UNIVERSITÄT
BIELEFELD
Fakultät für
Wirtschaftswissenschaften

Einführung Mikroökonometrie

310208 (VO)

Dietmar Bauer

Organisatorisches

Der Hauptteil dieser Vorlesung orientiert sich am Buch:

Jeff Wooldridge: Introductory Econometrics (fifth edition), Cengage Learning, 2013.

Die Theorie wird mit Datenbeispielen veranschaulicht. Dazu wird R verwendet.

- Eine sehr gute Einführung in R ist John Verzani: SimpleR.
- R ist frei erhältlich: CRAN project.
- Eine gute Oberfläche für R bietet das R-Studio: R-Studio Download

Ein Wort zu R: Empirie lebt davon, dass mit Daten gearbeitet wird. Nutzen Sie die angebotenen Materialien, experimentieren Sie selber.

In der Klausur: wieder nur die Interpretation von R-Output, kein Wissen über Coding nötig.

Organisatorisches
Vorlesung

- nötige Vorkenntnisse: Statistik I + II, Einführung in die Ökonometrie: in den nächsten Wochen gibt es einen Schnelldurchlauf als Wiederholung der nötigen Konzepte.
- Klausur: Modulprüfung (31-M23) ein Mal pro Semester. Sie wählen zwei der drei Vorlesungen 'Einführung in die Mikroökonometrie', 'Zeitreihenanalyse' und 'Multivariate Methoden' in der Klausur aus.
- findet wöchentlich Dienstags von 08:15-09:45 in X-E0-222.
- Infos und Folien finden Sie immer im Lernraum im eKVV.

Organisatorisches
Praktische Übung

- Die Vorlesung wird durch eine praktische Übung (310214) unterstützt.
- Sie brauchen eine Studienleistung zur Absolvierung des Moduls. In Zeitreihenanalyse nächstes Semester wir voraussichtlich keine PÜ angeboten werden können. Erledigen Sie die PÜ also dieses Semester.
- Die Übung findet zweiwöchentlich immer Mo 16-18 statt.
- Jede zweite Woche gibt es einen Übungszettel, der in der darauffolgenden Session besprochen wird.
- Die Übung wird geleitet von Herrn Lennart Oelschläger.
- Nähere Information im eKVV bzw. in der ersten Übungsstunde.

Organisatorisches

NOCH FRAGEN zur Organisation?

Was ist Ökonometrie?

Definition (Gabler Wirtschaftslexikon, Prof. Rottmann)

Die Ökonometrie ist ein Teilgebiet der Wirtschaftswissenschaften, das ökonomische Theorie, empirische Daten und statistische Methoden vereinigt.

Anders gesagt:

Ökonometrie handelt von

- der Verbindung von **mathematischen Modellen** mit empirischen Daten
- zur Beschreibung von **ökonomischen Vorgängen**
- mit Hilfe von **statistischen Methoden**.

Wortsinn: 'Vermessen der Wirtschaft'

Womit beschäftigt sich die Mikroökonometrie?

Die Ökonometrie kann grob in drei Teilbereiche unterteilt werden:

- **Makroökonometrie:** behandelt makroökonomische Prozesse, betrachtet also meist Aggregate: BIP, Einkommen, Konsum, Produktion, Zinsniveau, Wechselkurse, ...: vorwiegend Zeitreihenmethoden, teils raumzeitliche Prozesse.
- **Finance:** behandelt Aktien, Indices, Zinssätze: vor allem Zeitreihenmethoden.
- **Mikroökonometrie:** behandelt Beobachtungen von Charakteristika, Entscheidungen von Individuen (Personen, Haushalte, Unternehmen): vorwiegend Querschnittsdaten und Paneldaten

Mikroökonometrische Methoden kommen auch in der BWL (vor allem im Marketing) zum Einsatz.

Welche Daten kommen in der Mikroökonometrie vor?

