B^2.

Setzt man diese Varianzschätzer in den FGLS-Schätzer für β ein, so lautet dieser

$$\widehat{\beta}_{FGLS} = \left(\sum_{i \in A} s_A^{-2} x_i x_i' + \sum_{i \in B} s_B^{-2} x_i x_i' \right)^{-1} \left(\sum_{i \in A} s_A^{-2} x_i y_i + \sum_{i \in B} s_B^{-2} x_i y_i \right)$$

- Er lässt sich darstellen als ein gewichtetes arithmetisches Mittel der beiden OLS-Schätzungen

$$\widehat{\beta}_{FGLS} = W \widehat{\beta}_A + (I - W) \widehat{\beta}_B$$

mit der $K \times K$-Matrix

$$W = \left(\sum_{i \in A} s_A^{-2} x_i x_i' + \sum_{i \in B} s_B^{-2} x_i x_i' \right)^{-1} \sum_{i \in A} s_A^{-2} x_i x_i'$$

- Fazit: Die Teilgruppe mit kleinerem Schätzwert für die Varianz der Störgrößen (d.h. die Punktwolke “streut weniger” und die Regressionsbeziehung ist “sicherer”) bekommt das größere Gewicht.

Multiplikative Heteroskedastie

- Ausgangspunkt:

$$\begin{aligned} Var(\varepsilon_i|x_i) &= \sigma_i^2 = \sigma^2 \exp(\alpha_1 z_{i1} + \ldots + \alpha_J z_{iJ}) \\ &= \sigma^2 \exp(z_i'\alpha) \end{aligned}$$

- *Beachte*: Zumeist enthält z_i eine Teilmenge der Variablen, die in x_i enthalten sind. Ggf. werden die Variablen noch transformiert (quadriert oder logarithmiert).
- *Beachte*: Wenn $J = 1$ und z_{i1} eine Dummy-Variable ist, erhält man den Spezialfall zweier unbekannter Varianzen.
- EGLS-Schätzung verlangt konsistente Schätzer für α. Ausgangspunkt für diese Schätzer ist die Annahme, dass die OLS-Residuen $\widehat{\varepsilon}_i^2$ Informationen über σ_i^2 enthalten.

- Regressionsmodell:
$$\log \widehat{\varepsilon}_i^2 = \log \sigma^2 + z_i'\alpha + v_i$$
mit $v_i = \log(\widehat{\varepsilon}_i^2/\sigma_i^2)$ (asymptotisch) homoskedastisch und unkorreliert mit z_i. Beachte: $E(v_i) \neq 0$, womit per OLS der Achsenabschnitt $\log \sigma^2$ verzerrt geschätzt wird. Es interessiert aber nur α.

Arbeitsschritte der EGLS-Schätzung für β

1. Modellschätzung (d.h. Schätzung von β) mittels OLS liefert die OLS-Schätzer $\widehat{\beta}$.
2. Berechnung von $\log \widehat{\varepsilon}_i^2 = \log(y_i - x_i'\widehat{\beta})^2$ aus den OLS-Residuen $\widehat{\varepsilon}_i$.
3. Schätzung der Hilfsregression

$$\log \widehat{\varepsilon}_i^2 = \log \sigma^2 + z_i'\alpha + v_i$$

   mittels OLS liefert konsistenten Schätzer $\hat{\alpha}$ für α.
4. Berechne $\hat{h}_i^2 = \exp(z_i'\hat{\alpha})$ und transformiere die Beobachtungen, was zum transformierten Modell

$$y_i/\hat{h}_i = (x_i/\hat{h}_i)'\beta + (\varepsilon_i/\hat{h}_i)$$

   führt, in dem β erneut mit OLS geschätzt wird. Der EGLS-Schätzer wird mit $\hat{\beta}^*$ bezeichnet.
   Beachte: Auch der Achsenabschnitt wird transformiert.

Arbeitsschritte der EGLS-Schätzung für β

5. Konsistente Schätzung der skalaren Konstanten σ^2 mittels

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^{n} \frac{(y_i - x_i' \widehat{\beta}_{FGLS})^2}{\hat{h}_i^2}$$

6. Konsistente Schätzung der Varianz-Kovarianzmatrix des FGLS-Schätzers $\hat{\beta}^*$:

$$\hat{V}(\widehat{\beta}_{FGLS}) = \hat{\sigma}^2 \left(\sum_{i=1}^{n} \frac{x_i x_i'}{\hat{h}_i^2} \right)^{-1}$$

Beachte: Dies ist die Varianz-Kovarianzmatrix des OLS-Schätzers im transformierten Modell.

Testen auf Gleichheit zweier Varianzen

- Einfachste Situation: Zwei u.U. unterschiedliche Varianzen σ_A^2 und σ_B^2
- Nullhypothese: $\sigma_A^2 = \sigma_B^2$.
- Es gilt unter der Nullhypothese (d.h. Gauß-Markov-Annahmen sind erfüllt) für normalverteilte Residuen oder approximativ

$$(n_j - K)\frac{s_j^2}{\sigma_j^2} \sim \chi^2_{n_j - K} \text{ für } j = A, B$$

- Wegen der Unabhängigkeit der Teilstichproben sind s_A^2 und s_B^2 stochastisch unabhängig, so dass exakt oder approximativ

$$\frac{s_A^2/\sigma_A^2}{s_B^2/\sigma_B^2} \sim F^{n_A - K}_{n_B - K}$$

D.h. unter H_0 ist

$$\lambda = s_A^2/s_B^2 \sim F^{n_A - K}_{n_B - K}$$

- Entscheidungsregel:
  1. Lehne H_0 gegen $H_1 : \sigma_A^2 > \sigma_B^2$ ab, wenn λ sehr groß ist, d.h. z.B. größer als das 95%-Quantil der F-Verteilung.
  2. Dieselbe Entscheidungsregel wird für die Alternativhypothese $H_1 : \sigma_A^2 < \sigma_B^2$ angewandt, wenn die Rolle von A und B vertauscht werden. Dies schließt eine Vertauschung von Zähler- und Nennerfreiheitsgraden ein.
  3. Lehne H_0 gegen die zweiseitige Alternative $H_1 : \sigma_A^2 \neq \sigma_B^2$ ab, wenn λ sehr klein oder sehr groß ist, d.h. z.B. kleiner als das 2.5%-Quantil bzw. größer als das 97.5%-Quantil der F-Verteilung.
- Dieser Test heißt in der Literatur auch **Goldfeld-Quandt-Test**.

Testen von multiplikativer Heteroskedastie

- Ausgangspunkt:
$$\sigma_i^2 = \sigma^2 \exp(z_i'\alpha)$$
- Testhypothesen:
$$H_0 : \ \alpha = 0 \ \text{ gegen } \ H_1 : \alpha \neq 0.$$
- Möglichkeit: F-Test mit J Zähler- und $n - J - 1$ Nennerfreiheitsgraden auf $\alpha = 0$ für die Regression
$$\log \widehat{\varepsilon}_i^2 = \log \sigma^2 + z_i'\alpha + v_i$$
- *Beachte*: F-Verteilung gilt nur noch approximativ, da Gauß-Markov-Annahmen nicht erfüllt sind.

Breusch-Pagan-Test

- Ausgangspunkt:

$$\sigma_i^2 = \sigma^2 h(z_i'\alpha),\ h(.) > 0,\ h(0) = 1 \text{ unbekannt, stetig differenzierbar.}$$

- Testhypothesen (ohne Kenntnis von h):

$$H_0:\ \alpha = 0 \text{ gegen } H_1 : \alpha \neq 0$$

- Vorgehen:
  1. Hilfsregression von $\widehat{\varepsilon}_1^2, ..., \widehat{\varepsilon}_n^2$ auf z_i und eine Konstante,
  2. Berechnung des Bestimmtheitsmaßes für diese Regression und der Prüfgröße

$$\zeta = n \cdot R^2 \overset{asy}{\sim} \chi^2(J) \text{ unter } H_0.$$

- Es handelt sich um einen sog. LM- (Lagrange-Multiplikator-) Test.

White-Test

- Prinzip:
  1. Verzicht auf eine konkrete parametrisierte Form der Heteroskedastie in der Alternativhypothese.
  2. Verwendung des Heteroskedastie-konsistenten Schätzers der Varianz-Kovarianzmatrix des OLS-Schätzers.
- Ausgangspunkt: Wenn keine Heteroskedastie (Nullhypothese) vorliegt, liefert

$$\hat{V}(\widehat{\beta}) = s^2 \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1}$$

eine konsistente Schätzung der Varianz-Kovarianzmatrix des OLS-Schätzers $\widehat{\beta}$.

- Prozedur von White (ähnlich wie beim Breusch-Pagan-Test):
  Regressiere das Quadrat der OLS-Residuen $\widehat{\varepsilon}_i^2$ auf Achsenabschnitt und

  $$x_{ik}, x_{ik}^2, x_{ik}x_{jk}$$

  und berechne nR^2 für diese Regression. Unter H_0 ist nR^2 asymptotisch χ^2-verteilt mit P Freiheitsgraden, wobei P die Anzahl der Regressoren (ohne Achsenabschnitt) in der Hilfsregression angibt.
- *Beachte*: Der White-Test ist zwar allgemein, da er keine konkrete Alternativhypothese verlangt, besitzt aber nur geringe Power, wenn man tatsächlich spezifische Formen der Heteroskedastie betrachtet.

Welcher Test?

- Problem: Je spezifischer die Form der Heteroskedastie, desto besser der zugehörige Test.
- Kenntnis der Form u.U. aus dem Plot der Residuen.

Fallstudie: Erklärung der Arbeitsnachfrage

- Ziel: Erklärung der Arbeitsnachfrage belgischer Firmen.
- Daten: Querschnittsdatensatz von 569 Firmen im Jahre 1996.
- Erhobene Variablen:
  1. $labour$: Beschäftigtenzahl,
  2. $capital$: Vermögen (in Mill. Bef),
  3. $wage$: Gesamte Lohnkosten dividiert durch Anzahl der Beschäftigten (in Mill. Bef),
  4. $output$: Mehrwert (in Mill. Bef).
- Herleitung der ökonometrischen Spezifikation aus der ökonomischen Theorie (d.h. Ableitung der Arbeitsnachfrage aus einer Produktionsfunktion und Kostenminimierung).
  Ergebnis:

$$L = g(Q, r, w)$$

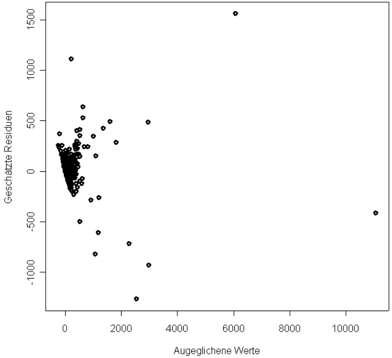
mit L Arbeitsnachfrage, Q Output, r Zinssatz und w Lohnsatz.

