universität wien

MAGISTERARBEIT

Titel der Magisterarbeit

„Politik und Finanz“

Eine Untersuchung der Verwendung von Methoden zur Verhinderung von Endogenität anhand wissenschaftlicher Arbeiten im Gebiet der politisch motivierten Kreditvergabe.

Verfasser

Simon Baumgartner BA

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften
(Mag. rer. soc. oec.)

Wien, im August 2011

Studienkennzahl lt. Studienblatt:	A 066 914
Dissertationsgebiet lt. Studienblatt:	Magisterstudium Internationale Betriebswirtschaft
Betreuerin / Betreuer:	Priv.-Doz. Dr. Alex Stomper

für Andreea

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Ein Skandal, wie der der Hypo Alpe Adria Bank ab 2007, verdeutlicht, wie weit Politik und Finanz miteinander verknüpft sind. Durch diese Art von Skandalen kommt aber nur die Spitze des Eisberges ans Licht. Daher geben diese Skandale, meiner Erfahrung nach, oft ein verzerrtes Bild der tatsächlichen Situation wieder. Eine kurze Recherche zeigte, dass der Anteil an staatlich kontrollierten Banken erstaunlich hoch ist: Eine internationale Studie ergab, dass im Jahr 1995 42% des Eigenkapitals der zehn größten Banken eines Landes in staatlichem Besitz waren.[LaPorta et al., 2002, S.267] Mich interessiert es, wie sich, abseits der Skandale, die Einflussnahme der Politik auf diese Banken auswirkt. Welches Muster zeigen die Daten? Gibt es systematische Effekte? Wirkt sich die Einflussnahme positiv, negativ oder gar nicht aus? Weitere Frage lauten: Wie verlässlich sind Studien über den Einfluss der Politik auf Banken? Was tun Forscher um unverzerrte Ergebnisse zu erhalten.

Erläuterung und Abgrenzung des zu untersuchenden Themas. Ziel dieser Arbeit ist es, die Verwendung von Methoden zur Verhinderung von Endogenität anhand von wissenschaftlichen Artikeln aus dem Bereich *Politik und Finanz* zu analysieren. Der Bereich *Politik und Finanz* ist sehr breit und es bedarf einer Eingrenzung des Themas. Die Arbeit konzentriert sich auf Zusammenhänge zwischen der Politik und Banken. Genauer werden Artikel bearbeitet, die die Kreditvergabe von staatlichen Banken untersuchen und überprüfen, inwieweit die Kreditvergabe politisch motiviert ist.

Gliederung. In Kapitel zwei wird dem Leser Grundlegendes aus der Ökonometrie vermittelt: die Methode der kleinsten Quadrate, die Ermittlung von Schätzern, die Eigenschaften von OLS-Schätzer, die Kriterien für Erwartungstreue und Konsistenz von Schätzern, Dummy Variablen und Panel Daten. Das dritte Kapitel erklärt wie Endogenität entsteht, was die Auswirkungen sind und mit welchen Methoden sie verhindert werden kann - drei Methoden werden dabei näher vorgestellt. In Kapitel vier wird zunächst ein kurzer Überblick über die Theorie der politisch motivierten Kreditvergabe gegeben. Anschließend wird die wissenschaftliche Literatur im Bereich der Kreditvergabe von staatseigenen Banken (i) inhaltlich analysiert und (ii) auf die Verwendung der, in Kapitel drei vorgestellten Methoden zur Verhinderung von Endogenität hin, untersucht. Dabei wird besonders auf zwei Artikel von Halling et al. [2010] und Cole [2007] eingegangen, da diese ein natürliches Experiment, bzw. eine Instrumentvariable benützen, um Endogenität zu verhindern. Im fünften Kapitel wird eine Schlussfolgerung gezogen.

2 Grundlegendes aus der Ökonometrie

Um Lesern mit wenigen Vorkenntnissen ein besseres Verständnis zu gewähren, werden in diesem Kapitel Grundlagen aus der Ökonometrie kurz wiederholt. Dazu gehören die Aufstellung eines linearen Regressionsmodelles für die Grundgesamtheit, die Methode der kleinsten Quadrate, das Schätzen eines Parameters und die zugrunde liegenden Annahmen, um unverzerrte und konsistente Schätzer für die interessierenden Parameter zu erhalten. Außerdem wird die Funktionsweise von Dummy Variablen erklärt.

2.1 OLS Methode bei einem linearen Regressionsmodell

Dieses Unterkapitel hat zum Ziel, jenen Lesern, die wenige Vorkenntnisse im Fach Ökonometrie besitzen, eine kurze Einführung in die Grundthematik zu geben. Jedes Regressionsmodell hat im Grunde zum Ziel, Zusammenhänge zwischen einer abhängigen Variable und, gegebenenfalls, mehreren erklärenden Variablen in der Grundgesamtheit zu messen oder zu schätzen. Die **Methode der Gewöhnlichen Kleinsten Quatrate (OLS)**, auf Englisch Ordinary Least Squares, ist eine Methode, um eben diese Zusammenhänge zu ermitteln. Damit diese Methode funktioniert und sinnvolle Messwerte ergibt, müssen jedoch bestimmte Bedingungen erfüllt sein. Dabei ist grundsätzlich zwischen einer Regression auf alle Daten einer Grundgesamtheit und einer Regression auf eine aus der Grundgesamtheit gezogene Stichprobe[1] zu unterscheiden. Das erste ist ein Element der deskriptiven Statistik,

[1] Oft wird nicht von einer Grundgesamtheit, sondern von einem datengenerierenden Prozess (DGP) gesprochen. Dieser ist der Mechanismus, welcher die beobachteten Daten generiert hat. [Davidson and MacKinnon, 2003, S.87]

das zweite ein Element der schließenden Statistik.[2] Die Ergebnisse für die gesuchten Parameter (Schätzer) werden sich mit hoher Wahrscheinlichkeit unterscheiden und die Interpretation der Ergebnisse ist eine andere.[Davidson and MacKinnon, 2003] Zunächst wird die Funktionsweise der OLS-Methode anhand einer Regression auf die *Grundgesamtheit* erläutert, später wird die gleiche Methode auf die Regression einer *Stichprobe* angewendet. Die Vorgehensweise ist ganz vereinfacht folgende:

1. Modellierung des Zusammenhanges
2. Erhebung der Daten
3. Ermittlung der gesuchten Parameter durch Auswertung der Daten

In den folgenden Unterkapiteln wird die Erklärung, zwecks Strukturierung, in diese drei Kategorien unterteilt.

2.1.1 OLS-Regression auf eine Grundgesamtheit.

Modellierung des Zusammenhanges. Die folgende Gleichung ist ein allgemeines Beispiel für ein lineares Regressionsmodell der Grundgesamtheit. Da sich dieses Modell auf die komplette Grundgesamtheit bezieht, ist es ein Element der deskriptiven Statistik und wird mit „Population Regression Function", kurz PRF, bezeichnet.

$$\begin{aligned} y_i &= f(x_{i,1}, x_{i,2}, ..., x_{i,K}) + \epsilon_i \;\text{ für alle } i \in \{1, 2, ..., N\} \quad (1) \\ y_1 &= \beta_1 x_{1,1} + \beta_2 x_{1,2} + ... + \beta_K x_{1,K} + \epsilon_1 \\ y_2 &= \beta_1 x_{2,1} + \beta_2 x_{2,2} + ... + \beta_K x_{2,K} + \epsilon_2 \\ &\vdots \\ y_N &= \beta_1 x_{N,1} + \beta_2 x_{N,2} + ... + \beta_K x_{N,K} + \epsilon_N \end{aligned}$$

[2] In der Schließenden Statistik, auf Englisch *Inference*, wird versucht von Eigenschaften einer Untermenge, z.B. einer Stichprobe, auf Eigenschaften der Obermenge N, der Grundgesamtheit, zu schließen. Dies wird auch induktiver Schluss genannt. Hierbei gilt es zu beachten, dass ein solcher Zusammenhang aus logischen Gesichtspunkten nicht verifiziert, sondern lediglich falsifiziert werden kann.

Das gleiche Modell in Matrixschreibweise:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ x_{2,1} & x_{2,2} & \dots & x_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,K} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

N bezeichnet die Anzahl der Elemente in der Grundgesamtheit. In Gleichung (1) wird die abhängige Variable y_i in zwei verschiedene Komponenten aufgeteilt: Die erste ist eine systematische Komponente $\hat{y}_i = f(x_{i,1}, x_{i,2}, ..., x_{i,K})$, in welcher ein Teil von y_i als eine Funktion der erklärenden Variablen $x_{i,j}$ dargestellt ist. In Gleichung (1) ist die systematische Komponente die Summe des Produktes von x_j und einem Faktor β_j für alle $j = \{1, ..., K\}$. Die zweite Komponente ist der Störterm ϵ_i, welcher als Differenz zwischen y_i und $f(x_{i,1}, x_{i,2}, ..., x_{i,K})$ definiert ist. Der Störterm existiert, da es nahezu unmöglich ist, jeden Einfluss auf y in den erklärenden Variablen x_j einzufangen. [Greene, 2002, S.8] Wie aus dem Namen schon zu entnehmen ist, wird beim Linearen Regressionsmodell ein linearer Kausalzusammenhang zwischen der abhängigen Variable und den erklärenden Variablen unterstellt. Dies wird durch Annahme OLS 1 impliziert.

Annahme OLS 1. *Das gewählte Modell stellt einen linearen Zusammenhang zwischen der abhängigen Variable und den erklärenden Variablen her. Die Funktion $f(x_{i,1}, x_{i,2}, ..., x_{i,K})$, welche die systematische Komponente von y_i darstellt ist linear in den Parametern β_j für alle $j = \{1, ..., K\}$ und $i = \{1, ..., N\}$.*

Welche Variable die abhängige Variable und welche die erklärenden Variablen sind, ergibt sich, genau wie die genaue Spezifikation der Funktion $f(x_1, x_2, ..., x_K)$, aus der dem Thema zugrundeliegenden Theorie. Grundsätzlich werden natürlich sinnvolle Kausalzusammenhänge zwischen der erklärten- und den erklärenden Variablen gesucht. Jedoch ist oft nicht klar abgrenzbar, welche Variable abhängig und welche erklärend ist. Bei einer falschen Spezifikation des Modells läuft man Gefahr auf Scheinkorrelationen[3] zu stoßen, welche die Ergebnisse verfälschen. Gleichung

[3]Scheinkorrelationen sind Korrelationen zwischen Variablen, welche *nicht* aufgrund eines Kau-

(2) stellt exakt das gleiche Modell in Matrixschreibweise dar. Dabei ist $\mathbf{X}$ eine $N \times K$ Matrix der erklärenden Variablen, $\mathbf{y}$ und $\boldsymbol{\epsilon}$ sind $N \times 1$ Vektoren für die abhängige Variable und den Störterm und $\boldsymbol{\beta}$ ist ein Vektor mit Dimension $K \times 1$ für die Parameter der Grundgesamtheit. Vektoren und Matrizen sind jeweils fett gedruckt dargestellt. Offensichtlich ist die Schreibweise in Gleichung (2) weitaus effizienter als die vorangegangenen. Daher wird ab sofort immer auf die Matrixschreibweise zurückgegriffen, wenn es möglich ist. Es wird zwischen Modellen mit und ohne Interzept unterschieden, dabei kann das Modell in den Gleichungen (1) und (2) sowohl das eine als auch das andere darstellen. Geht es um ein Modell mit Interzept gilt einzig $x_{i,1} = 1$ für alle $i = 1, ..., N$, d.h. die erste Spalte von $\mathbf{X}$ enthält nur Einsen. Daraus folgt die Interpretation von β_1 als Interzept. In einem Interzept sind die Effekte enthalten, welche (i) systematisch sind und (ii) nicht von einer der erklärenden Variable abhängen, also konstant sind.

Erhebung der Daten. Bei einer Regression auf die komplette Grundgesamtheit werden sämtliche Daten aller N Elemente der Grundgesamtheit(z.B. alle Frauen in Europa, alle Einwohner Österreichs, usw.) erhoben, welche zur Berechnung der gesuchten Parameter gebraucht werden. Im Falle von Gleichung (2) sind das $N \times (K + 1)$ Werte, d.h. alle Daten für y_i und $x_{i,j}$ der Grundgesamtheit. Im Falle der Regression auf die Grundgesamtheit sind alle Variablen deterministisch[4].

Ermittlung der gesuchten Parameter. Gesucht wird nach dem Vektor der Parameter $\boldsymbol{\beta}$. Die erklärenden Variablen $\mathbf{X}$ sowie der Vektor der abhängignen Variablen $\mathbf{y}$ sind bekannt. Jedes Element des Vektors der Störterme $\boldsymbol{\epsilon}$ ist eine Funktion der vorigen Variablen, $\epsilon_i = y_i - \sum_{j=1}^{K} \beta_j x_{i,j}$. In Matrixschreibweise lautet diese Funktion $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Die OLS-Methode ermittelt $\boldsymbol{\beta}$ indem die der Summe der quadrierten Elemente von $\boldsymbol{\epsilon}$ minimiert werden, d.h. $\min\limits_{\beta_j} \sum_{i=1}^{N} \epsilon_i^2$ für alle $i \in \{1, ..., N\}$.

salzusammenhanges bestehen.

[4] Dies ist ein großer Unterschied zur Regression einer zufälligen Stichprobe, die später behandelt wird.

In Matrixschreibweise lautet dieses Problem[5] wie folgt:

$$
\begin{aligned}
\min_{\boldsymbol{\beta}} \boldsymbol{\epsilon}^T\boldsymbol{\epsilon} &= \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \min_{\boldsymbol{\beta}} (\mathbf{y}^T - \boldsymbol{\beta}^T\mathbf{X}^T)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \min_{\boldsymbol{\beta}} \mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}
\end{aligned} \tag{3}
$$

Um zu minimieren, wird der letzte Ausdruck nach $\boldsymbol{\beta}$ abgeleitet und gleich null gesetzt. Daraus entsteht die Bedingung erster Ordnung (4).

$$
\begin{aligned}
\frac{\partial(\boldsymbol{\epsilon}^T\boldsymbol{\epsilon})}{\partial\boldsymbol{\beta}} &= \frac{\partial(\mathbf{y}^T\mathbf{y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \\
&= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{0}
\end{aligned} \tag{4}
$$

Daraus folgt für die Extremstelle von $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{5}$$

Damit $\boldsymbol{\beta}$ ermittelt werden kann, muss $\mathbf{X}^T\mathbf{X}$ invertierbar sein. Die Matrix $\mathbf{X}^T\mathbf{X}$ ist invertierbar, wenn und nur wenn die Determinante der Matrix ungleich null ist. Das ist äquivalent zu der Aussage, dass $\mathbf{X}^T\mathbf{X}$ vollen Rang hat. Für eine Matrix $\mathbf{A}$ gilt $\text{rang}(\mathbf{A}) = \text{rang}(\mathbf{A}^T\mathbf{A})$. Daher ist $\text{rang}(\mathbf{X}^T\mathbf{X}) = K$ äquivalent zu $\text{rang}(\mathbf{X}) = K$. Das ist wiederum äquivalent zu der Aussage, dass die $N \times K$ Matrix $\mathbf{X}$ vollen Spaltenrang hat. Der Spaltenrang ist die Anzahl an linear unabhängigen Spaltenvektoren in $\mathbf{X}$.[6] Zusammengefasst bedeutet dies:

$$
\begin{aligned}
\mathbf{X} \text{ hat vollen Spaltenrang} &\iff \text{rang}(\mathbf{X}) = K \iff \text{rang}(\mathbf{X}^T\mathbf{X}) = K \\
&\iff \mathbf{X}^T\mathbf{X} \text{ hat vollen Rang} \iff \det(\mathbf{X}^T\mathbf{X}) \neq 0 \\
&\iff \mathbf{X}^T\mathbf{X} \text{ ist invertierbar}
\end{aligned}
$$

[5] $\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \sum_{i=1}^{N} \epsilon_i^2$. $\boldsymbol{\epsilon}^T$ ist der transponierte Vektor von $\boldsymbol{\epsilon}$, d.h. aus dem $N \times 1$ Vektor ϵ wird der $1 \times N$ Vektor $\boldsymbol{\epsilon}^T$. Selbiges gilt für $\boldsymbol{\beta}^T$ und $\mathbf{y}^T$. Die Matrix $\mathbf{X}^T$ ist wiederum die transponierte Matrix von $\mathbf{X}$, d.h. für jedes Element aus $\mathbf{X}$ gilt $x_{i,j} = x_{j,i}$ in $\mathbf{X}^T$.

[6] Sämtliche Definition, welche für diese Beweiskizzen benutzt werden stammen aus Strang [1988] und Simon und Blume [1994].

Das führt zur zweiten Annahme OLS 2.

Annahme OLS 2. *Die Spaltenvektoren der Matrix* $\mathbf{X}$ *sind voneinander linear unabhängig. D.h.* $\mathbf{X}$ *hat vollen Spaltenrang.*

$\boldsymbol{\beta}$ ist, wie in (5) zu sehen, eine Funktion der erklärenden Variablen und der abhängigen Variable. Da sowohl $\mathbf{X}$ als auch $\mathbf{y}$ deterministisch sind ist in (5) auch $\boldsymbol{\beta}$ deterministisch. Damit $\boldsymbol{\beta}$ ein Minimum ist muss folgendes gelten:

$$\begin{aligned}\frac{\partial^2(\boldsymbol{\epsilon}^T\boldsymbol{\epsilon})}{\partial\boldsymbol{\beta}^2} &= \frac{\partial(-2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \\ &= 2\mathbf{X}^T\mathbf{X} \text{ ist positiv definit} \end{aligned} \tag{6}$$

(6) gilt, wenn und nur wenn $\mathbf{X}^T\mathbf{X}$ positiv definit ist. Das Kriterium für positiv definite Matrizen lautet $\mathbf{c}^T\mathbf{A}\mathbf{c} > 0$ für alle passend dimensionierten $\mathbf{c} \neq \mathbf{0}$. Aus dem Assoziativgesetz für der Matrixmultiplikation folgt $\mathbf{c}^T\mathbf{X}^T\mathbf{X}\mathbf{c} = (\mathbf{c}^T\mathbf{X}^T)(\mathbf{X}\mathbf{c})$. Durch Ausmultiplizieren erhält man $\sum_{i=1}^N(\sum_{j=1}^K c_j x_{i,j})^2$, wobei $(\sum_{j=1}^K c_j x_{i,j})^2$ nicht negativ sein kann. Daraus folgt dass $\mathbf{c}^T\mathbf{X}^T\mathbf{X}\mathbf{c} \geq 0$ gilt.
Der einzige Fall, in dem $\sum_{i=1}^N(\sum_{j=1}^K c_j x_{i,j})^2 = 0$ entsteht, ist wenn $\sum_{j=1}^K c_j x_{i,j}$ für jedes i null ergibt. Nehmen wir im folgenden Abschnitt an, $\mathbf{c}^T\mathbf{X}^T\mathbf{X}\mathbf{c} = 0$. Annahme OLS 2 besagt, dass die Spaltenvektoren von $\mathbf{X}$ linear unabhängig sind. Die lineare Unabhängigkeit der Spaltenvektoren von $\mathbf{X}$ sagt folgendes aus:

$$\mathbf{X}\mathbf{c} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ x_{2,1} & x_{2,2} & \dots & x_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \dots & x_{N,K} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^K c_j x_{1,j} \\ \sum_{j=1}^K c_j x_{2,j} \\ \vdots \\ \sum_{j=1}^K c_j x_{N,j} \end{pmatrix}$$

$$= c_1 \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{N,1} \end{pmatrix} + c_2 \begin{pmatrix} x_{1,2} \\ x_{2,2} \\ \vdots \\ x_{N,2} \end{pmatrix} + \ldots + c_K \begin{pmatrix} x_{1,K} \\ x_{2,K} \\ \vdots \\ x_{N,K} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \iff c_1 = c_2 = \ldots = c_K = 0$$

Dieser Fall ist durch die Annahme $\mathbf{c} = \mathbf{0}$ ausgeschlossen. Analog dazu gilt $\mathbf{c}^T\mathbf{X}^T = \mathbf{0}^T \iff \mathbf{c}^T = \mathbf{0}^T$. Für beide Faktoren besagt Annahme OLS 2, dass sie den Nullvektor ergeben, wenn und nur wenn $\mathbf{c} = \mathbf{0}$. Daraus folgt, dass kein $\mathbf{c} \neq \mathbf{0}$ existiert, bei dem $\mathbf{Xc} = \mathbf{0}$ gilt (bzw. kein $\mathbf{c}^T \neq \mathbf{0}^T$ sodass $\mathbf{c}^T\mathbf{X}^T = \mathbf{0}^T$). Daher folgt aus $\mathbf{c} \neq \mathbf{0}$ das Statement $\mathbf{c}^T\mathbf{X}^T\mathbf{Xc} \neq 0$. Folglich gilt $\mathbf{c}^T\mathbf{X}^T\mathbf{Xc} > 0$ und daher ist $\mathbf{X}^T\mathbf{X}$ positiv definit. Daher impliziert OLS 2, dass die in (5) ermittelte Extremstelle ein Minimum darstellt.