Die allermeisten Daten in der Mikroökonometrie kommen aus Befragungen und anderen Surveys von Individuen:

Beispiele:

- Volkszählungen
- Haushaltsbefragungen, wie etwa das SOEP (sozioäkonomische Panel) oder das MOP (Mobilitätspanel)
- Behördendaten (Arbeitslosen, Meldungen, Krankenkassen, ...)
- Kundendaten: E-Fahrzeuge, Strom aus Smart Meter, ...

Unterschieden wird zwischen

- **Querschnittsdaten:** jedes Individuum wird einmal beobachtet, typischerweise zu einem Zeitpunkt
- **Paneldaten:** jedes Individuum wird zu verschiedenen Zeiten beobachtet.

Typische mikroökonometrische Modelle

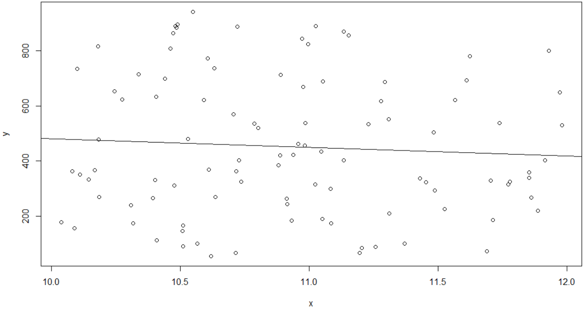
- lineare Modelle für Querschnittsdaten, gepoolte Querschnittsdaten und Paneldaten
- Modelle für qualitative Daten (nur wenige mögliche Antworten: diskrete Wahlmöglichkeiten: Logit- und Probit-Modelle) oder Zähldaten (etwa Zahl an gekauften Produkten; Poissonregression)
- zensierte Daten: Beobachtungen erfolgen nur unter bestimmten Gegebenheiten
  - etwa Lebensdaueruntersuchungen: beobachtet werden kann nur eine zeit lang, die Lebensdauer aller Überlebenden kann nicht erhoben werden
  - Einkommen: Angabe nur bis zu einer Maximalhöhe, danach nur Kategorie 'mehr als XY Euro im Jahr'

Für viele dieser Daten sind die linearen Regressionsmodelle aus der Einführung Ökonometrie nicht ideal geeignet.

Beispiel 1: Paneldaten

- Sie wollen den Zusammenhang zwischen Verkaufszahlen und Preis einer neuen Biersorte ermitteln.
- Dazu kaufen Sie von 100 Supermärkten in zwei Wochen jeweils die Verkaufszahlen, wobei jeweils 50 Supermärkte eine Preisreduktion anbieten und die restlichen 50 nicht.
- Man erwartet sich, dass ein niedrigerer Preis höhere Verkaufszahlen bedeutet.

Beispiel 1: Paneldaten (II)



Beispiel 1: Paneldaten (II)

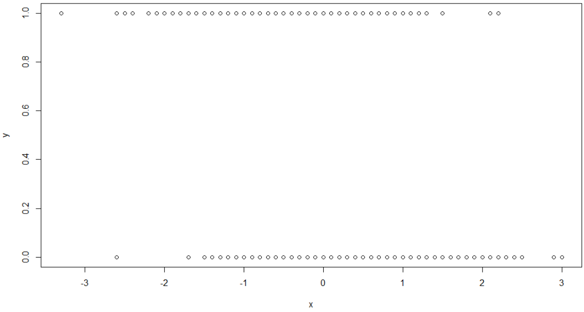
- Die Daten zeigen ein sehr unklares Bild.
- Aus den Querschnittsdaten alleine können wir zwei Effekte nicht auseinander halten:
  - Größere Supermärkte machen immer mehr Umsatz. Wir wissen nicht, ob in der zweiten Woche vielleicht die größeren Märkte die höheren Preise hatten?
  - Die Verkäufe können durch besondere Ereignisse in einer der Wochen systematisch variieren.
- Diese Effekte können wir entdecken und trennen, wenn wir die gleich Supermärkte jede Woche beobachten und kontrollieren können, in welchen Märkten verbilligt wird.
- Solche Datensätze heissen *Paneldaten*.