- Ökonometrische Spezifikation:
  Wähle die Funktion g linear und ersetze den Zinssatz durch den Kapitalstock.
- Ergebnis einer OLS-Schätzung:

Abhängige Variable: $labour$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	287.720	19.640	14.648
$wage$	-167.130	12.430	-13.446
$output$	0.382	0.009	43.304
$capital$	-0.114	0.007	-17.067
$s = 156.26$, $R^2 = 0.9352$, $\overline{R}^2 = 0.9348$, $F = 2716.02$			

- Interpretation der Ergebnisse:
  1. Die geschätzten Koeffizienten besitzen das zu erwartende Vorzeichen und sind signifikant (von Null verschieden).
  2. D.h. eine Erhöhung der Löhne führt c.p. zu einer Reduktion des Arbeitseinsatzes.
  3. D.h. eine Erhöhung der Produktion führt c.p. zu einer Erhöhung der Arbeitsnachfrage.

- Aber: Interpretation der Signifikanzen setzt Fehlen von Heteroskedastie voraus.
- Ein Plot der gefitteten Werte im linearen Modell ($\widehat{y_i}$) gegen die OLS-Residuen ($\widehat{\varepsilon_i}$) zeigt jedoch ein deutlich ausgeprägtes Heteroskedastieverhalten:



- Schätzergebnisse für die Hilfsregression des Breusch-Pagan-Testes ($z_i = x_i$):

Abhängige Variable: $\widehat{\varepsilon}_i^2$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-22719.51	11838.88	-1.919
$wage$	5673.13	7491.66	0.757
$output$	132.92	5.31	25.015
$capital$	-87.84	4.02	-21.858
$s = 94182$, $R^2 = 0.5818$, $\overline{R}^2 = 0.5796$, $F = 262.05$			

- Interpretation der Schätzergebnisse:
  1. Hohe t-Werte und das hohe R^2 lassen Heteroskedastie vermuten.
  2. Breusch-Pagan-Teststatistik: $nR^2 = 569 \cdot 0.5818 = 331.0$.
     95%-Quantil der χ^2-Verteilung mit 3 Freiheitsgraden: 7.81.
     Fazit: Ablehnung der Nullhypothese der Homoskedastie.
- Heteroskedastie ist normal, wenn man Beobachtungseinheiten sehr unterschiedlicher Größe vorliegen hat. Im Beispiel Firmen mit nur einem Mitarbeiter bis zu Firmen mit über 1000 Mitarbeitern.
  Es ist zu vermuten, dass sämtliche Variablen (inkl. dem Störterm) mit der Größe der Firmen wachsen.

- Lösung: Logarithmierung schwächt das Wachstum ab.
  D.h. es wird ein loglineares Modell betrachtet.
  Im Beispiel: Cobb-Douglas-Produktionsfunktion:

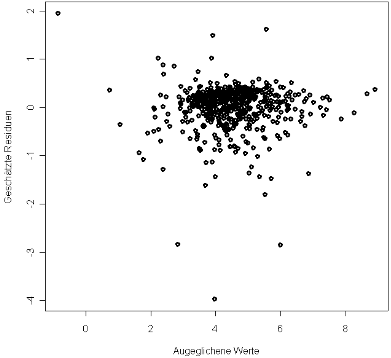
$$Q = AK^{\alpha}L^{\beta}$$

- Schätzergebnisse im loglinearen Modell:

Abhängige Variable: $\log(labour)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-0.448	0.093	-4.806
$\log(wage)$	-0.928	0.071	-12.993
$\log(output)$	0.990	0.026	37.487
$\log(capital)$	-0.004	0.019	-0.197

$s = 0.465,\ R^2 = 0.8430,\ \overline{R}^2 = 0.8421,\ F = 1011.02$

- Ein Plot der gefitteten Werte im logarithmierten Modell ($\widehat{\log y_i}$) gegen die OLS-Residuen ($\widehat{\varepsilon}_i$) zeigt ein deutlich weniger ausgeprägtes Heteroskedastieverhalten:



- Interpretation der Schätzergebnisse:
  1. Die Lohnelastizität der Arbeitsnachfrage ist -0.93 und relativ hoch. D.h. 1% Steigerung des Lohnes bewirkt c.p. ungefähr 1% Rückgang der Arbeitsnachfrage.
  2. Die Outputelastizität der Arbeitsnachfrage ist ungefähr 1.
  3. Breusch-Pagan-Teststatistik: $nR^2 = 569 \cdot 0.0136 = 7.74$.
     95%-Quantil der χ^2-Verteilung mit 3 Freiheitsgraden: 7.81.
     Fazit: Kein eindeutiges Ergebnis $\Longrightarrow$ zusätzlich White-Test.

Schätzergebnisse der Hilfsregression des White-Tests

Abhängige Variable: $\widehat{\varepsilon}_i^2$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	1.324	0.458	2.891
$\log(wage)$	0.359	0.556	0.646
$\log(output)$	-0.774	0.242	-3.194
$\log(capital)$	0.380	0.146	2.607
$\log^2(wage)$	0.193	0.259	0.744
$\log^2(output)$	0.138	0.036	3.877
$\log^2(capital)$	0.090	0.014	6.401
$\log(wage)\log(output)$	0.138	0.163	0.849
$\log(wage)\log(capital)$	-0.252	0.105	-2.399
$\log(output)\log(capital)$	-0.192	0.037	-5.197

$s = 0.851,\ R^2 = 0.1029,\ \overline{R}^2 = 0.0884,\ F = 7.12$

- Interpretation der Schätzergebnisse:
  1. White-Teststatistik: $nR^2 = 569 \cdot 0.1029 = 58.6$.
     95%-Quantil der χ^2-Verteilung mit 9 Freiheitsgraden: 16.91.
     Fazit: Nullhypothese der Homoskedastie ist abzulehnen.
  2. Ursache für die Heteroskedastie scheint die Abhängigkeit der Residuenvarianz von den Variablen $output$ und $capital$ zu sein, deren t-Werte in der Hilfsregression betragsmäßig relativ hoch sind.
- Fazit: HAC- (White-) Standardfehler, korrigierte t- und F-Werte.

Schätzergebnis im loglinearen Modell mit White-Standardfehlern, korrigierten t- und F-Werten

Abhängige Variable: $\log(labour)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-0.448	0.133	-3.362
$\log(wage)$	-0.928	0.087	-10.706
$\log(output)$	0.990	0.047	21.159
$\log(capital)$	-0.004	0.038	-0.098
$s = 0.465$, $R^2 = 0.8430$, $\overline{R}^2 = 0.8421$, $F = 544.73$			

- Interpretation der Schätzergebnisse:
  1. Die White-Standardfehler sind größer und damit die t-Werte betragsmäßig kleiner.
  2. Qualitativ ändert sich nichts am signifikanten Einfluss von $\log(wages)$ und $\log(output)$ auf die Arbeitsnachfrage. $\log(capital)$ bleibt insignifikant.

- Alternative: Multiplikative Modellierung der Heteroskedastie

$$V(\varepsilon_i|x_i) = \log \sigma_i^2 = \log \sigma^2 + x_i'\alpha$$

- Schätzergebnisse einer Hilfsregression für multiplikative Heteroskedastie:

Abhängige Variable: $\log(\widehat{\varepsilon}_i^2)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-3.214	0.449	-7.160
$\log(wage)$	-0.061	0.344	-0.178
$\log(output)$	0.267	0.127	2.099
$\log(capital)$	-0.331	0.090	-3.659
$s = 2.241,\ R^2 = 0.0245,\ \overline{R}^2 = 0.0193,\ F = 4.73$			

- Interpretation der Schätzergebnisse:
  1. $\log(output)$ und $\log(capital)$ besitzen einen signifikanten Einfluss auf die logarithmierte Residuenvarianz (bzw. stellvertretend den logarithmierten quadrierten OLS-Regressionsfehler).
  2. F-Statistik zeigt Erklärungskraft des Modells für die Residuenvarianz und damit Heteroskedastie an.
  3. Überprüfung, ob die Spezifikation der Heteroskedastie zu restriktiv ist, erfolgt mittels Aufnahme der quadrierten logarithmierten Werte der drei erklärenden Variablen in die Hilfsregression und Durchführung eines F-Testes, ob diese drei zusätzlichen Variablen gemeinsam signifikant sind.
     Ergebnis: $f = 1.85$ (p-Wert$= 0.137$), d.h. Nullhypothese kann nicht abgelehnt werden.
- *Beachte*: Bis auf den Achsenabschnitt sind sämtliche Schätzer konsistent.

EGLS-Schätzung der Parameter des loglinearen Modells

Abhängige Variable: $\log(labour)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-0.466	0.091	-5.145
$\log(wage)$	-0.856	0.072	-11.903
$\log(output)$	1.035	0.027	37.890
$\log(capital)$	-0.057	0.022	-2.636
$s = 2.509,\ R^2 = 0.9903,\ \overline{R}^2 = 0.9902,\ F = 14401.3$			

- Interpretation der Schätzergebnisse:
  1. Vergleich mit den HAC-Standardfehlern zeigt einen deutlichen (asymptotischen) Effizienzgewinn: Die EGLS-Standardfehler sind deutlich kleiner.
  2. Die EGLS-Schätzergebnisse für die Regressionskoeffizienten sind ähnlich den OLS-Schätzergebnissen.
     Ausnahme: Der Einfluss des Kapitals ist nun auf dem 5%-Niveau signifikant.
  3. t-Test auf die Nullhypothese einer Lohnelastizität der Arbeitsnachfrage von -1 ergibt $t = (-0.856 - 1)/0.072 = 2.01$, so dass die Nullhypothese auf dem 5%-Niveau gerade eben abgelehnt werden kann.

