

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

# Modul 32731

## Angewandte Ökonometrie

### Leseprobe

Einheit 1  
Theorie der Angewandten Ökonometrie

Fakultät für  
**Wirtschafts-**  
**wissenschaft**

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.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m<sup>2</sup>, weiß), hergestellt aus 100 % Altpapier.

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>5</b>
<b>Tabellenverzeichnis</b>	<b>11</b>
<b>I Theorie</b>	<b>13</b>
<b>1 Einführung</b>	<b>15</b>
1.1 Was ist Ökonometrie? . . . . .	15
1.2 Das ökonometrische Modell . . . . .	21
1.3 Literaturempfehlung . . . . .	26
<b>2 Multivariate Regression</b>	<b>27</b>
2.1 Herleitung des OLS Schätzers . . . . .	32
2.2 Beispiel 1: Empirische Anwendung des Solow Modells . . . . .	46
2.3 Hypothesentests . . . . .	54
2.4 Die Güte einer Schätzung . . . . .	58
2.5 Die Wirkung von Gewichten . . . . .	60

2.6	Beispiel 2: Cobb Douglas Produktion . . . . .	62
2.7	Mechanik des OLS Schätzers . . . . .	67
2.7.1	Deskriptive versus stochastische Regression . . . . .	68
2.7.2	Annahmen des OLS Schätzers . . . . .	69
2.7.3	Eigenschaften des OLS Schätzers . . . . .	73
2.7.4	Asymptotische Eigenschaften . . . . .	75
2.8	Effizienz des OLS Schätzers: Das Gauß-Markov Theorem . . . . .	76
2.9	Schätzer für die Varianz $\sigma^2$ . . . . .	80
2.10	Normalverteilung des OLS Schätzers . . . . .	83
<b>3</b>	<b>Erklärung von binären Variablen</b>	<b>85</b>
3.1	Schätzmethoden . . . . .	85
3.1.1	Das Logit-Modell . . . . .	89
3.1.2	Das Probit-Modell . . . . .	91
3.1.3	Das lineare Wahrscheinlichkeitsmodell . . . . .	92
3.2	Schätzung der Koeffizienten durch Maximum Likelihood . . . . .	93
3.3	Interpretation der marginalen Effekte im Probit/Logit Modell . . . . .	96
<b>4</b>	<b>Differenzierte Schätzgleichung</b>	<b>99</b>
4.1	Theoretische Fundierung . . . . .	100
4.2	Hypothetisches Beispiel . . . . .	102
<b>5</b>	<b>IV Schätzung</b>	<b>105</b>

5.1	Theoretische Fundierung . . . . .	106
5.2	Beispiel 3: Offenheit und Pro-Kopf-Einkommen . . . . .	111
<b>6</b>	<b>Fixed- vs. Random-Effects</b>	<b>121</b>
6.1	Theoretische Fundierung . . . . .	122
6.2	Fixed- und Random-Effects Modell . . . . .	128
6.2.1	Implementierung der Schätzer . . . . .	131
<b>7</b>	<b>Diff-in-Diff Schätzung</b>	<b>135</b>
7.1	Theoretische Fundierung . . . . .	137
7.2	Beispiel 4: Die Mariel Flüchtlingskrise . . . . .	144
7.3	Beispiel 5: Die Arbeitsmarkteffekte des Mindestlohns . . . . .	148
<b>8</b>	<b>GMM-Schätzer</b>	<b>155</b>
8.1	(G)MM am Beispiel einer IV-Schätzung . . . . .	158
8.2	Optimale Gewichtungsmatrix . . . . .	168
<b>9</b>	<b>Fazit zu Teil I</b>	<b>171</b>
<b>II</b>	<b>Praktische Anwendungen in R und STATA</b>	<b>173</b>
<b>10</b>	<b>Anwendungen in R</b>	<b>175</b>
10.1	OLS-Regressionen in R . . . . .	175
10.1.1	Die Entwicklungsumgebung RStudio . . . . .	178

