1. Einleitung

1.1 Phasen einer ökonometrischen Analyse

Empirische ökonomische bzw. ökonometrische Analyse:

Nutzung von Schätz- und Testmethoden zur Überprüfung ökonomischer Hypothesen oder Quantifizierung ökonomischer Zusammenhänge mit Daten

Einzelne Phasen der ökonometrischen Analyse:

- Formulierung eines Forschungsproblems
- Formulierung eines ökonomischen Modells oder ökonomischer Plausibilitäten und Intuition
- Übergang zu einem ökonometrischen Modell
- Formulierung von Hypothesen zu den unbekannten Parametern
- Schätzung des ökonometrischen Modells und Testen der Hypothesen mit entsprechenden Daten

---

Beispiel 1: Zusammenhang zwischen Bildung und Löhnen

Es soll der Effekt betrieblicher Weiterbildung auf individuelle Arbeitsproduktivität und damit auf den Lohn (wage) untersucht werden. Es wird angenommen, dass insbesondere Bildung (educ), Berufserfahrung (exper) und betriebliche Weiterbildung (training) einen Einfluss auf Löhne haben können.

Ökonomisches Modell:

$$\text{wage} = f(\text{educ}, \text{exper}, \text{training})$$

Dabei gilt z.B.:

wage: Stundenlohn, educ: Anzahl der Schuljahre, exper: Berufserfahrung in Jahren, training: Anzahl der Wochen in betrieblicher Weiterbildung

Ökonometrisches Modell:

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

Dabei stellen die Betas die Parameter und u den Störterm (der andere unbeobachtete Faktoren enthält) dar. Mit Hilfe von Individualdaten werden dann diese Parameter mit entsprechenden Schätzverfahren geschätzt und Hypothesen zu den Parametern mit entsprechenden Testverfahren statistisch überprüft.

Beispiel 2: Ökonomisches Modell der Kriminalität

Das Ausmaß krimineller Aktivitäten soll im Rahmen eines Nutzenmaximierungskalküls ökonomisch erklärt werden.

Ökonomisches Modell nach Becker:

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

Dabei gilt:

y: Verbrachte Zeit mit illegalen Aktivitäten
x_1: Stundenlohn für legale Beschäftigung
x_2: Einkommen aus anderen Quellen (z.B. Vermögen)
x_3: Wahrscheinlichkeit einer Festnahme
x_4: Wahrscheinlichkeit einer Verurteilung nach einer Festnahme
x_5: Erwartete Strafe nach einer Verurteilung
x_6: Alter
x_7: „Lohn“ für illegale Aktivitäten

Ökonometrisches Modell (das den „Lohn“ für illegale Aktivitäten ignoriert, da dieser praktisch nicht beobachtbar ist):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + u$$

1.2 Struktur ökonomischer Daten

Querschnittsdaten:

- Daten, die zu einem Zeitpunkt (der gelegentlich auch etwas variieren kann) an einer Untersuchungseinheit erhoben werden
- Die Anordnung der Einheiten im Datensatz spielt für die Analyse keine Rolle
- Ausgangspunkt ist meist die implizite Annahme, dass es sich bei der Erhebung um eine Zufallsstichprobe handelt
- Beispiele: Personen- oder Haushaltsdaten (z.B. Einkommen, Ausgaben), Unternehmensdaten (z.B. Umsätze, Produktion, Beschäftigung), Städte- oder Länderdaten (z.B. Arbeitslosigkeit, Bruttowertschöpfung)

Beispiel:

Beobachtungs-nummer	Monatliches Nettoeinkommen (in Euro)	Betriebszu-gehörigkeit (in Jahren)	Wöchentliche Arbeitsstunden	Alter (in Jahren)
1	848	18	20	60
2	1570	13	40	44
:	:	:	:	:
1456	2444	32	20	56

Zeitreihendaten:

- Daten, die bei einer Variablen oder verschiedenen Variablen über mehrere aufeinander folgende Zeitperioden erhoben werden
- Zeit ist hier eine wichtige Dimension (d.h. Beobachtungen sind meist über die Zeit korreliert), so dass die Anordnung der Beobachtungen im Datensatz potentiell wichtige Informationen enthält
- Die Häufigkeit der Datensammlung über die Zeit kann stark variieren, z.B. bei täglichen, wöchentlichen, monatlichen, Quartals- und Jahresdaten mit möglichen Saisoneffekten bei unterjährigen Daten
- Beispiele: Makroökomische Daten (z.B. Konsum, Investitionen, Geldangebot, Preisindex, Arbeitslosenraten), Finanzmarktdaten (z.B. Aktienkurse)

Beispiel:

Beobachtungsnummer	Jahr	Fertilitätsrate Westdeutschland	Arbeitslosenrate Westdeutschland (in %)
1	1955	2,11	5,6
2	1956	2,20	4,4
:	:	:	:
68	2022	1,51	5,4
69	2023	1,41	5,8

Aggregierte (gepoolte) Querschnittsdaten:

- Daten, die sowohl Querschnitts- als auch Zeitreiheneigenschaften aufweisen, da mehrere Querschnittsdatensätze unabhängig voneinander über verschiedene Perioden erhoben und zur Erhöhung des Stichprobenumfangs miteinander verknüpft werden
- Obwohl die Anordnung der Beobachtungen im Datensatz nicht wesentlich ist, wird die entsprechende Periode als wichtige Variable erfasst
- Daten werden meist wie herkömmliche Querschnittsdaten analysiert
- Beispiele: Personen- oder Haushaltsdaten (z.B. Einkommen, Ausgaben) in verschiedenen Jahren

---

Beispiel:

Beobachtungsnummer	Jahr	Monatliches Bruttoeinkommen (in Euro)	Wöchentliche Arbeitsstunden
1	2014	400	10
2	2014	4637	35
:	:	:	:
1262	2014	5003	40
1263	2018	1510	20
:	:	:	:
2524	2018	3535	20

Paneldaten:

- Daten, die sowohl eine Zeitreihen- als auch eine Querschnittsdimension haben, wobei hier im Unterschied zu aggregierten Querschnittsdaten dieselben Untersuchungseinheiten (z.B. Personen, Unternehmen, Länder) über mehrere Zeitperioden beobachtet werden
- Oft ist die Anzahl der Einheiten deutlich größer als die Zeitdimension
- Anordnung der Daten erfolgt oft erst nach Einheiten und dann Perioden
- Daten bieten die Möglichkeit für nicht beobachtbare Charakteristika der Einheiten zu kontrollieren sowie verzögerte Variablen zu untersuchen
- Beispiele: Personen- oder Haushaltspaneldaten (z.B. SOEP), Unternehmenspaneldaten (z.B. MIP), Länderpaneldaten

---

Beispiel:

Beobachtungsnummer	Individuum	Jahr	Monatliches Bruttoeinkommen (in Euro)	Geschlecht
1	1	2018	1567,21	Weiblich
2	1	2019	1649,59	Weiblich
3	2	2018	2362,25	Männlich
4	2	2019	3970,55	Männlich
:	:	:	:	:
2751	1376	2018	2195,89	Männlich
2752	1376	2019	3116,80	Männlich

1.3 Kausalität und ceteris paribus Begriff

→ In vielen empirischen Anwendungen besteht das Ziel einer Analyse im Auffinden, der Bestätigung und der Quantifizierung von kausalen Effekten

Kausalität:

Beschreibt die Abfolge aufeinander bezogener Ereignisse und Zustände, wobei die Wirkung (bei der abhängigen Variablen) immer auf die Ursache (bei der erklärenden Variablen) folgt

Ceteris paribus Begriff:

Alle anderen relevanten Faktoren bleiben konstant. Wenn nicht alle anderen Faktoren als konstant angenommen werden, kann kein kausaler Effekt einer einzelnen Variablen gemessen werden.

Messung:

Ideal wäre die Messung eines Effekts in einem randomisierten, kontrollierten Experiment. Dies ist aber bei vielen ökonomischen Fragestellungen nicht möglich.

→ Unter bestimmten Voraussetzungen können ökonometrische Methoden ein ceteris paribus Experiment simulieren

Beispiel: Bildungsrendite

Es soll untersucht werden, inwieweit sich der Lohn einer zufällig ausgewählten Person erhöht, wenn diese ein Jahr mehr Ausbildung genießt.

→ Ceteris paribus Betrachtung: Alle anderen Faktoren, die einen Effekt auf den Lohn haben (z.B. Erfahrung, Alter), bleiben konstant

Aufbau eines hypothetischen Experiments:

- Einer Gruppe von Personen wird zufällig eine bestimmte Ausbildungszeit zugeordnet
- Anschließend werden die Löhne der Personen gemessen
- Daraus resultiert die Unabhängigkeit zwischen Ausbildungszeit und den anderen Faktoren
- Eine empirische Analyse könnte dann die gestellte Frage beantworten

→ Allerdings ist das Experiment nur theoretisch denkbar, da die Durchführbarkeit natürlich insbesondere aufgrund ethischer Bedenken nicht möglich ist. Deshalb würden zur Untersuchung der Fragestellung nicht-experimentelle Daten verwendet werden, d.h. tatsächliche Lohn-, Ausbildungs- und andere Daten (z.B. zu Erfahrung und Alter)