- *Beachte*: Ein Vergleich der Bestimmtheitsmaße im originalen und transformierten Modell ist nicht zulässig, da
  1. im transformierten Modell das nicht-zentrierte Bestimmtheitsmaß berechnet wird (Achsenabschnitt fehlt),
  2. im transformierten Modell eine andere (transformierte) endogene Variable verwendet wird.
- Verwendet man im transformierten Modell als Gütemaß

$$R^2 = \text{corr}^2\{y_i, \hat{y}_i\},$$

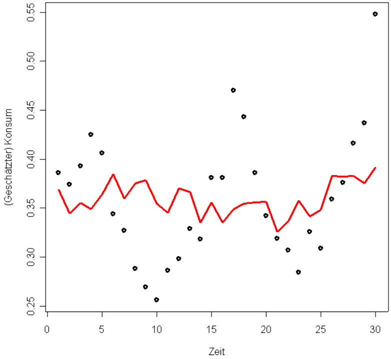
so erhält man $R^2 = 0.8403$, was nur geringfügig kleiner als der OLS-Wert ist.
Fazit: Das Bestimmtheitsmaß eignet sich nicht zum Vergleich unterschiedlicher Schätzer.
"Of course, there are more important things in an econometrician's life than a high R^2." (S. 90).

Vorbemerkungen

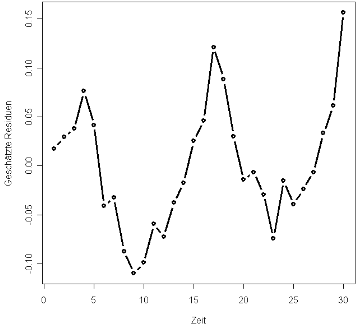
- Korrelation zwischen verschiedenen Störtermen führt zu nicht-skalarer Varianz-Kovarianzmatrix.
- Auch bei Autokorrelation bleibt der OLS-Schätzer unverzerrt und konsistent.
- Der OLS-Schätzer wird aber ineffizient, so dass die Standardfehler falsch geschätzt werden.
- Autokorrelation ist typisch bei Zeitreihendaten, da diese chronologisch geordnet vorliegen.
- Für Zeitreihendaten wird t (statt i) als Index der Beobachtungen und T (statt n) als Gesamtzahl der Beobachtungen verwendet.
- Der Störterm umfasst sämtliche nicht explizit im Modell berücksichtigten Variablen. Besitzen diese Variablen einen dauerhaften Einfluss, so führt dies zu positiver Autokorrelation.
- Wenn diese ausgeschlossenen Variablen prinzipiell beobachtbar sind und im Modell berücksichtigt werden könnten, ist Autokorrelation ein Indiz für Fehlspezifikation.
- Nicht korrekte Funktionsform, weggelassene Variablen und eine nicht adäquate Spezifikation der dynamischen Struktur können zu Autokorrelation führen.

Beispiel: Zeitreihe der Nachfrage nach Eiscreme

- Es werden monatliche Daten betrachtet.
- Die Nachfrage nach Eiscreme wird abhängig vom Einkommen und Preisindex modelliert. Das Wetter kommt als Regressor nicht vor und ist im Störterm enthalten.
- Vergleich des tatsächlichen Konsums von Eiscreme (Punkte) mit den ausgeglichenen Werten (rote Linie) einer OLS-Regression mit den erklärenden Variablen Einkommen und Preisindex.



- Ein Plot der OLS-Residuen ($\widehat{\varepsilon}_i$) gegen die Zeit zeigt ein deutlich ausgeprägtes Autokorrelationsverhalten:



Autokorrelation erster Ordnung

- Autokorrelation 1. Ordnung ist die einfachste und populärste Form der Modellierung von Autokorrelation.
- Modellgleichung für den Störterm:

$$y_t = x_t'\beta + \varepsilon_t$$

mit

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \text{ mit } v_t \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$$

und $|\rho| < 1$.

- Die Folge von Zufallsvariablen $(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ wird als **stochastischer Prozess** bezeichnet.

Autokorrelation erster Ordnung

- Setzt man die autoregressive Beziehung mehrfach ein, so ergibt sich:

$$\begin{aligned} \varepsilon_t &= \rho\varepsilon_{t-1} + v_t \\ &= \rho(\rho\varepsilon_{t-2} + v_{t-1}) + v_t \\ &= \rho^t\varepsilon_0 + v_t + \rho v_{t-1} + \ldots + \rho^{t-1} v_1 \end{aligned}$$

- Es ist nun sinnvoll anzunehmen, dass die Wirkung eines Schocks v_{t-i} in der Periode $t-i$ auf ε_t umso geringer ausfällt, je weiter die Periode $t-i$ in der Vergangenheit liegt. Dies bedingt aber $|\rho| < 1$.
- Konsequenzen von $|\rho| < 1$:
  1. $E(\varepsilon_t) = 0$ für alle t:

$$\begin{aligned} E(\varepsilon_t) &= \rho^i E(\varepsilon_{t-i}) + \sum_{j=0}^{i-1} \rho^j E(v_{t-j}) \\ &\rightarrow 0 \text{ für } i \rightarrow \infty. \end{aligned}$$

2. $V(\varepsilon_t)$ ist konstant:

$$\begin{aligned} V(\varepsilon_t) &= \rho^{2i} V(\varepsilon_{t-i}) + \sum_{j=0}^{t-i} \rho^{2j} V(v_{t-j}) \\ &\rightarrow \sigma_v^2/(1-\rho^2) \text{ für } i \rightarrow \infty. \end{aligned}$$

3. Die sog. Autokovarianz $cov(\varepsilon_t, \varepsilon_{t-l})$ hängt nur vom Zeitlag $l \neq 0$ ab:

$$E(\varepsilon_t \varepsilon_{t-l}) = \rho^l \frac{\sigma_v^2}{1-\rho^2}$$

wegen

$$\begin{aligned} \varepsilon_t &= \sum_{j=0}^{\infty} \rho^j v_{t-j} \text{ und} \\ \varepsilon_{t-l} &= \sum_{j'=0}^{\infty} \rho^{j'} v_{t-(l+j')}. \end{aligned}$$

- Man nennt die Eigenschaften 1 bis 3 **(schwache) Stationarität**. Hinreichend und notwendig für die Stationarität eines autoregressiven Prozesses 1. Ordnung ist $|\rho| < 1$.
- Aus den Autokovarianzen und Varianzen errechnet man die Autokorrelation

$$\text{corr}(\varepsilon_t, \varepsilon_{t-l}) = \rho^l$$

  für $l = 0, 1, 2, \ldots$. Die chronologische Aneinanderreihung der Autokorrelationen heißt **(theoretische) Autokorrelationsfolge**.
- Die Autokorrelationsfolge eines stationären autoregressiven Prozesses 1. Ordnung klingt geometrisch ab, d.h weit zurückliegende Störterme besitzen nur einen relativ kleinen Einfluss.

Cochrane-Orcutt-Schätzer

- Gesucht ist eine Transformation, die die Autokorrelation bereinigt.
  Vorschlag: $\varepsilon_t - \rho\varepsilon_{t-1} = v_t$, so dass v_t die Gauß-Markov-Annahmen erfüllt.
  D.h. transformiertes Modell

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + v_t, \quad t = 2, 3, \ldots, T.$$

- Ist ρ bekannt, so erhält man den GLS-Schätzer für β durch OLS-Schätzung des transformierten Modells.
- Problem: Zeitreihe beginnt irgendwann und y_0 bzw. x_0 sind unbekannt.
  Lösung 1: Verzicht auf die erste Beobachtung liefert einen approximativen GLS-Schätzer, der sicher nicht effizient ist. Er wird auch als **Cochrane-Orcutt** (**CORC**-) Schätzer bezeichnet.

Prais-Winston-Schätzer

- Lösung 2: Besondere homoskedastieerhaltende Transformation der ersten Beobachtung

  $$\sqrt{1-\rho^2}y_1 = \sqrt{1-\rho^2}x_1'\beta + \sqrt{1-\rho^2}\varepsilon_1$$

  Dann ist

  $$V(\sqrt{1-\rho^2}\varepsilon_1) = \sigma_v^2,$$

  so dass Homoskedastie im transformierten Modell vorliegt. Der OLS-Schätzer ist jetzt als GLS-Schätzer BLUE für β. Er wird als **Prais-Winsten**-Schätzer bezeichnet.
- Cochrane-Orcutt- und Prais-Winsten-Schätzer besitzen dieselben asymptotischen Eigenschaften.

Unbekanntes ρ

- Wenn ρ unbekannt ist, muss ρ konsistent geschätzt und anschließend eine (nicht mehr effiziente) EGLS-Schätzung durchgeführt werden. Asymptotisch sind GLS- und EGLS-Schätzer unter schwachen Regularitätsbedingungen äquivalent.
- Ausgehend von einer ersten OLS-Schätzung kann aus deren Regressionsfehlern ρ mittels

$$\hat{\rho} = \left(\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^{T} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-1} \right)$$

  konsistent geschätzt werden.
  Es handelt sich um den OLS-Schätzer einer homogenen Regression von $\widehat{\varepsilon}_t$ auf $\widehat{\varepsilon}_{t-1}$.

- Iterative Cochrane-Orcutt-Prozedur: Die Schätzung von ρ und β wird nicht nur jeweils einmal, sondern im Wechsel solange durchgeführt, bis Konvergenz eintritt, d.h. die Schätzwerte sich von einer zur nächsten Iteration um weniger als eine vorgegebene Genauigkeitsschwelle unterscheiden.
  Asymptotisch führt dieses iterative Vorgehen zu keinem Effizienzgewinn. In kleinen Stichproben ist dies jedoch möglich, wenn auch nicht garantiert.

Tests auf Autokorrelation erster Ordnung

- Zu testen

$$H_0: \ \rho = 0 \ \text{ gegen } \ H_1: \ \rho \neq (<,>)0.$$

1. Asymptotische Tests
2. Durbin-Watson-Test

Asymptotische Tests

- Wenn $|\rho| < 1$ ist, lässt sich zeigen, dass

$$\sqrt{T}(\hat{\rho} - \rho) \overset{asy}{\sim} \mathcal{N}(0, 1 - \rho^2)$$

d.h. $\hat{\rho}$ ist konsistent und asymptotisch normalverteilt.