10.1.2	OLS-Schätzung in R . . . . .	181
10.2	Probit und Logit Schätzungen in R . . . . .	199
10.2.1	Einfacher Schätzansatz . . . . .	202
10.2.2	Erweiterte Schätzgleichung . . . . .	218
10.3	First Difference Schätzung in R . . . . .	226
10.4	Schätzung mit Instrumentvariablen in R . . . . .	233
10.5	Fixed & Random Effects Schätzungen in R . . . . .	247
10.6	Diff-in-Diff Schätzung in R . . . . .	265
<b>11</b>	<b>Anwendungen in STATA</b>	<b>271</b>
11.1	OLS-Regressionen in STATA . . . . .	271
11.2	Probit und Logit Schätzungen in STATA . . . . .	294
11.2.1	Einfacher Schätzansatz . . . . .	295
11.2.2	Erweiterte Schätzgleichung . . . . .	313
11.3	First Difference Schätzung in STATA . . . . .	316
11.4	Schätzung mit Instrumentvariablen in STATA . . . . .	322
11.5	Fixed & Random Effects Schätzungen in STATA . . . . .	335
11.6	Diff-in-Diff Schätzung in STATA . . . . .	352
	Literaturverzeichnis . . . . .	361

## 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

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

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.<sup>1</sup> Außerdem enthält die Gleichung 1.2 einen zusätzlichen Störterm  $\varepsilon$ , 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.

---

<sup>1</sup> 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:

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

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  $\varepsilon$  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  $\varepsilon$  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  $\beta_3$  sind also die Parameter des theoretischen Modells und beschreiben die Richtung und die Intensität der Beziehung zwischen dem Lohn und dessen Determinanten. Variable  $\varepsilon$  erfasst hier alle *übrigen* Effekte auf den Lohn einer Person. Der Umgang mit diesem *Fehlerterm* oder *Störterm* ( $\varepsilon$ ) 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  $\beta$ , die den wahren Zusammenhang angeben und unbeobachtbar sind, werden durch die geschätzten Parameter  $b$  approximiert. Der wahre Störterm  $\varepsilon$  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.

Untergliederung der Vektoren in

$$x = [x'_1 \ x'_2] \quad (5.12)$$

$$z = [z'_1 \ x'_2] \quad (5.13)$$

wobei der Vektor  $z_1$  die Instrumente für die Regressoren in  $x_1$  enthält. Im obigen Beispiel wäre das jeweils eine Variable in  $x_1$  und das dazugehörige Instrument in  $z_1$ .<sup>1</sup> Für die übrigen Variablen wird angenommen, dass sie exogen sind. Diese in  $x_2$  zusammengefassten Variablen tauchen ebenfalls in der Instrumentenmatrix  $z$  auf. Die exogenen Regressoren werden als Instrument für sich selbst verwendet.

## 5.2 Beispiel 3: Offenheit und Pro-Kopf-Einkommen

Ein bekanntes Beispiel für eine IV Regression liefert das Papier von Frankel und Romer (1999). Ausgangspunkt dieser Studie ist die Frage nach dem Einfluss von Globalisierung auf das Pro-Kopf-Einkommen einer Ökonomie. Die klassische Handelsliteratur sieht einen positiven Einfluss einer Handelsliberalisierung auf das Durchschnittseinkommen vor. Länder, die sich auf den komparativen Vorteil spezialisieren, können Handelsgewinne realisieren. Die neuere Handelsliteratur zeigt außerdem, dass positive Skalenerträge ebenfalls mit einem Anstieg des Pro-Kopf-Einkommens in Verbindung gebracht werden können. Diesen Zusammenhang empirisch zu untersuchen ist Ziel der Studie von Frankel und Romer (1999). Dabei wird das Problem der Endogenität explizit in einem IV Regressionsschätzer berücksichtigt. Obwohl es viele Gründe für einen Einfluss von Handel auf das Pro-Kopf-Einkommen einer Ökonomie gibt, existieren mindestens genauso viele Gründe für einen signifikanten Einfluss vom Entwicklungsstadium einer Ökonomie auf den Handel. Reichere Länder handeln im Durchschnitt einen größeren Anteil ihres Bruttoinlandsprodukts. Frankel und Romer

---

<sup>1</sup> Das Modell könnte auch über-spezifiziert sein, so dass mehr Instrumente als endogene Variablen einbezogen werden. Dieser Fall wird zunächst allerdings nicht betrachtet.

adressieren dieses Problem durch die Konstruktion eines neuen Instruments basierend auf der folgenden Vorüberlegung.