Es ergibt sich aus der Mechanik der OLS-Methode (*mit* Interzept), dass der Mittelwert der abhängigen Variable gleich dem Mittelwert der systematischen Komponente $\hat{y}$ ist. Diese Aussage folgt aus folgendem Zusammenhang: $\hat{y}_i = y_i - \epsilon_i \implies \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i = \frac{1}{N}\sum_{i=1}^{N}y_i - \frac{1}{N}\sum_{i=1}^{N}\epsilon_i = \mu = \hat{\mu}$, wobei μ das arithmetische Mittel der Grundgesamtheit für y_i ist, und $\hat{\mu}$ das arithmetische Mittel der systematischen Komponente ist. Das gilt, weil $\sum_{i=1}^{N}\epsilon_i = 0$ aus (4) folgt. $2(\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} - \mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})) = \mathbf{0}$ impliziert $\mathbf{X}^T\boldsymbol{\epsilon} = \mathbf{0}$ und die Bedingung bedeutet, dass alle Spaltenvektoren in $\mathbf{X}$ orthogonal zum Vektor der Störterme liegen. Bei einer Regression mit Interzept ist eine Zeile j^* von $\mathbf{X}^T$ nur mit Einsen belegt, daher gilt für das j^*. Element des resultierenden Vektors $\sum_{i=1}^{N}\epsilon_i = 0$. Für die restlichen Elemente des Vektors ergeben sich die sonstigen Bedingungen erster Ordnung $\sum_{i=1}^{N}x_{i,j}\epsilon_i = 0$ für alle $j \in \{1, 2, ..., j^* - 1, j^* + 1, ..., K\}$.

Interpretation der Parameter. Dieser Paragraph dient zur Erläuterung der Interpretation von Ergebnissen der OLS Regression. Es wird bei Annahme OLS 1 gestartet und (1) wird als das richtig spezifizierte Modell angenommen. Für dieses Modell wird der bedingte Erwartungswert genommen.

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
\mathbb{E}(\mathbf{y} \mid \mathbf{X}) &= \mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \mid \mathbf{X}) \\
&= \mathbb{E}(\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}) + \mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})
\end{aligned}$$

Der Schritt $\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}) + \mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})$ ist aufgrund der Linearität des Operators des bedingten Erwartungswert möglch. An dieser Stelle wird die dritte Annahme für die OLS Methode eingeführt.

Annahme OLS 3. *Exogenität der erklärenden Variablen (i). Die Daten für* $\mathbf{X}$

liefern keine Information über den Erwartungswert von $\boldsymbol{\epsilon}$, bzw.

$$\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}$$

Diese Annahme impliziert, dass alle Elemente des Vektors $\boldsymbol{\epsilon}$ mit den erklärenden Variablen für alle Beobachtungen unkorreliert sind. Anders ausgedrückt stehen also der Vektor $\boldsymbol{\epsilon}$ und alle Zeilenvektoren der Matrix $\mathbf{X}$ orthogonal zueinander.

Da $\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \epsilon \mid \mathbf{X}) = \mathbb{E}(\mathbf{X})\boldsymbol{\beta} + \mathbb{E}(\epsilon \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, gilt für den bedingten Erwartungswert:

$$\mathbb{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \tag{7}$$

Daraus folgt die Interpretation der systematischen Komponente als bedingter Erwartungswert von y, wenn die die Matrix der erklärenden Variablen $\mathbf{X}$ gegeben ist. Auch wird klar wie die Parameter zu interpretieren sind. Durch Annahme OLS 1 liegen die Werte für den bedingten Erwartungswert auf einer Linie. Für ein Element aus der Stichprobe gilt der Zusammenhang:

$$\mathbb{E}(y \mid x_1, \dots, x_K) = \hat{y} = \sum_{j=1}^{K} \beta_j x_j$$

Das jeweilige β_j ist also die Steigung der systematischen Komponente, bzw. des bedingten Erwartungswertes für eine Veränderung von x_j oder:

$$\beta_j = \frac{\partial \hat{y}}{\partial x_j} = \frac{\partial \mathbb{E}(y \mid x_1, \dots, x_K)}{\partial x_j} \tag{8}$$

Durch die Definition der partiellen Ableitung ist klar, dass durch β_j nur die Auswirkung einer Veränderung in x_j gemessen wird. Alle anderen erklärenden Variablen sind konstant. Dies wird manchmal mit Ceteris Paribus-Annahme bezeichnet. Gilt Annahme OLS 1, dann ist β_j der kausale Effekt einer Veränderung in x_j auf $\mathbb{E}(y \mid x_1, \dots, x_K)$. Das bedeutet β_j gibt an, wie der Erwartungswert von y auf eine Bewegung in und nur in x_j reagiert. Hat die Regression ein Interzept, d.h. die j^*-ste Spalte der Matrix der erklärenden Variablen $\mathbf{X_{j^*}} = \mathbf{1}$, so ist der Parameter β_{j^*} hingegen als Schnittpunkt mit der Achse der abhängigen Variable zu interpretieren. Also jenen Wert für $\hat{y}$ an dem alle erklärenden Variablen, ausser x_{j^*}, den Wert null haben.

2.1.2 OLS-Regression auf eine Stichprobe.

Folgendes Unterkapitel erläutert die Theorie hinter einer Stichprobenregression. Bei der **Modellierung des Zusammenhanges** gibt es nur kleine Unterschiede zur Regression bei einer Grundgesamtheit. Die Spezifikation muss natürlich einem sinnvollen theoretischen Zusammenhang folgen. Die kleinen Unterschiede sind lediglich in der Notation zu finden, das wird in den folgenden Abschnitten deutlich wird. Bei der **Erhebung der Daten** ist zu beachten, dass im Gegensatz zur Regression bei einer Grundgesamtheit bei der Stichprobenregression nur eine *zufällig* ermittelte Untermenge der Grundgesamtheit untersucht wird. Wie bereits angesprochen, gibt es einen großen Unterschied zwischen (i) einer Regression auf eine Grundgesamtheit und (ii) einer Regression auf eine Stichprobe. Die erste beschreibt die Grundgesamtheit, indem die „wahren" Parameter $\boldsymbol{\beta}$ der Grundgesamtheit ermittelt werden. Diese Parameter sind deterministische Variablen und zumeist unbekannt. Die zweite schätzt dagegen die Parameter $\boldsymbol{\beta}$ der Grundgesamtheit, indem Elemente[7] *zufällig* aus der Grundgesamtheit gezogen[8] werden, und daraus eine Schätzung für β berechnet wird. Werden mehrere Stichproben untersucht, so erhält man mit sehr hoher Wahrscheinlichkeit unterschiedliche Schätzungen. Die erzeugten abhängigen Variablen sind dabei in jedem Fall stochastisch, während die erklärenden Variablen sowohl stochastisch, als auch deterministisch sein können. Bei einer zufälligen Stichprobe sind deterministische, erklärende Variablen üblicherweise nur durch eine streng kontrollierte Messung zu garantieren, z.B. ein Laborversuch bei dem die erklärenden Variablen vorab vom Forscher festgelegt werden.[Davidson and MacKinnon, 2003, S.90] Das ist in den Sozialwissenschaften jedoch unüblich [Greene, 2002, S.17], weswegen im folgenden von stochastischen, erklärenden Variablen ausgegangen wird. Alle Beweise und Herleitungen sind ebenfalls für deterministische, erklärende Variablen gültig. Die Funktion, die aus den Daten einer Stichprobe eine Schätzung für β_j erzeugt wird, Schätzer genannt. Die Schätzer für die „wahren" Parameter $\boldsymbol{\beta_j}$ werden mit

[7] Genauergesagt deren Ausprägungen für die abhängige- und die erklärenden Variablen.
[8] bzw. durch den DGP erzeugt werden

b_j für alle $j = 1, ..., K$ bezeichnet. Sie sind Funktionen von Zufallsvariablen und sind somit ebenfalls stochasitsch. Sie unterliegen einer Verteilung, der **Stichprobenkennwertverteilung**. Das Prinzip der Stichprobenkennwertverteilung beruht auf der sich wiederholenden Stichprobenziehung. Nimmt man also an, man generiert sehr viele Stichproben, so folgt aus dem Gesetz der großen Zahl, dass sich der Mittelwert der erhobenen Stichproben, bei steigender Anzahl an Stichproben, an den Mittelwert der Grundgesamtheit angleicht.[Stocker, 2011, S.60] Aus dem zentralen Grenzwertsatz folgt, dass die Summe der voneinander unabhängigen, identisch verteilten Schätzer normalverteilt ist, unabhängig von der Verteilung der einzelnen Schätzer $\mathbf{b}$ in der Grungesamtheit.[Stocker, 2011, S.60] Bevor eine Funktion zur Ermittlung des Schätzers hergeleitet wird, ist es wichtig zu verstehen, dass der Zusammenhang in der Grundgesamtheit unbeobachtbar ist. Folglich müssen Annahmen über die Grundgesamtheit, bzw. den datengenerierenden Prozess getroffen werden. So nimmt man an, dass die beobachtete Daten einer Stichprobe $\mathbf{X}$, $\mathbf{y}$ aus einem DGP stammen, bzw. von diesem erzeugt werden. Zusätzlich werden für den unbeobachtbaren stochastischen Störterm Annahmen getroffen. Für den DGP gilt Annahme OLS 1, d.h. der wahre Zusammenhang zwischen der erzeugten abhängigen und den erklärenden Variablen der Grundgesamtheit wird annahmegemäß durch den Prozess aus (2) korrekt beschrieben. Ein solcher DGP könnte z.B. wie folgt aussehen:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\epsilon} \sim iid(\mathbf{0}, \sigma^2\mathbf{I})$$

Wobei $\mathbf{I}$ die Identitätsmatrix ist und *iid* bedeutet, dass die Störterme annahmegemäß voneinander stochastisch unabhängig sind und aus der gleichen Verteilung gezogen werden.

Ermittlung der Schätzer b. Im folgenden Abschnitt wird die Schätzerfunktion ermittelt. Zunächst ist nochmals festzustellen, dass es sich bei einer Stichprobenregression nicht mehr um die Ermittlung der „wahren"Parameter $\boldsymbol{\beta}$ handelt und dass nicht die „wahren" Störterme $\boldsymbol{\epsilon}$ erfasst werden, sondern lediglich die Residuen $\mathbf{e}$. Daher werden in dem angenommenen Modell die Vektoren $\mathbf{b}$ und $\mathbf{e}$ für die

$\boldsymbol{\beta}$ bzw. $\boldsymbol{\epsilon}$ substituiert. Es entsteht folgendes System, welches Sample Regression Function genannt wird:

$$y_i = f(x_{i,1}, x_{i,2}, ..., x_{i,K}) + e_i \text{ für alle } i \in \{1, 2, ..., n\}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ x_{2,1} & x_{2,2} & \dots & x_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,K} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{9}$$

Wobei n die Anzahl der Elemente der Stichprobe, auch Stichprobengröße genannt, bezeichnet. Analog zur Berechnung der Parameter bei der Regression auf die Daten einer Grundgesamtheit, wird der Schätzer für diese Parameter durch die Minimierung der quadrierten Residuen hergeleitet. Man erhält folgenden Term für die Bedingungen erster Ordnung.

$$\begin{aligned} \frac{\partial(\mathbf{e}^T\mathbf{e})}{\partial\mathbf{b}} &= \frac{\partial(\mathbf{y}^T\mathbf{y} - 2\mathbf{b}^T\mathbf{X}^T\mathbf{y} + \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{b})}{\partial\mathbf{b}} \\ &= -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{0} \end{aligned} \tag{10}$$

und somit folgende Gleichung für den Schätzer

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{11}$$

Wie bei bereits gezeigt genügt Annahme OLS 2, um zu garantieren, dass die die ermittelte Extremstelle eindeutig- und ein ein Minimum ist. Außerdem ist der Schätzer in (11) eine Zufallsvariable, welche der Stichprobenkennwertverteilung unterliegt, für die ein Erwartungswert und ein Maß für die Streuung der Schätzer, der Standardfehler, existiert. Die Streuung der Schätzer, hängt maßgeblich von der Streuung in der Störterme in der Grundgesamtheit ab. Diese ist jedoch eine unbeobachtete Größe und kann, wie auch der Parameter selbst, nur geschätzt werden. Im folgenden Unterkapitel wird hergeleitet unter welchen Bedingungen ein

Schätzer erwartungstreu oder konsistent ist. Warum diese beiden Eigenschaften so wichtig sind liegt auf der Hand. Soll z.B. anhand einer t-Statistik geprüft werden, ob ein Schätzer signifikant [9] ist, so spielt es eine Rolle wie die Stichprobenkennwertverteilung aussieht, wie groß der Standartfehler des Schätzers ist und welchen Erwartungswert der Schätzer hat. Ist der Schätzer nicht erwartungstreu, das heißt verzerrt, so gleicht der Erwartungswert des Schätzers nicht dem wahren Wert der Grundgesamtheit. Geht man von einem Wert für den echten Parameter von null aus, folgt für einen verzerrten Schätzer, dass dieser im Erwartungswert nicht null sein darf. Daraus folgt, dass man solche Schätzer kaum sinnvoll interpretieren-, bzw. Schlüsse auf die Grundgesamtheit ziehen kann, da eine systematische Abweichung vom wahren Parameter besteht. Das gilt selbst wenn die Gestalt der Stichprobenkennwertverteilung und der Standartfehler korrekt geschätzt werden.

Interpretation der Parameter. Dieser Paragraph dient zur Erläuterung der Interpretation von Ergebnissen der OLS Regression. Es wird bei Annahme OLS 1 gestartet und (1) wird als das richtig spezifizierte Modell angenommen. Für dieses Modell wird der bedingte Erwartungswert genommen.

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \mathbb{E}(\mathbf{y} \mid \mathbf{X}) &= \mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \mid \mathbf{X}) \\ &= \mathbb{E}(\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}) + \mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})\end{aligned}$$

Der Schritt $\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbb{E}(\mathbf{X}\boldsymbol{\beta} \mid \mathbf{X}) + \mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})$ ist aufgrund der Linearität des Operators des bedingten Erwartungswert möglch. An dieser Stelle wird die dritte Annahme für die OLS Methode eingeführt.

Annahme OLS 4. *Exogenität der erklärenden Variablen (i). Die Daten für* $\mathbf{X}$ *liefern keine Information über den Erwartungswert von* $\boldsymbol{\epsilon}$*, bzw.*

$$\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}$$

Diese Annahme impliziert, dass alle Elemente des Vektors $\boldsymbol{\epsilon}$ *mit den erklärenden*

[9] Das heißt ob die Nullhypothese, dass der wahre Parameter null beträgt, bei einem gegebenen Konfidenzintervall durch den Schätzer abgelehnt werden kann.

Variablen für alle Beobachtungen unkorreliert sind. Anders ausgedrückt stehen also der Vektor $\boldsymbol{\epsilon}$ und alle Zeilenvektoren der Matrix $\mathbf{X}$ orthogonal zueinander.

Da $\mathbb{E}(\mathbf{X}\boldsymbol{\beta} + \epsilon \mid \mathbf{X}) = \mathbb{E}(\mathbf{X})\boldsymbol{\beta} + \mathbb{E}(\epsilon \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$, gilt für den bedingten Erwartungswert:

$$\mathbb{E}(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \tag{12}$$

Daraus folgt die Interpretation der systematischen Komponente als bedingter Erwartungswert von y, wenn die die Matrix der erklärenden Variablen $\mathbf{X}$ gegeben ist. Auch wird klar wie die Parameter zu interpretieren sind. Durch Annahme OLS 1 liegen die Werte für den bedingten Erwartungswert auf einer Linie. Für ein Element aus der Stichprobe gilt der Zusammenhang:

$$\mathbb{E}(y \mid x_1, \ldots, x_K) = \hat{y} = \sum_{j=1}^{K} \beta_j x_j$$

Das jeweilige β_j ist also die Steigung der systematischen Komponente, bzw. des bedingten Erwartungswertes für eine Veränderung von x_j oder:

$$\beta_j = \frac{\partial \hat{y}}{\partial x_j} = \frac{\partial \mathbb{E}(y \mid x_1, \ldots, x_K)}{\partial x_j} \tag{13}$$

Durch die Definition der partiellen Ableitung ist klar, dass durch β_j nur die Auswirkung einer Veränderung in x_j gemessen wird. Alle anderen erklärenden Variablen sind konstant. Dies wird manchmal mit Ceteris Paribus-Annahme bezeichnet. Gilt Annahme OLS 1, dann ist β_j der kausale Effekt einer Veränderung in x_j auf $\mathbb{E}(y \mid x_1, \ldots, x_K)$. Das bedeutet β_j gibt an, wie der Erwartungswert von y auf eine Bewegung in und nur in x_j reagiert. Hat die Regression ein Interzept, d.h. die j^*-ste Spalte der Matrix der erklärenden Variablen $\mathbf{X_{j^*}} = \mathbf{1}$, so ist der Parameter β_{j^*} hingegen als Schnittpunkt mit der Achse der abhängigen Variable zu interpretieren. Also jenen Wert für $\hat{y}$ an dem alle erklärenden Variablen, ausser x_{j^*}, den Wert null haben.

2.2 Unverzerrte und konsistente Schätzer

Wie bereits erwähnt, ist es das Ziel einer Regressionsanalyse Zusammenhänge zwischen der erklärenden- und den abhängigen Variablen zu ermitteln. Unter bestimmten Bedingungen sind die sogenannten Schätzer, in (1) die Variablen $\beta_j \; j \in \{1, ..., K\}$, als dieser Zusammenhang zu interpretieren.

Erwartungstreue eines Schätzers. Bislang wurde der Ausdruck für den OLS Schätzer (11) hergeleitet. Eine wichtige Frage in der schließenden Statistik ist, ob dieser Schätzer eine verzerrte- oder ob er eine erwartungstreue Schätzung von $\boldsymbol{\beta}$ ergibt. Die Erwartungstreue ist eine sogenannte „small-sample" Eigenschaft, da sie von der Stichprobengröße unabhängig ist und folglich auch für kleine Stichproben gilt. Das Kriterium für Erwartungstreue lautet $\mathbb{E}(\mathbf{b}) = \boldsymbol{\beta}$. Um dieses Kriterium für den OLS Schätzer zu untersuchen wird dieser zunächst umgeformt. Dazu wird $(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$ für $\mathbf{y}$ substituiert.

$$\begin{aligned}
\mathbf{b} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \qquad (14)
\end{aligned}$$

Wird nun der bedingte Erwartungswert gebildet, erhält man:

$$\begin{aligned}
\mathbb{E}(\mathbf{b} \mid \mathbf{X}) &= \mathbb{E}(\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \mid \mathbf{X}) \\
&= \mathbb{E}(\boldsymbol{\beta} \mid \mathbf{X}) + \mathbb{E}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \mid \mathbf{X}) \\
&= \boldsymbol{\beta} + \mathbb{E}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \mid \mathbf{X}) \\
&= \boldsymbol{\beta} + ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})
\end{aligned}$$

Der Schritt $\mathbb{E}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \mid \mathbf{X}) = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})$ ist gültig, da $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ eine Funktion von $\mathbf{X}$ ist und für Funktionen h und den Operator des bedingten Erwartungswertes folgendes gilt: $\mathbb{E}(h(\mathbf{X})\boldsymbol{\epsilon} \mid \mathbf{X}) = h(\mathbf{X})\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})$.[Davidson and

MacKinnon, 2003, S.16] Gilt Annahme OLS 3, so fahren wir fort:

$$
\begin{aligned}
\mathbb{E}(\mathbf{b} \mid \mathbf{X}) &= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{0} = \boldsymbol{\beta} \\
\mathbb{E}(\mathbb{E}(\mathbf{b} \mid \mathbf{X})) &= \mathbb{E}(\boldsymbol{\beta}) \\
\mathbb{E}(\mathbf{b}) &= \boldsymbol{\beta} \qquad (15)
\end{aligned}
$$

Das gilt, da $\boldsymbol{\beta}$ eine Konstante ist und da $\mathbb{E}(\mathbb{E}(\mathbf{b} \mid \mathbf{X})) = \mathbb{E}(\mathbf{b})$ durch das Gesetz des iterativen Erwartungen gilt.[10]. Das heißt, das Kriterium für Erwartungstreue ist erfüllt, wenn Annahmen OLS 1, 2 und 3 gelten. Wird jedoch z.B. Annahme OLS 3 verletzt, so ist der Schätzer für die Parameter systematisch verzerrt. Das bedeutet, dass $\mathbf{b}$ nicht den wahren Kausalzusammenhang zwischen $\mathbf{X}$ und $\mathbf{y}$ misst. Damit stoßen wir zum ersten Mal auf das Grundproblem, welches hinter der Endogenität steht, stellen eine genaue Definition des Begriffs jedoch für später zurück.

Konsistenz eines Schätzers. Nach der Erwartungstreue gibt es ein zweites, häufig benutztes Kriterium für einen Schätzer, die Konsistenz. Sie ist eine asymptotische Eigenschaft, was bedeutet, dass sie das Verhalten von Schätzern untersucht, wenn die Größe einer zufällig ermittelten Stichprobengröße gegen unendlich strebt. Ein Vektor von Schätzern $\hat{\theta}$ gilt als konsistent wenn folgende Bedingung[Davidson and MacKinnon, 2003, S.94] erfüllt ist:

$$
\lim_{n\to\infty} \mathbb{P}(||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}|| < \sigma) = 1 \text{ für alle } \sigma > 0
$$

Der Vektor $\hat{\boldsymbol{\theta}}_n$ ist das n-te Glied einer Folge von Schätzern, deren Index die Stichprobengröße n ist. $\mathbb{P}$ gibt dabei die Wahrscheinlichkeit an, mit der die Länge des Vektors $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$ geringer als ein beliebiges positives σ ist. Ist $\hat{\boldsymbol{\theta}}$ ein Skalar, so gilt ein Schätzer als konsistent, wenn für alle $n$groß genug, die Wahrscheinlichkeit eins beträgt, mit der die absolute Differenz zwischen n-ten Glied einer Folge von Schätzern und dem wahren Parameter kleiner ist als jedes beliebige positive σ. Ein äquivalenter Ausdruck hierzu ist, dass eine Folge von Schätzern in Wahrscheinlich-

[10] Dieses Gesetz besagt, dass für zwei Zufallsvariablen X_1 und X_2 gilt: $\mathbb{E}_{X_2}(X_1) = \mathbb{E}_{X_2}(\mathbb{E}(X_1 \mid X_2))$, wobei $\mathbb{E}_{X_2}$ der Erwartungswertoperater für die Verteilung von X_2 ist.[Davidson and MacKinnon, 2003, S.16]

keit gegen den Wert der wahren Parameter θ konvergiert oder kürzer:

$$\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$$

Der plim-Operator steht dabei für das Probability Limit. Das Probability Limit gibt den Wert, gegen den die Folge der Schätzer konvergiert. Die Idee der Konsistenz ist eng verknüpft mit dem Gesetz der großen Zahl, welches in seiner schwachen Ausführung besagt, dass der arithmetische Mittelwert einer Zufallsvariable unter bestimmten Bedinigungen[11]in Wahrscheinlichkeit gegen den Erwartungswert der Zufallsvariable strebt, wenn die Anzahl der Zufallsvariablen gegen unendlich strebt.[Hansen, 2011, S.11] Für eine genauere Bearbeitung des Probability Limit wird der Leser z.B. an [Davidson and MacKinnon, 2003, Kapitel 3] verwiesen. Nach der Definition der Konsistenz eines Schätzers wird nun geprüft, ob, bzw. unter welchen Annahmen der in (11) ermittelte Schätzer konsistent ist.

