

Univ.-Prof. Dr. Hans-Jörg Schmerer/Dr. Steffen Sirries

Angewandte Ökonometrie

Kurs 42270

Leseprobe

Fakultät für
**Wirtschafts-
wissenschaft**

Inhaltsverzeichnis

Abbildungsverzeichnis	5
Tabellenverzeichnis	9
1 Einführung	11
1.1 Was ist Ökonometrie?	11
1.2 Das ökonometrische Modell	17
1.3 Literaturempfehlung	22
2 Multivariate Regression	23
2.1 Herleitung des OLS Schätzers	28
2.2 Beispiel 1: Empirische Anwendung des Solow Modells	42
2.3 Hypothesentests	50
2.4 Beispiel 2: Cobb Douglas Produktion	54
2.5 Mechanik des OLS Schätzers	59
2.5.1 Deskriptive versus stochastische Regression	59
2.5.2 Annahmen des OLS Schätzers	61

2.5.3	Eigenschaften des OLS Schätzers	65
2.5.4	Kurzer Exkurs vorab: Asymptotische Eigenschaften	67
2.6	Effizienz des OLS Schätzers: Das Gauß-Markov Theorem	68
2.7	Schätzer für die Varianz σ^2	72
2.8	Normalverteilung des OLS Schätzers	76
2.9	Beispiel 3: Vorhersagen auf Basis einer OLS-Regression	77
2.9.1	Schätzung in STATA	79
2.9.2	Die Güte der Schätzung	98
3	Erklärung von binären Variablen	101
3.1	Schätzmethoden	101
3.1.1	Schätzung der Koeffizienten durch Maximum Likelihood	109
3.1.2	Interpretation der marginalen Effekte im Probit/Logit Modell	112
3.2	Beispiel 4: Gesundheitsökonomische Anwendung für Probit	113
3.2.1	Exkurs zur Wirkung von Gewichten	132
4	Differenzierte Schätzgleichung	135
4.1	Theoretische Fundierung	136
4.2	(Hypothetisches) Beispiel 5: Preis und Nachfrage	140
5	IV Schätzung	147
5.1	Theoretische Überlegungen zur IV Schätzung	148

5.2	Beispiel 6: Globalisierung und das Pro-Kopf-Einkommen einer Volkswirtschaft	153
5.3	Beispiel 7: IV Regression aus Cameron und Trivedi (2005)	161
6	Fixed- vs. Random-Effects	175
6.1	Theoretische Überlegungen	176
6.1.1	Fixed- und Random-Effects Modell	182
6.2	Beispiel 8: Random- und Fixed-Effects Schätzer in STATA	185
7	Diff-in-Diff Schätzung	205
7.1	Theorie	207
7.2	Beispiel 9: Die Mariel Flüchtlingskrise	214
7.3	Beispiel 10: Die Arbeitsmarkteffekte des Mindestlohns	218
7.4	Beispiel 11: Anwendung für Diff-in-Diff in STATA	223
8	Fazit	233
	Literaturverzeichnis	234

Abbildungsverzeichnis

2.1	Interpretation der Regressionslinie	45
2.2	Berechnung der Fehlerterme	47
2.3	Zweiseitiger Hypothesentest	52
2.4	Schätzung einer Produktionsfunktion	57
2.5	Asymptotische Eigenschaften des Schätzers	68
2.6	Öffnen des do-file Editors	80
2.7	Starten des do-files	81
2.8	Import der Datensätze	82
2.9	Einlesen von Excel Daten	84
2.10	Datenaufbereitung (Teil 1)	88
2.11	Datenaufbereitung (Teil 2)	90
2.12	Datenaufbereitung (Teil 3)	93
2.13	Schätzung der Koeffizienten mit OLS	94
2.14	Schätzung der Koeffizienten mit OLS (Teil 2)	95
2.15	Darstellung der Ergebnisse	96

2.16	Scatterplot mit Regressionslinie	98
3.1	Gegenüberstellung von Logit, Probit und OLS	103
3.2	Grafische Illustrierung des ML Verfahrens	111
3.3	STATA Beispiel für Probit Schätzung	117
3.4	STATA Beispiel für Probit Schätzung (Teil 2)	119
3.5	Codierung der Variable <i>RIDRETH1</i>	121
3.6	Erstellung einer Dummy-Variable	122
3.7	Berechnung des marginalen Effekts	123
3.8	Berechnung des marginalen Effekts	126
3.9	Implementierung einer Probit Schätzung in STATA	128
3.10	Berechnung des marginalen Effekts	129
3.11	Wirkungsweise von Gewichten	133
3.12	Datenstruktur mit und ohne Gewichten	134
4.1	(Hypothetisches) Datenstruktur im Panel	138
4.2	Beispiel für eine Schätzung in Differenzen	140
4.3	Implementierung der Schätzung in STATA	141
4.4	Paneldaten	143
4.5	Datenaufbereitung	143
4.6	Schätzung des Modells	144
5.1	Datenaufbereitung für die IV Anwendung in STATA	166

5.2	OLS Regression in STATA	168
5.3	IV Regression in STATA	171
5.4	Shea's Partial R2: Teilschritt 1	172
5.5	Shea's Partial R2: Teilschritt 2	172
5.6	First stage F-Stat	173
6.1	Datenaufbereitung für FE/RE Beispiel in STATA	189
6.2	Datenstruktur im FE/RE Beispiel in STATA	190
6.3	Generierung einer Indexvariable	194
6.4	OLS Regression in STATA	199
6.5	RE Regression in STATA	200
6.6	FE Regression in STATA	202
6.7	Übersichtstabelle der Ergebnisse	204
7.1	Annahme des gemeinsamen Trends	208
7.2	Treatment Effekt	211
7.3	Schätzergebnisse aus Card (1990)	215
7.4	Weitere Schätzergebnisse aus Card (1990)	218
7.5	Arbeitsmarkteffekte eines Mindestlohns	219
7.6	Übersichtliche Darstellung der Schätzergebnisse	221
7.7	Weitere Schätzergebnisse aus Card and Krüger (1994)	222
7.8	DiD Beispiel: Datenaufbereitung	224

7.9	DiD Beispiel: Ergebnis Regression	225
7.10	DiD Beispiel: Regression mit Kontrollvariablen	226
7.11	DiD Beispiel: Berechnung über Durchschnitte	227
7.12	DiD Beispiel: Datenstruktur der Durchschnitte	228
7.13	DiD Beispiel: Collapse	230
7.14	Endergebnis der Durchschnittsbildung	230
7.15	DiD Beispiel	231

Tabellenverzeichnis

2.1	Schätzung der Koeffizienten des Solow Modells	44
3.1	Gesundheitsökonomische Schätzung im NHNASE	131
5.1	Schätzergebnisse für bilaterale Handelsgleichung	157
5.2	Handel und Pro-Kopf-Einkommen	160
5.3	IV Regression für Lohn und Ausbildungsniveau	163
6.1	Darstellung der Regressionsergebnisse	196

Kapitel 1

Einführung

1.1 Was ist Ökonometrie?

Die Ökonometrie wird als die Schnittmenge der ökonomischen Theorie, der Mathematik und der Statistik verstanden.