- Für endliche Stichproben ergibt sich daraus die Approximation

$$z = \frac{\sqrt{T}(\hat{\rho} - \rho)}{\sqrt{1 - \rho^2}} \overset{approx}{\sim} \mathcal{N}(0, 1)$$

- Unter H_0 ist $\rho = 0$ und $z = \sqrt{T}\hat{\rho}$ approximativ standardnormalverteilt.
- Entscheidungsregel: Lehne die Nullhypothese zugunsten der zweiseitigen Alternativhypothese mit einer Irrtumswahrscheinlichkeit von 5% ab, wenn

$$|\sqrt{T}\hat{\rho}| > 1.96.$$

- Alternativer asymptotischer Test: **Breusch-Godfrey**- (LM-) Test
  Regressiere die OLS-Residuen $\widehat{\varepsilon}_t$ auf $\widehat{\varepsilon}_{t-1}$, berechne den Wert des Bestimmtheitsmaßes R^2 und teste mittels der unter H_0 asymptotisch mit einem Freiheitsgrad χ^2-verteilten Prüfgröße $(T-1)R^2$.
  Vorteil: Erweiterbar auf Autokorrelationen höherer Ordnung.
- Nachteil der asymptotischen Tests: Approximationen können für kleine Stichproben (Zeitreihenlängen) schlecht sein.

Durbin-Watson-Test

- Wichtige Annahme:
  Alle erklärenden Variablen sind als deterministisch aufzufassen, so dass der Störterm von allen erklärenden Variablen unabhängig ist.
- Durbin-Watson-Statistik:

$$dw = \frac{\sum_{t=2}^{T} (\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \widehat{\varepsilon}_t^2}$$

- Beziehung zwischen Durbin-Watson-Statistik und dem empirischen Autokorrelationskoeffizienten:

$$\begin{aligned} dw &= 2 - 2\hat{\rho}\left(\frac{\sum_{t=2}^{T}\widehat{\varepsilon}_{t-1}^2}{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}\right) - \left(\frac{\widehat{\varepsilon}_1^2 + \widehat{\varepsilon}_T^2}{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2}\right) \\ &\approx 2(1-\hat{\rho}), \end{aligned}$$

  wegen

$$\sum_{t=2}^{T}(\widehat{\varepsilon}_t - \widehat{\varepsilon}_{t-1})^2 = \sum_{t=1}^{T}\widehat{\varepsilon}_t^2 - \widehat{\varepsilon}_1^2 - 2\sum_{t=2}^{T}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1} + \sum_{t=2}^{T+1}\widehat{\varepsilon}_{t-1}^2 - \widehat{\varepsilon}_T^2$$

  und für T groß

$$\frac{\sum_{t=2}^{T}\widehat{\varepsilon}_t^2}{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2} \approx 1 \;\text{ und }\; \frac{\widehat{\varepsilon}_1^2 + \widehat{\varepsilon}_T^2}{\sum_{t=1}^{T}\widehat{\varepsilon}_t^2} \approx 0$$

- Alternativer (konsistenter) Schätzer des Autokorrelationskoeffizienten

$$\tilde{\rho} = 1 - 1/2dw$$

- Unter H_0 (d.h. fehlender Autokorrelation) kann man zeigen, dass die Verteilung von dw symmetrisch um 2 ist, wobei
  - Werte in der Nähe von 2 fehlende Autokorrelation,
  - Werte kleiner als 2 positive Autokorrelation,
  - Werte größer als 2 negative Autokorrelation anzeigen.
- Problem: Selbst unter H_0 hängt die Verteilung von dw nicht nur von der Zeitreihenlänge T, der Anzahl der Regressoren K, sondern auch noch von den Werten x_t der Regressoren ab.
  D.h. Tabellierung der kritischen Werte ist nicht möglich.

- **Lösung**: Tabellierung von Ober- d_U und Untergrenzen d_L von dw, die nur noch von T und K abhängen. Der "wahre" kritische Wert d_{crit} liegt zwischen diesen Grenzen, so dass z.B. für eine Irrtumswahrscheinlichkeit von 5% gilt:

$$P(dw < d_L) \leq P(dw \leq d_{crit}) = 0.05 \leq P(dw < d_U).$$

- Ober- und Untergrenzen der kritischen Werte der Durbin-Watson-Statistik (5%-Niveau)

Anzahl der Beobachtungen	Anzahl der Regressoren (inklusive Achsenabschnitt)							
	$K=3$		$K=5$		$K=7$		$K=9$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
$T=25$	1.206	1.550	1.038	1.767	0.868	2.012	0.702	2.280
$T=50$	1.462	1.628	1.378	1.721	1.291	1.822	1.201	1.930
$T=75$	1.571	1.680	1.515	1.739	1.458	1.801	1.399	1.867
$T=100$	1.634	1.715	1.592	1.758	1.550	1.803	1.506	1.850
$T=200$	1.748	1.789	1.728	1.810	1.707	1.831	1.686	1.852

- Mögliche Entscheidungen bei einem Test mit der Alternativhypothese der positiven Autokorrelation:
  1. Wenn $dw < d_L$, dann ist $dw < d_{crit}$ und die Nullhypothese ist abzulehnen.
  2. Wenn $dw > d_U$, dann ist $dw > d_{crit}$ und die Nullhypothese ist nicht abzulehnen.
  3. Wenn $d_L \leq dw \leq d_U$, dann kann dw kleiner oder größer dem wahren, unbekannten kritischen Wert sein und es ist keine Entscheidung möglich.
- *Beachte*: Dieser Indifferenz- oder Inkonklusionsbereich wird mit wachsender Zeitreihenlänge immer kleiner.
- Gelegentlich werden in der Literatur Approximationen der Verteilung der Durbin-Watson-Statistik diskutiert, die sich in der Praxis (speziell: Programmpakete) nicht durchgesetzt haben.
  Prinzipiell ist es bei der heutigen Rechnerleistung auch möglich, für jeden Datensatz der Regressoren die wahre Verteilung zu simulieren.
- *Beachte*: Bei einem Test auf negative Autokorrelation werden wegen der Symmetrie der Verteilung von dw unter H_0 die Grenzen $4 - d_L$ und $4 - d_U$ verwendet.

Fallstudie: Nachfrage nach Eiscreme

- Originaldatensatz von Hildreth & Lu (1960).
- Daten umfassen 30 Monate vom 18.3.1951 bis 11.7.1953.
- Variablen:
  - $cons$: Konsum von Eiscreme pro Kopf (in pints),
  - $income$: durchschnittliches Familieneinkommen pro Woche (in USD),
  - $price$: Preis von Eiscreme (pro pint),
  - $temp$: durchschnittliche Temperatur (in Fahrenheit).

- Verlauf von $cons$, $price$ und $temp/100$ über die Zeit:

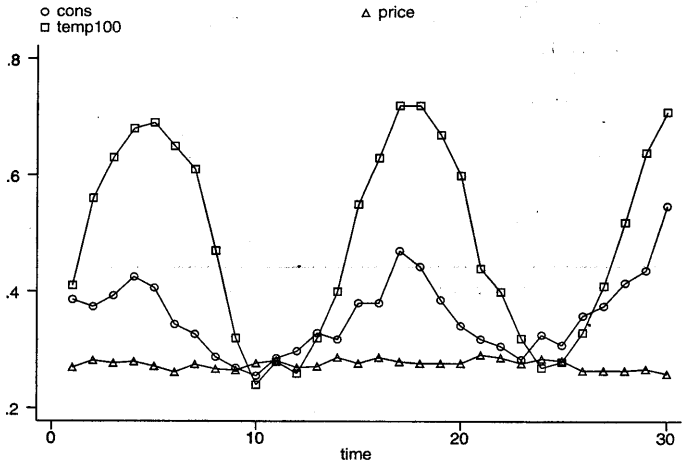


Figure 4.2 Ice cream consumption, price and temperature/100

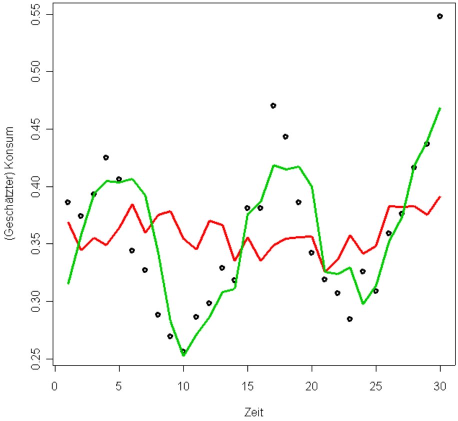
OLS-Schätzergebnisse

Abhängige Variable: *cons*			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	0.197	0.270	0.730
price	-1.044	0.834	-1.252
income	0.00331	0.00117	2.824
temp	0.00345	0.00045	7.762

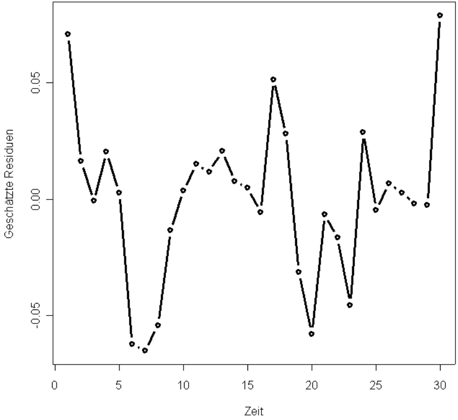
$s = 0.0368,\ R^2 = 0.7190,\ \overline{R}^2 = 0.6866,\ F = 22.175,\ dw = 1.0212$

- Interpretation der Schätzergebnisse:
  1. Die Vorzeichen der geschätzten Koeffizienten sind erwartungsgemäß.
  2. Durbin-Watson-Test: $dw = 1.0212$.
     Für $T = 30$ und $K = 4$ sind $d_L = 1.21$ und $d_U = 1.65$.
     Testentscheidung: Da $dw < d_L$ ist, kann mit einer Irrtumswahrscheinlichkeit von 5% die Nullhypothese fehlender Autokorrelation zugunsten der Alternative positiver Autokorrelation verworfen werden.