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \mathbf{b} &= \operatorname*{plim}_{n\to\infty}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\ &= \operatorname*{plim}_{n\to\infty}(\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}) \\ &= \boldsymbol{\beta} + \operatorname*{plim}_{n\to\infty}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}) \end{aligned} \tag{16}$$

Die Bedingung der Konsistenz $\operatorname{plim} \mathbf{b} = \boldsymbol{\beta}$ gilt also, wenn und nur wenn der rechte Teil der Summe in (16) ein Nullvektor ist. Dieser Term ist das Produkt zweier Matrizen. $\mathbf{X}^T\mathbf{X}$ und $\mathbf{X}^T\boldsymbol{\epsilon}$ besitzen beide kein Probability Limit.[Davidson and MacKinnon, 2003, S.96] Werden sie jedoch durch die Stichprobengröße n dividiert, ändert sich das Produkt in (16) nicht[12] und beide Faktoren haben ein

[11]Wenn $\mathbb{E}|y| < \infty$, $n \to \infty$ und $y \sim iid(\mathbb{E}(y), \sigma^2)$ dann folgt

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}(y)$$

[12]Weil $(\frac{1}{n})^{-1}\frac{1}{n} = n\frac{1}{n} = 1$.

Probability Limit.

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}((\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\frac{1}{n}\mathbf{X}^T\boldsymbol{\epsilon}) &= \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\ \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\mathbf{X}^T\boldsymbol{\epsilon})\\ &= (\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\ \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\mathbf{X}^T\boldsymbol{\epsilon}) \end{aligned} \tag{17}$$

Der rechte Teil der Gleichung in der ersten Zeile gilt, da $\operatorname{plim}(u\,v) = \operatorname{plim} u \operatorname{plim} v$. Der Schritt in der zweiten Zeile folgt aus Slutsky's Theorem[Wooldridge, 2002b, S.37], welches besagt, dass $\operatorname{plim} g(u) = g(\operatorname{plim} u)$ für stetige Funktionen g gilt. Der linke Teil des Produktes ist eine $K \times K$ Matrix für die angenommen wird, dass das Gesetz der großen Zahl auf sie anwendbar ist. Dann konvergiert die Matrix gegen eine nicht-stochastische finite Matrix, wenn n gegen unendlich strebt, was bedeutet, dass ein Probability Limit für diese Matrix existiert.[Davidson and MacKinnon, 2003, S.96] Eine zusätzliche Annahme über die Exogenität der erklärenden Variablen wird benötigt, um den rechten Teil des Produktes aus (17) zu vereinfachen.

Annahme OLS 5. *Exogenität der erklärenden Variablen (ii). Der bedingte Erwartungswert des Störterms ϵ_i einer Beobachtung basierend auf den erklärenden Variablen dieser Beobachtung $\mathbf{X}_i$ ist null. Außerdem wird Annahme OLS 2 über den vollen Spaltenrang von $\mathbf{X}$ weiterhin angenommen.*

$$\mathbb{E}(\epsilon_i|\mathbf{X}_i) = 0 \tag{18}$$

Wobei $\mathbf{X}_i$ der i-te Zeilenvektor der Matrix $\mathbf{X}$ ist. Diese Annahme impliziert folgendes:

$$\begin{aligned}\mathbb{E}(\mathbf{X}_i^T\epsilon_i|\mathbf{X}_i) &= \mathbf{X}_i^T\mathbb{E}(\epsilon_i|\mathbf{X}_i)\\ &= \mathbf{X}_i^T\ 0 = \mathbf{0}\end{aligned}$$

Durch das Gesetz des iterativen Erwartungswertes folgt, dass $\mathbb{E}(\mathbf{X}_i^T\epsilon_i|\mathbf{X}_i) = \mathbb{E}(\mathbb{E}(\mathbf{X}_i^T\epsilon_i|\mathbf{X}_i)) = \mathbb{E}(\mathbf{X}_i^T\epsilon_i) = 0$. Der zweite Teil des Produktes in (17) kann auch geschrieben werden als:

$$\operatorname*{plim}_{n\to\infty}(\frac{1}{n}\mathbf{X}^T\boldsymbol{\epsilon}) = \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\epsilon_i) \tag{19}$$

Nehmen wir nun an, das Gesetz der großen Zahl in seiner schwachen Ausführung[13] gilt, das heißt $\mathbb{E}|\mathbf{X}_{i,j}^T\boldsymbol{\epsilon}_i| < \infty$ und $\mathbf{X}_{i,j}^T\boldsymbol{\epsilon}_i \sim iid(\mathbb{E}(\mathbf{X}_{i,j}^T\boldsymbol{\epsilon}_i), \sigma^2)$ für alle j. Dann folgt daraus, dass $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\epsilon_i$ gegen den Erwartungswert konvergiert.

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\epsilon_i \xrightarrow{p} \mathbb{E}(\mathbf{X}_i^T\epsilon_i)$$

Da der Erwartungswert null beträgt gilt also:

$$\operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\epsilon_i) = 0 \tag{20}$$

Aus diesem Ergebnis, der Aussage in (19) und der Annahme dass der linke Teil des Produktes in (17) eine nicht-stochastische $K \times K$ Matrix ist, ergibt dieses Produkt den Nullvektor. Die Bedingung für die Konsistenz ist folglich für den Schätzer aus (11) erfüllt, da der zweite Teil der Summe in (16) gegen den Nullvektor konvergiert.

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty}\mathbf{b} &= \boldsymbol{\beta} + \operatorname*{plim}_{n\to\infty}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\epsilon_i)) \\
&= \boldsymbol{\beta} + (\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{0} \\
\operatorname*{plim}_{n\to\infty}\mathbf{b} &= \boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta}
\end{aligned}$$

Man beachte, dass für die Erwartungstreue und die Konsistenz der Schätzer die Annahmen OLS 3, bzw. OLS 4 benötigt werden. Beide Annahmen sichern die Exogenität der erklärenden Variablen im Modell. Annahme OLS 3 stärker als OLS 4.

[13] Für einen Beweis der Konsistenz des OLS-Schätzers bei der starken Form des Gesetzes der großen Zahl wird der Leser an das Lehrbuch von Greene aus dem Jahr 2002 verwiesen.

Während bei OLS 4 nur die Störterme mit den erklärenden Variablen für das jeweilige Element i einer Stichprobe stochastisch unabhängig sind, impliziert OLS 3 stochastische Unabhängigkeit zwischen allen Störtermen und allen erklärenden Variablen. Jedoch ist die Erwartungstreue nicht unbedingt ein strikteres Kriterium für einen Schätzer. Erwartungstreue Schätzer sind nicht automatisch auch konsistent, der Umkehrschluss ist auch nicht korrekt.[Davidson and MacKinnon, 2003, S.97]

2.3 Dummy Variablen

Durch das folgende Unterkapitel wird versucht, dem Leser Dummy Variablen näher zu bringen. Sie sind ein äußerst nützliches und oft verwendetes Werkzeug in der Ökonometrie. So werden Dummy Variablen dann verwendet, wenn Unterschiede zwischen z.B. verschiedenen Gruppen aufgezeigt werden sollen. Die Dummy Variable, sie wird mit D bezeichnet, ist eine erklärende Variable und kann, anders als die bisher verwendeten erklärenden Variablen, nur zwei Werte annehmen, null und eins.

$$D_i = \begin{cases} 1 & \text{wenn Kriterium „wahr" für Beobachtung } i \\ 0 & \text{wenn Kriterium „falsch" für Beobachtung } i \end{cases}$$

So kann ein Dummy z.B. die Zugehörigkeit zu einer Gruppe signalisieren, z.B. kann ein Dummy so definiert sein, dass er eins ergibt, wenn die Beobachtung eine Frau ist und null, wenn dem nicht so ist. In einer Stichprobe ergeben sich so N Ausprägungen für den Dummy. Natürlich können auch mehrere Dummy Variablen in ein System integriert werden, einzig sollte man besonders darauf achten, dass diese untereinander nicht linear abhängig sind, da ansonsten die Spaltenrang-Annahme OLS 2 verletzt würde. Alle $N \times K_D$ Beobachtungen für alle K_D Dummies können effizient in einer Matrix $\mathbf{D}$ geordnet werden. Dummies können einfach in ein System wie (2) eingebaut werden, z.B. als Spalte(n) in der Matrix der erklärenden Variblen $\mathbf{X}$ oder als separate Matrix aller Dummy Variablen. Ein Modell könnte dann so aussehen:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Wenn Annahmen OLS 1 und 2 für dieses Modell gelten, ist $\boldsymbol{\gamma}$, analog zu nicht-Dummy Variablen, der Vektor der „wahren" Parameter der Grundgesamtheit. Einzig die Interpetation der Parameter unterscheidet sich. So ist γ nicht wie bisher als Steigung zu interpretieren, da die partielle Ableitung nur für „kleine" Veränderungen in D definiert ist, D aber nur eins und null annehmen kann. Vielmehr gibt der Parameter die Differenz des bedingten Erwartungswertes zwischen zwei Systemen, bei denen sich *einzig* die Dummy Variable unterscheidet. Dies wird in (21) dargestellt. [Wooldridge, 2002a, S.213]

$$\begin{aligned}
&\gamma \neq \frac{\partial \mathbb{E}(y \mid D)}{\partial D} \\
&\gamma = \mathbb{E}(y \mid D = 1, x_1, \ldots, x_K) - \mathbb{E}(y \mid D = 0, x_1, \ldots, x_K) \qquad (21)
\end{aligned}$$

Diese Differenz ist nicht von den sonstigen erklärenden Variablen abhängig. Daher ist γ eine Konstante und verändert das Interzept, also den Punkt an dem die errechnete Regressionsfunktion die Achse der abhängigen Variable schneidet.

Interaktionseffekte. Um neben Veränderungen im Interzept auch Veränderungen in der Steigung des bedingten Erwartungswertes festzustellen, dient ein Interaktionsterm.[Stocker, 2011, S.275] Lässt man z.B. eine erklärende Variable des Modells mit einem Dummy interagieren, also $(x_j \times D)$ so wird damit exakt die Veränderung der Steigung gemessen. Wird der Interaktionsterm durch einem herkömmlichen Dummy ergänzt, dann wird sowohl der Unterschied in der Steigung, als auch der Unterschied im Interzept gemessen. Ein solches Modell ist z.B. das folgende mit nur einer erklärenden Variable und einem Dummy.[Stocker, 2011, S.275]

$$\begin{aligned}
y &= b_1 + \hat{\gamma} D + b_2 x + b_3(D \times x) + e \\
\mathbb{E}(y \mid D = 1, x) &= \mathbb{E}(b_1) + \mathbb{E}(\hat{\gamma}) + (\mathbb{E}(b_1) + \mathbb{E}(b_2))x = \beta_1 + \gamma + (\beta_2 + \beta_3)x \\
\mathbb{E}(y \mid D = 0, x) &= \beta_1 + \beta_2 x
\end{aligned}$$

Wobei $\hat{\gamma}$ für den Schätzer von γ steht. Aus den letzten beiden oberen Gleichungen ist leicht ersichtlich, dass durch γ die Differenz im Interzept- und durch $\beta_2 x$ die Differenz in der Steigung von $\mathbb{E}(y \mid D, x)$ gemessen wird. Ebenfalls ist es möglich Interaktionseffekte zwischen Dummy Variablen zu messen und die Vorgehensweise ist identisch zu der eben gezeigten. Nur wird dann nicht mehr der Unterschied in der Steigung gemessen, sondern ein konstanter Unterschied. Die Formeln für den Parameter, Schätzer und die Kriterien für Erwartungstreue und Konsistenz gelten auch, wenn Dummy Variablen für x eingesetzt werden. Daher braucht keines dieser Ergebnisse für Dummy Variablen angepasst werden.

2.4 Panel Daten

Daten für mehrere Individuen l, $l \in \{1, \dots, L\}$ über mehrere Perioden t, $t \in \{1, \dots, L\}$ werden Panel Daten genannt. Sie haben die Eigenschaft, dass die Individuen über die Perioden hinweg gleich bleiben. Paneldaten können z.B. gestapelt dargestellt werden. Das bedeutet, dass Vektoren gebildet werden, bei denen für jedes Individuum l die Beobachtungen aller Zeitpunkte t übereinander gestapelt werden.[Stocker, 2011, 282ff.] Ein gestapeltes Modell mit K erklärenden Variablen schaut z.B. wie folgt aus:

$$y_{lt} = \mathbf{x}_{lt}\mathbf{b} + e_{lt} \text{ für alle } l \in \{1, 2, ..., L\} \text{ und alle } t \in \{1, 2, ..., T\}$$

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{L1} \\ \vdots \\ y_{LT} \end{pmatrix} = \begin{pmatrix} x_{11,1} & x_{11,2} & \dots & x_{11,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T,1} & x_{1T,2} & \dots & x_{1T,K} \\ x_{21,1} & x_{21,2} & \dots & x_{21,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{2T,1} & x_{2T,2} & \dots & x_{2T,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L1,1} & x_{L1,2} & \dots & x_{L1,K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{LT,1} & x_{LT,2} & \dots & x_{LT,K} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{pmatrix} + \begin{pmatrix} e_{11} \\ \vdots \\ e_{1T} \\ e_{21} \\ \vdots \\ e_{2T} \\ \vdots \\ e_{L1} \\ \vdots \\ e_{LT} \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{22}$$

$\mathbf{x}_{it}$ ist ein $1 \times K$ Zeilenvektor und β ist ein $K \times 1$ Spaltenvektor, $\mathbf{X}$ ist eine Matrix mit Dimension $LT \times K$, die anderen Variablen sind Skalare. Der Schätzer

b kann wie in (11) ermittelt werden. Dazu definieren wir den Index i als $i = lt \in \{11, \ldots, 1T, 21, \ldots, 2T, \ldots, L1, \ldots, LT\}$. Man sollte sich folglich nicht vom Doppelindex lt verwirren lassen, da die Daten sehr einfach gestapelt dargestellt und bearbeitet werden können.

3 Endogenität

Das folgende Kapitel hat zum Ziel dem Leser den Begriff der Endogenität in der Ökonometrie näher zu bringen. Die Vorgehensweise ist wie folgt: Als erstes wird der Begriff Endogenität definiert. Zweitens wird analysiert, wann diese auftritt und, drittens, mit welchen Methoden sie verhindert werden kann. Daraufhin werden kurz die Methoden der Instrumentvariablen, der randomisierten Experimente und der festen Effekte vorgestellt.

3.1 Was ist Endogenität?

Endogenität tritt dann auf, wenn erklärende Variablen mit den Störtermen der Grundgesamtheit ϵ korrelieren. Die betreffenden erklärenden Variablen werden endogene, erklärende Variablen genannt. Folgendes Beispiel ist aus Stocker von 2002 übernommen.[Stocker, 2011, S.331] Intuitiv kann man sich folgende Auswirkung der Endogenität auf die Schätzer vorstellen: Korreliert der Störterm mit den erklärenden Variablen, so kann, vereinfacht, der Störterm als eine Funktion der erklärenden Variable dargestellt werden. Ein Beispiel für ein solches System wäre z.B. folgendes:

$$y = b_1 + b_2 x + \epsilon(x)$$

$$\frac{\partial y}{\partial x} = b_2 + \frac{\partial \epsilon(x)}{\partial x} = b_2^*$$

Die zweite Zeile verdeutlicht die Auswirkung einer kleinen Bewegung in der erklärenden Variable auf die abhängige Variable. Durch die Korrelation zwischen ϵ und x taucht der marginale Effekt von x auf ϵ in der partiellen Ableitung auf. Dieser Effekt „stört" den Parameter b_2, der eigentlich den marginalen Effekt von x auf y

messen sollte, es entsteht ein anderer, verzerrter Parameter, b_2^*. Das folgende, kleine Beispiel soll das Grundproblem erläutern: Nehmen wir an, dass bei einer Regression mit Interzept eine Dummy Variable positiv mit dem Störterm korreliert. Beträgt die Dummy z.B. eins, dann ist auch der Störterm im Erwartungswert positiv, oder $\mathbb{E}(\epsilon \mid D = 1) > 0$. Folglich wird der bedingte Erwartungswert von y, $\mathbb{E}(y \mid D = 1)$ systematisch überschätzt.

Zur Veranschaulichung der Effekte, welche von endogenen, erklärenden Variablen auf die Schätzer ausgehen, dient folgende Abbildung, welche von Stocker übernommen wurde. [Stocker, 2011, S.332] Abbildung 1 zeigt vereinfacht dargestellt den Effekt, den die Endogenität auf die Schätzung hat. Sind x und ϵ unkorreliert, wird nur der direkte Effekt von x auf y gemessen. Besteht jedoch Korrelation zwischen x und ϵ, so wird der direkte Effekt b durch einen indirekten Effekt von ϵ über x zu y gestört. Der neue, verzerrte Schätzer heißt b^*.

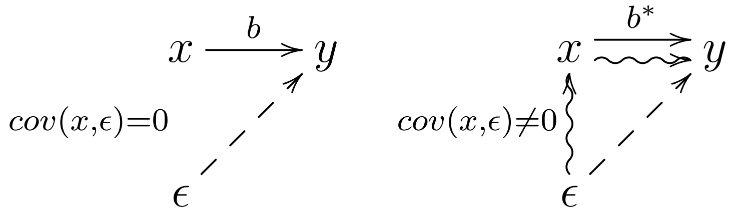

Abbildung 1: Auswirkung der Endogenität

Erwartungstreue und Konsistenz von Schätzern bei Endogenität. Im folgenden Abschnitt wird genauer untersucht, wie der OLS-Schätzer aus (11) auf Endogenität reagiert. Nehmen wir an Annahme OLS 3 ist erfüllt. Dann gilt für das Produkt der transponierten Matrix der erklärenden Variablen und dem Vektor der Störterme:

$$\begin{aligned}
\mathbb{E}(\mathbf{X}^T\boldsymbol{\epsilon}|\mathbf{X}) &= \mathbf{X}^T\mathbb{E}(\boldsymbol{\epsilon}|\mathbf{X}) = \mathbf{0} \\
\mathbb{E}(\mathbb{E}(\mathbf{X}^T\boldsymbol{\epsilon}|\mathbf{X}) &= \mathbb{E}(\mathbf{0}) \\
\mathbb{E}(\mathbf{X}^T\boldsymbol{\epsilon}) &= \mathbf{0}
\end{aligned}$$

Annahme OLS 3 impliziert also Orthogonalität zwischen allen Spaltenvektoren von $\mathbf{X}$ und dem Vektor der Störterme. Die Orthogonalität impliziert wiederum, dass keine Korrelation zwischen dem Störterm und x_j für alle $j = \{1, \dots K\}$ besteht. Zwei Zufallsvariablen ϵ und x_j sind unkorreliert, wenn die Kovarianz zwischen beiden null ergibt. Es kann gezeigt werden, dass das n-fache der Formel der Kovarianz gleich der Formel des inneren Produktes der zwei Vektoren ist, wenn annahmegemäß $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbb{E}(\mathbb{E}(\boldsymbol{\epsilon} \mid \mathbf{X})) = \mathbf{0}$ gilt. Die letzte Aussage wird etwas ausführlicher beleuchtet. Die Kovarianz zwischen dem Störterm und einer erklärenden Variable x_j ist gegeben als:

$$\begin{aligned}
cov(\epsilon, x_j) &= \mathbb{E}((\epsilon - \mathbb{E}(\epsilon))(x_j - \mathbb{E}(x_j))) \\
&= \mathbb{E}(\epsilon(x_j - \mathbb{E}(x_j))) = \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i x_{i,j} - \epsilon_i \mathbb{E}(x_j)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\epsilon_i x_{i,j} - \mathbb{E}(x_j)\frac{1}{n}\sum_{i=1}^{n}\epsilon_i = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i x_{i,j} - \mathbb{E}(x_j) n \mathbb{E}(\epsilon) \\
cov(\epsilon, x_j) &= \frac{1}{n}\sum_{i=1}^{n}\epsilon_i x_{i,j} - 0 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i x_{i,j} \qquad (23)
\end{aligned}$$

Das innere Produkt eines Spaltenvektors $\mathbf{X_j}$ aus $\mathbf{X}$ und dem Vektor der Störterme ist definiert als:

$$\boldsymbol{\epsilon}\mathbf{X_j} = \sum_{i=1}^{n} \epsilon_i x_{i,j}$$

Nimmt man also an, Annahme OLS 3 sei erfüllt, so folgt daraus, dass $\sum_{i=1}^{n} \epsilon_i x_{i,j} = 0$ für alle j. Das impliziert durch (23), dass die Kovarianz zwischen dem Störterm und allen K erklärenden Variablen null sein muss. Im Umkehrschluss bedeutet dies: Ist die Kovarianz nicht null, kann Annahme OLS 3 nicht gelten. Ist Annahme OLS 3 aber verletzt, so ist der rechte Teil der Summe beim Schätzer aus (14) $\mathbf{b} = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}$ im Erwartungswert nicht der Nullvektor. Daher ist der Schätzer verzerrt, wenn endogene Variablen im Modell existieren.