Dieser Sachverhalt sei exemplarisch veranschaulicht an der Hypothese, dass der Preis eines Gutes negativ mit dessen Nachfrage in Verbindung steht. Geleitet durch Intuition und theoretisches Wissen über die Grundlagen der Wirtschaftswissenschaft, ist man dazu gewillt dieser Hypothese Glauben zu schenken. Unter der Prämisse, dass ein normales Gut betrachtet wird, führt ein Preisanstieg für gewöhnlich zu einer reduzierten Nachfrage nach eben diesem Gut. Dieser Zusammenhang ist so offensichtlich, dass eine empirische Überprüfung redundant erscheinen mag, und dennoch wird dieser Zusammenhang noch immer sehr häufig in ökonometrischen Studien analysiert. Wir werden im Verlauf dieses Kurses sehen, dass die Ökonometrie sehr praxisrelevante Anwendungen der ökonomischen Modelle erlaubt, die über einen simplen Hypothesentest - also die Bestätigung oder Ablehnung dieser Hypothese - hinausgehen. Es kann eben nicht nur die Existenz eines beliebigen Zusammenhangs überprüft, sondern auch die Stärke und mögliche Interaktionen mit anderen Variablen quantifiziert

werden.

Um all dies leisten zu können, muss der postulierte Zusammenhang zunächst einmal in einem Modell dargestellt werden. Die Theorie hilft uns also mögliche Erklärungen für beobachtbare Ereignisketten zu formulieren, wobei der Ausdruck "mögliche Erklärungen" hier ganz bewusst gewählt wurde. Eine theoretische Begründung ist eben noch kein Beweis für ein Zusammenhang. Erst die ökonometrische Analyse kann die Theorie durch eine geschickte Anwendung der zur Verfügung stehenden Methoden empirisch validieren. Die theoretisch motivierte Hypothese kann sich im Verlauf der empirischen Überprüfung als wahr oder falsch herausstellen. Es reicht eben nicht aus einen Zusammenhang zu glauben, in der Wissenschaft müssen Sachverhalte bewiesen werden. Aber die geschickte Verknüpfung von theoretischem Modell und empirischer Überprüfung kann nicht nur die Existenz bestimmter Zusammenhänge nachweisen. Auch die Stärke des Effekts lässt sich in einer solchen Analyse quantitativ bestimmen, was den sehr starken Anwendungsbezug der Ökonometrie unterstreicht. Die Kombination aus Theorie und Empirie liefert wichtige Vorhersagen für die reale Welt und erlaubt das wirtschaftliche Handeln von Individuen zu prognostizieren.

Es ist also nicht verwunderlich, dass die Ökonometrie auch in der Privatwirtschaft häufig zur Messung und Analyse wirtschaftlichen Handelns eingesetzt wird. Mit der Rechenkapazität moderner Computer und der zunehmenden Verfügbarkeit von Daten können empirische Methoden weit über die *klassische* Ökonomie hinaus angewandt werden.

Zu guter Letzt ist auch der Staat auf verlässliche Prognosen angewiesen, um Politiken entsprechend zu gestalten. Voraussetzung für eine gute Handlungsempfehlung für die Politik ist aber natürlich immer eine gründliche Analyse mit Berücksichtigung ihrer inhärenten Probleme. Diese Probleme hängen ab von der gestellten Frage und müssen entsprechend analysiert, beschrieben und auf Validität hin überprüft werden. Die Statistik hat eine Vielzahl an Methoden parat, die eingesetzt werden können, um diese Probleme zu umgehen und verlässliche Antworten auf die gestellten Fragen zu

finden.

Zusammenfassend kann also gesagt werden, dass die Mathematik und die Statistik dem Forscher das Methodenwissen an die Hand geben, das zur Überprüfung von wirtschaftlichen Zusammenhängen benötigt wird. Die Wirtschaftstheorie basiert auf der Mathematik und liefert vereinfachte Zusammenhänge, die mathematisch formuliert anhand statistischer Methoden geschätzt werden können. Die Verbindung aus Wirtschaftstheorie und Statistik ist Gegenstand der Ökonometrie.

In der Realität klaffen Theorie und Empirie allerdings sehr häufig auseinander, da oftmals diese logische Verknüpfung der beiden Disziplinen nicht ausreichend berücksichtigt wird. Vielfach fallen die postulierten Schätzgleichungen sprichwörtlich vom Himmel und ökonometrische Probleme werden häufig eher stiefmütterlich behandelt. Der gefundene Zusammenhang zwischen zwei beliebigen Variablen kann zwar noch immer statistisch signifikant sein, komplexere Zusammenhänge lassen sich ohne Modell allerdings nur sehr schwer deuten und es ist nicht klar, ob der gefundene Zusammenhang dem tatsächlichen Zusammenhang entspricht. Man spricht hier häufig vom "wahren Zusammenhang", der unbeobachtbar ist und durch ökonometrische Modelle ausfindig gemacht werden soll.

Ökonometrische Modelle sind geeignet für:

- Systematische Datenexploration,
- Falsifikation von Hypothesen,
- Prognosen über zukünftige Entwicklungen und
- Evaluation vergangener Entscheidungen.

Bei der systematischen Datenexploration soll zunächst ein Zusammenhang identifi-

ziert werden. Die Theorie liefert uns eine Hypothese über den Zusammenhang zweier Variablen und dieser Zusammenhang kann in der systematischen Datenexploration zunächst einmal über deskriptive Statistiken und einfache Korrelationsanalysen veranschaulicht werden.

Dieser erste Schritt der Analyse ist allerdings zu einfach, um weitergehende Rückschlüsse über Kausalität und/oder die Stärke des Effekts zu ziehen.

Im zweiten Schritt kann eine Hypothese aus der Theorie aufgestellt werden, die dann anhand eines geeigneten Tests bestätigt oder abgelehnt wird. Ablehnen bedeutet in diesem Zusammenhang, dass eine Hypothese durch Aufzeigen eines Widerspruchs zwischen Hypothese und Testergebnissen widerlegt wird. Komplexe Probleme müssen dafür geschickt vereinfacht werden, um überhaupt erst eine eindeutige, testbare Aussage treffen zu können. Auch für diesen Schritt kann ein Modell hilfreich sein, um die Komplexität bestimmter Vorgänge so zu reduzieren, dass bestimmte Hypothesen in einfachen wahr/falsch Aussagen formuliert werden können. Eine Anwendung haben wir bereits kennengelernt. Wir erwarten, dass ein Preisanstieg zu einer Reduktion der Nachfrage führt. Im einfachsten Fall kann diese Hypothese bestätigt oder abgelehnt werden. Die Stärke des Effekts ist bei der Überprüfung der Hypothese zunächst einmal nebensächlich.

Die dritte Form der ökonometrischen Analyse ist die Prognose. Man versucht aus dem Erkenntnisgewinn der Vergangenheit einen Rückschluss über die Zukunft zu gewinnen. Diese Art der Analyse ist nicht unumstritten und muss mit Vorsicht verwendet werden. Zunächst einmal ist nicht sicher, ob alle Einflussfaktoren des *“wahren Zusammenhangs“* in der Analyse der zugänglichen Daten ausreichend berücksichtigt wurden. Dieses Problem des *“omitted variable bias“*, also dem Problem der Verzerrung aufgrund von ausgelassenen Variablen, sollte nicht unterschätzt werden. Eine Aussage zu validieren reicht für die Prognose in der Regel nicht aus, da die Stärke des Zusammenhangs für viele Fragen quantifiziert werden muss. In der Praxis geschieht dies häufig über eine Regressionsanalyse. Daten werden verglichen, um so

eine Aussage über die Stärke des Zusammenhangs zu treffen. Im weiteren Verlauf dieses Kurses werden wir allerdings sehen, dass die Stärke des Effekts, gemessen durch den sogenannten Koeffizienten der erklärenden Variable in einer Regression, mitunter stark verzerrt sein kann. Auf verzerrten Schätzern basierte Prognosen sind ungenau und selbst der einfache Hypothesentest, der ja nur auf Annahme oder Ablehnung der Hypothese testet, kann durch verzerrte Schätzer ein falsches Ergebnis liefern.