Beispiel 2: binäre Daten

- In vielen Fällen beobachten wir für Individuen, ob etwas zutrifft oder nicht.
  - Jemand hat ein Auto oder nicht.
  - Jemand kauft, oder nicht.
  - Jemand macht Gewinn oder nicht.
  - Jemand nimmt an einem Gewinnspiel teil oder nicht.
- Wir kennen solche Dummy-Variablen aus der Einführung Ökonometrie, dort aber als Regressoren.

Beispiel 2: binäre Daten (II)

Scatterplots geben dann ein sehr schlechtes Bild mit wenig direkt ersichtlicher Information:



Beispiel 2: binäre Daten (III)

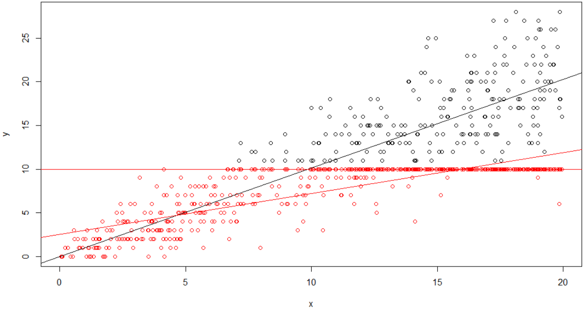
- Neben den Scatterplots sind auch die Modelle nicht immer sinnvoll.
- Ein lineares Modell kann in Prognosen münden, die kleiner als null oder größer als 1 sind.
- Bei einer binären Variable ist die Varianz an den Erwartungswert gekoppelt: Alternativ verteilte Variablen mit $\mathbb{P}(\mathbf{y} = 1) = p$ haben Erwartungswert p und Varianz $p(1 - p)$.
- Das hat Auswirkungen auf die Schätzung: Heteroskedastizität!

Beispiel 3: Zähldaten mit Zensurierung

- Sie kennen vielleicht die Zählart 'Eins, zwei, drei, viele ...'.
- In manchen Fällen trifft das auf echte Daten zu.
- Zählungen von Fußgängern, die durch visuelle Beobachtung gemacht werden.

Beispiel 3: Zähldaten mit Zensurierung (II)

In solchen Fällen muss die Verfälschung der Zählungen in der Modellierung berücksichtigt werden.



Beispiel 3: Zähldaten mit Zensurierung (III)

Die Zensurierung kann dabei verschieden Auswirkungen haben:

- Niedrige Daten werden nicht gemessen: Spielgewinne.
- Hohe Daten werden nicht gemessen: siehe oben.
- Daten fallen auf andere Art systematisch aus: z.B. Nichtkäufe.

Inhaltsangabe dieser Vorlesung

Die Vorlesung hält sich eng an das Buch

Jeff Wooldridge: Introductory Econometrics, Cengage Learning, Part III: Advanced Topics.

1. lineare Modelle für gepoolte Querschnittsdaten (Chapter 13)
2. Paneldatenmodelle (Chapter 14)
3. Instrumental Variables Schätzung (Chapter 15)
4. Modelle für diskrete Auswahl (qualitative Daten): binäre Auswahl (Section 17.1)
5. Modelle für Zähldaten (Section 17.3)
6. Modelle für zensierte Daten (Section 17.2, 17.4)

Wiederholung benötigter Konzepte

- Im Folgenden wiederhole ich einige Konzepte aus Mathematik, Wahrscheinlichkeitstheorie und Statistik, die wir das Semester über brauchen werden.
- Eine ausführlichere Erklärung findet sich in den Appendices A-D von Wooldridge.
- Ich habe ein PDF Dokument erstellt, in dem ich alle Grundlagen gesammelt habe. Dieses finden Sie im Lernraum.
- Einen Teil davon werde ich in der Vorlesung wiederholen, die Grundlagen aus der Statistik aber nicht.
- In der gesamten Vorlesung werde ich teils aus Zwecken der mathematischen Stringenz zusätzliche Formeln und Definitionen bringen,
  - die nicht geprüft werden,
  - deren Verständnis für das Verständnis der Vorlesung hilfreich jedoch nicht notwendig ist
  - die zum Teil einen starken mathematischen Hintergrund brauchen
- Diese zusätzlichen Dinge werden in blau geschrieben, ich werde in der Vorlesung teils auch nicht extra darauf eingehen.