Ein ähnliches Problem ergibt sich bei Annahme OLS 4. Diese sagt folgendes aus: $\mathbb{E}(\epsilon_i|\mathbf{X}_i) = 0$. Das bedeutet, dass die erklärenden Variablen einer Beobachtung[14] keinerlei Einfluss auf den Erwartungswert des Störterms dieser Beobachtung

[14] Die i-te Zeile aus X.

haben. Sie impliziert ebenfalls, dass $\mathbb{E}(u_i\mathbf{X}_i) = \mathbb{E}(\mathbb{E}(u_i\mathbf{X}_i|\mathbf{X}_i)) = 0$. In (23) eingesetzt impliziert dies wiederum eine Kovarianz von null zwischen dem Störterm und den erklärenden Variablen für diese Beobachtung. Die Kovarianz von null ist folglich eine notwendige Bedingung, damit Annahme OLS 4 gelten kann. Wenn jedoch Annahme OLS 4 verletzt ist, konvergiert der rechte Teil der Summe in (16) nicht gegen den Nullvektor und der Schätzer ist inkonsistent. Mit den Störtermen korrelierende, erklärende Variablen implizieren somit inkonsistente Schätzer. Wie demonstriert wurde, ist der Schätzer in (11) sowohl verzerrt, als auch inkonsistent wenn die Störterme mit den erklärenden Variablen korrelieren.

3.2 Wodurch wird Endogenität verursacht?

Die Literatur nennt verschiedene Quellen für Endogenität. Die wichtigsten sind in der folgenden Liste aufgezählt und werden in diesem Unterkapitel erläutert.[Wooldridge, 2002b, S. 51ff.]

1. Nicht berücksichtigte, relevante Variablen
2. Simultane Kausalität
3. Messfehler

Es ist nicht immer einfach, bzw. nicht möglich, die Quelle der Endogenität zu ermitteln. So kann es sein, dass in bestimmten Fällen die Endogenität aus allen drei Quellen gleichzeitig stammt.

Nicht berückstigte, relevante Variablen (Omitted Variables). Das Problem der in einem Modell nicht berücksichtigten relevante Variablen tritt dann auf, wenn relevante Variablen, normalerweise aufgrund der Nichtverfügbarkeit von Daten, nicht in ein Regressionsmodell, wie z.B. (9) eingebaut werden können. Falls diese nicht berücksichtigten Variablen mit zumindest einer erklärenden Variable x_j korrelieren, so ist x_j ein endogene Variable.[Wooldridge, 2002b, S. 51]

Im folgenden Abschnitt wird die Auswirkung dieses Problems auf den Schätzer etwas ausführlicher bearbeitet. Nehmen wir an, das „wahre" Modell der Grundgesamtheit wird durch (24) beschrieben. In der dritten Zeile wird dieses Modell in

Vektorschreibweise angegeben.

$$\begin{aligned}
&y = \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \gamma q + \epsilon \\
&\mathbb{E}(\epsilon | x_1, \ldots x_k, q) = 0 \\
&\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{q}\gamma + \boldsymbol{\epsilon} \qquad (24)
\end{aligned}$$

Für den wahren Störterm ϵ gilt die übliche Annahme, dass die erklärenden Variablen keinen Einfluss auf den Erwartungswert des Störterms haben. Wird jedoch anstatt (24) ein anderes Modell geschätzt, in welchem die erklärende Variable q nicht berücksichtigt wurde, so wurde q implizit in das Residuum e „verschoben".

$$\begin{aligned}
&y = b_1 + b_2 x_2 + \ldots + b_K x_K + e \\
&e = \gamma q + \epsilon \\
&\mathbf{y} = \mathbf{X}\boldsymbol{b} + \boldsymbol{e} \qquad (25)
\end{aligned}$$

Per Definition ist ϵ mit allen erklärenden Variablen unkorreliert, das Residuum e ist jedoch nur dann mit den erklärenden Variablen unkorreliert, wenn und nur wenn *keine* der erklärenden Variablen mit q korreliert sind. Ist dies nicht der Fall, also liegt Korrelation zwischen zumindest einer berücksichtigten, erklärenden Variable und dem unberücksichtigten q vor, dann entsteht Endogenität mit Auswirkungen auf die Schätzer. Die Auswirkungen werden im folgenden Abschnitt genauer analysiert. (11) ist der errechnete Schätzer für das Modell aus (25). Die Erwartungstreue ergibt sich durch Einsetzen von (24) in (11). Es wird angenommen, dass Annahme OLS 1, 2 und 3 gelten.

$$\begin{aligned}
\mathbf{b} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{q}\gamma + \boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{q}\gamma + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \\
\mathbb{E}(\mathbf{b} \mid \mathbf{X}) &= \mathbb{E}(\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{q}\gamma + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon} \mid \mathbf{X}) \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\gamma\mathbb{E}(\mathbf{q} \mid \mathbf{X}) \qquad (26)
\end{aligned}$$

Der Schätzer ist nur dann erwartungstreu, wenn der rechte Teil in in (26) den Nullvektor ergibt. Das ist nur der Fall wenn $\mathbb{E}(\mathbf{q} \mid \mathbf{X}) = \mathbf{0}$. Eine Kovarianz von null zwischen q und allen x_j ist eine notwendige Bedingung, dass dies gelten kann.

Korrelieren beide Variablen, so ist $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\gamma\mathbb{E}(\mathbf{q} \mid \mathbf{X}) \neq \mathbf{0}$ und damit ist der Schätzer systematisch verzerrt.

Der Schätzer ist auch inkonsistent, da eine Korrelation zwischen zumindest einem x_j und q bewirkt, dass der zweite Teil der Summe in (27) nicht gegen den Nullvektor konvergiert. Das gilt, da die Korrelation impliziert, dass $\mathbb{E}(\mathbf{q} \mid \mathbf{X}_i) \neq \mathbf{0}$, was wiederum $\mathbb{E}(\mathbf{X}_i^T\mathbf{q}) = \mathbb{E}(\mathbb{E}(\mathbf{X}_i^T\mathbf{q} \mid \mathbf{X}_i)) \neq \mathbf{0}$ impliziert.

$$
\begin{aligned}
\operatorname*{plim}_{n\to\infty} \mathbf{b} &= \boldsymbol{\beta} + \operatorname*{plim}_{n\to\infty}((\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}\frac{1}{n}\mathbf{X}^T\mathbf{q}\gamma) + \operatorname*{plim}_{n\to\infty}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (\operatorname*{plim}_{n\to\infty}\frac{1}{n}(\mathbf{X}^T\mathbf{X}))^{-1}\gamma \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\mathbf{X}^T\mathbf{q}) \\
&= \boldsymbol{\beta} + (\operatorname*{plim}_{n\to\infty}\frac{1}{n}(\mathbf{X}^T\mathbf{X}))^{-1}\gamma \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T q_i)
\end{aligned} \tag{27}
$$

Messfehler. Das Problem des Messfehlers ist recht intuitiv. Anstatt der eigentlich erwünschten, erklärenden Variable x_j wird tatsächlich x_j^* gemessen. Die Messung beinhaltet also einen Messfehler. Korreliert dieser Messfehler mit der erwünschten Variable x_j, dann ist x_j eine endogene Variable.

Das Problem wird anhand eines Beispiels[Davidson and MacKinnon, 2003, S.310ff.] erläutert. Nehmen wir an, das Regressionsmodell in (28) beschreibt den tatsächlichen Zusammenhang in der Grundgesamtheit. Jedoch tritt bei der Datenerhebung ein Messfehler auf und es wird anstatt der erklärenden Variable x_K aus (28), die Variable aus (29) gemessen.

$$
\begin{aligned}
y &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \epsilon \\
\mathbb{E}(\epsilon | x_1, \ldots, x_K) &= 0
\end{aligned} \tag{28}
$$

Der Störterm erfüllt die übliche Exogenitätsannahme. Die beobachtete erklärende Variable x_K^* enthält einen Messfehler v.

$$
x_K^* = x_K + v \tag{29}
$$

Nehmen wir zudem an, dass der Störterm aus (28) mit x_j^* unkorreliert ist und dass der Erwartungswert des Messfehlers null beträgt. Wird die Variable aus (29) in (28) eingesetzt, so ergibt sich (30) durch Verschieben des Term für den Messfehler in den Störterm.

$$\begin{aligned} y &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K (x_K - v) + \epsilon \\ &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \epsilon - \beta_K v_K \\ &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \epsilon^* \qquad (30) \end{aligned}$$

Wobei $\epsilon^* = \epsilon - \beta_K v$. Es wird angenommen, dass der Messfehler v mit der unbeobachteten Variable x_K unkorreliert ist ist. Diese Annahme ist bekannt als „classical errors-in-variables" Annahme[Wooldridge, 2002b, S.74] und impliziert:

$$\begin{aligned} cov(x_K, v) &= cov(x_K^* - v, v) = cov(x_K^*, v) - \sigma_v^2 = 0 \\ &\iff cov(x_K^*, v) = \sigma_v^2 \\ \implies cov(x_K^*, \epsilon^*) &= cov(x_K^*, \epsilon - \beta_K v) \\ &= cov(x_K^*, \epsilon) - cov(x_K^*, -\beta_K v) = -\beta_K cov(x_K^*, v) \\ cov(x_K^*, \epsilon^*) &= -\beta_K \sigma_v^2 \qquad (31) \end{aligned}$$

Das folgt, da laut Annahme $cov(x_K^*, \epsilon) = 0$ gilt. Sowohl ϵ^* als auch die erklärenden Variablen x_j^* hängen von den Messfehlern v_j ab. Daher korrelieren ϵ^* und x_K^* wie in (31) gezeigt, wenn $\beta_K \neq 0$. Wie im vorigen Abschnitt über das Problem bei nicht berücksichtigten, relevanten Variablen bereits ausführlich erläutert wurde werden dadurch die Annahmen OLS 3 und OLS 4 verletzt, was zu verzerrten und inkonsistenten Schätzern bei Messfehlern führt. Messfehler, welche die „classical errors-in-variables" erfüllen werden als klassische Messfehler bezeichnet.

Es kann gezeigt werden, dass der absolute Wert des Schätzers bei einem Messfehler geringer ist, als der des wahren Parameters. Zur Vereinfachung wird dies anhand des folgenden Beispiels aus Johnston und Dinardo [1997, S.154ff.] erläutert. In einer Regression mit einer erklärenden Variable x wird diese fehlerhaft gemessen. Die beobachtete Variable wird mit x^* bezeichnet und ist wie folgt definiert: $x^* = x + v$. Der wahre Zusammenhang lautet $y = \beta x + \epsilon$. Aus (11) folgt für den

Schätzer b:

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} y_i x_i^*}{\sum_{i=1}^{n} x_i^{*2}} \\ &= \frac{\sum_{i=1}^{n} x_i^*(\beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^{*2}} \\ &= \beta \frac{\sum_{i=1}^{n} x_i^* x_i}{\sum_{i=1}^{n} x_i^{*2}} + \frac{\sum_{i=1}^{n} x_i^* \epsilon_i}{\sum_{i=1}^{n} x_i^{*2}} \end{aligned} \tag{32}$$

Wird angenommen, dass (i) x, ϵ und v voneinander unabhängig sind und dass (ii) die Momente zweiter Ordnung sowie die Probability Limits existieren, dann folgt daraus:

$$\begin{aligned} \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n} x_i^{*2}) &= \sigma_x^2 + \sigma_\epsilon^2 \\ \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n} x_i^* x_i) &= \sigma_x^2 \\ \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n} x_i^* \epsilon_i) &= 0 \end{aligned}$$

Wird das Probability Limit von b aus (32) gebildet, so erhält man durch einsetzen:

$$\operatorname*{plim}_{n\to\infty} b = \beta(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_\epsilon^2}) \tag{33}$$

Da β in (33) mit einem Faktor multipliziert wird, der immer positiv und kleiner als eins ist, unterschätzt b systematisch den absoluten Wert für den wahren Parameter. Der Schätzer ist folglich näher an null als der wahre Parameter. Dieser Effekt wird als *Attenuation Bias* bezeichnet.

Ist die „classical errors-in-variables" Annahme nicht erfüllt, korrelieren also die unbeobachteten Variablen x_K mit den Messfehlern, so spricht man von nicht-klassischen Messfehlern. Wooldridge [2002, S.76] nennt die Frequenz des Marihuana Konsum als Beispiel für eine Variable mit nicht-klassischem Messfehler. Während Befragte, die kein Marihuana konsumieren dies wahrscheinlich auch korrekt angeben, ist bei Befragten, welche Marihuana konsumieren eine (selbst unbeabsichtigte) falsche Angabe wahrscheinlich. Der Messfehler ist folglich mit dem tatsächlichen Marihuana Konsum korreliert.

Ein nicht-klassischer Messfehler führt nicht zwingend zu einem Attenuation Bias. Das folgende Beispiel, bei dem angenommen wird, dass der Erwartungswert des Messfehlers null ist, soll dies verdeutlichen:

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n} x_i^* x_i) &= \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}(x_i + v)x_i) \\ &= \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}(x_i^2) + \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\sum_{i=1}^{n}(v_i x_i)) \\ &= \sigma_x^2 + cov(x_i, v_i) = \sigma_x^2 + z\end{aligned}$$

Aus (33) wird $\operatorname*{plim}_{n\to\infty} b = \beta(\sigma_x^2 + z)(\sigma_x^2 + \sigma_\epsilon^2)^{-1}$. Daraus folgt, dass nur dann ein Attenuation Bias vorliegt, wenn $-\sigma_x^2 < z < \sigma_\epsilon^2$. Für den Fall, dass $z = \sigma_\epsilon^2$ ist der Schätzer konsistent, für alle anderen Fälle wird der absolute Wert des Parameters überschätzt.

Simultane Kausalität. Simultanität entsteht, wenn zumindest eine der erklärenden Variablen x_j zum Teil von der abhängigen Variable y bestimmt wird. Ist dies der Fall, korreliert x_j mit dem Störterm. Will man z.B. eine Regression auf das Modell in (34) durchführen, bei dem die K-te erklärende Variable eine Funktion[15] von unter anderem y ist, dann wird man auf notwendigerweise auf Korrelation zwischen ϵ und x_K stoßen.

$$\begin{aligned} y &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \epsilon \qquad (34) \\ x_K &= f(y) \\ \frac{dx_k}{dy} &\neq 0 \end{aligned}$$

In diesem Beispiel ist die erklärende Variable x_K auf mehr als eine Weise mit y verbunden. Die Kausalbeziehung besteht also nicht nur in einer Richtung, von x_K nach y, sondern auch in der entgegengesetzten Richtung. Entsteht durch den

[15] Wobei x_K nicht konstant in y sein darf. Außerdem darf x_K natürlich auch von anderen Faktoren als y abhängen.

Störterm eine exogene Variation in y, dann wirkt sich das auch auf x_K aus, da y und x_K über eine Funktion vebunden sind und somit x_K von y determiniert wird. Indirekt wirkt somit eine Variation des Störterms auch auf x_K, wodurch beide Variablen korrelieren. Wie schon bei den vorigen Beispielen impliziert Korrelation zwischen zumindest einer erklärenden Variable und dem Störterm verzerrte und inkonsistente Schätzer. Das Problem der simultanen Kausalität taucht in der Praxis oft auf, da in den Sozialwissenschaften interdependente Systeme häufig sind.[Stocker, 2011, S.341]

3.3 Wie kann Endogenität verhindert werden?

Dieses Unterkapitel behandelt Methoden welche eingesetzt werden, um die Probleme von endogenen, erklärenden Variablen zu umgehen. Zunächst wird die Methode der Instrumentvariablen kurz in einem Unterkapitel analysiert. Danach handelt ein ausführlicheres Unterkapitel von *Natürlichen Experimenten*, welche eine weitere Methode darstellen.

3.3.1 Instrumentvariablen

Im folgenden Unterkapitel wird die in Lehrbüchern wohl weitverbreitetste Variante zur Verhinderung von Endogenität behandelt: Instrumentvariablen. Gehen wir von folgendem Modell aus:

$$\begin{aligned} y &= \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \epsilon \qquad (35) \\ \mathbb{E}(\epsilon \mid x_j) &= 0, \quad j = \{1, \ldots, K-1\} \\ \mathbb{E}(\epsilon) &= 0 \end{aligned}$$

Die Exogenitätsannahme wurde nicht für x_K getroffen. x_K kann folglich mit ϵ korreliert sein, daher ist diese erklärende Variable womöglich endogen. Um trotz dieses Problemes brauchbare Schätzer ermitteln zu können, kann eine Instrumentvariable, oder kurz ein Instrument, für x_K benützt werden. Die Idee hinter einer Instrumentvariablen ist folgende: Es wird versucht die Streuung der endogenen Variable in zwei Teile zu zerlegen, wovon ein Teil mit dem Störterm korreliert und der zweite Teil rein exogene Streuung enthält. Dazu dienen Instrumentvariablen und die Methode der zweistufigen kleinsten Quadrate, auch 2SLS genannt.[Stocker, 2011, S.342ff.] Dabei kann natürlich nicht jede Variable als Instrument für x_K dienen. Die Variable muss bestimmte Bedingungen erfüllen, welche im folgenden Abschnitt erläutert werden.

Bedingungen für Instrumentvariablen Eine Instrumentvariable (IV) muss folgende zwei Bedingungen erfüllen:

Bedingung IV 1. *Die potenzielle Instrumentvariable z darf nicht mit dem Störterm der ursprünglichen Regressionsgleichung korrelieren, das heißt $cov(z, \epsilon) = 0$.*

Bedingung IV 2. *Die Instrumentvariable muss mit dem endogenen Regressor partiell korreliert sein. Partielle Korrelation liegt dann vor, wenn bei der linearen Projektion*[16] *von dem endogenen Regressor auf alle exogenen Variablen der geschätzte Koeffizient für die Instrumentvariable ungleich null ist.*

$$\begin{aligned} x_K &= \delta_1 + \delta_2 x_2 + \ldots + \delta_{K-1} x_{K-1} + \theta z + r_K \\ \mathbb{E}(r_K \mid x_j, z) &= 0, \quad j = \{1, \ldots, K-1\} \end{aligned} \tag{36}$$

Im Fall von Modell (35) *bedeutet das, dass für den geschätzten Parameter $\theta \neq 0$ in* (36) *gelten muss.*

In anderen Worten, der Regressionskoeffizient der IV θz darf nicht null sein, nachdem für alle anderen exogenen Variablen kontrolliert wurde.[Wooldridge, 2002b, S.83ff.] Sind beide Bedingungen für eine Variable erfüllt, so ist diese Variable ein potenzieller Kandidat einer Instrumentvariable für x_K. Gibt es mehrere Variablen, welche die beiden Bedingungen erfüllen, so wird die Variable, bzw. die Kombination an Variablen gewählt, welche am stärksten mit x_K partiell korrelieren.

Intrumentvariablenschätzer Der Schätzer für das Modell (35) wird nicht mit der OLS-Methode, sondern anhand der Methode der Momente hergeleitet. Bei dieser Methode wird angenommen, dass die Mittelwerte aus einer zufällig ermittelten Stichprobe als Schätzer für den Erwartungswert der Grundgesamtheit die-

[16] Die lineare Projektion ist definiert als $L(y \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$ für jeden $(1 \times K)$ Vektor $\mathbf{x}$ für den gilt, dass die $(K \times K)$ Kovarianzmatrix von $\mathbf{x}$ vollen Rang besitzt.[Wooldridge, 2002b, S.25]

nen.[Stocker, 2011, S.345] $\mathbf{x}$ bezeichnet den Vektor der erklärenden Variablen inklusive dem endogenen x_K. $\mathbf{z}$ ist der Vektor, in welchem das Instrument z für die endogene Variable x_K substituiert wurde. $\mathbf{x}, \mathbf{z}$ sind $1 \times K$ Zeilenvektoren. Es wird angenommen, dass das Instrument im kommenden Beispiel die Bedingung IV1 erfüllt. Daraus und aus $\mathbb{E}(\epsilon) = 0$ folgt $\mathbb{E}(\mathbf{z}\epsilon) = \mathbf{0}$. Außerdem wird angenommen, dass die Matrix $\mathbb{E}(\mathbf{z}\mathbf{x}^T)$ vollen Spaltenrang hat und daher invertierbar ist. Die Herleitung folgt Wooldridge [2002, S.85ff.]

$$\begin{aligned}
y &= \mathbf{x}\boldsymbol{\beta} + \epsilon \\
\mathbf{z}^T y &= \mathbf{z}^T(\mathbf{x}\boldsymbol{\beta} + \epsilon) \\
\mathbb{E}(\mathbf{z}^T y) &= \mathbb{E}(\mathbf{z}^T(\mathbf{x}\boldsymbol{\beta} + \epsilon)) = \mathbb{E}(\mathbf{z}^T\mathbf{x})\boldsymbol{\beta} \\
\iff \boldsymbol{\beta} &= \mathbb{E}(\mathbf{z}^T\mathbf{x})^{-1}\mathbb{E}(\mathbf{z}^T y) \\
\implies \boldsymbol{b_{IV}} &= (\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i^T \mathbf{x}_i)^{-1}(\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i^T y_i) \\
&= (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{y} \qquad (37)
\end{aligned}$$

Wobei $\mathbf{z}_i^T\mathbf{x}_i$ das äußere Produkt der beiden Vektoren und daher eine $K \times K$ Matrix ist. Das sollte nicht mit dem inneren Produkt verwechselt werden. Bevor gezeigt wird, dass der Schätzer aus (37) konsistent ist, wird eine weitere Methode zur Ermittlung des Schätzers vorgestellt.