- Plot der tatsächlichen (Punkte), den ausgeglichenen Werte von Modell 1 (rote Linie) und Modell 2 mit Berücksichtigung der Temperatur (grüne Linie) für den Konsum von Eiscreme:

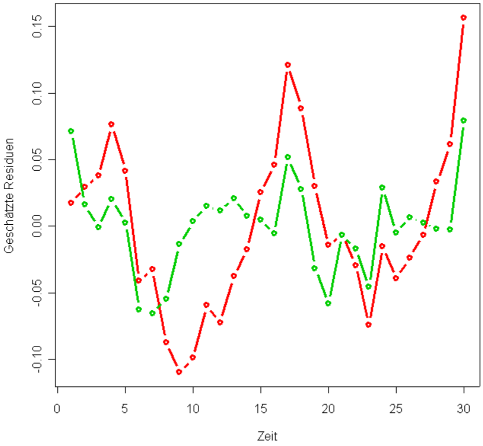


- Ein Plot der OLS-Residuen ($\widehat{\varepsilon}_i$) gegen die Zeit zeigt ausgeprägtes Autokorrelationsverhalten:



Fazit: Positive Regressionsfehler werden von positiven gefolgt. Die Temperaturvariable ist nicht in der Lage, die saisonalen Schwankungen ausreichend zu erfassen.

- Ein Plot der OLS-Residuen von Modell 1 (rote Linie) und Modell 2 (grüne Linie) gegen die Zeit zeigt zwar eine Verbesserung aber nach wie vor ein ausgeprägtes Autokorrelationsverhalten:



- Asymptotische Tests:
    1. Als Schätzwert für ρ erhält man aus den Residuen der OLS-Regression $\hat{\rho} = 0.401$. Das Bestimmtheitsmaß dieser Hilfsregression ist $R^2 = 0.149$.
    2. Asymptotischer Test: $\sqrt{T}\hat{\rho} = 2.19 > 1.96$, d.h. Nullhypothese fehlender Autokorrelation ist wieder auf dem 5%-Niveau zugunsten (jetzt aber positiver oder negativer) Autokorrelation zu verworfen.
    3. Asymptotischer Breusch-Godfrey-Test: $(T-1)R^2 = 4.32 > 3.84$, was zu demselben Resultat führt. 3.84 ist das 95%-Quantil einer χ^2-Verteilung mit einem Freiheitsgrad.
- Fazit: Es ist von Autokorrelation auszugehen, so dass die OLS-Schätzer nicht mehr effizient und die Standardfehler verzerrt sind.

- EGLS-Schätzergebnisse nach CORC (voll-iteriert) für Autokorrelation 1. Ordnung:

Abhängige Variable: $cons$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	0.157	0.300	0.524
$price$	-0.892	0.830	-1.076
$income$	0.00320	0.00159	2.005
$temp$	0.00356	0.00061	5.800

$$s = 0.0326^*,\ R^2 = 0.7961^*,\ \overline{R}^2 = 0.7621^*,\ F = 23.419,\ dw = 1.5486^*$$

Beachte: Sämtliche mit * markierten Ergebnisse beziehen sich auf das transformierte Modell und sind nicht vergleichbar mit den Werten der OLS-Regression.

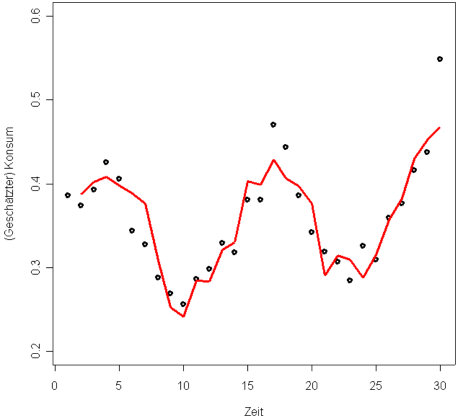
- Interpretation der Schätzergebnisse:
  1. Durbin-Watson-Statistik zeigt im transformierten Modell keine Autokorrelation mehr an.
  2. Einkommen und Temperatur sind weiterhin die bestimmenden Einflussfaktoren. Der Preis ist von untergeordneter Bedeutung.
- Wenn Autokorrelation ein Zeichen für Fehlspezifikation ist, sollte die Modellspezifikation erweitert werden.
  Dies geschieht durch Aufnahme der um eine Periode verzögerten Temperatur.

- OLS-Schätzergebnisse mit verzögerter Temperatur als Regressor:

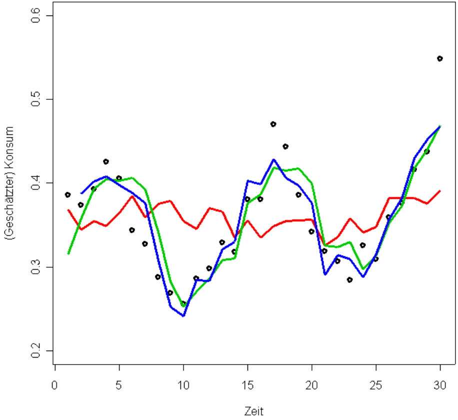
Abhängige Variable: *cons*			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	0.189	0.232	0.816
$price$	-0.838	0.688	-1.218
$income$	0.00287	0.00105	2.722
$temp$	0.00533	0.00067	7.953
$temp_{t-1}$	-0.00220	0.00073	-3.016
$s = 0.0299,\ R^2 = 0.8285,\ \overline{R}^2 = 0.7999,\ F = 28.979,\ dw = 1.5822$			

- Interpretation der Schätzergebnisse:
  1. Die verzögerte Temperatur besitzt einen signifikant negativen Einfluss. Wenn die Temperatur steigt, erhöhen sich c.p. die Ausgaben für Speiseeis. Da dieses nicht sofort voll konsumiert wird, sinken die Ausgaben in der nächsten Periode.
  2. Der Wert der Durbin-Watson-Statistik (1.5822) liegt nun im Inkonklusionsbereich $(1.14, 1.74)$.

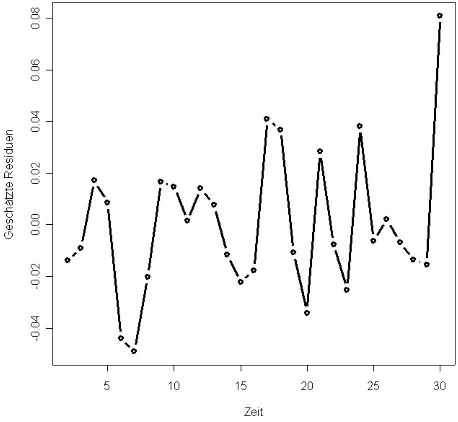
- Plot der tatsächlichen (Punkte) und ausgeglichenen Werte (rote Linie) für den Konsum von Eiscreme:



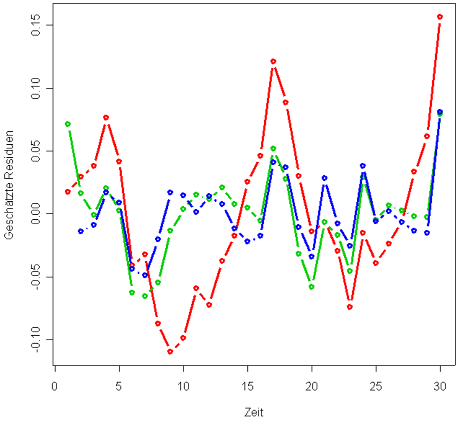
- Plot der tatsächlichen (Punkte) und den Prognosewerten von Modell 1 (rot), Modell 2 (grün) und Modell 3 mit zusätzlich verzögerter Temperatur als exogene Variable (blau) für den Konsum von Eiscreme:



- Ein Plot der OLS-Residuen ($\widehat{\varepsilon}_i$) gegen die Zeit zeigt ein geringes Autokorrelationsverhalten:



- Ein Plot der OLS-Residuen von Modell 1 (rot), Modell 2 (grün) und Modell 3 (blau) zeigt eine deutliche Abnahme der Autokorrelation:



Autokorrelation höherer Ordnung

- In (statischen) Modellen mit trendbehafteten ökonomischen Daten ist häufig Autokorrelation 1. Ordnung festzustellen.
- Autokorrelation höherer (4. oder 12.) Ordnung ist von Bedeutung, wenn Quartals- oder Monatsdaten betrachtet werden und ein Saisoneffekt vorliegt.
- Autokorrelation 4. Ordnung:

$$\varepsilon_t = \gamma_1 \varepsilon_{t-1} + \ldots + \gamma_4 \varepsilon_{t-4} + v_t$$

mit dem Spezialfall

$$\varepsilon_t = \gamma \varepsilon_{t-4} + v_t$$

Autokorrelation höherer Ordnung

- Unter der Bedingung der Stationarität ist das Vorgehen analog zur EGLS-Schätzung im Falle Autokorrelation 1. Ordnung.
- OLS-Regression und Schätzung der γ_i aus einer OLS-Regression der OLS-Residuen auf ihre Vergangenheit (bis Lag 4). Anschließend Transformation des Modells und OLS-Regression des transformierten Modells.

Was ist zu tun bei Autokorrelation?

Fehlspezifikation

- Hinweis auf Fehlspezifikation macht Änderung des Modells statt Änderung des Schätzers nötig.
- 3 Typen von Fehlspezifikation, die zur Autokorrelation führen können:
  1. Fehlspezifikation der funktionalen Form,
  2. weggelassene Variablen,
  3. dynamische Fehlspezifikation.
- *Beachte*: Selbst wenn OLS-Schätzer konsistent ist, kann GLS-Schätzer inkonsistent sein (siehe Fallstudie).

- **Fehlspezifikation der funktionalen Form**:
  Als wahres Modell wird
  $$y_t = \beta_1 + \beta_2 \log x_t + \varepsilon_t$$
  betrachtet.
  Besitzt x_t einen positiven Zeittrend und wird eine Gerade angepasst, so werden negative Regressionsfehler zu Beginn und am Ende des Beobachtungszeitraums und positive Fehler in der Mitte auftreten, so dass eine zeitliche Abhängigkeit besteht, obwohl im wahren Modell gar keine vorhanden ist.

- Simulationsexperiment: Es wird $x_t = t$ gesetzt und $y_t = 0.5 \log t$ um eine zufällige Störung ε_t überlagert. Dadurch ergibt sich eine Punktwolke.
  Durch diese Punktwolke wird eine Gerade gelegt, die autokorrelierte Regressionsfehler produziert.