Zu guter Letzt findet man ökonometrische Analysen im Bereich der Evaluation. Die Politik ist häufig an einer Einschätzung des Wirkungsgrads einer durchgeführten Politikmaßnahme interessiert. Im Anschluss an eine bestimmte Maßnahme können Daten über die betroffenen Personen herangezogen werden, um den Erfolg des jeweiligen Programms einschätzen zu können. Sehr häufig verwendet man eine experimentelle Herangehensweise. In einem Vergleich der Situation vor Einführung einer Maßnahme mit der Situation nach der Einführung der Maßnahme kann die Wirkung eingeschätzt werden. Die Veränderung in der Gruppe der betroffenen Individuen (Treatment group) wird mit den Veränderungen in einer Kontrollgruppe verglichen. Auf diese Weise sollen Einflussfaktoren kontrolliert werden, die nichts mit der eigentlichen Maßnahme zu tun haben. Stellen Sie sich vor, dass zeitgleich zur politischen Maßnahme ein Schock auftritt, der alle Individuen (sowohl in der Treatment Gruppe als auch in der Kontrollgruppe) gleichermaßen beeinträchtigt. Durch einen Vergleich der Änderungen in der Gruppe der tatsächlich betroffenen Individuen mit den Veränderungen in der Kontrollgruppe lassen sich solche Effekte herausrechnen.

Für jede dieser Hauptanwendungen stellt die Ökonometrie also mehr oder weniger eigene Methoden zur Verfügung.

- Einige Beispiele:
 - Was ist die Nachfrageelastizität eines Gutes oder einer Dienstleistung (qualitativ und quantitativ)?

- Hat die individuelle Investition in Humankapital (Bildung) einen Effekt auf den Lohn?
- Welche Auswirkungen hat eine Weiterbildungsmaßnahme auf die Produktivität von Mitarbeitern?
- Hat die Einführung eines Mindestlohns negative Effekte auf die Beschäftigung?
- Sind exportierende Unternehmen produktiver als andere?
- Führt die Verlagerung von wirtschaftlichen Aktivitäten ins Ausland zu negativen Beschäftigungseffekten im Inland?
- Wachsen arme Länder schneller als reiche (Konvergenzhypothese des Solow Modells)?
- Was sind die Determinanten für das Ausfallrisiko eines Kredits?
- Wie entwickelt sich die Inflationsrate im nächsten Jahr?
- ...

Grundsätzlich verschreibt sich die Wissenschaft bei der Beantwortung solcher Fragen der *Objektivität!* Das heißt, dass die Analyse stets ergebnisoffen erfolgen muss und die Regeln der guten wissenschaftlichen Praxis eingehalten werden sollten. Die veröffentlichten Ergebnisse müssen valide und nachvollziehbar sein, was wiederum durch fest vorgeschriebene Regeln erreicht werden kann. Diese Regeln werden aus der Wahrscheinlichkeitstheorie und der Mathematik beziehungsweise der Statistik abgeleitet. Ein solch formaler Zugang erlaubt es dem Forscher seine *Annahmen* offen zu legen, wodurch andere Wissenschaftler die Analysen nachvollziehen können. Jeder Wissenschaftler kann basierend auf den gleichen Daten zu exakt der gleichen Schlussfolgerung kommen.

1.2 Das ökonomische Modell

Wie kommt man von einem theoretischen (ökonomischen) Modell zu einem ökonomischen Modell? In der Regel beginnt die empirische Analyse immer mit der Diskussion des der Analyse zugrunde liegenden *theoretischen Modells*. Nehmen wir die folgende Modellgleichung als einfaches Beispiel

$$y = f(x_1, x_2, \dots, x_k) . \quad (1.1)$$

Ohne den funktionalen Zusammenhang genauer zu spezifizieren, nehmen wir an, es gebe eine endogene Variable y , die durch die exogenen Variablen x_1, x_2 bis x_k bestimmt wird. Wichtig ist, dass diese Variablen veränderbar sind und dass sich die endogene Variable gemäß dieser exogenen Variablen sowie einem bislang noch nicht weiter charakterisierten Prozess anpassen. Dieser Anpassungsprozess sei zunächst einmal allgemein durch die Funktion $f()$ charakterisiert. In der Regel werden bestimmte Eigenschaften der Funktion f unterstellt, um den Prozess zumindest charakterisieren zu können.

Die funktionalen Eigenschaften des Prozesses können aus modell-theoretischen Vorüberlegungen abgeleitet werden.

Das theoretische Modell muss dann zu einem *ökonomischen Modell*

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1.2)$$

transformiert werden, welches mit Daten *geschätzt* werden kann. Für die Ermittlung der Koeffizienten, die das angenommene Modell quantifizieren, werden für y und x_i (mit $i = 1, \dots, k$) reale Werte eingesetzt. Die Koeffizienten können dann nach bestimmten Regeln, die wir sogleich noch diskutieren werden, gesetzt werden. Im Unterschied zu Gleichung 1.1 wird in der Schätzgleichung bereits ein bestimmter Prozess unterstellt. Wir gehen zunächst einmal davon aus, dass ein linearer Zusammen-

hang zwischen y und den k unabhängigen Variablen auf der rechten Seite existiert. Es sei angemerkt, dass wir den wahren Zusammenhang noch nicht kennen und dass der Übergang von Gleichung 1.1 zu Gleichung 1.2 zunächst einmal willkürlich ist. Tatsächlich wird in der Regel zunächst immer ein stark vereinfachtes, lineares Modell geschätzt. Aufbauend auf diesen ersten Ergebnissen können dann komplexere Zusammenhänge in weiterführenden Analysen untersucht werden. Ein solches *“trial and error“* Vorgehen, das verschiedene Schätzmethoden miteinander vergleicht, kann uns dabei helfen, den wahren Zusammenhang näherungsweise zu ermitteln. Je robuster die Ergebnisse, desto belastbarer die aus der Analyse gewonnenen Rückschlüsse. In seriösen Analysen wird also in der Regel nicht nur ein Modell gezeigt, sondern viele Robustheitsanalysen präsentiert, um die Allgemeingültigkeit der Ergebnisse zu untermauern und den Leser davon zu überzeugen, dass der wahre Zusammenhang ausfindig gemacht wurde.¹ Außerdem enthält die Gleichung 1.2 einen zusätzlichen Störterm ε , der ebenfalls etwas später in diesem Kapitel behandelt wird.