Methode der zweistufigen kleinsten Quadrate. Anhand der Methode der zweistufigen, kleinsten Quadrate kann recht anschaulich erklärt werden, was Instrumentvariablen bewirken. Um aus der Streuung der endogenen, erklärenden Variable x_K den exogenen Part „herauszufiltern” wird zunächst eine erste Regression durchgeführt, bei der die systematische Komponente von x_K, $\hat{x}_K$ ermittelt wird.

$$\begin{aligned}
x_K &= \delta_1 + \delta_2 x_2 + \ldots + \delta_{K-1}x_{K-1} + \theta z + r_K \\
\mathbb{E}(x_K \mid x_1, \ldots, x_{K-1}, z) &= \hat{x}_K = \delta_1 + \delta_2 x_2 + \ldots + \delta_{K-1}x_{K-1} + \theta z \qquad (38)
\end{aligned}$$

Im zweiten Schritt wird mit dem im ersten Schritt ermittelten Wert für $\hat{x}_K$, anstatt x_K die ursprüngliche Regression aus (35), mit der OLS Methode durchgeführt. Die veränderte Matrix der erklärenden Variablen wird mit $\hat{\mathbf{X}}$ anstatt $\mathbf{X}$ bezeichnet. In dieser Matrix werden die Werte für x_K durch $\hat{x}_K$ ersetzt. Analog zur Herleitung des OLS Schätzers in (11) entsteht der Schätzer in (39).

$$\mathbf{b_{2SLS}} = (\hat{\mathbf{X}}^T\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}^T\mathbf{y} \tag{39}$$

Erwartungstreue des IV-Schätzers. Die Schätzer in (37) und (39) sind allgemein nicht erwartungstreu. Beide Schätzer sind äquivalent, wenn $\hat{x}_K$ als Instrumentvariable z in (37) benützt wird.[17]

$$\begin{aligned}
\mathbf{b_{IV,2SLS}} &= (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T\mathbf{X}\boldsymbol{\beta} + (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T\boldsymbol{\epsilon} \\
&= \boldsymbol{\beta} + (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T\boldsymbol{\epsilon} \\
\mathbb{E}(\mathbf{b_{IV,2SLS}} \mid \hat{\mathbf{X}}) &= \boldsymbol{\beta} + \mathbb{E}((\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T\boldsymbol{\epsilon} \mid \hat{\mathbf{X}}) \qquad (40)
\end{aligned}$$

Selbst wenn Annahme OLS 3 für dieses Problem erfüllt ist, ist der ermittelte Schätzer nicht erwartungstreu, da der rechte Term in (40) nicht notwendigerweise den Nullvektor ergibt. Dies folgt aus der Annahme, dass $\mathbf{X}$ eine Matrix aus Zufallsvariablen sein kann.

Konsistenz des IV-Schätzers. In diesem Paragraph wird gezeigt, dass die in (37) und (39) ermittelten Schätzer konsistent sind, wenn in (41) $n^{-1}\hat{\mathbf{X}}^T\mathbf{X}$ in Wahrscheinlichkeit gegen eine deterministische und invertierbare Matrix konvergiert und $n^{-1}\hat{\mathbf{X}}^T\boldsymbol{\epsilon}$ gegen den Nullvektor konvergiert. Das ist der Fall, wenn $\mathbb{E}(\hat{\mathbf{X}}^T\epsilon) = \mathbf{0}$ was

[17] Da $\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{X} = \mathbf{P}_Z\mathbf{X}$, wobei $\mathbf{P}$ eine idempodente, symetrische Matrix ist, gilt $\hat{\mathbf{X}}^T\mathbf{X} = (\mathbf{P}_Z\mathbf{X})^T\mathbf{X} = \mathbf{X}^T\mathbf{P}\mathbf{X} = \mathbf{X}^T\mathbf{P}\mathbf{P}\mathbf{X} = (\mathbf{P}\mathbf{X})^T\mathbf{P}\mathbf{X} = \hat{\mathbf{X}}^T\hat{\mathbf{X}}$.[Wooldridge, 2002b, S.91]

durch Bedingung IV1 und der Annahme $\mathbb{E}(\epsilon) = 0$ impliziert wird.

$$
\begin{aligned}
\mathbf{b_{IV,2SLS}} &= (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (\hat{\mathbf{X}}^T\mathbf{X})^{-1}\hat{\mathbf{X}}^T\boldsymbol{\epsilon} \\
&= \boldsymbol{\beta} + (\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{X})^{-1}\frac{1}{n}\hat{\mathbf{X}}^T\boldsymbol{\epsilon} \\
\operatorname*{plim}_{n\to\infty} \mathbf{b_{IV,2SLS}} &= \operatorname*{plim}_{n\to\infty}(\boldsymbol{\beta} + (\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{X})^{-1}\frac{1}{n}\hat{\mathbf{X}}^T\boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + \operatorname*{plim}_{n\to\infty}(\frac{1}{n}\hat{\mathbf{X}}^T\mathbf{X})^{-1}\operatorname*{plim}_{n\to\infty}(\frac{1}{n}\hat{\mathbf{X}}^T\boldsymbol{\epsilon}) \\
\operatorname*{plim}_{n\to\infty} \mathbf{b_{IV,2SLS}} &= \boldsymbol{\beta} \qquad (41)
\end{aligned}
$$

Kritik Ein Problem, das auftreten kann, ist der Fall in dem ein Instrument nur schwach mit der endogenen Variable korreliert und dazu das Instrument die Exogenitätsannahme verletzt. Dann ist der Schätzer schon bei einer schwachen Korrelation zwischen dem Instrument und dem Störterm sehr stark inkonsistent. Um intuitiv zu verstehen, warum dieses Problem so schwerwiegend ist, nehmen wir den Fall von einer x Variable und einem Instrument z. Dafür ergibt sich folgende Gleichung:

$$
\begin{aligned}
\operatorname*{plim}_{n\to\infty} b_{IV,2SLS} &= \beta + \operatorname*{plim}_{n\to\infty}(n^{-1}zx)^{-1}\operatorname*{plim}_{n\to\infty}(n^{-1}z\epsilon) \\
&= \beta + \frac{\mathbb{E}(z\epsilon)}{\mathbb{E}(zx)} = \beta + \frac{cov(z\epsilon)}{cov(zx)} \qquad (42)
\end{aligned}
$$

In (42) wird deutlich, dass plim $b_{IV,2SLS}$ stark steigt oder fällt wenn die Kovarianz zwischen z und x klein ist. Auf eine Korrelation zwischen z und x kann und sollte in jedem Fall getestet werden. Tritt dieses Problem dennoch auf, so sind die resultierenden Schätzer unbrauchbar, um Schlüsse über β zu treffen.

3.3.2 Natürliche Experimente

Nachdem im letzten Unterkapitel die Methode der Instrumentvariablen erklärt wurde, folgt in diesem Unterkapitel eine zweite Methode, um Endogenität zu umgehen. Möchte man die Streuung in der abhängigen Variable durch die Streuung in einer erklärenden Variable erklären, so gilt es zu beachten, dass die erklärende

Variable nicht endogen ist. Ansonsten sind, wie bereits ausführlich beschrieben, die Schätzer, welche zur Lösung verzerrt und inkonsistent, also unbrauchbar. Möchte man trotz Endogenität ein Modell schätzen, so gibt es bestimmte Methoden, um trotzdem brauchbare Schätzer zu erhalten. Eine davon ist die Methode der natürlichen Experimente.

Die in ökonometrischen Studien oft als „natürliche Experimente" bezeichnete Methode basiert konzeptuell auf randomisierten Experimenten. Anders als bei klassischen Experimenten ist bei randomisierten Experimenten die Zuordnung der Beobachtungen in eine Treatment Gruppe und eine Kontrollgruppe nicht durch die Leiterin des Experiments kontrolliert. Vielmehr gibt das natürliche Experiment auf „natürliche" Weise den Impuls zur Einteilung der Gruppen. Geht man davon aus, dass bei einem Experiment die Einteilung der Beobachtungen in Gruppen anhand eines zufälligen Algorithmus geschieht, dann sollten die einzelnen Elemente der Gruppen vor dem Experiment unabhängig voneinander sein und es sollte keine systematischen Unterschiede zwischen den Gruppen geben. Bei natürlichen Experimenten wird vorausgesetzt, dass es exogen ist, das heißt das Eintreten des natürlichen Experimentes ist unabhängig von den zu untersuchenden erklärenden Variablen. Daher wird spekuliert, dass Daten aus randomisierten Experimenten sehr nahe an Daten aus Experimenten herankommen, bzw. dass die Daten Eigenschaften besitzen, als ob sie aus einem Experiment stammen. Beispielsweise eignen sich als nätürliches Experimente in der Ökonomie laut Meyer Veränderungen in Gesetzen, Einzugsmechanismen des Militärs („Draft Mechanismen") und Veränderungen in Sozialversicherungsbestimmungen.[Meyer, 1995, S.151ff.] Diese Liste hat aber nicht den Anspruch komplett zu sein.

Im Fokus dieser Methode steht es, ähnlich wie bei dem Ansatz der Instrumentvariable, durch das nätürliche Experiment exogene Streuung in ansonsten endogenen, erklärenden Variablen zu erzeugen. Eine Kernidee dieser Methode ist, dass eine Gruppe, genannt Treatment-Gruppe, dem exogenen Treatment, in diesem Fall dem natürlichen Experiment, dem Treatment, ausgesetzt ist, während eine Kontrollgruppe diesem Treatment nicht ausgesetzt ist.

Natürliche Experimente können als Instrumentvariable genützt werden. Diese

Methode werden anhand eines kurzen Beispiels erläutert. Angrist hat die Auswirkung von geleistetem Militärdienst auf das Einkommen untersucht.[Angrist, 1990] Eine einfache Regression zwischen dem Einkommen und z.B. einem Dummy für geleisteten Militärdienst, Arbeitsjahren im Militärdienst oder militärischem Rang könnte aus mehreren Gründen Endogenität hervorrufen und somit ein verzerrtes Ergebnis produzieren. So könnte man argumentieren, dass sich vor allem Personen zum Militärdienst verpflichten lassen, welche von vornherein schlechtere Arbeitsmarktaussichten haben, als jene, die sich nicht verpflichten. Dieser Effekt kann von unbeobachteten Variablen, wie eventuell Begabung, hervorgerufen werden, welche ebenfalls für das Einkommen relevant sind. Ebenfalls ist es möglich, dass der z.B. durch den Prozess der militärischen Ausbildung, Personen mit schlechten gesundheitlichen Eigenschaften aussortiert werden. In Summe scheint es nicht unplausibel, dass die Variable *Arbeitsjahre im Militär* mit nicht berücksichtigten, relevanten Variablen korrelieren, was Endogenität mit allen bekannten Problemen verursacht. Angrist benutzt den Draftmechanismus während des Vietnamkrieges, um dieses Problem zu vermindern. Der Draft war eine Art Lotterie in welcher ausgelost wurde, wer Militärdienst zu leisten hat. Angrist zeigt, dass Draft-Teilnehmer mit niedrigen Zahlen eine höhere Wahrscheinlichkeit hatten gezogen zu werden. Die endogene Variable für den Militärdienst ist also mit dem Draftmechanismus korreliert, jedoch ist es unwahrscheinlich, dass der Draftmechanismus mit einer der vorhin angesprochenen, unbeobachteten Variablen korreliert. [Wooldridge, 2002b, S.88] Daraus folgt, dass eine Variable für den Draftmechanismus rein exogene Streuung in der bislang endogenen, erklärenden Variable bewirkt. Beide Bedingungen IV1 und IV2 für eine potentielle Instrumentvariable sind folglich erfüllt.

Es gibt verschiedene Forschungsdesigns, die benützt werden, um Studien mit natürlichen Experimenten auszuwerten. Eine beliebte Methode wird im nächsten Paragraphen vorgestellt.

Difference-in-Differences Methode. Die Difference-in-Differences Methode, ab diesem Punkt auch DD genannt, ist ein mögliches Forschungsdesign um Studien zu untersuchen, die ein natürliches Experiment ausnützen, um exogene Streuung zu generieren. Dabei verhindert die Methode als solches nicht unbedingt Endogenität, sie dient vorallem der Analyse von natürlichen Experimenten, welche Endogenität verhindern können. Das Design der DD Methode sieht im einfachsten Fall vor, dass Daten für (i) zwei Zeitpunkte und (ii) zwei Gruppen gemessen werden. Die beiden Zeitpunkte sind vor und nach dem Treatment und die beiden Gruppen sind die bereits erwähnte Treatment- und Kontrollgruppe. Da es Daten zu mindestens zwei Zeitpunkten für zwei Gruppen geben muss, werden für die DD-Methode Paneldaten benötigt.

$$y_t^j = \beta + \beta_1 D_t + \beta^1 D^j + \beta_1^1 D_t^j + \epsilon_t^j \tag{43}$$
$$\mathbb{E}(\epsilon_t^j \mid D_t^j) = \mathbb{E}(\epsilon_{ti}^j \mid D_t) = \mathbb{E}(\epsilon_{ti}^j \mid D^j) = 0$$

Wobei y_t^j die abhängige Variable ist. $t \in \{0, 1\}$ ist ein Zeitindex, der vor dem Treatment null und nach dem Treatment eins beträgt. ϵ_t^j ist der Störterm der jeweiligen Gruppe j zum Zeitpunkt t. Der Dummy D_t beträgt eins, wenn die Beobachtung zeitlich nach dem Treatment ist, $t = 1$, und null, wenn davor. $j \in \{0, 1\}$ indiziert die Gruppen, $j = 1$ wenn das beobachtete Element in der Treatment-Gruppe ist und $j = 0$ wenn nicht. $D^j = 1$ wenn $j = 1$ und $D^j = 0$ wenn $j = 0$. D_t^j ist ein Dummy der eins beträgt, wenn das beobachtete Element zur Treatment-Gruppe zählt und die Beobachtung nach dem Treatment erfolgt, das heißt $D_t^j = 1$ wenn $j = 1$ und $t = 1$ und $D_t^j = 0$ wenn $j \neq 1$ oder $j \neq 1$. Ist das Modell richtig spezifiziert und entspricht dem wahren DGP der Grundgesamtheit, dann ist β_1^1 der Kausalzusammenhang zwischen dem Treatment und der abhängigen Variable. [18] Der interessierende Parameter ist β_1^1. Der Schätzer für β_1^1, nennen wir ihn b_1^1 ist definiert als:

$$b_1^1 = \bar{y}_1^1 - \bar{y}_0^1 - (\bar{y}_1^0 - \bar{y}_0^0) \tag{44}$$

$\bar{y}_1^1$ ist z.B. der Stichprobenmittelwert der abhängigen Variable der Treatment-Gruppe nach dem Treatment, $\bar{y}_0^0$ ist der Mittelwert der Kontrollgruppe vor dem Treatment, usw.

Erwartungstreue. Wenn davon ausgegangen wird, dass das Stichprobenmittel ein unverzerrter Schätzer für den wahren Wert von y_t^j ist, gilt $\mathbb{E}(\bar{y}_t^j) = \bar{y}_t^j =$

[18] Die Notation wurde ungefähr aus Mayer [1995] übernommen.

$\mathbb{E}(y_t^j)$. Für diese Annahme ist es wichtig, dass Treatment- und Kontrollgruppe zufällig eingeteilt werden, sodass bei der Einteilung keine Verzerrung entstehen kann. Außerdem gilt $\mathbb{E}(\epsilon_t^j) = \mathbb{E}(\mathbb{E}(\epsilon_t^j \mid D_t^j)) = 0$. Folglich lautet der Schätzer aus (44):

$$\begin{aligned}\mathbb{E}(b_1^1) &= \mathbb{E}(\bar{y}_1^1 - \bar{y}_0^1 - (\bar{y}_1^0 - \bar{y}_0^0)) \\ &= \mathbb{E}(y_1^1) - \mathbb{E}(y_0^1) - (\mathbb{E}(y_1^0) - \mathbb{E}(y_0^0))\end{aligned}$$

Für die bedingten Erwartungswerte $\mathbb{E}(b_t^j)$ ergeben sich aus (43) folgende Werte:

$$\begin{aligned}&\mathbb{E}(y \mid D_t = 0, D^j = 1) = \beta + \beta^1 \\ &\mathbb{E}(y \mid D_t = 0, D^j = 0) = \beta \\ &\mathbb{E}(y \mid D_t = 1, D^j = 1) = \beta + \beta_1 + \beta^1 + \beta_1^1 \\ &\mathbb{E}(y \mid D_t = 1, D^j = 0) = \beta + \beta_1\end{aligned}$$

Also lautet der Erwartungswert für den Schätzer wie folgt:

$$\begin{aligned}\mathbb{E}(b_1^1) &= \beta + \beta_1 + \beta^1 + \beta_1^1 - (\beta + \beta^1) - (\beta + \beta_1 - \beta) \\ &= \beta_1^1\end{aligned}$$

Unter den Annahmen aus (43) und $\bar{y}_t^j = \mathbb{E}(y_t^j)$ ist der DD Schätzer also unverzerrt. Als nächstes wird die Konsistenz des Schätzers untersucht.

Konsistenz. Die Konsistenz folgt aus der Annahme, dass (i) das Modell richtig spezifiziert ist, also dass außer β_1^1 keine weitere Interaktion zwischen $t = 1$ und $j = 1$ herrscht und (ii) dass die abhängigen Variablen in den Gruppen identisch und unabhängig verteilt sind. n_t^j ist dabei die Stichprobengröße der jeweiligen Gruppe j zu Zeitpunkt t. Strebt n_t^j gegen unendlich, dann konvergiert der Schätzer gegen den wahren Wert des Parameter.

$$
\begin{aligned}
b_1^1 &= \bar{y}_1^1 - \bar{y}_0^1 - (\bar{y}_1^0 - \bar{y}_0^0) \\
\operatorname*{plim}_{n_t^j \to \infty} (b_1^1) &= \operatorname*{plim}_{n_t^j \to \infty} (\bar{y}_1^1 - \bar{y}_0^1 - (\bar{y}_1^0 - \bar{y}_0^0)) \\
&= \operatorname*{plim}_{n_1^1 \to \infty} ((n_1^1)^{-1} \sum_{i=1}^{n_1^1} y_{1i}^1) - \operatorname*{plim}_{n_0^1 \to \infty} ((n_0^1)^{-1} \sum_{i=1}^{n_0^1} y_{0i}^1) \\
&\quad - (\operatorname*{plim}_{n_1^0 \to \infty} ((n_1^0)^{-1} \sum_{i=1}^{n_1^0} y_{1i}^0) - \operatorname*{plim}_{n_0^0 \to \infty} ((n_0^0)^{-1} \sum_{i=1}^{n_0^0} y_{0i}^0)) \\
&= \mathbb{E}(y_1^1) - \mathbb{E}(y_0^1) - (\mathbb{E}(y_1^0) - \mathbb{E}(y_0^0)) \\
\operatorname*{plim}_{n \to \infty} (b_1^1) &= \beta + \beta_1 + \beta^1 + \beta_1^1 - (\beta + \beta^1) - (\beta + \beta_1 - \beta) = \beta_1^1
\end{aligned} \tag{45}
$$

β_1 kontrolliert für einen generellen Zeittrend der beide Gruppen in gleichem Ausmaß betrifft. β^1 macht das gleiche für zeitunabhängige Unterschiede in den Mittelwerten der beiden Gruppen. Beide Effekte würden, würde man sie nicht beachten, Endogenität im interessierten Schätzer verursachen. β_1^1 misst damit die Differenz der Differenzen der Mittelwerte vor und nach dem Treatment, zwischen beiden Gruppen - daher der Name der Methode. Die dabei entscheidende Annahme, die einen unverzerrten Schätzer ergibt, ist $\mathbb{E}(\epsilon_{ti}^j \mid D_t^j) = 0$. Das heißt die Störterme der jeweiligen Gruppe in Periode t korrelieren nicht mit D_t^j und es ist genauso wahrscheinlich, dass der Störterm positiv wie negativ ist. Es darf außer dem Treatment keine weitere Beeinflussung von $t = 1$ auf die Treatment-Gruppe geben. Es sollte also keinen zeitabhängigen Trend geben, der die Gruppen unproportional beeinflusst. Im Modell (44) wird angenommen, dass sich die Treatment- und die Kontrollgruppe nur hinsichtlich des Treatments unterscheiden. Würde das Treatment ausbleiben, wäre der Erwartungswert für y identisch. Ist diese Annahme nicht erfüllt, ist der Schätzer nicht mehr unverzerrt. Ein Beispiel, wann das der Fall wäre, ist ein Trend, der Kontroll- und Treatmentgruppe ungleichmäßig beeinflusst. Daraus ergibt sich, dass es bei der DD-Methode von großem Vorteil ist, wenn sich Kontroll- und Treatmentgruppe ähneln.[Meyer, 1995, S.155]

Kritik. Es gibt eine ganze Reihe an Literatur, die sich mit potentiellen Problemen bei der DD Methode beschäftigen. In seinem 1978 erschienenen Artikel über den Effekt von staatlichen Trainingsprogrammen auf das Einkommen der Teilnehmer beschreibt Ashenfelter, dass das Durchschnittseinkommen der Teilnehmer in der Periode vor Beginn des Trainingsprogramms merklich sinkt. Ein Effekt, der auch in weiteren Studien auftritt (siehe Heckman und Smith [1999] für eine kurze Auflistung) und als *Ashenfelter's Dip* oder als *pre-programme dip* bezeichnet wird. Demnach ist es wichtig zu wissen, ob dieser Rückgang permanenter- oder vorübergehender Natur ist, um die Validität einer DD Schätzung zu garantieren. [Heckman and Smith, 1999, S.314] Die Frage nach der Natur des Rückgangs hängt maßgeblich damit zusammen, wie sich das Einkommen einer Teilnehmerin entwickelt hätte, hätte sie das Programm *nicht* besucht. Folglich ist die Interpretation der Auswirkung des Trainingsprogramms abhängig von der Natur des *pre-programme dip*. Zum Beispiel kann eine Steigerung des Einkommens nach einem vorübergehenden pre-programme dip fälschlicherweise dem Trainingsprogramm zugeordnet werden, weil das Kontrafaktual nicht korrekt ermittelt wurde. Der Forscher könnte meinen, bei dem pre-programme dip handelt es sich um eine permanente Veränderung, obwohl diese in Wahrheit nur vorübergehend ist. Eine weitere prominente Kritik stammt von Bertrand et al. [2004] die kritisieren, dass die herkömmliche OLS-Methode wegen autokorrelierten Daten zu geringe Standardfehler für den DD Schätzer produziert, was zu überschätzten t-Statistiken führt und daher die Wahrscheinlichkeit eines Typ I Fehlers erhöht. Bertrand et. al schlagen einige Methoden vor um dieses Problem in den Griff zu bekommen, u.a. Bootstrapping oder die Einteilung der Daten in prä- und post-Treatment Phasen. Außerdem wird angeregt, DD Modelle mit anderen Schätztechniken als OLS zu bearbeiten, dazu zählen die Authoren Generalized Least Squares oder die Generalized Method of Moments.[Bertrand et al., 2004]

3.3.3 Fixed Effects (feste Effekte)

Eine, in der Analyse von Paneldaten, sehr beliebte Methode ist es, feste Effekte zuzulassen. Dadurch kann Endogenität, die bei zeitunabhängigen, nicht berücksichtigten, jedoch relevanten Variablen entsteht, wirkungsvoll und einfach verhindert werden. Eine Querschnittsbeobachtung i aus einem Panel, bei einem Modell mit nicht berücksichtigten Variablen, kann wie folgt aussehen:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \tag{46}$$

Wobei $\mathbf{x}_{it}$ ein $1 \times K$ Zeilenvektor- und $\boldsymbol{\beta}$ ein $K \times 1$ Spaltenvektor ist. In diesem Modell taucht eine unbeobachtete Variable c_i auf, die konstant in t ist. Alle anderen Variablen variieren über beide Indizes i und t. Nun stellt sich die Frage, wie c_i behandelt werden soll. Ist c_i eine Zufallsvariable, die für alle i eine gemeinsame Wahrscheinlichkeitsverteilung besitzt[19], oder ist c_i ein fester Effekt, also ein Parameter, der unabhängig von t, einen fixen Wert für jedes i ergibt? Behandelt man c_i als festen Effekt, dann wird c_i als Parameter betrachtet und für jede Beobachtung i geschätzt. Diesen Ansatz kann man sich wie einen Dummy vorstellen wenn, der eins ergibt, wenn $i = i^*$ und null, wenn $i \neq i^*$. Das bedeutet, dass das unbeobachtete, aber geschätzte c_i mit den erklärenden Variablen $\mathbf{x}_{it}$ korrelieren darf, ohne dass Endogenität entsteht.[Wooldridge, 2002b, S.251ff.] Somit ist c_i in der Lage, für systematische Unterschiede zwischen verschiedenen Beobachtungen i zu kontrollieren.