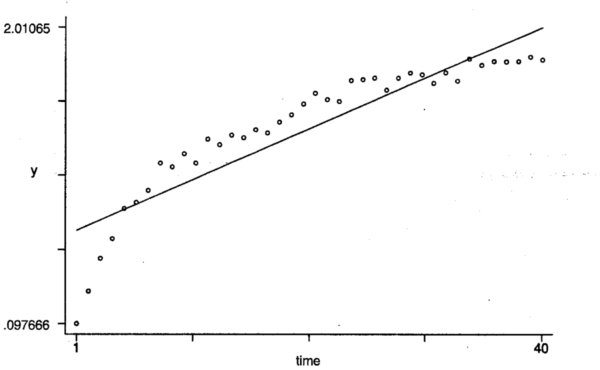


Figure 4.4 Actual and fitted values when true model is $y_t = 0.5 \log t + \varepsilon_t$

- **Weggelassene Variablen**: Besitzt die erklärte Variable einen Zeittrend und verfügt keine der erklärenden Variablen über einen entsprechenden Zeittrend, muss der Fehlerterm autokorreliert sein.
  Weggelassene Variable: Zeittrend oder Saisondummies.
- **Dynamische Fehlspezifikation**:
  Wenn das (statische) Modell

$$y_t = x_t'\beta + \varepsilon_t \quad \text{mit } \varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

betrachtet wird, dann ist wegen $\varepsilon_{t-1} = y_{t-1} - x_{t-1}'\beta$

$$\begin{aligned} y_t &= x_t'\beta + \rho(y_{t-1} - x_{t-1}'\beta) + v_t \\ &= \rho y_{t-1} + (x_t - \rho x_{t-1})'\beta + v_t \end{aligned}$$

ein dynamisches Regressionsmodell mit nicht-autokorrelierter Störgröße v_t.

- Mögliche Ursache der Autokorrelation: Verzicht auf die zeitverzögerten Regressoren. Behebung der Autokorrelation: Dynamische Spezifikation.
- Beachte: In dynamischen Modellen mit zeitverzögerter erklärter Variable als Regressor ist der Durbin-Watson-Test ungeeignet.

HAC-Standardfehler für OLS-Schätzer

- 2 Strategien zur Berücksichtigung von Heteroskedastie und Autokorrelation:
  1. GLS- (bzw. EGLS-) Schätzung,
  2. OLS-Schätzung mit angepassten Standardfehlern (z.B. wenn GLS inkonsistent).
- Wenn die Regressoren und der Störterm unkorreliert und die Störterme ab einem bestimmten Lag H nicht mehr autokorreliert sind, d.h.

$$E(x_t \varepsilon_t) = 0, \ ; E(\varepsilon_t \varepsilon_{t-s}) = 0 \text{ für } s = H, H+1, \ldots,$$

dann ist der OLS-Schätzer konsistent mit Varianz-Kovarianzmatrix

$$\widehat{V}^*(\widehat{\beta}) = (X'X)^{-1} T S^* (X'X)^{-1}$$

mit

$$S^* = \underbrace{\frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t x_t'}_{Heteroskedastie} + \underbrace{\frac{1}{T}\sum_{j=1}^{H-1} w_j \sum_{s=j+1}^{T} e_s e_{s-j}(x_s x_{s-j}' + x_{s-j} x_s')}_{Autokorrelation}$$

- Beachte: Für $w_j = 0$ ergibt sich die White-Varianz-Kovarianzmatrix.
- Problem: $w_j = 1$ führt zu einem u.U. nicht positiv definiten Schätzer der Varianz-Kovarianzmatrix.
  Lösung: Bartlett-Gewichte: $w_j = 1 - \frac{j}{H}$ nach Newey & West, die mit zunehmendem Lag j linear abnehmen.
- Die Wurzel der Hauptdiagonalelemente von $\hat{V}^*(b)$ unter Verwendung der Bartlett-Gewichte liefert die **HAC**- (Heteroskedastie- und autokorrelationskonsistenten) oder kurz **Newey-West**-Standardfehler des OLS-Schätzers.

- HAC-Standardfehler werden gelegentlich auch verwendet, wenn H (wie bei einem autoregressiven Fehlerterm) nicht endlich ist.
  Dies lässt sich bestenfalls über ein asymptotisches Argument rechtfertigen,wenn H mit T wächst.
- Intuition der Schätzformel:
  Zu schätzen ist die Varianz von

$$1/T \sum_{t=1} \varepsilon_t x_t$$

in der Formel für die Varianz-Kovarianzmatrix des OLS-Schätzers bei allgemeiner Varianz-Kovarianzmatrix des Fehlerterms:

$$\begin{aligned} V(\widehat{\beta}) &= (X'X)^{-1} X' V(\varepsilon) X (X'X)^{-1} \\ &= \left(1/T \sum_{t=1}^{T} x_t x_t'\right)^{-1} V(1/T \sum_{t=1}^{T} \varepsilon_t x_t) \left(1/T \sum_{t=1}^{T} x_t x_t'\right)^{-1} \end{aligned}$$

- Problem: Da nur eine Beobachtung pro Zeitpunkt vorliegt, ist

$$1/T \sum_{s=1}^{T} \sum_{t=1}^{T} e_s e_t x_t x_s'$$

kein konsistenter Schätzer für $V(1/T \sum_{t=1}^{T} \varepsilon_t x_t)$. Die Kovarianz zwischen z.B. $x_1 \varepsilon_1$ und $x_T \varepsilon_T$ müsste aus einem Datenpunkt geschätzt werden.
Lösung: Annahme, dass die Autokorrelationen für Lags $|s-t| \geq H$ gleich 0 sind, reduziert die Anzahl der zu schätzenden Kovarianzen.

Gliederung Kapitel 5

1. Fälle, in denen der OLS-Schätzer nicht mehr zu retten ist
2. Instrumentvariablenschätzung (IV)
3. Fallstudie: Returns to schooling

Korrelation der Störterme mit den erklärenden Variablen

- $E(\varepsilon_t x_t) = 0$ ist eine notwendige Bedingung für die Unverzerrtheit und Konsistenz des OLS-Schätzers.
- $E(\varepsilon_t x_t) = 0$ heißt, daß Störterm und erklärende Variablen nicht kontemporär korreliert sind.
- Statistische und ökonomische Gründe, warum Störterm und erklärende Variablen kontemporär korreliert sein können, sind z.B.
  1. Autokorrelation und verzögert abhängige Variable,
  2. Meßfehler in den Regressoren,
  3. Simultanität oder Endogenität der Regressoren.

Autokorrelation und verzögert abhängige Variable

- Modell:
$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t$$
Wenn $E(\varepsilon_t x_t) = E(\varepsilon_t y_{t-1}) = 0$ sind, ist der OLS-Schätzer konsistent.
- Autokorrelation 1. Ordnung für den Störterm:
$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$$
mit $v_t \sim$ IID$(0, \sigma_v^2)$ führt zu
$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \rho\varepsilon_{t-1} + v_t$$
Wegen
$$y_{t-1} = \beta_1 + \beta_2 x_{t-1} + \beta_3 y_{t-2} + \varepsilon_{t-1}$$
sind y_{t-1} und ε_{t-1} und damit y_{t-1} und ε_t korreliert.

- D.h.: Für $\rho \neq 0$ ist der OLS-Schätzer verzerrt und inkonsistent.
- Alternative Argumentation: Bei Autokorrelation 1. Ordnung gilt

$$E(y_t|x_t, y_{t-1}) = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + E(\varepsilon_t|x_t, y_{t-1})$$

Da OLS immer dann konsistent ist, wenn ein bedingter Erwartungswert zu schätzen ist und $\widehat{\beta}_1 + \widehat{\beta}_2 x_t + \widehat{\beta}_3 y_{t-1}$ kein Schätzer von

$$E(y_t|x_t, y_{t-1})$$

ist, folgt notwendig die Inkonsistenz des OLS-Schätzers.

Meßfehler

- Modell:

$$y_t = \beta_1 + \beta_2 w_t + v_t$$

mit der Meßfehlergleichung

$$x_t = w_t + u_t,$$

wobei x_t der beobachtete und w_t der wahre Wert des Regressors ist.

- Annahmen:
  1. $u_t \sim$ IID$(0, \sigma_u^2)$,
  2. u_t, v_t sind stochastisch unabhängig,
  3. u_t und w_t sind stochastisch unabhängig.

- Eingesetzt ist

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t - \beta_2 u_t + v_t \\ &= \beta_1 + \beta_2 x_t + \varepsilon_t, \end{aligned}$$

  so daß x_t und ε_t via u_t korreliert sind und damit der OLS-Schätzer inkonsistent ist.
- *Beachte*: Wenn $\beta_2 > 0$ ist, sind x_t und ε_t wegen $\varepsilon_t = v_t - \beta_2 u_t$ negativ korreliert.

- Direkter Nachweis der Inkonsistenz des OLS-Schätzers für β_2:

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$$

Für y_t eingesetzt ergibt sich

$$\widehat{\beta}_2 = \beta_2 + \frac{(1/T)\sum_{t=1}^{T}(x_t - \overline{x})(\varepsilon_t - \overline{\varepsilon})}{(1/T)\sum_{t=1}^{T}(x_t - \overline{x})^2},$$

so daß

$$\text{plim}\ \widehat{\beta}_2 = \beta_2 + \frac{\text{plim}(1/T\sum_{t=1}^{T}(x_t - \overline{x})(\varepsilon_t - \overline{\varepsilon}))}{\text{plim}(1/T\sum_{t=1}^{T}(x_t - \overline{x})^2)} = \beta_2 + \frac{E(x_t\varepsilon_t)}{V(x_t)}$$

Mit

$$E(x_t \varepsilon_t) = E((w_t + u_t)(v_t - \beta_2 u_t)) = -\beta_2 \sigma_u^2$$

und

$$V(x_t) = \sigma_w^2 + \sigma_u^2$$

ist

$$\text{plim } \widehat{\beta}_2 = \beta_2 \left(1 - \frac{\sigma_u^2}{\sigma_w^2 + \sigma_u^2}\right)$$

- D.h.
  - $\widehat{\beta}_2$ ist konsistent, falls $\sigma_u^2 = 0$ ist, womit kein Meßfehler vorliegt.
  - $\widehat{\beta}_2$ ist asymptotisch in Richtung 0 verzerrt.
  - Die Verzerrung ist umso größer, je größer das "noise-to-signal"-Verhältnis σ_u^2/σ_w^2 ist.