Um zu sehen, wie sich die Modellgleichung in die Schätzgleichung überträgt, schauen wir uns zunächst einmal ein paar Eigenschaften der beiden Gleichungen an. Wie wirkt sich eine Änderung der erklärenden Variable x_i mit $i \in \{1, \dots, k\}$ auf die abhängige Variable aus? Dazu bilden wir zunächst die erste Ableitung der beiden Gleichungen 1.1 und 1.2. Beachten Sie, dass es sich hierbei um eine *ceteris paribus Betrachtung* handelt. Wir schauen uns an, wie sich die Variable y ändern muss, wenn sich eine der x Variablen ändert. Das heißt, dass mit Ausnahme der betrachteten Variable x_i , alle anderen Variablen $x_{j \neq i}$ konstant bleiben. An dieser Stelle reicht also die Betrachtung der partiellen Ableitung völlig aus, da zunächst ein lineares Modell unterstellt wird.

¹ Es sei noch einmal darauf hingewiesen, dass der wahre Zusammenhang niemals aufgedeckt werden wird. Auch wenn man die Zahl der Robustheitsanalysen gegen unendlich gehen lassen würde, könnte dadurch nicht das wahre Modell identifiziert werden. Allerdings sinkt mit der Anzahl an durchgeführten Robustheitsanalysen die Wahrscheinlichkeit, dass der gezeigte Zusammenhang rein zufällig ist.

Aus Gleichung 1.1 lässt sich exemplarisch die erste Ableitung nach x_1 bilden, um folgenden Zusammenhang zu bekommen:

$$dy = \frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_1} dx_1 . \quad (1.3)$$

Gemäß des Modells führt eine Änderung der Variable x_1 um dx_1 Einheiten zu einer Anpassung der endogenen Variable y um insgesamt dy Einheiten. Es sei noch einmal angemerkt, dass die Änderung von x_1 isoliert betrachtet und alle anderen x Variablen als konstant angenommen wurden. Die erste Ableitung der Funktion f bestimmt die Stärke und die Richtung des Effekts. Wie sieht dieser Zusammenhang in der zugehörigen Schätzgleichung aus?

Aus Gleichung 1.2 bekommen wir

$$dy = \beta_1 dx_1 \quad (1.4)$$

Der funktionale Zusammenhang f wurde also implizit so gewählt, dass

$$\frac{\partial f(x_1, x_2, \dots, x_k)}{\partial x_1} = \beta_1 \quad (1.5)$$

gilt. Diese Herangehensweise ist zwar sehr willkürlich, hilft aber zunächst einmal dabei ein Gespür für die Daten zu bekommen. Es sei noch erwähnt, dass die Variable ε nicht durch das Modell erklärbare Einflussfaktoren auffängt. Ein Modell ist immer eine Abstraktion von der Wirklichkeit und kann nicht alle wichtigen Einflussfaktoren beinhalten. In der Wirklichkeit hingegen ist die tatsächliche Zahl der x Variablen, die berücksichtigt werden müssten, sehr wahrscheinlich extrem groß. Um diese nicht einbezogenen Einflussfaktoren in der Schätzgleichung zu berücksichtigen, wird der Fehlerterm ε eingeführt, in dem alle weiteren vom Modell nicht erklärten Einflussfaktoren zu einem Wert verdichtet werden (also alle Einflussfaktoren mit Ausnahme von x_1 bis x_k).

Warum werden dann nicht einfach alle erdenkbaren Einflussfaktoren im Modell berücksichtigt? Dies hat zwei Gründe: Daten stehen nicht zu allen erdenklichen Einflussfaktoren zur Verfügung und die Anzahl an Freiheitsgraden ist im Modell begrenzt. Es können nicht unendlich viele Parameter geschätzt werden, da das zu lösende Gleichungssystem auch lösbar sein muss. Zu viele unbekannte Variablen in einem Gleichungssystem führen zwangsläufig dazu, dass die einzelnen Parameter nicht mehr identifiziert werden können. Wir werden sehen, dass ein ökonometrisches Modell letztendlich eben "nur" ein einfaches, mathematisches Optimierungsproblem ist und die Gesetze der Mathematik entsprechend gelten.

Beispiel 0: Vom theoretischen (ökonomischen) zum ökonometrischen Modell

- Nehmen wir an, wir haben ein theoretisches Modell, das den Lohn einer Person beschreibt

$$\text{Lohn} = f(\text{Bildung}, \text{Berufserfahrung}, \text{Alter})$$

- Ein mögliches ökonometrisches Modell könnte sein:

$$\text{Lohn} = \beta_0 + \beta_1 \text{Bildung} + \beta_2 \text{Berufserfahrung} + \beta_3 \text{Alter} + \varepsilon$$

- In dieser ökonometrischen Darstellung des Modells nehmen wir also eine lineare Form der Beziehung zwischen den *erklärenden Variablen* und der *zu erklärenden Variablen* an!
- Der Fehlerterm berücksichtigt weitere Einflussfaktoren, die zwar einen Einfluss haben auf die abhängige Variable, jedoch nicht im Modell berücksichtigt wurden.

Auch für dieses Beispiel lassen sich ganz einfach die marginalen Effekte berechnen. Aus dieser Berechnung lässt sich erkennen, dass das postulierte ökonometrische Modell auch tatsächlich linear ist. Egal wie alt die Person bereits ist, ein weiteres Jahr ($d\text{Alter} = 1$) erhöht den Lohn um $d\text{Lohn} = \beta_3$ Einheiten, wobei d für Veränderung

(difference) steht.

Die Konstanten $\beta_0, \beta_1, \beta_2$ und β_3 sind also die Parameter des theoretischen Modells und beschreiben die Richtung und die Intensität der Beziehung zwischen dem Lohn und dessen Determinanten. Variable ε erfasst hier alle *übrigen* Effekte auf den Lohn einer Person. Der Umgang mit diesem *Fehlerterm* oder *Störterm* (ε) ist extrem wichtig für die Anfertigung einer validen Analyse. Das ökonometrische Modell wird oft als das *wahre* Modell bezeichnet, das für den Forscher in der Realität auf Grund der Abstraktheit nicht so ohne weiteres beobachtbar ist. Durch die Anwendung der statistischen (empirischen) Methode auf die Daten, wird das *wahre* Modell dann *geschätzt*. Diese Unterscheidung ist wichtig und wird in diesem Kapitel durch die Verwendung unterschiedlicher Buchstaben hervorgehoben. Die Koeffizienten im Vektor β , die den wahren Zusammenhang angeben und unbeobachtbar sind, werden durch die geschätzten Parameter b approximiert. Der wahre Störterm ε wird durch den geschätzten Störterm e approximiert. Wir können eine Vermutung über den Zusammenhang aufstellen und diesen dann anhand von Daten schätzen. Dennoch wird der wahre Zusammenhang auch durch die Schätzung niemals entdeckt werden. Lediglich die geschätzten Werte können anhand von Hypothesentests mit einer bestimmten Sicherheit validiert oder abgelehnt werden.

Im Modell unterstellen wir, dass der Zusammenhang zwischen Lohn, Bildung, Berufserfahrung und Alter positiv ist. Dies wird durch das positive Vorzeichen der Koeffizienten angezeigt. Ob die Daten diesen postulierten Zusammenhang unterstützen, zeigt sich an den geschätzten b Koeffizienten, die ein entsprechendes Vorzeichen haben sollten. Eine Vielzahl an ökonometrischen Problemen, die wir im Verlauf dieses Moduls noch kennenlernen werden, kann zu entsprechenden Abweichungen zwischen tatsächlichem und geschätztem Koeffizienten führen. Im Extremfall kann sich durch diese Art von Problem nicht nur die Stärke des Effekts ändern, sondern sogar das Vorzeichen umkehren.