4 Politik und Banken

In dieser Arbeit werden Artikel über das Verhalten bei der Kreditvergabe von Banken untersucht, bei denen der Staat ein Anteilseigner ist, welche also von Politikern zumeist indirekt kontrolliert oder zumindest beeinflusst werden. Die Studien in diesen Artikeln haben oft das Problem, dass unklar ist, wie erfolgreich Banken gewesen wären, wenn Politiker keinen Einfluss auf sie gehabt hätten.[Halling et al., 2010, S.10] Es ist denkbar, dass relevante Effekte, die in den Studien nicht berücksichtigt werden, mit den erklärenden Variablen korrelieren und dadurch Endogenität verursachen. Diese macht die Interpretation der Schätzer aber unmöglich. Zunächst behandelt das erste Unterkapitel ausgewählte, wissenschaftliche Artikel aus dem Bereich Politik und Banken. Daraufhin gibt es im zweiten Unterkapitel einen kurzen Überblick über die Theorie, die hinter dem Verhalten bei der Kreditvergabe

[19]Dann würde man von einem Random Effects Modell sprechen.

von öffentlichen Banken steht. Im dritten und vierten Unterkapitel werden zwei wissenschaftliche Artikel, mit besonderem Augenmerk auf die verwendeten Methoden zur Verhinderung der Endogenität, analysiert. Im ersten Artikel werden Instrumentvariablen verwendet, während im zweiten Artikel auf ein natürliches Experiment zurückgegriffen wird.

4.1 Empirische Arbeiten aus dem Feld Politik und Banken

Es existiert eine große Anzahl von Publikationen, noch nicht veröffentlichten Artikeln und Arbeitspapieren, welche sich mit dem Thema Politik und Banken beschäftigen. Dieses Unterkapitel soll dem Leser einen kurzen Überblick über einige dieser Arbeiten verschaffen. Bezüglich der Methoden zur Verhinderung der Endogenität gilt es anzumerken: Alle Autoren verwenden Paneldaten mit festen Effekten, um Endogenität zu vermeiden. Daher wird diese Methode in diesem Unterkapitel nicht jedesmal angesprochen.

LaPorta et al. [2002] untersuchen in einer internationalen Studie das Ausmaß von staatlichen Beteiligungen an Banken. Sie finden heraus, dass diese um 1990 weitverbreitet sind. So waren im Jahr 1995 durchschnittlich 42% des Eigenkapitals der zehn größten Banken eines Landes in staatlichem Besitz.[LaPorta et al., 2002, S.267] Außerdem zeigen die Autoren, dass in ihrer Studie ein hohes Maß an staatlichen Beteiligungen negativ mit dem folgenden Wachstum des Durchschnittseinkommens und dem Wachstum der Produktivität korrelieren. [LaPorta et al., 2002, S.285]

Micco et al. [2005] untersuchen den Zusammenhang zwischen der Eigentümerschaft einer Bank und der Profitabilität der Bank anhand einem internationalen Datensatz auf Bankenebene. Die Autoren finden für Entwicklungsländer eine Tendenz, die besagt, dass staatseigene Banken niedrigere Profitabilität als Privatbanken haben. In entwickelten Ländern ist diese Tendenz laut Micco et al. [2005] nicht zu finden. Außerdem untersuchen sie, ob die Differenz der Profitabilität zwischen staatseigenen und privaten Banken in Wahljahren steigt, was auf einen Einfluss von politischen Erwägungen auf die Profitabilität der Staatsbanken hindeuten würde.

Sie finden, dass die Differenz in Wahljahren steigt (dieser Effekt ist bei 5% Typ I Fehler signifikant). Micco et al. [2005] setzen weder Instrumentvariablen ein, noch nützen sie ein natürliches Experiment, obwohl die Autoren Zweifel über die Exogenität zweier Kontrollvariablen hegen - es fehlen jedoch, so die Autoren, gute Instrumente. Anstelle dessen zeigen sie, dass diese Variablen in ihren Modellen keinen großen Einfluss auf die Profitabilität haben.

Cole [2009] untersucht in seiner Studie den Effekt, den die Eigentümerstruktur von Banken auf die Entwicklung des Finanzsystems, der Zinsraten, die Qualität der Finanzdienstleistungen, der Arbeitslosigkeit und die Investitionen haben. Die Daten für diese Studie stammen aus Indien und beschreiben Märte auf Städteebene. Insgesamt werden 2400 Märkte in Indien von 1981 bis 2000 untersucht. Ähnlich wie Halling et al. [2010] benützt Cole [2009] ein natürliches Experiment um Endogenität zu vermeiden und Kausalzusammenhänge in den bereits genannten Effekten zu identifizieren. Als natürliches Experiment zieht der Autor die Verstaatlichung der Banken in Indien im Jahr 1980 heran. Diese Verstaatlichung hat für eine exogene Veränderung in der Eigentümerstruktur der Banken auf den Märkten auf Städteebene gesorgt. Das heißt, in verschiedenen Städten oder Dörfern kam es durch die Verstaatlichung zu unterschiedlichen Veränderungen in der Eigentümerstruktur: in manchen Märkten wurden alle operierenden Banken (nicht) verstaatlicht, in anderen nur ein Teil. Um zu zeigen, dass die, durch die Verstaatlichung erzeugte, Veränderung exogen ist, überprüft Cole [2009] ob sich die Gruppen der verstaatlichten und der nicht verstaatlichten Banken grundsätzlich unterscheiden. Wenn sich beide Gruppen zu sehr unterscheiden, dann ist es schwer, Schlüsse aus der Entwicklung nach der Verstaatlichung zu ziehen, da nie klar ist, ob ein Resultat der Eigentümerstruktur der Bank oder einer anderen, sich unterscheidenden Eigenschaft zuzuordnen ist. Der Autor argumentiert, dass sich beide Gruppen (bis auf die Bankgröße) ähneln und, dass somit die Auswertung der Studie mit dem natürlichen Experiments nicht zu verzerrten Ergebnissen führen sollte. Die Studie besagt, dass Märkte mit verstaatlichten Banken in den 1980er Jahren ein schnelleres Wachstum an ausgegebenen Krediten vorzuweisen hatten. Dabei war ein großer Teil der Kredite in ländliche Bereiche und an den Agrarsektor gerichtet. Trotz des Wachstums an Krediten sind die Investitionen im Agrarsektor nicht gestiegen. Dagegen stieg der Anteil an unprofitablen Krediten in diesem Bereich stark an. Die Verstaatlichung führte auch zu niedrigeren Zinssätzen. Außerdem ist eine niedrigere Qualität der geleisteten Finanzdienstleistungen ein weiteres Ergebnis der Verstaatlichung.

Kwaja und Mian [2005] untersuchen in ihrem Artikel das Ausmaß und die Kosten von Related Lending, wobei der Kreditnehmer politisch[20] mit der Bank verbunden ist. Der Datensatz besteht aus Krediten von über 90.000 Unternehmen in Paksitan zwischen 1996 und 2002. In der Studie werden kein natürliches Experiment und keine Instrumentvariable verwendet. Die Ergebnisse der Studie sind, dass politisch verbundene Firmen 45% mehr Kreditvolumen bei einer um 50% höheren Ausfallrate haben. Dabei beschränkt sich dieser Zusammenhang nur auf staatseigene Banken. Bei privaten Banken ist dieses Muster nicht zu erkennen. Analog dazu steigt der Effekt (i) mit der Macht des, mit dem Unternehmen verbundenen Politikers an und (ii) ebenso, wenn die Partei des Politikers zu diesem Zeitpunkt regiert. Die Autoren weisen darauf hin, dass sich dieser Zusammenhang maßgeblich auf die Struktur einer Industrie auswirken kann. So kann es für Firmen vorteilhaft sein, Ressourcen auf das Kreieren von politischen Kontakten zu verwenden. Alles in allem schätzen Kwaja und Mian [2005] die jährlichen Kosten dieser Form von Related Lending auf 0,3% bis 1,9% des BIP.

Sapienza [2004] untersucht den Einfluss der Eigentümerschaft einer Bank auf deren Kreditvergabeverhalten. Dazu benützt sie einen Datensatz auf Kreditebene für Banken in Italien[21], welcher die Eigenschaften von Krediten an Unternehmen dokumentiert. Sie analysiert, dass staatliche Banken für Kredite an gleiche oder ähnliche Unternehmen niedrigere Zinssätze verlangen, als private Banken. Zusätzlich findet sie einen Zusammenhang zwischen der Größe der Schuldner, ihrer geografischen Lage und den Zinssätzen von staatlichen Banken: Die Zinssätze sind niedriger (i) je größer der Schuldner und (ii) je ärmer die Region ist, in welcher der Schuldner sich befindet. Außerdem sind die Zinssätze in jenen Regionen niedriger,

[20] Ein kreditnehmendes Unternehmen wird als politisch klassifiziert, wenn der Direktor des Unternehmens in einer Wahl antritt.
[21] Alle Kredite über etwa 41.300 €scheinen im Datensatz auf.

in denen die Partei, welche die Bank kontolliert, besonders stark ist. Ein Grund hierfür könnte taktische Umverteilung der Politiker an ihre Unterstützer darstellen.

Dinc [2005] untersucht die Auswirkung von politischer Einflussnahme auf das Volumen von neuausgegebenen Krediten anhand eines Datensatzes mit internationalen Daten auf der Banken-Ebene. Der Autor benützt eine Art Difference-in-Differences Modell und verwendet Instumentvariablen, um Endogenität in sequentiell exogenen Variablen (in diesem Fall ist das die Größe der Banken und die Eigenkapitalquote) zu verhindern. Er benützt für diese Variablen frühere Werte als Instrument, da diese annahmegemäß nicht mit den Störtermen korrelieren. Er findet, dass in aufstrebenden Märkten das Wachstum des Volumen an neuen Krediten in Wahljahren bei staatseigenen Banken signifikant höher ist, als das von privaten Banken. Anders ausgedrückt, erhöhen staatseigene Banken das Volumen an neuen Krediten im Vergleich zum Vorjahr stärker als private Banken. Umgekehrt ist dieses Muster nicht in den Daten für entwickelte Länder zu beobachten.

4.2 Die zugrunde liegende Theorie

In diesem Artikel wird dem Leser ein kurzer Überblick über die zugrunde liegende Theorie verschafft. Diese kann in die folgenden Bereiche unterteilt werden.

Related Lending. Related Lending tritt dann auf, wenn Banken an diejenigen Kredite gewähren, die die Bank kontrollieren. Zu den Partein, die Kontrolle ausüben zählen Aktieninhaber der Bank, deren Freunde und Familie, sowie Firmen, die unter ihrer Kontrolle stehen. Grundsätzlich gibt es zwei theoretische Erklärungsansätze für das Related Lending: die *Information View* und das *Looting*. Die *Information View* besagt, dass Banken die Risiken eines Kredites an einen, in Beziehung stehenden Kreditnehmer besser einschätzen können, als an eine außenstehende Kreditnehmerin. Banken liegen sozusagen mehr Informationen beim Related Lending vor. Gleichzeitig können gewisse Hold-up Probleme verringert werden, wenn der Kreditnehmer einen Anteil am Kreditgeber hält. Die konträre Theorie dazu heißt *Looting*. Beim Looting wird davon ausgegangen, dass der Kreditnehmer versucht, durch die Vergabe von Krediten an sich selbst, Ressourcen der Bank von anderen Anteilseignern der Bank zu sich zu transferieren. So können z.B. Kredite zu vergünstigten Bedingungen vergeben werden oder Kredite können, ohne Absicht auf Rückzahlung, genehmigt werden, usw.[LaPorta et al., 2003, S.231]

Politisch Motivierte Umverteilung Geht es um die Umverteilung von Ressourcen in einer Gesellschaft, wird zwischen programmatischer- und „Pork Barrel"-Umverteilung unterschieden. Programmatisch ist eine Umverteilung dann, wenn es das Ziel ist, ein von der Gesellschaft angestrebtes Maß an Gleichheit herzustellen. Beispiele hierfür sind die Sozialversicherung, Rentenkassen, öffentliche Bildungssysteme, usw. „Pork Barrel"-Umverteilung liegt vor, wenn durch Politiker bestimmte Projekte aus öffentlicher Hand finanziert werden werden, welche gezielt den Politikern nützen.[Cole, 2007, S.10] Diese Art von Umverteilung folgt zumeist einem oder beiden dieser zwei Motiven: (i) Die Politikerin möchte „Pork Barrel"-Umverteilung nützen, um eine für sie wünschenswerte Allokation von Ressourcen herzustellen, z.B. Unterstützer besonders zu belohnen, und (ii) sie möchte „Pork Barrel"-Umverteilung dahingehend einsetzen, um die Chancen einer Wiederwahl zu erhöhen, dieses Motiv wird *taktische Umverteilung* bezeichnet.[Cole, 2007, S.10ff.] überwiegt das erste Motiv, dann könnten logischerweise „Pork Barrel"-Umverteilung vor allem dorthin ausgerichtet sein, wo der Politiker überdurchschnittlich viele Anhänger hat. Ist das zweite Motiv entscheidend, so sollten vor allem politisch umkämpfte Teile oder die Anhängerschaft von den Umverteilungen profitieren.

Der Einfluss von Wahlzyklen auf die Wirtschaft. Die Frage, wie sich politische Zyklen auf wirtschaftliche Ergebnisse auswirkt, wurde bereits von mehreren Autoren bearbeitet. Nordhaus [1975] hat ein Modell entworfen, in dem Regierungen ihre Chance erhöhen können, wiedergewählt zu werden, indem sie die zeitübergreifende Allokation von Ressourcen zu ihren Gunsten nutzen. Dadurch können „kurzsichtige" Wähler, welche aktuelle wirtschaftliche Ergebnisse als ein Indiz für die Leistung der aktuellen Regierung sehen, getäuscht werden. Da sich Regierungen Budgetbeschränkungen ausgesetzt sehen, beinhaltet dies einen zeitübergreifenden Trade-off bei der Allokation der Ressourcen. Viele Modelle über den Einfluss politischer Zyklen auf eine Volkswirtschaft liefern dieselbe Prognose: die Ergebnisse politischer Maßnahmen sind mit Wahlzyklen korreliert. Modelle wie z.B. das von Nordhaus [1975] prognostizieren konjunkturfördernde, fiskale Maßnahmen vor Wahlen, gefolgt von wirtschaftlichem Abschwung und Steuererhöhungen nach Wahlen. Die Korrelation zwischen Ergebnissen politischer Maßnahmen muss jedoch nicht zwangsläufig durch opportunistische Politiker erklärt werden. Ein anderer Erklärungsansatz lautet, dass Politiker über ihre Amtszeit hinweg lernen, Maßnahmen durchzusetzen und am Ende ihrer Amtszeit am effektivsten Maßnahmen durchsetzen.[Cole, 2007, S.8]

4.3 Eine Studie zum Einfluss von Politikern auf Banken in Indien

Das folgende Unterkapitel fasst die wichtigsten Erkenntnisse einer Studie von Shawn Cole [2007] zusammen und analysiert dabei Coles Einsatz von Instrumentvariablen zur Verhinderung von Endogenität.

4.3.1 Der Einfluss von Wahlzyklen auf Banken.

Cole [2007] untersucht die Auswirkungen der Einflussnahme durch Regierungen auf öffentliche Banken anhand einer Studie über den Kreditmarkt in Indien. Bei dieser Studie konzentriert sich der Autor auf Kredite an Kreditnehmer aus dem Agrarsektor. Agrarkredite sind laut Cole [2007, S.10] ein Mittel der Umverteilung. Sie haben von Gesetzes wegen her deutlich niedrigere Zinssätze als nicht-Agrarkredite bei gleichzeitig sehr hoher Ausfallwahrscheinlichkeit.

Der Fokus auf Agrarkredite macht im Beispiel von Indien Sinn, da im Jahr 1996 ca. 40% aller öffentlichen Kredite an diesen Sektor vergeben wurden, während der Wert dieser Kredite nur 17% des gesamten Kreditportfolios öffentlicher Banken ausmacht. [22] Der Fokus auf Kredite aus öffentlichen Banken ist daher sinnvoll,

[22] Das gibt eine Intuition, warum Agrarkredite ein effektives Mittel sein könnten um Wähler zu gewinnen. Die eben genannten Ziffern implizieren, dass bei Agrarkredite durchschnittlich eine niedrigere Summe je Kredit verliehen wird als bei nicht Agrarkrediten. Geht man davon aus, dass der Kreditnehmer jeweils eine wahlberechtigte Person ist, ist der Kehrwert der verliehenen Durchschnittssumme pro Kredit die Anzahl der wahlberechtigten Personen, welche

da in Indien Politiker über gewisse Gremien indirekt Einfluss auf das Management von Banken haben und da viele Banken in Indien durch Verstaatlichungen in öffentlicher Hand sind.

Aus den Abschnitten über politisch motivierte Umverteilung und den Einfluss von Wahlzyklen auf die Wirtschaft ergibt sich eine Prognose, dass Politiker, sollten sie opportunistisch handeln und versuchen durch taktische Umverteilung ihre Gewinnchancen vor einer Wahl zu maximieren, versuchen Banken, auf die sie Einfluss ausüben können, dazu zu bewegen vor Wahlen überdurchschnittlich viele Kredite zu vergeben, damit Wähler wirtschaftlich zufriedengestellt werden.

Geht man davon aus, dass Politiker nicht opportunistisch handeln, dann besagt die Intuition, dass keine Korrelation zwischen Wahlzyklen und dem Agrarkrediten existieren sollte. Es gibt von Seiten des Agrarsektors intuitiv keinen kausalen Zusammenhang warum z.B. in Wahljahren die Nachfrage nach Agrarkrediten signifikant höher oder niedriger sein sollte, als in nicht-Wahljahren. Sollte dennoch eine positive Korrelation zwischen Agrarkrediten und dem Wahlzyklus erkennbar sein, dann ist es recht wahrscheinlich, dass diese auf Einfussnahme von Politikern auf Banken zurückzuführen ist. Falls die Theorie stimmt, dass Politiker öffentliche Banken dazu benützen, taktische Umverteilung zu praktizieren, dann sollte eine positive Korrelation zwischen Agrarkrediten und Wahljahren, bzw. den Jahren kurz vor einer Wahl, herrschen. Ist die taktische Umverteilung vor allem auf politisch umkämpfte Gebiete gerichtet, dann sollte das Ausmaß an öffentlichen Krediten in diesen Gebieten höher sein als in unumkämpften Gebieten.

Die Daten. Alle Daten aus Coles Studie werden so zusammengefasst, dass sie auf der Ebene von Distrikten[23] verglichen werden können. Insgesamt untersucht Cole 412 Distrikte mit 3,296 Beobachtungen.

Empirische Untersuchung. Cole untersucht die folgenden Themen empirisch:

mit einer Währungseinheit Kredit bedient werden. Ein ansonsten identisches Kreditportfolio, erreicht folglich überdurchschnittlich viele wahlberechtigte Personen wird es im Agrarsektor vergeben.