Zunächst wird der geschätzte Zusammenhang als

$$\ln Y_i = \alpha + \beta T_i + \gamma W_i + \epsilon_i \quad (5.14)$$

beschrieben. Die abhängige Variable ist das logarithmierte Pro-Kopf-BIP in Land  $i$ . An dieser Stelle sei angemerkt, dass die Zeit vernachlässigt wird, so dass der Zusammenhang im Querschnitt über die verschiedenen Länder hinweg geschätzt wird. Das logarithmierte Pro-Kopf-Einkommen in einem Land wird durch den Globalisierungsgrad in Land  $i$ ,  $T_i$  (Trade), dem nationalen Handelsvolumen  $W_i$  (within country trade) und weiteren Faktoren erklärt. Sofern diese weiteren Faktoren nicht als Kontrollvariablen berücksichtigt werden, werden sie über den Fehlerterm  $\epsilon_i$  aufgefangen.

Frankel und Romer (1999) argumentieren, dass die Variablen  $T$  und  $W$  ihrerseits wiederum über die Variablen Proximität,  $P$ , und Ländergröße,  $S$ , erklärt werden. Proximität wird über die Variable Distanz approximiert und die Ländergröße wahlweise durch Bevölkerung und/oder der Fläche eines Landes erklärt:

$$T_i = \psi + \phi P_i + \delta_i \quad (5.15)$$

$$W_i = \eta + \lambda S_i + \nu_i \quad (5.16)$$

Die erste Gleichung besagt, dass der Handel einer Ökonomie unter anderem auch von der Distanz zu ihren Handelspartnern abhängt. Es sei vorweggenommen, dass diese Gleichung in einem späteren Schritt bilateral für verschiedene Handelspartner geschätzt wird, da die Distanz vom jeweils betrachteten Handelspartner abhängt. Diese erste Gleichung ist also streng genommen nicht schätzbar und dient nur der Veranschaulichung des Zusammenhangs zwischen Distanz und Handel.

Die zweite Gleichung besagt, dass der interne Handel eines Landes von der Größe der Ökonomie abhängt. Dieser Zusammenhang ist einleuchtend. Je größer ein Land, desto

mehr Güter werden national zwischen verschiedenen Regionen gehandelt. Streng genommen ist auch hier die Distanz zwischen den Regionen eine wichtige Determinante des bilateralen Handels zweier Regionen. In der praktischen Umsetzung lässt sich dieser Zusammenhang jedoch nicht ohne weiteres schätzen. Es gibt keine Distanz auf Länderebene, sondern nur bilateral zwischen zwei Handelspartnern. Der Warenverkehr über die Grenzen eines Landes hinweg kann eindeutig durch den Zoll bestimmt werden, innerhalb eines Landes gibt es keine Institution, die diese Warenströme dokumentiert. Frankel und Romer schlagen vor, die Gleichungen zusammenzufassen zu

$$\ln Y_i = \alpha + \beta T_i + \gamma[\eta + \lambda S_i + v_i] + \epsilon_i \quad (5.17)$$

$$= \alpha + \beta T_i + \gamma\eta + \gamma\lambda S_i + [\gamma v_i + \epsilon_i] \quad (5.18)$$

Auf diese Weise wird der Einfluss des internen Handels einer Ökonomie auf das Pro-Kopf-Einkommen bereits über die Größe des Landes berücksichtigt. Es fällt auf, dass verschiedene Parameter nicht mehr getrennt voneinander identifiziert werden können. Die Konstante besteht aus mehreren Parametern und wird nur noch gemeinsam als  $K = \alpha + \gamma\eta$  geschätzt. Für den Fehlerterm gilt  $e_i = [\gamma v_i + \epsilon_i]$  und der Koeffizient der Variable Ländergröße ist abhängig von  $\gamma$  und  $\lambda$ . Frankel und Romer (1999) argumentieren, dass der Fehlerterm  $e_i$  nicht mit den Regressoren  $S_i$  oder  $P_i$  korreliert. Tatsächlich gibt es bei dieser Annahme keinen offensichtlichen Grund zur Sorge. Die Tatsache, dass außer  $\beta$  keiner der Koeffizienten isoliert identifiziert werden kann, ist ebenfalls nicht schlimm, da uns nur der Koeffizient  $\beta$  interessiert. Die anderen Parameter gehören zu den Kontrollvariablen, die wir im weiteren Verlauf nicht weiter interpretieren werden. Dennoch bleibt das Endogenitätsproblem. Wir wissen, dass  $\beta$  nicht einwandfrei identifiziert werden kann ohne Berücksichtigung des Endogenitätsproblems. Hierfür hilft uns die Gleichung 5.15, die einen Zusammenhang zwischen Distanz und Handelsvolumen vorhersagt. Allerdings ist der Zusammenhang nicht so