Bevor wir aber auf diese Probleme näher eingehen, wird im nächsten Kapitel zunächst

gezeigt, wie die Koeffizienten b , und damit einhergehend auch der Fehlerterm e , basierend auf beobachteten Daten in einer multivariaten Regressionsanalyse bestimmt werden können.

1.3 Literaturempfehlung

Das Fach "Ökonometrie" kann in einem so kompakten Kurs natürlich nicht erschöpfend abgehandelt werden. Statt alle Themen mit sämtlichen Erweiterungen zu besprechen, werden sehr viele Hintergrundinformationen ausgiebig besprochen und vertieft. Basierend auf diesem Kurs empfehlen wir für die eigene Analyse folgende Basisliteratur:

- Greene, W.H. (2012): "Econometric Analysis".
- Cameron, A.C. und Trivedi, P.K. (2005): "Microeconometrics: Methods and Applications".

Beide Bücher bieten weiterführende Erklärungen, die für die Bearbeitung einer eigenständigen Datenanalyse hilfreich sein sollten.

Kapitel 4

Differenzierte Schätzgleichung

Bislang wurde in den Beispielen immer zwischen der abhängigen Variable, den interessierenden Variablen und weiteren Kontrollvariablen unterschieden. Obwohl wir uns häufig nur für den Effekt einer einzigen Variable interessieren, werden in der Regel weitere Variablen im ökonometrischen Modell berücksichtigt. Es gibt Grund zur Annahme, dass die berücksichtigten Kontrollvariablen einen Einfluss auf die abhängige Variablen haben und dass diese Effekte entsprechend gefiltert werden müssen. Würde man diese Einflussgrößen vernachlässigen, dann könnte der Zusammenhang zwischen den Variablen, deren Zusammenhang uns interessiert, verfälscht sein. In einem solchen Fall wären die getroffenen Annahmen nicht mehr erfüllt und der Schätzer auch nicht mehr *BLUE*.

In der Praxis können diese Einflussfaktoren im Modell jedoch nur dann berücksichtigt werden, wenn die entsprechende Variablen auch tatsächlich beobachtbar sind. Auf dieses Problem wurde im Verlauf dieses Kurses bereits häufig hingewiesen. Viele der wesentlichen Einflussfaktoren sind nicht direkt messbar und führen entsprechend dazu, so dass diese Effekte dann automatisch vom Fehlerterm aufgefangen werden. Die Schätzung ist entsprechend verzerrt.

In diesem Kapitel besprechen wir einen ersten Ansatz, der verspricht dieses Problem

zu adressieren.

4.1 Theoretische Fundierung

Folgende Schätzgleichung wird untersucht:

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, \dots, T, \quad (4.1)$$

Der Zusammenhang zwischen der abhängigen und den unabhängigen Variablen sei wieder linear. Im Vergleich zu den vorangegangenen Modellen wird nun aber zusätzlich zum Querschnitt auch die Zeitdimension der Daten über den Index t berücksichtigt. Es werden also wie zuvor bestimmte Eigenschaften der Individuen im Querschnitt betrachtet, jedoch liegen die Beobachtungen im Panel mehrfach über die Zeit hinweg für das gleiche Individuum vor. In der Regel werden wiederholt Jahres- oder Monatsdurchschnitte erhoben, denkbar sind aber auch sehr viel genauere Erhebungen auf Tages-, Minuten-, oder sogar Sekundenbasis. Beispielsweise lässt sich das Alter einer Arbeitskraft i in jeder Periode t beobachten und der Effekt dieser Variable über den entsprechenden Koeffizient β_k bestimmen. An der Notation erkennen wir jedoch auch, dass der Koeffizient nicht zeitspezifisch ist, sondern ein Koeffizient für alle Perioden gleichermaßen gilt. Die Koeffizienten für die unterschiedlichen Variablen werden auch für diesen Schätzer wieder in einem Vektor β zusammengefasst. Man könnte also kritisch hinterfragen, welchen Mehrwert die zusätzliche Zeitdimension bringt?

Neu hinzugekommen ist die Modellierung des "Fixed-Effects", c_i , der unabhängig von der Zeit t einen konstanten Effekt auf der Individualebene misst. Hiermit ist gemeint, dass für jedes Individuum im Querschnitt ein spezifischer Effekt gemessen wird, der über die Zeit hinweg konstant ist. Dies führt dazu, dass der Fehlerterm nun in zwei Komponenten unterteilt werden kann, nämlich einer fixen und einer variablen Komponente. Streng genommen war das in den vorangegangenen Kapiteln

natürlich auch der Fall. Die Unterteilung wurde nur nicht vorgenommen, da es ohne Zeitdimension keine Möglichkeit gibt, diesen Effekt zu adressieren und eine solche Unterteilung dementsprechend kaum einen Mehrwert gebracht hätte.

Aber wie realistisch ist die Annahme, dass diese unbeobachtbaren Effekte nicht auch für die Zeit spezifisch sind? Häufig ist dies natürlich nicht sehr realistisch, da sich diese Einflussgrößen in der Regel dem Zeitverlauf anpassen. Beispielsweise variiert auch die Fähigkeit einer Arbeitskraft über die Zeit hinweg. Tatsächlich bleibt kaum eine Variable über die Zeit hinweg vollkommen konstant. Ausnahmen könnten Handicaps oder der IQ der Arbeitskräfte sein, über die aus Datenschutzgründen keine Informationen vorliegen, die aber die Fähigkeit der Arbeitskräfte in jeder Periode gleichermaßen beeinflussen. Solche konstanten Faktoren werden für jedes Individuum in der Variable c_i zusammengefasst. Im Panel besteht dann die Möglichkeit diese Variable durch eine einfache Transformation der Daten in der Schätzung zu eliminieren. Hierfür schätzen wir folgenden Zusammenhang

$$\Delta y_{it} = y_{it} - y_{it-1} \quad , \quad (4.2)$$

wobei wir die Gleichungen

$$y_{it} = x_{it}\beta + c_i + u_{it} \quad \text{und} \quad (4.3)$$

$$y_{it-1} = x_{it-1}\beta + c_i + u_{it-1} \quad (4.4)$$

einsetzen können, um daraus das Ergebnis

$$\begin{aligned} \Delta y_{it} &= x_{it}\beta + c_i + u_{it} - (x_{it-1}\beta + c_i + u_{it-1}) \\ &= (x_{it} - x_{it-1})\beta + c_i - c_i + (u_{it} - u_{it-1}) \\ &= (x_{it} - x_{it-1})\beta + (u_{it} - u_{it-1}) \end{aligned} \quad (4.5)$$

zu erhalten. Die Transformation bildet also die erste Differenz der abhängigen Variable und setzt dann entsprechend die Gleichungen zur Transformation ein. Der