- Beachte: Auch der OLS-Schätzer für b_1 ist inkonsistent, da

$$\text{plim}(\widehat{\beta}_1 - \beta_1) = -\text{plim}(\widehat{\beta}_2 - \beta_2)E(x_t)$$

wegen

$$\widehat{\beta}_1 = \overline{y} - \widehat{\beta}_2\overline{x}$$

und

$$\beta_1 = E(y_t - \beta_2 x_t)$$

- Wiederum korrespondiert das interessierende Modell nicht mit dem bedingten Erwartungswert, da

$$E(y_t|x_t) = \beta_1 + \beta_2 x_t - \beta_2 E(u_t|x_t)$$

Simultanität im Keynesianischen Modell

- Simultanes makroökonomisches Gleichungssystem in struktureller Form:
  Konsumgleichung:

  $$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$$

  Identität für die Einkommensgleichung

  $$Y_t = C_t + I_t$$

- Annahme: I_t und ε_t stochastisch unabhängig, d.h. I_t ist exogen.
- Problem: Y_t und ε_t sind korreliert.

- Reduzierte Form:

$$\begin{aligned} C_t &= \beta_1 + \beta_2(C_t + I_t) + \varepsilon_t \\ &= \frac{\beta_1}{1-\beta_2} + \frac{\beta_2}{1-\beta_2} I_t + \frac{1}{1-\beta_2}\varepsilon_t, \\ Y_t &= \frac{\beta_1}{1-\beta_2} + \frac{1}{1-\beta_2} I_t + \frac{1}{1-\beta_2}\varepsilon_t \end{aligned}$$

Damit ist

$$\mathrm{cov}(Y_t, \varepsilon_t) = \frac{1}{1-\beta_2}\mathrm{cov}(I_t, \varepsilon_t) + \frac{1}{1-\beta_2} V(\varepsilon_t) = \frac{\sigma^2}{1-\beta_2} \neq 0$$

Wiederum gilt

$$\text{plim } \widehat{\beta}_2 = \beta_2 + \frac{\text{cov}(Y_t, \varepsilon_t)}{V(Y_t)},$$

wobei

$$V(Y_t) = \frac{1}{(1-\beta_2)^2}(V(I_t) + \sigma^2),$$

so daß

$$\text{plim } \widehat{\beta}_2 = \beta_2 + (1-\beta_2)\frac{\sigma^2}{V(I_t) + \sigma^2}$$

- D.h. wenn $0 < \beta_2 < 1$ und $\sigma^2 > 0$ überschätzt der OLS-Schätzer die wahre marginale Konsumneigung β_2.

Vorbemerkungen

- Spezialfall:
$$y_i = x'_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$$
- Notwendig für die Konsistenz des OLS-Schätzers:
$$E(\varepsilon_i x_{i1}) = 0 \text{ und } E(\varepsilon_i x_{i2}) = 0$$
- Endogenitätsproblem: "..., when we interpret the model as a conditional expectation, then ceteris paribus condition only refers to the included variables, while for a causal interpretation it also includes the observables (omitted variables) in the error term."
- Definition: x_{2i} heißt **endogen**, wenn $E(\varepsilon_i x_{2i}) \neq 0$.
- Aufgabe: Lösung der Inkonsistenz der OLS-Schätzung durch Modifikation der Schätzmethode.

Schätzung mit einem einzigen endogenen Regressor und einem Instrument

- Wenn
$$E(\varepsilon_i x_{2i}) = E((y_i - x'_{1i}\beta_1 - x_{2i}\beta_2)x_{2i}) \neq 0,$$
suche sog. Instrumentvariablen z_{2i} mit
$$E(\varepsilon_i z_{2i}) = E((y_i - x'_{1i}\beta_1 - x_{2i}\beta_2)z_{2i}) = 0$$
- Der OLS-Schätzer ist Lösung von
$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(y_i - x'_{1i}\widehat{\beta}_1 + x_{2i}\widehat{\beta}_2)x_{1i} &= 0 \\ \frac{1}{n}\sum_{i=1}^{n}(y_i - x'_{1i}\widehat{\beta}_1 + x_{2i}\widehat{\beta}_2)x_{2i} &= 0 \end{aligned}$$

Dies sind die zu

$$E(\varepsilon_i x_{1i}) = 0 \text{ und } E(\varepsilon_i x_{2i}) = 0$$

korrespondierenden Bedingungen für die Stichprobenmomente.

- Der **Instrumentvariablenschätzer** ergibt sich als Lösung von

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(y_i - x_{1i}'\hat{\beta}_{1,IV} + x_{2i}\hat{\beta}_{2,IV})x_{1i} &= 0 \\ \frac{1}{n}\sum_{i=1}^{n}(y_i - x_{1i}'\hat{\beta}_{1,IV} + x_{2i}\hat{\beta}_{2,IV})z_{2i} &= 0, \end{aligned}$$

was zu

$$E(\varepsilon_i x_{1i}) = 0 \text{ und } E(\varepsilon_i z_{2i}) = 0$$

gehört.

Die Auflösung nach $\hat{\beta}_{IV} = (\hat{\beta}'_{1,IV}, \hat{\beta}_{2,IV})'$ ergibt

$$\hat{\beta}_{IV} = \left(\sum_{i=1}^{n} z_i x_i' \right)^{-1} \sum_{i=1}^{n} z_i y_i$$

mit $x_i = (x'_{1i}, x_{2i})'$ und $z_i = (x'_{1i}, z_{2i})$.

- In Matrixschreibweise ist

$$\hat{\beta}_{IV} = (Z'X)^{-1} Z'y$$

- Unter der Bedingung $E(\varepsilon_i z_i) = 0$ und mit

$$\Sigma_{zx} \equiv \text{plim} \frac{1}{n} \sum_{i=1}^{n} z_i x_i'$$

finite und invertierbare $(K \times K)$-Matrix, ist der Instrumentvariablenschätzer konsistent.

- Dies setzt voraus, daß die Instrumente z_i korreliert mit x_i sind. z_{2i} darf jedoch nicht Linearkombination von x_{1i} sein.
- Asymptotische Normalverteilung des Instrumentvariablenschätzers ergibt sich für $\varepsilon_i \sim IID(0, \sigma^2)$, ε_i, z_i stochastisch unabhängig und

$$\Sigma_{zz} = \text{plim } \frac{1}{n} \sum_{i=1}^{n} z_i z_i'$$

symmetrische, invertierbare $(K \times K)$-Matrix als

$$\hat{\beta}_{IV} \overset{approx}{\sim} \mathcal{N}(\beta, \frac{1}{n} \sigma^2 (\Sigma_{zx}' \Sigma_{zz}^{-1} \Sigma_{zx})^{-1}).$$

- In endlichen Stichproben lassen sich mit

$$\hat{V}(\hat{\beta}_{IV}) = \hat{\sigma}^2 \left(\left(\sum_{i=1}^{n} x_i z_i' \right) \left(\sum_{i=1}^{n} z_i z_i' \right)^{-1} \left(\sum_{i=1}^{n} z_i x_i' \right) \right)^{-1}$$

$V(\hat{\beta}_{IV})$ und mit

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i' \hat{\beta}_{IV})^2$$

σ^2 konsistent schätzen.

- Problem der Wahl der Instrumente: Bedingung $E(\varepsilon_i z_i) = 0$ dient der Identifikation des bedingten Erwartungswertes und ist nicht mittels Daten testbar.

Zurück zum Keynesianischen Modell

- Lösung des Problems der Instrumentenwahl:
  Jede exogene Variable des Gleichungssystems, die nicht in der interessierenden Gleichung enthalten ist, kann als Instrument benutzt werden.
- Beispiel des Keynesianischen Modells:

$$\begin{aligned} C_t &= \beta_1 + \beta_2 Y_t + \varepsilon_t \\ Y_t &= C_t + I_t \end{aligned}$$

  D.h. I_t ist als Instrument geeignet.
- Instrumentvariablenschätzer für $\beta = (\beta_1, \beta_2)'$:

$$\hat{\beta}_{IV} = \left(\sum_{t=1}^{T} \begin{pmatrix} 1 \\ I_t \end{pmatrix} (1, Y_t) \right)^{-1} \sum_{t=1}^{T} \begin{pmatrix} 1 \\ I_t \end{pmatrix} C_t$$

D.h.

$$\hat{\beta}_{2,IV} = \frac{\sum_{t=1}^{T}(I_t - \overline{I})(C_t - \overline{C})}{\sum_{t=1}^{T}(I_t - \overline{I})(Y_t - \overline{Y})}$$

- Wegen

$$\text{cov}(C_t, I_t) = \beta_2 \text{cov}(Y_t, I_t) + \text{cov}(\varepsilon_t, I_t)$$

und der Annahme der Exogenität $E(\varepsilon_t I_t) = 0$ folgt

$$\beta_2 = \frac{\text{cov}(C_t, I_t)}{\text{cov}(Y_t, I_t)},$$

so daß $\hat{\beta}_{2,IV}$ das Stichprobenäquivalent von β_2 unter der Annahme $E(\varepsilon_t I_t) = 0$ ist.

Zurück zum Problem des Messfehlers

- Modell:

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \text{ mit } \varepsilon_t = v_t - \beta_2 u_t$$

- *Beachte*: Instrument muß mit x_t korrelieren, darf aber nicht mit u_t und v_t und damit nicht mit ε_t korrelieren.
  Fazit: "Mainly due to the problem of finding suitable instruments, the problem of measurement errors is often ignored in empirical work."

Multiple endogene Regressoren

- Anzahl der im Gleichungssystem vorhandenen exogenen Variablen (potentielle Instrumente) bestimmt die Zahl der endogenen Regressoren.
- Ist die Zahl der exogenen größer als die Zahl der endogenen Regressoren, so stehen alternative Instrumente zur Auswahl.
  In diesem Fall können aber auch alle Instrumente gemeinsam verwendet werden (siehe verallgemeinerter Instrumentvariablenschätzer).

Returns to Schooling

- Offensichtlich verdienen Personen mit "mehr" Ausbildung im Durchschnitt auch mehr Geld.
  Unklar ist aber die **Kausalitätsrichtung**: Verdienen sie mehr Geld wegen der Ausbildung oder haben Personen, mit dem Potential für einen höheren Lohn, einen Hang zu längerer Ausbildung?
  Wenn die zweite Kausalrichtung richtig ist, mißt der OLS-Schätzwert für den Einfluß der Ausbildung auf die Entlohnung nur unbeobachtete Merkmale der arbeitenden Personen, so daß sich ein exogener Schock auf die Ausbildungsdauer nicht auf den Lohn auswirkt.