Was brauchen wir und wozu?

Ökonometrie behandelt mathematische Modelle mit statistischen Methoden:

- Mathematik zur Beschreibung der Modelle
- Mathematik zur Herleitung von 'optimalen' Parametern und deren Eigenschaften.
- Stochastik zur Beschreibung der Daten: viele Phänomene in der Ökonomie beinhalten Unsicherheiten, die mittels Stochastik modelliert werden.
- Statistik zum Abgleich von Modellen und Daten durch Anpassung von 'Parametern'.

Neben den mathematischen und statistischen Grundvorlesungen liefert die Vorlesung 'Einführung in die Ökonometrie' die Grundlagen, diese werden auch wiederholt.

Mathematik: Lineare Algebra

- Vektor: $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ (n-Tupel): Vektor immer Spaltenvektor.
- Transposition: $x' = [x_1, ..., x_n]$ Zeilenvektor. Sprechweise: x-Strich.
- Transpositionsregel: $(AB)' = B'A'$.
- inneres Vektorenprodukt: $x'y = \sum_{i=1}^n x_i y_i = y'x$ für $x, y \in \mathbb{R}^n$.
- Einheitsmatrix $I_n = \text{diag}(1, ..., 1)$
- Einsvektor der Dimension n: $\iota_n = [1, 1, \ldots, 1]' \in \mathbb{R}^n$.
- Lösen von linearen Gleichungen: $Ax = b \Rightarrow x = A^{-1}b$, wenn $A \in \mathbb{R}^{n \times n}$ (quadratisch) invertierbar ist (die Spalten sind linear unabhängig).

Mathematik: Funktionen und Ableitungen

Funktion: Zuordnung: $x = [x_i] \mapsto f(x_1, x_2, \dots, x_n) = f(x) \in \mathbb{R}$.

Ableitung: $\frac{\partial f(x_0)}{\partial x} = \begin{pmatrix} \frac{\partial f(x_0)}{\partial x_1} \\ \frac{\partial f(x_0)}{\partial x_2} \\ \dots \\ \frac{\partial f(x_0)}{\partial x_n} \end{pmatrix}$: zuerst abgeleitet, dann ausgewertet an x_0

Beispiel 1: $f(x) = \beta' x = \sum_{i=1}^n \beta_i x_i \Rightarrow \frac{\partial f(x_0)}{\partial x_i} = \partial(\sum_{i=1}^n \beta_i x_i)/\partial x_i(x_0) = \beta_i \Rightarrow$
$\frac{\partial(\beta' x)(x_0)}{\partial x} = \beta \in \mathbb{R}^n$.

Beispiel 2: $f(x) = x' Q x = \sum_{i=1}^n \sum_{j=1}^n x_i Q_{i,j} x_j$ für symmetrische Matrix $Q = Q'$:

$$\Rightarrow \frac{\partial f(x_0)}{\partial x} = \sum_{j=1}^n Q_{i,j} x_{j,0} + \sum_{i=1}^n x_{i,0} Q_{i,j} = 2Qx_0 \in \mathbb{R}^n .$$

Stochastik zur Beschreibung der Daten

Wo kommen die Unsicherheiten und Zufälle in den Daten her?
Beispiel Stundenlohn einer Person: auf Basis des Consumer Panel Surveys (CPS)

- **Stichprobenziehung:** die Daten stammen aus einer Stichprobe der Grundgesamtheit, die zufällig bestimmt wurde.
- **Unkontrollierbare Faktoren:** ob jemand gerade vor oder nach einer Gehaltserhöhung steht, kann nicht kontrolliert werden
- **Nicht beobachtete Einflußgrößen:** ein Fragebogen kann nur manche Merkmale abfragen, aber nicht umfassend alles. Zusatzausbildungen und Kurse werden zum Beispiel selten erfragt.
- **Nicht beobachtbare Einflußgrößen:** individuelle Fähigkeiten, die sich nicht hart quantifizieren lassen, können nicht abgefragt werden. Etwa handwerkliches Geschick oder Intelligenz.
- **Fehlende Modellkomplexität:** Jedes Modell ist eine Vereinfachung der Realität. Wie sich diese im Einzelfall auswirkt, ist nicht immer vollständig kontrollierbar.