Term x_{it} ist ein Zeilenvektor, der die Beobachtungen für den Fall it beinhaltet. Der Spaltenvektor β beinhaltet die gemeinsamen Koeffizienten für die jeweils beobachtete Variable. Für die Identifikation ist wichtig, dass jede Beobachtung eindeutig über die beiden Indexvariablen i und t identifiziert werden kann. Das heißt, dass für kein Individuum wiederholt Beobachtungen für den gleichen Zeitpunkt vorliegen dürfen. Abbildung 4.1 illustriert die Datenstruktur eines geeigneten Datensatzes und veranschaulicht den dazugehörigen Vektor der Regressionskoeffizienten β . Es werden insgesamt vier Variablen berücksichtigt und jede Zeile kann eindeutig über den Index it identifiziert werden, da keine doppelten Einträge bestehen. Der Vektor x_{it} in

i	t	Y_{it}	X_{it1}	X_{it2}	X_{it3}	X_{it4}
1	1	234	12	33	545	54
1	2	233	14	36	565	67
2	1	345	42	25	965	64
2	2	377	46	30	885	65

β_1
β_2
β_3
β_4

Abbildung 4.1: (Hypothetisches) Datenstruktur im Panel

Gleichung 4.5 soll einmal exemplarisch für den Fall $it = 11$ veranschaulicht werden als

$$x_{11} = \begin{pmatrix} 12 & 33 & 545 & 54 \end{pmatrix}$$

Multipliziert mit dem Spaltenvektor β ergibt sich der Zusammenhang

$$234 = 12 \times \beta_1 + 33 \times \beta_2 + 545 \times \beta_3 + 54 \times \beta_4 + c_1 + u_{11}$$

Da eine zweite Beobachtung für das Individuum i vorliegt, nämlich für den Zeitpunkt $t = 2$, kann der differenzierte Zusammenhang betrachtet werden als

$$\begin{aligned} 233 - 234 &= (14 - 12) \times \beta_1 + (36 - 33) \times \beta_2 + (565 - 545) \times \beta_3 + (67 - 54) \times \beta_4 \\ &\quad + (u_{12} - u_{11}) \\ \Rightarrow \Delta y_1 &= -1 = 2 \times \beta_1 + 3 \times \beta_2 + 20 \times \beta_3 + 13 \times \beta_4 + \Delta u_1 \end{aligned}$$

Diese Prozedur kann für jede Zeile wiederholt werden, um dann im letzten Schritt alle Δy_{it} in einem Vektor y und alle Differenzen der vier X -Variablen in einer Matrix X zusammenzufassen. Diese Matrix enthält in der ersten Spalte lauter 1en, um zusätzlich die Konstante β_0 schätzen zu können. Der Vektor der Koeffizienten β muss entsprechend um diesen einen Koeffizienten erweitert werden. Durch die Differenzbildung reduziert sich die Zahl an Beobachtungen, da eine Periode verloren geht. In unserem Beispiel hatten wir ursprünglich zwei Perioden aus denen wir für jedes Individuum i eine Differenz bilden konnten. Die Zahl der Beobachtungen reduziert sich also von $2 \times I$ auf I , wobei I die Gesamtzahl an Beobachtungen im Querschnitt bezeichnet ($i \in 1, \dots, I$).

Nach Transformation und Zusammenfassen der Daten in einer Matrix kann die Lösung für den Vektor an Koeffizienten, β , nun sehr einfach über den OLS Schätzer

$$\hat{\beta} = (\Delta X' \Delta X)^{-1} \Delta X' \Delta y \quad (4.6)$$

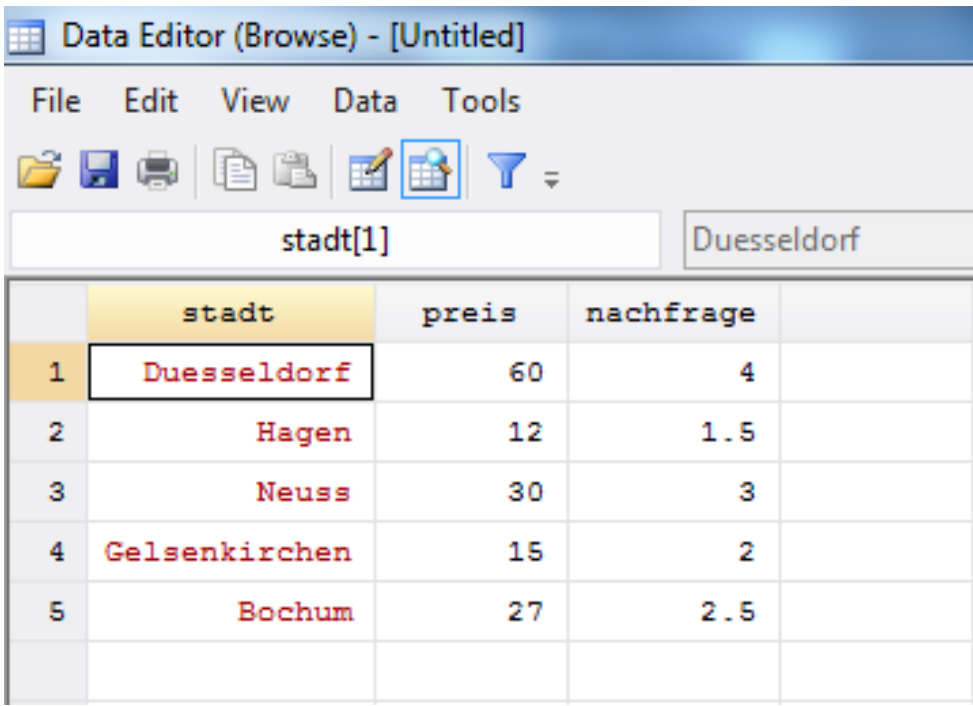
bestimmt werden. Der einzige Unterschied zu dem zu Beginn des Kurses besprochenen Modell ist die Verwendung der Änderungsraten, ΔX und Δy , anstatt der absoluten Werte. Es werden also nur die Daten transformiert und dann mit dem bereits ausführlich besprochenen Kleinste-Quadrate-Schätzer berechnet. Alle über die Zeit hinweg konstanten Faktoren, die die Schätzung verzerren würden, sind durch die Datentransformation im Vorfeld der Schätzung bereits herausgerechnet worden.

Häufig werden die erklärenden Variablen auf der rechten Seite der Schätzgleichung verzögert in die Schätzgleichung aufgenommen, um eine kausale Interpretation aus der Schätzung ableiten zu können. Ist die Veränderung einer erklärenden Variablen mit der Veränderung einer abhängigen Variable korreliert und liegt die Veränderung der erklärenden Variablen in der Vergangenheit, dann ist der identifizierte Effekt mit hoher Wahrscheinlichkeit auch kausal. Allerdings muss auch diese Aussage kritisch überprüft werden, da nur die über die Zeit hinweg konstanten Effekte in der Schätzung kontrolliert wurden. Bei einem vermeintlich konstanten Zusammenhang könnte

trotzdem eine Scheinkorrelation bestehen, wenn die unbeobachteten Einflussfaktoren sich über die Zeit hinweg anpassen, also nicht vollständig über die Variable c_i aufgefangen werden.