ohne weiteres schätzbar. Aus diesem Grund wird der Zusammenhang in einer bilateralen Gravitationsgleichung als

$$\ln(\tau_{ij}/GDP_j) = a_0 + a_1 \ln D_{ij} + a_2 \ln S_i + a_3 \ln S_j + E_{ij} \quad (5.19)$$

formuliert. Das bilaterale Handelsvolumen  $\tau_{ij}$  wird mit der Größe des Herkunftslands  $GDP_j$  relativiert und nach Logarithmierung über die bilaterale Distanz  $D_{ij}$ , die Größe des Herkunftslands  $S_i$  und die Größe des Ziellands  $S_j$  erklärt. Auch hier gilt wieder, dass weitere Einflussgrößen zunächst nicht weiter berücksichtigt werden und entsprechend im Fehlerterm  $E_{ij}$  enthalten sind. Die eigentliche Schätzgleichung lautet

$$\ln\left(\frac{\tau_{ij}}{GDP_j}\right) = a_0 + a_1 \ln D_{ij} + a_2 \ln N_i + a_3 \ln A_i + a_4 \ln N_j + a_5 \ln A_j + a_6(L_i + L_j) + E_{ij} \quad (5.20)$$

- Die Variable  $a_0$  ist die Konstante in der Schätzung
- Die Variable  $D_{ij}$  ist ein Maß für die Distanz zwischen den Ländern  $i$  (Herkunftsland) und  $j$  (Zielland)
- Die Variablen  $N_i$  und  $N_j$  sind Maße für die Bevölkerungsgrößen der Länder  $i$  und  $j$
- Die Variablen  $A_i$  und  $A_j$  sind Maße für die flächenmäßigen Größen der Länder  $i$  und  $j$
- Die Variablen  $L_i$  und  $L_j$  sind Dummies für die Isolation des Landes. Der Koeffizient  $a_6$  wird gemeinsam für Ziel und Herkunftsländer geschätzt, so dass es keinen ziel- und herkunftslandspezifischen Effekt gibt.

Die Koeffizienten dieser Schätzung werden in Tabelle 11.1 präsentiert. Die Tabelle ist dem Papier entnommen und beinhaltet sowohl die Koeffizienten des in Gleichung 5.21 beschriebenen Modells (Spalte 2), als auch eine im Papier beschriebene Spezifikation mit zusätzlichen Interaktionen (Spalte 3). Wir werden die Diskussion im weiteren

Tabelle 5.1: Schätzergebnisse für bilaterale Handelsgleichung

Konstante [ $a_0$ ]	-6.38 (0.42)	5.10 (1.78)
Distanz (log) [ $a_1$ ]	-0.85 (0.04)	0.15 (0.30)
Bevölkerungsgröße (log) [ $a_2$ ]	-0.24 (0.03)	-0.29 (0.18)
Ländergröße (log) [ $a_3$ ]	-0.12 (0.02)	-0.06 (0.15)
Bevölkerungsgröße (log) [ $a_4$ ]	0.61 (0.03)	-0.14 (0.18)
Ländergröße (log) [ $a_5$ ]	-0.19 (0.02)	-0.07 (0.15)
Isolation (dummy) [ $a_6$ ]	-0.36 (0.08)	0.33 (0.33)

Notes: Die abhängige Variable ist  $\ln(\tau_{ij}/GDP_i)$ . In der zweiten und dritten Spalte werden Koeffizienten und Standardfehler (Wert in Klammern) zu den in der ersten Spalte genannten Variablen berichtet. Die Koeffizienten der zweiten Spalte sind die direkten Effekte ohne Interaktion. In der dritten Spalte werden Koeffizienten der Interaktionen zwischen der jeweiligen Variable und einem Dummy für eine gemeinsame Grenze zwischen dem jeweils betrachteten Länderpaar berichtet.