[23] Indische Distrikte sind laut Autor vergleichbar mit U.S. Amerikanischen Counties.

1. Der Zusammenang zwischen dem Wert von ausgegeben Agrarkrediten und Wahljahren.

2. Wer profitiert besonders von Agrarkrediten?

3. Ist die Umverteilung mittels Agrarkrediten teuer?

4.3.2 Der Zusammenang zwischen Agrarkrediten und Wahljahren.

Cole untersucht den Kausalzusammenhang zwischen dem Volumen an Agrarkrediten und Wahljahren, indem er den Wert von neu ausgegebenen Agrarkrediten in Wahljahren mit demselben Wert in Nicht-Wahljahren vergleicht. Um die Heterogenität zwischen verschiedenen Zeitpunkten und Distrikten zu kontrollieren, berücksichtigt Cole in seinem Modell (i) feste Effekte α_d, die für zeitunabhängige Faktoren kontrollieren, (ii) feste Effekte ϕ_{rt}, welche für einen zeitabhängigen makroökonomischen Trend in Regionen[24] kontrollieren und (iii) den durchschnittlichen Regenfall $Rain_{dst}$ in einem Distrikt in der jeweiligen Beobachtungsperiode, der für Unterschiede in dieser Kategorie zwischen verschiedenen Distrikten und Zeitpunkten kontrolliert. Die Miteinbeziehung dieser Variablen kann Endogenität verhindern, denn falls sie mit den sonstigen abhängigen Variablen korrelieren, würden sie im Störterm eine nichtberücksichtigte Variable darstellen. Folgendes Modell[Cole, 2007, S.15] entsteht:

$$\log y_{dst} = \alpha_d + \phi_{rt} + \delta Rain_{dst} + \beta E_{st} + \epsilon_{dst} \tag{47}$$

$\log y_{dst}$ ist der Logarithmus der neu ausgegebenen Kredite, bzw. Agrarkredite zum Zeitpunkt t in Distrikt d. E_{st} ist ein Dummy, der eins beträgt, wenn im Staat s zum Zeitpunkt t eine Wahl stattfindet.

Einsatz einer Instrumentvariable. Man könnte argumentieren, dass ein Dummy für Wahljahre nicht exogen ist, da regierende Parteien einen Anreiz haben, in wirtschaftlich besonders erfolgreichen Jahren Frühwahlen auszurufen. Wahlen

[24] Es bleibt unklar, was der Autor mit Regionen meint. Meine Vermutung ist, dass mit Regionen Distrikte gemeint sind oder dass Distrikte Mitglieder einer und nur einer Region sind. Jedem Distrikt kann somit eine Region zugeordnet werden.

geschehen aber nicht auf Regionalebene, sondern auf Staatsebene. Wenn mit Region etwas anderes als ein Staat gemeint ist, dann wird für den ökonomischen Trend in Staaten nur unzureichend kontrolliert. Es ist aber dennoch wahrscheinlich, dass der ökonomische Trend im Staat mit dem Wert von neu ausgegebenen Krediten in Wahlkreisen korreliert. Ein Teil des ökonomischen Trends auf Staatsebene wäre somit im Störterm ϵ_{dst} enthalten. Korreliert ein Dummy für Wahljahre mit dem Störterm entsteht Endogenität mit allen bislang angesprochenen Problemen. Daher schätzt Cole (47) mit einem Instrument für Wahljahre S^0_{st}, welches eins ist, wenn zum Zeitpunkt t die letzte Wahl fünf Jahre zurückliegt und null, wenn nicht. Die hochgestellte 0 signalisiert die Restdauer $k, k = -4, -3, -3, -1, 0$ bis zur nächsten geplanten Wahl. Während der Zeitpunkt einer Wahl zum Teil in der Hand der amtierenden Regierung liegt, ist der Abstand zur nächsten geplanten Wahl exogen. Daher ist die erste Bedingung IV1 für gültige Instrumentvariablen erfüllt.

Wie im Unterkapitel über Instrumentvariablen beschrieben wird, wie bei der Methode der zweistufigen kleinsten Quadrate beschrieben, zunächst die potentiell endogene erklärende Variable E_{st} auf alle exogenen erklärende Variablen, inklusive dem Instrument, projiziert. Dazu wird E_{st} wie in (48) in reduzierter Form dargestellt.

$$E_{st} = \alpha^*_d + \phi^*_{rt} + \delta^* Rain_{dst} + \beta^0 S^0_{st} + r_E \tag{48}$$

Der daraus ermittelte systematische Teil der Schwankung in E_{st}, β^0 ist signifikant mit Wert 0,99 und erfüllt damit Bedingung IV2. Somit kann $\hat{E}_{st}$ in (47) eingesetzt werden. Das Instrument ist mit $R^2 = 86\%$ ein starkes Instrument und erklärt den Hauptteil in der Schwankung von E_{st}. Im zweiten Schritt der Methode der zweistufigen, kleinsten Quadrate wird somit folgendes Modell geschätzt:

$$\log y_{dst} = \alpha_d + \phi_{rt} + \delta Rain_{dst} + \beta \hat{E}_{st} + \epsilon_{dst} \tag{49}$$

In reduzierter Form angeschrieben lautet die Regressionsgleichung wie folgt:

$$\begin{aligned}\log y_{dst} &= \alpha_d + \phi_{rt} + \delta Rain_{dst} + \beta(\alpha^*_d + \phi^*_{rt} + \delta^* Rain_{dst} + \beta^0 S^0_{st} + r_E) + \epsilon_{dst} \\ &= (\alpha_d + \beta\alpha^*_d) + (\phi_{rt} + \beta\phi^*_{rt}) + (\delta + \beta\delta^*) Rain_{dst} + (\beta\beta^0) S^0_{st} + (\epsilon_{dst} + \beta r_E) \\ &= \gamma_1 + \gamma_2 + \gamma_3 Rain_{dst} + \gamma_4 S^0_{st} + v \qquad (50)\end{aligned}$$

γ_i sind die geschätzten Parameter für die reduzierte Form in (50). Da Cole eine hochsignifikante Schätzung für β^0 von 0,99[25] erhält, umgeht er den zweiten Schritt der Methode der zweistufigen, kleinsten Quadrate häufig in dieser Arbeit und stellt y_{dst} in der reduzierten Form dar, da sich die Schätzungen beider Methoden kaum unterscheiden.[26] Die reduzierte Form hat den Vorteil, dass per Definition einer linearen Projektion der v mit γ_i für alle i unkorreliert ist. Jedoch muss das geschätzte Modell in (50) wenig mit dem eigentlich interessierenden Modell zu tun haben. Für Coles Datensatz scheint das zwar kein Problem zu sein, da die Schätzungen für β^0 nahe eins und für den Rest der Parameter nahe null sind, aber dies ist nicht notwendigerweise so.

Ergebnisse. Cole [2007] schätzt (49) für y_{dst} = Agrarkredite und y_{dst} = Kredite. Für von öffentlichen Banken vergebene Agrarkredite wird ein β^0 von 0,06 bei einem Standartfehler von 0,02 geschätzt. Diese Schätzung ist bei einem 99% Konfidenzintervall signifikant. Für die Daten aller Banken ist diese Schätzung 0,046 und immer noch bei 1% Signifikant. Für private Banken ist der Schätzer mit -0,021 unsignifikant negativ. Da in (49) die neuausgegebenen Agrarkredite im Logarithmus sind, bedeutet dies eine Steigerung der ausgegeben Agrarkredite von öffentlichen Banken (privaten Banken) um durchschnittlich ca. 6%[27], (4,6%) in Wahljahren, verglichen zu nicht-Wahljahren. Das spricht für einen positiven Kausalzusammenhang zwischen Agrarkrediten aus öffentlichen Banken und Wahljahren. Zudem ist der Punktschätzer für Agrarkredite von privaten Banken zwar unsignifikant aber dennoch negativ, also ein gegenteiliges Ergebnis. Cole [2007, S.16] betont aber, dass sich private Banken in diesem Fall wegen großer struktureller Unterschiede

[25] Standartfehler = 0,01 und $R^2 = 86\%$

[26] $\gamma_4 = \beta \times \beta^0 = \beta \times 0,99$

[27] β^0 ist als Veränderung in Prozent zu interpretieren da $\frac{\partial \log y}{\partial x} = \beta \iff \frac{\partial \log y}{\partial y} = \frac{\beta \partial x}{\partial y} \iff \frac{1}{y}\frac{\partial y}{\partial x} = \beta$

nicht sehr gut als Kontrollgruppe eignen würden. Cole [2007, S.15ff.] erweitert (49) noch um Dummy Variablen für den gesamten Wahlzyklus und entfernt die Instrumentvariable für das Wahljahr. Damit sind die Schätzungen der Parameter für die Dummies als prozentuelle Unterschiede zu den Wahljahren zu interpretieren. Er erhält als Ergebnis, dass Agrarkredite im in allen vier Jahren vor einer Wahl niedriger als am Wahljahr geschätzt werden. Im Jahr zwei vor einer Wahl sind die Agrarkredite sogar um 8,1% niedriger als im Wahljahr. Ein Wert der für ein 99 prozentiges Kofidenzintervall die Hypothese ablehnt, dass der wahre Parameter null ist.

Diese Ergebnisse bestätigen die Theorie, dass Politiker zeitlich vor Wahlen gezielt taktische Umverteilungen einsetzen, um die Chancen ihrer Wiederwahl zu ehöhen

4.3.3 Wer profitiert besonders von Agrarkrediten?

Um die Frage zu klären, ob politisch umkämpfte Distrikte stärker von taktischer Umverteilung profitieren als unumkäpfte Distrikte, definiert Cole [2007, S.18ff.] eine Variable für den Stimmenvorsprung einer Regierung in einem Distrikt, M_{dst}. Diese Variable ist abhängig von t. Kurz vor einer Wahl ist M_{dst} definiert als Proxy[28] für den durchschnittlichen, erwarteten Stimmenvorsprung einer amtierenden Regierung bei der darauffolgenden Wahl in einem Distrikt. In den Jahren nach der Wahl ist M_{dst} der realisierte Stimmenvorsprung in der vorherigen Wahl.

Das Modell. Cole [2007, S.20] erweitert (49) um Dummy Variablen für den gesamten Wahlzyklus und um M^+_{dst}[29], M^-_{dst}[30] sowie Interaktionterme für M^+_{dst} und M^-_{dst} mit den Dummy Variablen für den gesamten Wahlzyklus und entfernt die Instrumentvariable für das Wahljahr.

[28] M_{dst} ist definiert als der durchschnittliche Stimmenvorsprung den die amtierende Regierung in der folgenden Wahl realisiert hat.

[29] Eine nicht-lineare Funktion in M_{dst}, die M_{dst} ergibt, wenn $M_{dst} > 0$ und null wenn $M_{dst} < 0$

[30] Wiederum eine nicht-lineare Funktion in M_{dst}, die M_{dst} ergibt, wenn $M_{dst} < 0$ und null wenn $M_{dst} > 0$

Dadurch entsteht das folgende Modell[Cole, 2007, S.20]:

$$\log y_{dst} = \alpha_d + \phi_{rt} + \delta Rain_{dst} + \sum_{k=-4}^{-1} \beta^k S_{st}^k + \pi^+ M_{dst}^+ + \pi^- M_{dst}^- \\ + \sum_{k=-4}^{-1} \theta_k^+ (M_{dst}^+ \times S_{st}^k) + \sum_{k=-4}^{-1} \theta_k^- (M_{dst}^- \times S_{st}^k) + \epsilon_{dst} \tag{51}$$

Wenn die Theorie zutrifft, dass Politiker besonders die Distrikte durch taktische Umverteilung belohnen, die sie besonders stark unterstützt haben, dann sollte gelten: $\frac{\partial(\log y_{dst})}{\partial M_{dst}^+} > 0$ und $\frac{\partial(\log y_{dst})}{\partial M_{dst}^+} > 0$. Das bedeutet, der Wert von neuvergebenen Agrarkrediten sollte mit zunehmendem Stimmenvorsprung steigen. Gilt jedoch die Theorie, dass die taktische Umverteilung besonders stark auf umkämpfte Distrikte ausgerichtet ist, dann müsste in den Daten folgendes Muster erkennbar sein: $\frac{\partial(\log y_{dst})}{\partial M_{dst}^+} = \pi^+ + \theta_k^+ < 0$ und $\frac{\partial(\log y_{dst})}{\partial M_{dst}^-} \pi^- + \theta_k^- > 0$, also der der Wert der neuvergebenen Agrarkredite sollte im übergang von M_{dst}^+ zu M_{dst}^- ein Maximum haben. Intuitiv bedeutet das, dass Distrikte mit geringem Stimmenvorsprung am stärksten von den Agrarkrediten profitieren sollten.

Empirische Ergebnisse. In Wahljahren sind die Interaktionsterme null, daher gilt $\frac{\partial(\log y_{dst})}{\partial M_{dst}^+} = \pi^+$ und $\frac{\partial(\log y_{dst})}{\partial M_{dst}^-} = \pi^-$. Für die Daten der öffentliche Banken ist $\pi^- = -0,34$ und $\pi^+ = 0,428$. Beide Schätzungen sind bei 1% Typ I Fehler signifikant. Das bedeutet, in Wahljahren folgen die Daten dem Muster, welches auf taktische Umverteilung in politisch besonders umkämpften Distrikten hindeutet.[Cole, 2007, Table 6]

In Nicht-Wahljahren sind die Interaktionsterme ungleich null, daher sind nun $\pi^+ + \theta_k^+$ und $\pi^- + \theta_k^-$ die interessierenden Zusammenhänge. Ein Test der acht Hypothesen (i) $\pi^+ + \theta_k^+ = 0$ oder (ii) $\pi^- + \theta_k^- = 0$ für alle $k \in \{-4, -3, -2, -1\}$ konnte für keine der Hypothesen die Nullhypothese verwerfen. $\frac{\partial(\log y_{dst})}{\partial M_{dst}^+}$ und $\frac{\partial(\log y_{dst})}{\partial M_{dst}^-}$ sind damit statistisch nicht von null zu unterscheiden. Folglich deuten die Daten für öffentliche Banken in Nicht-Wahljahren nicht auf taktische Umverteilung hin, also weder an Unterstützer, noch an umkämpfte Distrikte.

4.3.4 Ist die Umverteilung mittels Agrarkrediten teuer?

Das dritte Thema, das Cole [2007, S.22ff.] untersucht, sind die Kosten der Agrarkredite für die Allgemeinheit. Dazu wird (i) die Ausfallwahrscheinlichkeit dieser Art von Krediten und (ii) die Auswirkungen von Agrarkrediten auf den landwirtschaftlichen Output getestet.

Ausfallwahrscheinlichkeit von Agrarkrediten. Cole [2007, S.23] testet ein Modell, das die Änderung des Volumens an überfälligen Agrarkrediten über den kompletten Wahlzyklus hinweg untersucht. Er stellt fest, dass bei öffentlichen Banken zwei Jahre vor einer Wahl 16,2% weniger überfällige Agrarkredite[31] existieren, als an Wahljahren. Diese Schätzung ist mit 5% Typ I Fehler signifikant. Ansonsten sind die Punktschätzer zwar insignifikant, aber deuten dennoch auf ein niedrigeres Volumen an überfälligen Agrarkrediten an Nicht-Wahljahren hin. Bei privaten Banken ergibt sich ein anderes Bild. Drei von Vier Punktschätzern deuten auf eine höhere Zahl an überfälligen Krediten in Nicht-Wahljahren, verglichen zu Wahljahren hin. Einer davon ist bei 5% signifikant. Als potentielle Gründe für diesen Trend nennt Cole (i) dass Politiker Druck ausüben, um Schulden weniger streng einzutreiben und (ii) dass der durchschnittliche Schuldner in Wahljahren weniger kreditwürdig ist, als in Nicht-Wahljahren. Der zweite Punkt ist darauf zurückzuführen, dass, wie bereits gezeigt, in Wahljahren - aus rein politischen Aspekten - signifikant mehr Kredite gegeben werden und daher Schuldner an Kredite kommen können, die ansonsten als nicht kreditwürdig eingeschätzt werden.

Die Auswirkung von Agrarkrediten auf den Output. Um die Auswirkungen von Agrarkrediten[32] auf den landwirtschaftlichen Output zu messen, bedient sich Cole wiederum einer Instrumentvariable, da nicht berücksichtigte Variablen wie z.B. Produktivität, Preise oder idiosynkratische Erschütterungen zu Endogenität in einer erklärenden Variable „Agrarkredit” führen würden.

[31] Es ist nicht klar definiert, was Cole in Table 7 mit 'Bad Loans" meint. Aus dem Text interpretiere ich, dass dies der Logarithmus des Volumens der überfälligen Kredite ist.

[32] Hierbei handelt es sich Agrarkredite von öffentlichen- und privaten Banken, da Daten auf niedrigerer Ebene nicht vorhanden sind.

Einsatz einer Instrumentvariable. Er benutzt den Wahlzyklus und den Stimmenunterschied in Distrikten, um den Kausalzusammenhang zwischen Agrarkrediten und Output zu messen. Bereits in (51) wurde gezeigt, dass sowohl der Wahlzyklus als auch der Stimmenunterschied mit dem Volumen an Agrarkrediten korrelieren: In Wahljahren z.B. ist das Volumen an Agrarkrediten in politisch umkämpften Distrikten signifikant höher, als in Nicht-Wahljahren in unumkämpften Distrikten. Gleichzeitig gibt es keinen Grund anzunehmen, dass der Wahlzyklus oder der Stimmenunterschied mit den vorher genannten nicht berücksichtigten Variablen korrelieren würden. Daher ist Bedingungen IV1 für gültige Instrumentvariablen erfüllt. Cole [2007, S.25] schätzt folgendes Modell.

$$y_{dt} = \alpha_d + \delta Rain_{dt} + \beta \times credit_{dt} + \phi_t + \epsilon_{dt} \tag{52}$$

Wobei α_i und ϕ_t fixe Effekte sind, die für systematische Unterschiede zwischen Distrikten-, bzw. für einen zeitabhängigen Trend in der gesamten Stichprobe kontrollieren. y_{dt} ist (i) die durchschnittliche Ernte pro Anbaufläche im Distrikt d in Periode t oder (ii) der Logarithmus des Umsatzes in Distrikt d zum Zeitpunkt t. $credit_{dt}$ ist das Volumen ausgegebener Agrarkredite im Distrikt d zur Zeit t. (53) gibt die $credit_{dt}$ in reduzierter Form an und damit den ersten Schritt der Instrumentvariablenschätzung.

$$\begin{aligned} credit_{dt} = \alpha_d^* + \phi_t^* + \delta^* Rain_{dt} + \sum_{k=-4}^{-1} \beta_k^* S_{st}^k \\ + \pi^A M_{dt}^A + \sum_{k=-4}^{-1} (M_{dt}^A \times S_{st}^k) + r_c \end{aligned} \tag{53}$$

Als Instrumente dienen die Variablen für den Wahlzyklus, S_{st}^k, der absolute Wert des Stimmenvorsprungs in Distrikt d, welcher das arithmetische Mittel der Stimmenvorsprünge aller Wahlkreise des Distrikts ist, und ein Interaktionsterm zwischen diesen Variablen.[Cole, 2007, S.25 und Table 8] Die Nullhypothese, dass die Parameter aller Instrumente in (53) $credit_{dt}$ null sind, kann mit 1% Typ I Fehler verworfen werden. Damit ist auch Bedingung IV2 für gültige Instrumentvariablen erfüllt. Die Schätzung für $credit_{dt}$ wird in (52) substituiert und das resultierende

Modell wird geschätzt. Die Schätzungen für β bei beiden abhängigen Variablen, Umsatz und Ernte, sind unsignifikant positiv. Daraus folgt, dass in dieser Studie kein kausaler Zusammenhang zwischen Agrarkrediten und Output hergestellt werden kann.