Stochastik zur Beschreibung der Daten

Beispiel für ein Modell

Reales Phänomen: Würfeln mit zwei Würfeln.

Zufallsexperiment: mathematische Repräsentation eines realen Phänomens; beschrieben mittels einer bivariaten Zufallsvariablen $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, wobei $\mathbf{x}_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, 2$.

Zufallsvariable: $\mathbf{x}_i$ ist eine Variable, deren Wert durch ein Zufallsexperiment bestimmt wird.

Verteilung von $\mathbf{x}_i$:

- Wahrscheinlichkeit $\mathbb{P}(\mathbf{x}_i = k) = p_k$ bezeichnet mit Parameter $p_k, k = 1, 2, ..., 6$ wobei $0 \leq p_k \leq 1, \sum_{k=1}^{6} p_k = 1$.
- $\mathbf{x}_1$ und $\mathbf{x}_2$ unabhängig voneinander: $\mathbb{P}(\mathbf{x}_1 = j, \mathbf{x}_2 = k) = \mathbb{P}(\mathbf{x}_1 = j)\,\mathbb{P}(\mathbf{x}_2 = k) = p_j p_k$

Stochastik zur Beschreibung der Daten

Zufallsvariable

Zufallsexperiment: mathematisches Pendant zu realer Beobachtung.

Zufallsvariable: mathematische Beschreibung der Ergebnisse eines Zufallsexperiments.

Zwei Grundtypen werden meist verwendet:

1. diskrete Zufallsvariable:
   - können nur eine abzählbare (mit natürlichen Zahlen indizierbare) Anzahl an Ausprägungen haben.
   - werden beschrieben durch Angabe einer pmf (probability mass function): $\mathbb{P}(\mathbf{y} = S_j)$.
   - **Bekannteste diskrete Verteilungen:** Bernoulli (auch binomial, Alternativverteilung), multinomial, diskret gleichverteilt, Poisson.
2. kontinuierliche Zufallsvariable
   - mögliche Ausprägungen bilden eine offene Teilmenge eines $\mathbb{R}^k$.
   - werden beschrieben durch Angabe einer pdf (probability density function)
   - **Bekannteste kontinuierliche Verteilungen:** Normalverteilung, Gleichverteilung, t-Vert., F-Vert., χ^2, Exponentialverteilung.

Stochastik zur Beschreibung der Daten

Zufallsvariable

Zufallsexperiment: mathematisches Pendant zu realer Beobachtung.

Zufallsvariable: mathematische Beschreibung der Ergebnisse eines Zufallsexperiments.

Zwei Grundtypen werden meist verwendet:

1. diskrete Zufallsvariable:
    - können nur eine abzählbare (mit natürlichen Zahlen indizierbare) Anzahl an Ausprägungen haben.
    - werden beschrieben durch Angabe einer pmf (probability mass function): $\mathbb{P}(\mathbf{y} = S_j)$.
    - **Bekannteste diskrete Verteilungen:** Bernoulli (auch binomial, Alternativverteilung), multinomial, diskret gleichverteilt, Poisson.
2. kontinuierliche Zufallsvariable
    - mögliche Ausprägungen bilden eine offene Teilmenge eines $\mathbb{R}^k$.
    - werden beschrieben durch Angabe einer pdf (probability density function)
    - **Bekannteste kontinuierliche Verteilungen:** Normalverteilung, Gleichverteilung, t-Vert., F-Vert., χ^2, Exponentialverteilung.

1. Grundlagen stochastischer Modelle