4.2 (Hypothetisches) Beispiel 5: Preis und Nachfrage

Das folgende Beispiel ist frei erfunden und kann über das do-file "Fixed_Effects.do" geladen werden. Führt man zunächst die Zeilen 1 bis 5 aus, dann öffnet sich der Browser und zeigt die folgenden Daten an:



The screenshot shows a software window titled "Data Editor (Browse) - [Untitled]". It has a menu bar with "File", "Edit", "View", "Data", and "Tools". Below the menu is a toolbar with various icons. The main area displays a table with the following data:

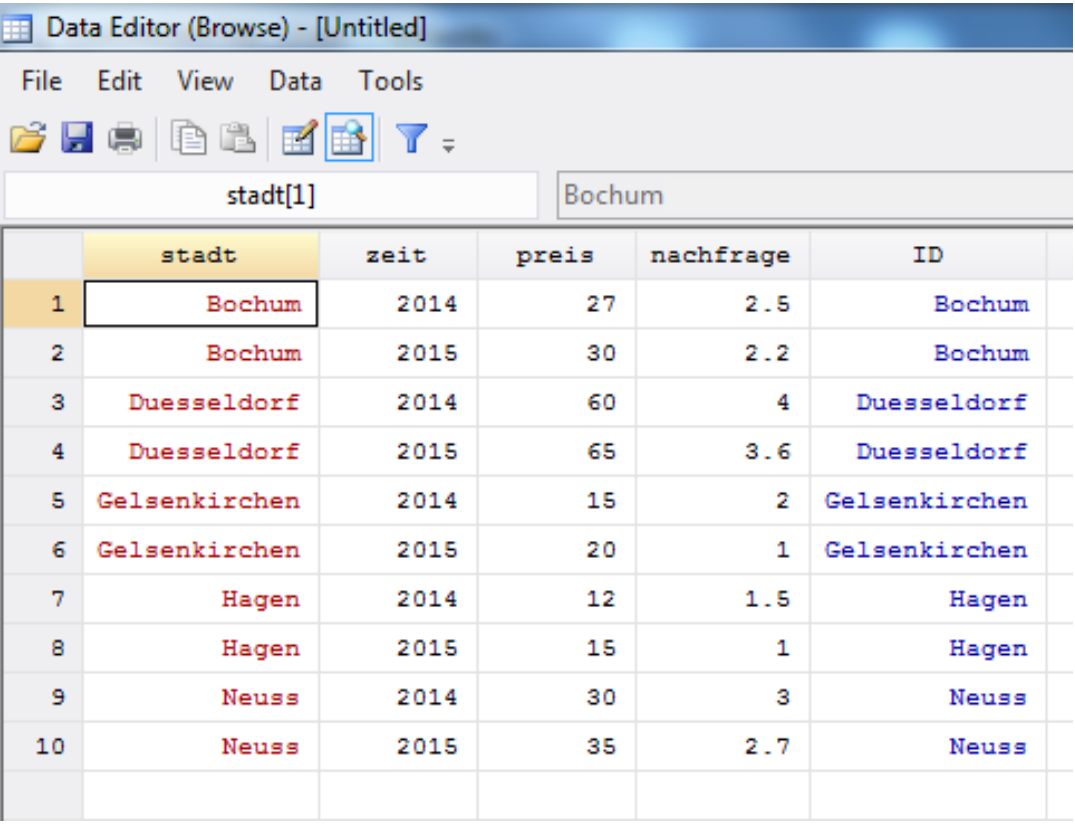
	stadt	preis	nachfrage
1	Duesseldorf	60	4
2	Hagen	12	1.5
3	Neuss	30	3
4	Gelsenkirchen	15	2
5	Bochum	27	2.5

Abbildung 4.2: Beispiel für eine Schätzung in Differenzen

Für fünf Städte in Nordrheinwestfalen wurden Daten zu den durchschnittlichen Preisen eines Restaurantbesuchs und der durchschnittlichen Nachfrage nach Mahlzeiten im Restaurant erfunden. Ein Blick auf die Daten zeigt, dass die Relationen etwas übertrieben dargestellt wurden. Eine Mahlzeit in Düsseldorf ist mehr als viermal so teuer wie in Gelsenkirchen oder Hagen und doppelt so teuer wie in Bochum oder Neuss.

Auch die Nachfrage wurde etwas überzeichnet dargestellt, sollte aber zumindest die Größenordnung widerspiegeln.

Mit bloßem Blick auf die Daten erkennt man eine positive Korrelation zwischen Preis und Nachfrage. Je teurer die Mahlzeiten, desto größer die Nachfrage. Auch die einfache OLS Regression ($Nachfrage_i = Cons + \alpha Preis_i + \epsilon_i$) bestätigt diesen kontraintuitiven Zusammenhang. Der Koeffizient der Variable $Preis_i$, α , ist signifikant positiv. Diesen Zusammenhang würden wir so nicht erwarten, da dieses Ergebnis allen Erkenntnissen der Mikroökonomie widerspricht.



	stadt	zeit	preis	nachfrage	ID
1	Bochum	2014	27	2.5	Bochum
2	Bochum	2015	30	2.2	Bochum
3	Duesseldorf	2014	60	4	Duesseldorf
4	Duesseldorf	2015	65	3.6	Duesseldorf
5	Gelsenkirchen	2014	15	2	Gelsenkirchen
6	Gelsenkirchen	2015	20	1	Gelsenkirchen
7	Hagen	2014	12	1.5	Hagen
8	Hagen	2015	15	1	Hagen
9	Neuss	2014	30	3	Neuss
10	Neuss	2015	35	2.7	Neuss

Abbildung 4.3: Implementierung der Schätzung in STATA

Grafik 4.3 zeigt das Ergebnis der Analyse. Der Koeffizient der Variable "preis" ist trotz der geringen Zahl an Beobachtungen signifikant positiv. Denkt man allerdings noch einmal genauer nach, dann wird man zu der Schlussfolgerung kommen, dass die Städte nicht so ohne weiteres vergleichbar sind. Ein wesentlicher Unterschied liegt

in der Qualität der Restaurants. Düsseldorf ist eine attraktive Stadt und sollte problemlos herausragende Köche aus ganz Deutschland anziehen. Entsprechend höher ist auch die durchschnittliche Qualität der angebotenen Mahlzeiten, was ebenfalls einen Einfluss auf die Nachfrage haben sollte. Ohne Berücksichtigung dieses Effekts ist nicht klar, ob nun der Preis oder die Qualität die Veränderung der Nachfrage bestimmt. Der Koeffizient der Variable *preis* könnte natürlich auch den Effekt der Qualität auffangen. Letztere Variable müsste entsprechend im Modell berücksichtigt werden, was allerdings nicht geht, da die Qualität nicht beobachtbar ist. Gehen wir davon aus, dass jede Stadt einen über die Zeit hinweg konstanten Qualitätsstandard aufweist. Beispielsweise ist das Qualitätsniveau, das ein Restaurant zum Überleben in Düsseldorf braucht höher als der Qualitätsstandard der in Hagen oder Gelsenkirchen erreicht werden müsste. Dieser unbeobachtete Mindeststandard gilt für alle Restaurants und könnte über die Bildung von Differenzen herausgerechnet werden.

Hierfür bräuchten wir allerdings ein Panel mit mindestens zwei Beobachtungen für jede Stadt. Wir haben einen solchen Datensatz generiert. Für jede Stadt liegen nun Beobachtungen für die Jahre 2014 und 2015 vor. Bevor wir den Zusammenhang schätzen können, müssen wir STATA allerdings zunächst beibringen, wie zwischen den Dimensionen *Querschnitt* und *Zeit* unterschieden werden kann. Der Querschnitt wird über die Variable *stadt* identifiziert und die Zeit über die Variable *zeit*.