Quelle: Frankel und Romer (1999)

Verlauf auf die Spezifikation 5.21 und die dazugehörigen Ergebnisse in Spalte 2 beschränken.

Basierend auf diesen Schätzergebnissen lässt sich nun eine neue Variable konstruieren, die den bilateralen Handel basierend auf den geografischen Eigenschaften eines Landes bestimmt als

$$\ln \hat{T}_{ij} = a_0 + a_1 \ln D_{ij} + a_2 \ln N_i + a_3 \ln A_i + a_4 \ln N_j + a_5 \ln A_j + a_6(L_i + L_j) \quad (5.21)$$

wobei  $\hat{T}$  den konstruierten Handelsanteil für ein bestimmtes Länderpaar angibt. Für jedes Länderpaar lässt sich auf Grundlage der Distanz und der Größen prognostizieren, wie viel diese Länder handeln. Da wir davon ausgehen, dass Distanz und Größe exogen sind, sollte dieser prognostizierte Handel nicht mit dem Pro-Kopf-Einkommen korreliert sein. Weder die Distanz, noch die Ländergröße sollten durch den Handel oder das Pro-Kopf-Einkommen beeinflusst werden, so dass die Kausalität für diese konstruierte Variable eindeutig ist. Es gibt keinen Grund zur Sorge, dass das Residuum mit dem Instrument, nämlich dem konstruierten Handelsanteil, korreliert ist. Auch für den nationalen Handel wurde kontrolliert, so dass es auch über diesen Kanal nicht zu einem indirekten Einfluss kommen kann. Frankel und Romer (1999) zeigen eine Vielzahl an Tests auf Korrelation zwischen dem Instrument und den zu instrumentierenden Variablen.

Gemäß dieser Tests sind die Voraussetzungen für eine IV-Regression erfüllt. Es wurde eine Variable konstruiert, die hoch korreliert ist mit der interessierenden Variable. Größere Länder, die näher beieinander liegen, handeln mehr miteinander, so dass der konstruierte Handel und der tatsächliche Handel hoch korreliert sind. Da der konstruierte Handel ausschließlich über geografische Variablen konstruiert wurde, die nicht durch die abhängige Variable beeinflusst werden können, gibt es keinen Kanal durch den das Pro-Kopf-Einkommen das Instrument in der IV Regression beeinflusst.

Bislang haben wir gesehen, dass die Ländergrößen als Proxies für den nationalen Handel herangezogen werden können, da diese Variablen nicht mit dem Fehlerterm korreliert sein sollten. Außerdem haben wir mit der Variable  $\hat{T}$  ein mögliches Instrument, das zumindest schon einmal hoch mit der zu instrumentierenden Variable korreliert ist. Ob das Instrument tatsächlich für die Anwendung geeignet ist, wird anhand weiterer Teststatistiken überprüft sobald die Schätzergebnisse für das Hauptmodell (Gleichung 5.18) vorliegen. Diese Ergebnisse werden in Tabelle 11.1 präsen-

tiert.

Zunächst fällt auf, dass die Autoren vier verschiedene Versionen des Modells schätzen, nämlich zwei OLS und zwei IV Regressionen. Im Zusatztext unter der Tabelle wird beschrieben, worin sich die Regressionen unterscheiden. Die ersten beiden Regressionen sind für ein Sample von insgesamt 150 Ländern geschätzt worden und die Regressionen in Spalte 4 bis 5 wurden für ein verkleinertes Sample in Anlehnung an das Beispiel von Mankiw et al. (1992) geschätzt. Für beide Samples werden jeweils Regressionen ohne und mit Berücksichtigung des Endogenitätsproblems miteinander verglichen.