- "Human capital earning's function":

$$w_i = \beta_1 + \beta_2 S_i + \beta_3 E_i + \beta_4 E_i^2 + \varepsilon_i$$

  mit
  - w_i: logarithmierter individueller Lohn,
  - S_i: Anzahl der Schuljahre,
  - E_i: Anzahl an Jahren der Berufserfahrung.

  Da die aktuelle Berufserfahrung nicht bekannt ist, wird manchmal die potentielle Berufserfahrung $age_i - S_i - 6$ betrachtet.
  Zusätzlich werden noch Variablen wie Region, Geschlecht etc. aufgenommen, "für die" man kontrollieren möchte.
  Schließlich darf der Einfluß Schulausbildung von Individuum zu Individuum schwanken.

- Spezifikation:

$$\begin{aligned} w_i &= z_i'\beta + \gamma_i S_i + u_i \\ &= z_i'\beta + \gamma S_i + \underbrace{u_i + (\gamma_i - \gamma)S_i}_{\varepsilon_i}, \end{aligned}$$

  wobei z_i alle Regressoren außer der Ausbildungsdauer S_i umfaßt.
- Annahme: $E(\varepsilon_i z_i) = 0$
- γ gibt c.p. den durchschnittlichen Ertrag eines zusätzlichen Ausbildungsjahres an.

- Modell der reduzierten Form für S_i:

$$S_i = z_i'\pi + v_i$$

  mit $E(v_i z_i) = 0$
  Beachte: Dieses Modell stellt lediglich eine lineare Approximation von S_i durch eine Linearkombination der in z_i enthaltenen Variablen dar und muß keine ökonomische Interpretation besitzen.

- Konsistente OLS-Schätzung von β und γ ist nur dann möglich, wenn $E(\varepsilon_i S_i) = 0$ ist. Dies führt aber $E(\varepsilon_i v_i) = 0$ nach sich.
  In diesem Fall gibt es keine unbeobachteten Merkmale, die gemeinsam die Schulwahl (d.h. die Schuldauer) und die späteren Verdienste beeinflussen.

- Gründe, warum ε_i und v_i korreliert sein können:
  1. "Ability bias": Wenn Personen über eine allgemeine unbeobachtete Disposition verfügen, sowohl länger zur Schule zu gehen als auch höhere Einkommen zu verdienen, so sind ε_i und v_i positiv korreliert und der OLS-Schätzer für γ nach oben verzerrt.
  2. ε_i und v_i können aufgrund von Meßfehlern in S_i negativ korreliert sein, was zu einer Verzerrung nach unten führt.
  3. Wenn γ_i (individueller c.p.-Effekt der Schulbildung auf das individuelle Einkommen) höher ist für Personen mit niedrigem Niveau der Schulbildung, dann ist $(\gamma_i - \gamma)S_i$ negativ korreliert mit S_i und der OLS-Schätzer wiederum negativ verzerrt.
- Problem: Da alle verfügbaren Variablen in z_i enthalten sind, ist kein Instrument für S_i verfügbar.

- Problem: Es stehen mit

$$E(\varepsilon_i z_i) = E((w_i - z_i'\beta - \gamma S_i)z_i) = 0$$

  eine Momentenbedingungen weniger zur Identifikation von β und γ zur Verfügung als Parameter zu schätzen sind.

- Mögliche Lösung: Wenn z_i mindestens eine Variable z_{2i} enthält, die Ausbildung aber nicht die Löhne beeinflußt, kann diese aus z_i als Regressor herausgenommen und als Instrument für S_i verwendet werden.
  Die Anzahl der Momentenbedingungen ändert sich dadurch nicht, aber die Anzahl der zu schätzenden Parameter verringert sich um 1.

- Problem in der Arbeitsökonomie: Welche Variable kann als Instrument verwendet werden?
  Ideen:
  1. Als Instrument kann eine Variable dienen, die die Kosten der Ausbildung, aber nicht den potentiellen Verdienst betrifft.
     Card (1995) verwendet als Instrument die Tatsache, daß sich eine "High School" in unmittelbarer Nähe befindet.
  2. Eine andere Variable bezieht sich auf den familiären Hintergrund, wie das Ausbildungsniveau der Eltern.
  3. Institutionelle Faktoren des Schulsystems können auch als Instrument in Frage kommen.
  4. Angrist & Krueger (1991) haben festgestellt, daß das Quartal, in dem die Individuen geboren wurden, einen Einfluß auf das Ausbildungsniveau besitzt. Früh im Jahr geborene Personen neigen zu geringerer Schuldauer. Das Geburtsquartal könnte als Instrument verwendet werden.

- Datensatz: 3010 Männer des "US National Longitudinal Survey of Young Men".
  Spezifikum: Paneldatensatz, in dem seit 1966 eine Gruppe von Personen im Alter von 14 bis 24 Jahre über eine Reihe von Jahren regelmäßig befragt wurde.
  Hier wird die Welle des Jahres 1976 verwendet.
- Deskriptive Charakteristiken:
  Durchschnittliche Schuldauer: ca. 13 Jahre,
  Maximale Schuldauer: 18 Jahre,
  Durchschnittliche Berufserfahrung: 8.86 Jahre (Personen sind zwischen 24 und 34 Jahre alt),
  Durchschnittlicher Stundenlohn: 5.77 USD.
- Zusätzliche Variablen:
  age: Lebensalter,
  $black = 1$: schwarze Hautfarbe,
  $smsa = 1$: Großstadt,
  $south = 1$: Süden der USA,
  $lived\ near\ college = 1$: Wohnort nahe College.

Ergebnisse einer OLS-Schätzung der Einkommensfunktion

Abhängige Variable: $\log(wage)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	4.7337	0.0676	70.022
$schooling$	0.0740	0.0035	21.113
$exper$	0.0836	0.0066	12.575
$exper^2$	-0.0022	0.0003	-7.050
$black$	-0.1896	0.0176	-10.758
$smsa$	0.1614	0.0156	10.365
$south$	-0.1249	0.0151	-8.259

$s = 0.374,\ R^2 = 0.2905,\ \overline{R}^2 = 0.2891,\ F = 204.93$

- Problem der OLS-Schätzung: Per Konstruktion 3 endogene Regressoren: $schooling$, $exper$, $exper^2$.
  Alter ist als exogen anzusehen, so daß age und age^2 als Instrumente für $exper$ und $exper^2$ dienen können.
  Vorschlag für Instrument für $schooling$: $lived\ near\ college$.
- Wenn $lived\ near\ college$ ein geeignetes Instrument sein soll, muß diese Variable mit der Variablen $schooling$ (gegeben die anderen exogenen Variablen) gut korrelieren.
  Überprüfung durch Schätzung der reduzierten Form für die Variable $schooling$.

OLS-Schätzung der reduzierten Form für die Variable $schooling$

Abhängige Variable: $schooling$

Variable	Schätzwert	Standardfehler	t-Wert
Konstante	-1.8695	4.2984	-0.435
age	1.0614	0.3014	3.522
age^2	-0.0188	0.0052	-3.386
$black$	-1.4684	0.1154	-12.719
$smsa$	0.8354	0.1093	7.647
$south$	-0.4597	0.1024	-4.488
$lived\ near\ college$	0.3471	0.1070	3.244

$s = 2.5158,\ R^2 = 0.1185,\ \overline{R}^2 = 0.1168,\ F = 67.29$

- Fazit:
  Das Instrument hat ein signifikanten Einfluß.
  D.h. c.p. hat eine Person, die in der Nähe eines College lebt, im Durchschnitt eine um 0.35 Jahre längere Ausbildungsdauer.
- Problem des Tests, ob das Instrument keine Linearkombination der anderen Modellvariablen ist, läßt sich nicht lösen, da eine Bedingung dafür ist, daß es unkorreliert mit dem Störterm der Lohngleichung ist. Dann wären aber β und γ mit OLS konsistent schätzbar.
- **Lösung**: Da es keinen formalen Test für ein Instrument gibt, wird auf ökonomische statt statistische Gründe vertraut.

Instrumentvariablenschätzung der Lohngleichung

Abhängige Variable: $\log(wage)$			
Variable	Schätzwert	Standardfehler	t-Wert
Konstante	4.0656	0.6085	6.682
$schooling$	0.1329	0.0514	2.588
$exper$	0.0560	0.0260	2.153
$exper^2$	-0.0008	0.0013	-0.594
$black$	-0.1031	0.0774	-1.333
$smsa$	0.1080	0.0050	2.171
$south$	-0.0982	0.0288	-3.413

Instrumente: age, age^2, $lived\ near\ college$
für: $exper$, $exper^2$, $schooling$

Interpretation der Schätzergebnisse

1. Die erwarteten Erträge der Schulbildung sind c.p. 13%.
   Der Standardfehler der Schätzung ist allerdings relativ groß (verglichen mit der OLS-Schätzung) und der t-Wert nicht allzu deutlich über den Signifikanzschranken.
2. Ursache für die hohen Standardfehler ist, daß die Instrumente mit dem endogenen Regressor $wage$ relativ schwach korrelieren.
   Dies zeigt sich auch am niedrigen R^2 von 0.1185 der OLS-Schätzung der reduzierten Form.
3. Der IV-Schätzer ist im allg. weniger effizient als der OLS-Schätzer; dafür ist dieser u.U. verzerrt und inkonsistent, womit die Varianz als Effizienzmaß unbrauchbar wird.
4. Für die IV-Schätzung wurde kein Wert für das Bestimmtheitsmaß ausgewiesen.
   Wenn nicht OLS geschätzt wird, ist der Nutzen des Bestimmtheitsmaßes fragwürdig.
   Die IV-Schätzung hat ohnehin nicht das Ziel einer guten Anpassung. Das Ziel ist die Konsistenz der Schätzung.

6. Die Brauchbarkeit der Instrumente (speziell: *lived near college*) ist diskussionswürdig.
   So können sich Familien, die eine höhere Schulbildung für ihre Kinder anstreben bewußt für einen Wohnort in der Nähe eines College entscheiden.
7. Der IV-Schätzwert für *schooling* ist deutlich größer als der OLS-Schätzwert. Diese Unterschätzung durch OLS steht im Widerspruch zum "ability bias", ist aber konform mit der Meßfehlerhypothese und der Heterogenität der indivduellen Ausbildungseffekte.