Es gibt einen Befehl, der bewirkt, dass STATA die beiden Dimensionen sauber trennen kann. Zwei Indexvariablen werden hierfür benötigt. Außerdem müssen diese Variablen numerisch sein. Dies trifft auf die Variable *zeit* zu, gilt jedoch nicht für die Variable *stadt*, die als STRING Variable kodiert wurde. Zunächst verwenden wir die Befehlsfolge *encode stadt, gen(ID)*, um eine neue Variable *ID* mit der numerisch kodierten Information aus der Variable *stadt* zu generieren. Der Name der Stadt wird im Browser anschließend noch als String angezeigt, klickt man allerdings auf einen beliebigen Wert dieser Variable stellt man fest, dass die Information zusätzlich als Zahl kodiert wurde.

stadt	zeit	preis	nachfrage	ID
Bochum	2014	27	2.5	Bochum
Bochum	2015	30	2.2	Bochum
Dueseldorf	2014	60	4	Dueseldorf
Dueseldorf	2015	65	3.6	Dueseldorf
Gelsenkirchen	2014	15	2	Gelsenkirchen
Gelsenkirchen	2015	20	1	Gelsenkirchen
Hagen	2014	12	1.5	Hagen
Hagen	2015	15	1	Hagen
Neuss	2014	30	3	Neuss
Neuss	2015	35	2.7	Neuss

Abbildung 4.4: Paneldaten

Führt man nun den Befehl `xtset` aus und gibt zusätzlich die beiden Indexvariablen an, dann erkennt STATA automatisch die Datenstruktur. Als Identifikator für den Querschnitt wird die Variable `ID` verwendet und als Zeitindex die Variable `zeit`. Die Information `delta` gibt uns an, in welchen Zeitabständen die Daten erhoben wurden. In unserem Beispiel mit $\Delta \text{zeit} = 1$, also mit Abstand von einem Jahr, was auch als Information im Hauptfenster angezeigt wird (vergleiche Abbildung 4.5).

```

. encode stadt, gen(ID)

.
. xtset ID zeit
    panel variable:  ID (strongly balanced)
    time variable:   zeit, 2014 to 2015
                   delta: 1 unit

```

Abbildung 4.5: Datenaufbereitung

Der Vorteil dieser Vorgehensweise ist die Möglichkeit einfache Zeitreihenoperatoren

zu verwenden. Setzt man den Zusatz *D.* vor den Namen einer Variable nachdem man die Dimension entsprechend gesetzt hat, dann wird der jeweilige Befehl für die differenzierte Variable ausgeführt. Im nächsten Beispiel regressieren wir die erste Differenz der abhängigen Variable auf die erste Differenz der unabhängigen Variable mittels *“reg D.nachfrage D.preis“*. Die Änderung der Nachfrage nach Restaurantbesuchen wird also mit der Änderung der Preise über die Zeit hinweg korreliert. Als Ergebnis bekommen wir den wie erwartet negativen Zusammenhang. Jedoch ist dieser Effekt nicht statistisch signifikant. Interessant ist das Bestimmtheitsmaß bzw. die Teststatistik *R – squared*. Dieses ist im Vergleich zur vorangegangenen Regression nämlich sehr viel kleiner. Der *fixe Effekt* hat einen sehr großen Teil der Daten erklärt.

. reg D.nachfrage D.preis						
Source	SS	df	MS	Number of obs	=	5
Model	.033333349	1	.033333349	F(1, 3)	=	0.33
Residual	.30666667	3	.102222223	Prob > F	=	0.6079
Total	.340000019	4	.085000005	R-squared	=	0.0980
				Adj R-squared	=	-0.2026
				Root MSE	=	.31972
D.nachfrage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
preis						
D1.	-.0833334	.1459325	-0.57	0.608	-.5477557	.381089
_cons	-.1499999	.6293736	-0.24	0.827	-2.152948	1.852948

Abbildung 4.6: Schätzung des Modells

Ob der ermittelte Effekt von Preis auf Nachfrage nun kausal ist oder nicht, das lässt sich aus dieser Regression nicht ableiten. Selbst wenn der Koeffizient signifikant wäre, könnte ja noch immer eine Änderung der Qualität zwischen den Jahren 2014 und 2015 zu einer Scheinkausalität geführt haben. In der Regression wurde eben nur der über die Zeit hinweg konstante Effekt kontrolliert und eventuelle Änderungen außer Acht gelassen. Gibt es Effekte, die über die Zeit hinweg konstante und über die Zeit hinweg veränderliche Komponenten umfassen, dann würde nur der konstante Teil bereinigt werden und der variable Teil hingegen weiterhin die Regression verzer-

ren. Beispiel: Ein Absinken der Qualität könnte zu einem Rückgang der Nachfrage führen und dieser Effekt würde nicht über die Differenzenbildung gefiltert werden. Um diesem Rückgang entgegenzuwirken, könnten die Restaurants den Preis entsprechend angepasst haben. Dann wäre noch immer die Qualität und nicht der Preis die eigentliche Ursache für den Rückgang der Nachfrage gewesen. Die Regression würde allerdings noch immer die Preisänderung als Schlüsselvariable für die Reduktion der Nachfrage interpretieren. Für kausale Interpretationen brauchen wir etwas aufwendigere Schätzmethoden, wie beispielsweise die im folgenden Kapitel besprochene "IV Regression".

Literaturverzeichnis

- Cameron, A. C. und Trivedi, P. K. (2005) *Microeconometrics: methods and Applications*, Cambridge University Press.
- Card, D. (1990) *The Impact of the Mariel Boatlift on the Miami Labor Market*, *Industrial and Labor Relations Review*, 43(2), S. 245-257.
- Card, D. und Krueger, A. B. (1990) *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*, *The American Economic Review*, 84(4), S. 772-793.
- Cornwell, C. und Rupert, P. (1998) *Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators*, *Journal of Applied Econometrics*, 3, S. 149-155.
- Dollar, D. und Kraay, A. (2003) *Institutions, trade, and growth*, *Journal of Monetary Economics*, 50(1), S. 133-162.
- Frankel, J. (2005) *On the Renminbi: The Choice between Adjustment under a Fixed Exchange Rate and Adjustment under a Flexible Rate*, NBER Working Paper 11274.
- Frankel, J. und Romer, D. (1999) *Does Trade Cause Growth?*, *American Economic Review*, 89(3), S. 379-399.
- Greene, W.H. (2012) *Econometric Analysis*, Pearson Education Limited.
- Kling, R. (2001) *Interpreting Instrumental Variables Estimates of the Return to Schooling*, *Journal of Business and Economic Statistics*, 19, S. 358-364.
- Mankiw, N. G., Romer, D. und Weil, D. N. (1992) *A Contribution to the Empirics of Economic Growth*, *Quarterly Journal of Economics*, 107(2), S. 407-37.
- Olley, S. G. und Pakes, A. (1996) *The dynamics of Productivity in the Telecommunications Equipment Industry*, *Econometrica*, 64(6), S. 1263-1297.

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

000 000 000 (00/18)

00000-0-00-S1

Alle Rechte vorbehalten
© 2018 FernUniversität in Hagen
Fakultät für Wirtschaftswissenschaften