Die Koeffizienten in der ersten Reihe zeigen einen signifikant positiven Einfluss der Offenheit und dem Pro-Kopf-Einkommen einer Ökonomie. Gemäß der Daumenregel ergibt der Vergleich der Koeffizienten mit den Standardfehlern, dass die Koeffizienten zumindest schwach signifikant sind. Beim Vergleich von OLS und IV fällt auf, dass die T-Werte durch die Instrumentierung sinken. Für die OLS Regressionen bekommen wir T-Werte, die deutlich größer als 2 sind. Die T-Werte für die IV-Regressionen sind hingegen sehr nahe am kritischen Wert 2. Doch woran sieht man, ob die Instrumente auch tatsächlich geeignet sind? Hierfür wird in der Regel die F-Statistik der ersten Regressionsstufe interpretiert. In der ersten Stufe wird zunächst einmal das Instrument auf die zu instrumentierende Variable regressiert. Als Daumenregel gilt, dass die F-Statistik dieser Regression größer 10 sein sollte. Für das große Sample ist diese Voraussetzung erfüllt, für das kleinere Sample finden wir allerdings einen Wert kleiner 10. Da beide Fälle sehr nahe am Grenzwert dieser Daumenregel liegen, ist es ratsam hier einen genaueren Test durchzuführen.

Zumindest im großen Sample spricht nichts gegen die Validität des Instruments, so dass wir uns nun an die Interpretation der Koeffizienten machen können. Es ist auffällig, dass der Koeffizient für die IV Regression im Vergleich zum Koeffizienten der einfachen OLS Regression fast doppelt so groß ist. Dies ist ein erstaunliches Ergebnis, da es uns zeigt, dass wir ohne die Berücksichtigung des Endogenitätsproblems

Tabelle 5.2: Handel und Pro-Kopf-Einkommen

	OLS	IV	OLS	IV
Offenheit	0.85 (0.25)	1.97 (0.99)	0.82 (0.32)	2.96 (1.49)
Bevölkerungsgröße (log)	0.12 (0.06)	0.19 (0.09)	0.21 (0.10)	0.35 (0.15)
Ländergröße (log)	-0.01 (0.06)	0.09 (0.10)	-0.05 (0.08)	0.20 (0.19)
Konstante	7.40 (0.66)	4.96 (2.20)	6.95 (1.12)	1.62 (3.85)
Samplegröße	150	150	98	98
$R^2$	0.09	0.09	0.11	0.09
F-Statistik (erste Regressionsstufe)		13.13		8.45

Die abhängige Variable ist das Pro-Kopf-Einkommen im Jahr 1985. Länder-sample in Spalte 2 und 3 (volles Sample) unterscheidet sich vom Sample in Spalte 4 und 5 (Mankiw et al. (1992)). Standardfehler in Klammern.

Quelle: Frankel und Romer (1999)

den Einfluss von Handel auf das Einkommen einer Ökonomie erheblich unterschätzt hätten. Für das kleinere Sample wird dieser Unterschied sogar noch gravierender.

**Erweiterung der Schätzung um Institutionen.** Als weiteres Beispiel für eine Anwendung des IV Schätzers schauen wir uns die Ergebnisse aus Dollar und Kray (2003) an. Die Autoren schätzen den Zusammenhang zwischen dem Pro-Kopf-Einkommen einer Ökonomie, deren Institutionen und der Offenheit einer Ökonomie.

Die Schätzgleichung wird wie folgt motiviert. Länder mit besseren Institutionen sollen stärker wachsen können als Länder mit schlechten Institutionen. Außerdem wurde sehr häufig gezeigt, dass internationaler Handel zu mehr ökonomischem Wachs-

tum führt. Die zu testende Hypothese lautet: Beide Einflussfaktoren müssen in einer empirischen Analyse gleichermaßen berücksichtigt werden. Zudem wird argumentiert, dass diese beiden Variablen höchst endogen sind. Für Institutionen gilt, dass mit zunehmendem ökonomischen Wachstum auch das Verlangen nach besseren Institutionen zunimmt. Dieser Einfluss muss also entsprechend in der Regression berücksichtigt werden, doch es gilt wieder, dass das Maß für die Qualität der Institutionen schwer zu messen ist. Besonders die Vergleichbarkeit über die Grenzen der Länder hinweg muss gewährleistet sein. Die Autoren verwenden Einschätzungen von Experten, um diese Variable zu erstellen.

## 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.
- Kleiber, C. und Zeileis, A. (2008) *Applied Econometrics with R*, Springer-Verlag, New York
- 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.

002664909  
(10/21)

**32731-6-01-S 1**



Alle Rechte vorbehalten  
© 2021 FernUniversität in Hagen  
Fakultät für Wirtschaftswissenschaft