4.4 Eine Studie zum Einfluss von Politikern auf Banken in Österreich

4.4.1 Theorie des Looting.

Das folgende Unterkapitel behandelt einen Artikel von Halling et al. [2010] , der den Effekt von Related Lending auf die Performance von Sparkassen in Österreich untersucht. Die Gemeinden sind Anteilseigner an den untersuchten Banken und die Bürgermeister der Gemeinden dienen als Agent und sind Entscheidungsträger in den Banken. Wenn Gemeinden von ihren Sparkassen Kredite aufnehmen, in diesem Unterkapitel wird das mit Gemeindekredit bezeichnet, sind sie zugleich Gläubiger und der Schuldner. Als Agent fällt der Bürgermeister der Gemeinde die Entscheidung über die Aufnahme von Krediten und sitzt zugleich im Aufsichtsrat der Sparkasse. Daraus entsteht ein typisches Prinzipal-Agent Problem. Der Prinzipal, die Einwohner der Gemeinde, beauftragt den Agent, im Interesse des Prinzipals zu handeln. Es ist jedoch nicht garantiert, dass der Bürgermeister, der Agent, dies auch wirklich tut und nicht opportunistisch eigene Interessen verfolgt. Die Quelle des Opportunismus, welches Halling et al. [2010] analysieren, ist das Streben des Bürgermeisters wiedergewählt zu werden. Analog zur politisch motivierten Umverteilung im Artikel von Cole [2007] kann der Bürgermeister im Modell von Halling et al. die Sparkassen benützen, um Ressourcen der Bank an die Wähler zu transferieren, z.B. durch vergünstigte Kredite an die Wähler, von der Bank finanzierte Projekte in der Gemeinde, usw. Im folgenden wird dieser Prozess mit Looting bezeichnet. Durch Looting könnte die Wahrscheinlichkeit einer Wiederwahl erhöht werden, wenn Bürger die, von der Bank zusätzlich zur Verfügung gestellten, Ressourcen als ein Zeichen wirtschaftlicher Kompetenz des Bürgermeisters interpretieren würden. Halling et al. [2010, S.8] benützen ein zwei-Perioden-Modell, welches vorsieht, dass eine zufällige finanziellen Erschütterung, die die Gemeinde mit einer Wahrscheinlichkeit von γ[33] zwischen der ersten- und der zweiten Periode heimsucht. Hat der Bürgermeister in der ersten Periode Looting betrieben, sind nicht genug Reserven in der Sparkasse vorhanden um die finanzielle Erschütterung abzufedern und die Wähler leiden unter der Erschütterung. Damit würde der regierende Bürgermeister mit Sicherheit abgewählt werden, das heißt $p = 0$. Umgekehrt, betreibt der Bürgermeister in der ersten Periode kein Looting, dann hat seine Wiederwahl eine Wahrscheinlichkeit $p = p_0$. Hat der Bürgermeister Looting betrieben und die zufällige Erschütterung tritt nicht ein, dann erhöht der Bürgermeister die Wahrscheinlichkeit seiner Wiederwahl auf $p = p_0 + p_0(1 - p_o) > p_0$. Diese Wahrscheinlichkeit wurde gewählt, weil sie bestimmte Eigenschaften in p_0 hat. So steigt $p = p_0 + p_0(1 - p_o) > p_0$ in p_0 da $\frac{\partial(p_0+p_0(1-p_o))}{\partial p_0} = 2(1 - p_0) > 0 \iff 0 < p_0 < 1$. Es wäre recht wiedersinnig wenn die Wahrscheinlichkeit mit Looting wiedergewählt zu werden fallen würde, während die Wahrscheinlichkeit wiedergewählt zu werden steigt. Außerdem ist die Wahrscheinlichkeit auch konkav in p_0 weil $\frac{\partial^2(p_0+p_0(1-p_o))}{\partial {p_0}^2} = -2$, das heißt die Steigung nimmt ab, je größer p_0 wird.Zudem hat der Nutzen des Looting $p_0(1 - p_o)$ eine Extremstelle bei $p_0^* = 0,5$ da $\frac{\partial(p_0(1-p_o))}{\partial p_0} = 1 - 2p_0^* = 0 \iff p_0^* = 0,5$. Das heißt der Nutzen des Looting ist am höchsten, wenn die Wahrscheinlichkeit einer Wiederwahl genau so hoch wie die Wahrscheinlichkeit einer Wahlniederlage ist, oder in anderen Worten, wenn erwartet wird, dass die Wahl besonders knapp ausgeht. Die Kosten des Looting entstehen durch die zufälligen Erschütterung und hängen in diesem Modell einzig von der Wahrscheinlichkeit des Eintretens eines solchen Schocks ab.Wenn ein Bürgermeister wiedergewählt wird, zieht er daraus einen Nutzen B.

Abbildung 2 veranschaulicht die Entscheidung, vor der ein Bürgermeister steht: Im Punkt $\square$ muss entschieden werden, ob Looting angewandt wird oder nicht. Wendet er Looting an, schreitet er in Abbildung 2 nach rechts oben zum Punkt L voran, die Wahrscheinlichkeit im Punkt L wiedergewählt zu werden beträgt $p_o + p_0(1 - p_0) = p_0(2 - p_0)$ und die Wahrscheinlichkeit die Wahl zu verlieren beträgt $1 - p_0(2 - p_0) = p_0^2 - 2p_0 + 1 = (p_0 - 1)^2$. Geht die Wahl verloren, in der Abbildung ist das der Punkt $\neg W$, ist der Nutzen des Bürgermeisters null. Ge-

[33] γ ist eine Wahrscheinlichkeit, das heißt sie kann einen Wert im Intervall zwischen null und eins annehmen, oder genauer $\gamma \in \{\mathbb{R} \mid \gamma \in]0,1[\}$. Dasselbe gilt auch für die Wahrscheinlichkeit p.

winnt er die Wahl hingegen (W), erhält er Nutzen B. In dieser Amtsperiode tritt mit Wahrscheinlichkeit γ die zufällige Erschütterung S ein, bzw. mit Wahrscheinlichkeit $1-\gamma$ nicht ein, $\neg S$. Im Punkt S wird der Bürgermeister mit Sicherheit nicht wiedergewählt und hat somit auch keine Chance auf Nutzen in Periode zwei. Im Punkt $\neg S$ wird der Bürgermeister mit Wahrscheinlichkeit $p_0(2-p_0)$ ein zweites Mal wiedergewählt. In der letzten Periode ist das Looting immer optimal, da $p_0(1-p_o) > 0$ und da kein zufälliger Schock die Chancen in der nächsten Wahl beeinflussen kann. Daher ist in Abbildung 2 die zweite Periode nur sehr eingeschränkt dargestellt. Wird Looting eingesetzt, lautet der erwartete Nutzen des Bürgermeisters wie folgt:

$$\begin{aligned}
\mathbb{E}(\pi \mid L) &= p_0(2-p_0)B + p_0(2-p_0)(1-\gamma)(p_0(2-p_0))B \\
&= B(p_0(2-p_0) + (1-\gamma)(p_0(2-p_0))^2) \\
&= B(p_0(2-p_0) + (1-\gamma)p_0^2(2-p_0)^2) \qquad (54)
\end{aligned}$$

Wenn sich der Bürgermeister gegen Looting entscheidet, ist eine mögliche, finanzielle Erschütterung kein Problem mehr, da γ aus (55) herausfällt:

$$\begin{aligned}
\mathbb{E}(\pi \mid \neg L) &= p_0B + \gamma p_0(p_0(2-p_0)B + (1-\gamma)p_0(p_0(2-p_0)B \\
&= p_0B + p_0^2(2-p_0)B \qquad (55)
\end{aligned}$$

Halling et al. [2010, S.9] definieren den relativen erwarteten Nettonutzen von Looting als $\Delta := B^{-1}(\mathbb{E}(\pi \mid L) - \mathbb{E}(\pi \mid \neg L)$. Aus (54) und (55) folgt:

$$\begin{aligned}
\Delta &= B^{-1}(B(p_0(2-p_0) + (1-\gamma)p_0^2(2-p_0)^2) - (p_0B + p_0^2(2-p_0)B) \\
&= B^{-1}Bp_0(2-p_0 + p_0(4-4p_0+p^2) - 1 - 2b_0 + b_0^2) - \gamma p_0^2(2-p_0)^2 \\
&= p_0(1+p_0-3p_0^2+p_0^3) - \gamma p_0^2(2-p_0)^2 \\
&= p_0(1+2p_0-p_0^2-p_0-2p_0^2+p_0^3) - \gamma p_0^2(2-p_0)^2 \\
&= p_0(1-p_0)(1+2p_0-p_0^2) - \gamma p_0^2(2-p_0)^2 \qquad (56)
\end{aligned}$$

Der linke Term der Differenz in (56) ist der Vorteil des Looting und der rechte sind die Kosten. Ein Analyse der Kosten ergibt, was auch intuitiv in Abbildung 2 zu erkennen ist. Die Kosten sind $\gamma p_0^2(2-p_0)^2 = \gamma(p_0(2-p_0))^2$, $\gamma p_0(2-p_0)$

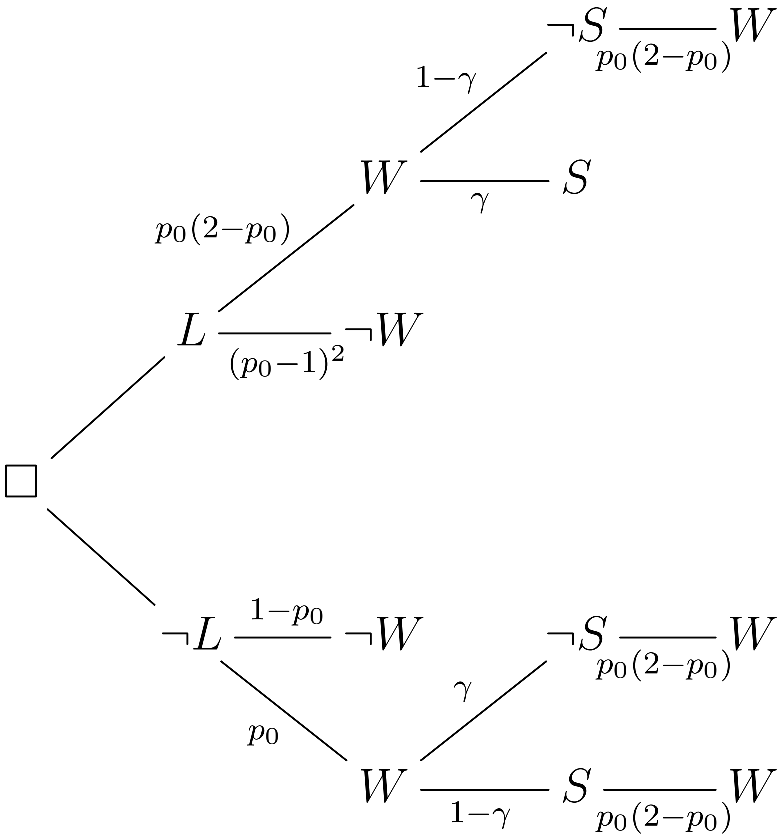

Abbildung 2: Looting Entscheidungsproblem

ist die Wahrscheinlichkeit, dass in Punkt W die zufällige Erschütterung S eintritt, nachdem sich der Bürgermeister für Looting entschieden hat. Im Punkt S ist $p_0(2-p_0)$ die Wahrscheinlichkeit, mit der der Bürgermeister eine zweite Wiederwahl gewinnen würde, wenn er durch die zufällige Erschütterung und Looting nicht verloren hätte. Die Kosten des Looting tauchen somit erst in der zweiten Periode auf, wenn eine Erschütterung die sichere Abwahl bedeutet. Der Vorteil durch Looting entsteht schon zu Beginn in der ersten Periode, indem die Wahrscheinlichkeiten der ersten Wiederwahl erhöht werden. Das Modell von Halling et al. [2010] sieht keine Zeitpräferenzen vor. Es wäre meines Erachtens eine interessante Erweiterung des Modells, wenn Zeitpräferenzen in das Modell miteinbezogen würden, da die Looting-Entscheidung womöglich von Zeitpräferenzen der Agenten beinflusst wird. Der Bürgermeister wird solange Looting betreiben, bis

$p_0(1-p_0)(1+2p_0-p_0^2) = p_0^2(2-p_0)^2$. An diesem Punkt wiegen die Kosten schwerer als der Nutzen. Klar ist, dass der Breakeven von γ abhängt. Die Intuition sagt, dass der Punkt p_0^* an dem $p_0(1-p_0)(1+2p_0-p_0^2) = p_0^2(2-p_0)^2$ gilt, niedriger werden sollte, wenn γ steigt und vice versa.

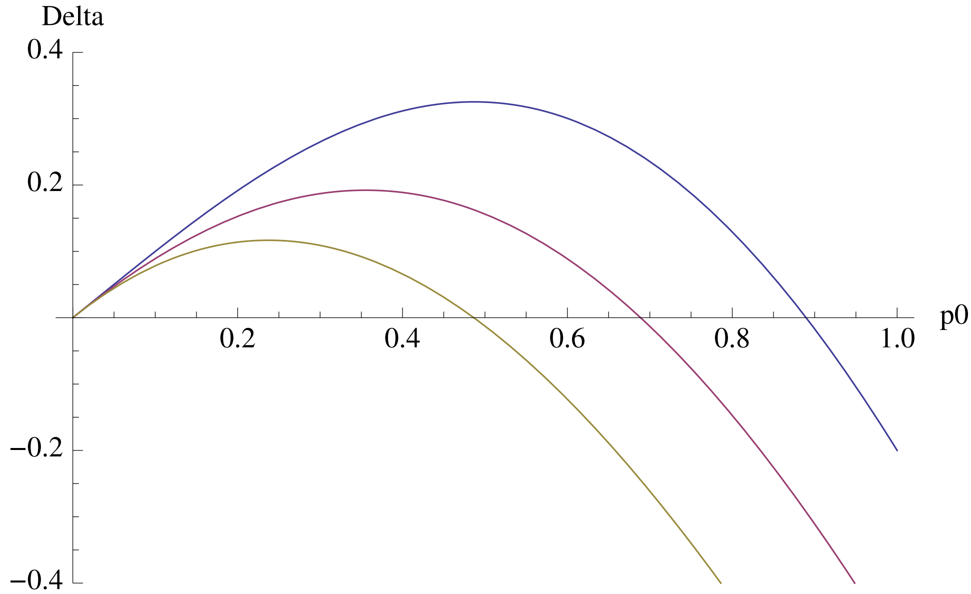

Abbildung 3: Looting: Der erwartete Nettonutzen bei $\gamma = 0.2, 0.5, 0.8$

Lässt man wie in Abbildung 3 den Computer die Funktion aus (56) zeichnen, dann wird die Intuition bestätigt: Für die gelbe Kurve ($\gamma = 0{,}8$) wird Looting am schnellsten nachteilig, während für die blaue Kurve ($\gamma = 0{,}2$) Looting noch bis $p_0 = 0{,}9$ vorteilhaft ist. Abbildung 3 veranschaulicht, dass Looting für Bürgermeister mit niedrigen p_0, also geringen Chancen, die Wahl ohne Looting zu gewinnen, selbst für hohe Werte von γ vorteilhaft sein kann. Für Bürgermeister mit hohen p_0 ist Looting, abhängig von γ, eventuell nicht mehr von Vorteil. Looting ist also tendenziell interessanter für Bürgermeister mit niedrigen p_0, als für jene mit hohen p_0.

Die Daten. Die Datenmenge von Halling et al. [2010] umfasst Daten für 53 Banken über die Jahre 1990 bis 1999. Sie beinhaltet für jede Bank zwischen drei bis fünf Beobachtungen für den Anteil an Gemeindekrediten, Gesamtkapitalrendite, usw. vor dem EU-Beitritt, sowie drei bis fünf nach dem EU-Beitritt. Aus diesen Daten wird jeweils der Median für die Daten vor- und nach dem EU-Beitritt berechnet und in den Analysen verwendet.

4.4.2 Einsatz eines natürlichen Experiments.

Wird der kausale Effekt von Related Lending auf die Performance von Banken untersucht, ist nicht klar in welche Richtung die Kausalität zeigt: Werden durch gute (schlechte) Performance mehr (weniger) Kredite an Gemeinden vergeben, die als Related Lending einzuschätzen sind, oder wirkt die Performance der Gemeindekredite sind in die auf die Performance von Banken. Treffen beide Verbindungen zu, spricht man von simultaner Kausalität. ein Problem welches im Kapitel über Endogenität beschrieben ist. Gleichzeitig können nicht berücksichtigte Variablen, wie Trends oder strukturelle Unterschiede zwischen Banken, leicht die Schätzung eines Kausalzusammenhanges verschmutzen. Um diese Probleme zu vermeiden benützen Halling et al. [2010] ein natürliches Experiment. Als natürliches Experiment dient den Autoren Österreichs EU-Beitritt im Jahr 1995, also zirka in der Mitte des Beobachtungszeitraumes. Durch den EU-Beitritt waren österreichische Gemeinden unter anderem gezwungen, sich EU-Richtlinien bezüglich der öffentlichen Beschaffung von bestimmten Bankdienstleistungen unterzuordnen. Diese Richtlinien sehen bei der Aufnahme von Krediten über 1,5 Millionen € durch Gemeinden vor, dass (i) alle interessierten Finanzdienstleister ein Angebot abgeben dürfen, (ii) der Wettbewerb durch ausreichend viele Anbieter gesichert wird, (iii) der Preis das einzige Kriterium bei der Auswahl des Anbieters sein darf und (iv) dass die Transparenz bei der Auswahl erhöht wird.

Es ist keine besonders starke Annahme, zu behaupten, dass der EU-Beitritt exogen war. Das heißt, die erklärenden Variablen der Bankperformance, z.B. das Volumen der Gemeindekredite, sollten keinen Einfluss auf die Entscheidung, der EU beizutreten, gehabt haben. Außerdem war es für Gemeinden natürlich nicht möglich nicht der EU beizutreten - dies lag außerhalb ihres Entscheidungsspielraumes. Der EU-Beitritt war folglich exogen und betraf alle Gemeinden. Ohne das natürliche Experiment ist unklar, inwiefern die Schwankung in den Gemeindekrediten exogen ist. Der EU-Beitritt hat jedoch durch die Änderungen der Regularien mit großer Wahrscheinlichkeit Auswirkungen auf das Looting durch Gemeindekredite gehabt und damit für exogene Schwankung in dieser Variablen gesorgt. Dabei muss ein hohes Maß an Gemeindekrediten nicht unbedingt ein Indiz für Looting sein. Interessanterweise steigt der durchschnittliche Anteil von Gemeindekrediten an den gesamten Vermögenswerten der Sparkassen um 13,6% an, nachdem Österreich der EU beigetreten ist. Für diese Veränderung sind, den Autoren zufolge, andere Faktoren als Looting verantwortlich.[Halling et al., 2010, S.12ff.] So mussten Gemeinden nach dem EU-Beitritt einen hohen Anteil der Kosten von Österreichs EU-Mitgliedschaft schultern. Außerdem ist es denkbar, dass sich Entwicklungen im tertiären Finanzausgleich[34] und bei gemeindeeigenen Steuern grundsätzlich auf die Nachfrage der Gemeinden nach Krediten ausgewirkt hat und diese erhöht hat.[für Finanz-und Handelspolitik, 2007, S.14] Auch muss eine höhere Gesamtkapitalrendite in der Phase nach dem EU-Beitritt nicht zwingend auf geringeres Looting hindeuten. Andere Faktoren können der Grund hierfür sein. Wichtig ist, dass diese Effekte nicht endogen sind. Die Intuition für die Auswirkung des EU-Beitrittes auf Looting ist folgende: durch höhere Transparenz und gestiegenen Wettbewerb war es nach dem EU-Beitritt schwerer für Gemeinden Looting zu betreiben als vor dem EU-Beitritt.

4.4.3 Der EU-Beitritt und die Auswirkungen auf Gemeindekredite.

Halling et al. [2010] benützen ein Difference-in-Differences Setting um den Kausalzusammenhang zwischen Related Lending und der der Gesamtkapitalrendite zu ermitteln. Anhand zweier Modelle, die von Halling et al. [2010] getestet wurden, kann man recht gut verstehen, wie die Difference-in-Difference Methode dabei hilft, kausale Effekte aus Korrelationen herauszufiltern.

Modell (44) behandelt die Unterschiede zwischen den zwei Zeitpunkten, vor und nach Österreichs EU-Beitritt.

$$RoA_{it} = u_i + a_{LM}LM_{it} + a_E E_t + a_{LM}^E(LM_{it}E_t) + a_x X_{it} + \epsilon_{it} \tag{57}$$

RoA_{it} ist die Gesamtkapitalrendite, LM_{it} ist der Anteil von Gemeindekrediten an den gesamten Aktiva der Bank, X_{it} ist ein Vektor von Kontrollvariablen[35] und

[34]Der tertiäre Finanzausgleich regelt den Ausgleich zwischen Ländern und Gemeinden.

[35]Es wird für den Logarithmus der gesamten Vermögenswerte der Bank, den Anteil an nicht-

ϵ_{it} ist der Störterm für Bank i zum Zeitpunkt t. u_i sind Individueneffekte die für zeitlich unabhängige Faktoren bei Bank i kontrollieren.[36] a_{LM}^E misst den Unterschied im Koeffizienten für LM_{it} vor und nach dem EU-Beitritt. Der Koeffizient ist eine normierte Kovarianz, folglich ist a_{LM}^E die Veränderung der Korrelation zwischen RoA_{it} und LM_{it} nach dem EU-Beitritt.

Die Ergebnisse. a_{LM}^E ist signifikant positiv (5% Typ I Fehler). Das bedeutet, dass der Parameter zwischen RoA_{it} und LM_{it} mit dem EU-Beitritt positv gestiegen ist, bzw. dass Related Loans nach dem EU-Beitritt im Schnitt profitabler wurden als vorher. Dieses Ergebnis kann, muss aber nicht von einer Reduktion im Looting nach dem EU-Beitritt erklärt werden. Gleichzeitig ist bei einer anderen Spezifikation von (57), bei der $LnoM_{it}E_t$ für den Interaktionsterm substituiert wird, der Parameter für $LnoM_{it}E_t$ signifikant negativ[37] (5% Typ I Fehler). $LnoM_{it}E_t$ ist ein Interaktionsterm, welcher den Unterschied in der Profitabilität von Nicht-Gemeindekrediten vor und nach EU-Beitritt ermittelt. Die Profitabilität von Nicht-Gemeindekrediten ist folglich gefallen. Es könnte durch den erhöhten Wettbewerb welcher von den EU-Richtlinien gefordert wird, erklärt werden.[Halling et al., 2010, S.17]

Difference-in-Differences Spezifikation. Um zu testen ob eine Reduktion des Looting der Grund für die gestiegene Profitabilität der Gemeindekredite ist, ändert sich die Spezifikation von (57) dahingehend, dass eine Variable für politischen Wettbewerb eingebaut wird.

Um das Niveau an politischem Wettbewerb zu messen definieren Halling et al. [2010, S.13ff.] drei Dummy Variablen, $Pol_i, i \in \{1, 2, 3\}$. Erstens, $Pol1 = 0$ wenn in allen sechs Wahlen zwischen 1975 und 1994 die gleiche Partei mit einem Vorsprung von mindestens 10% gewonnen hat. $Pol1 = 1$ wenn dem nicht so ist. Zweitens gilt: $Pol2 = 0$ wenn eine Partei im Durchschnitt über die Wahlen hinweg zumindest die Hälfte aller Stimmen erhalten hat. Ist das nicht der Fall gilt $Pol2 = 1$. Drittens,

Gemeindekrediten in der Bank, die Verschuldung der Gemeinde pro Kopf, regionales pro Kopf-BIPund regionales BIP Wachstum kontrolliert.

[36] Somit können diese Effekte Endogenität verhindern, falls ansonsten ein Problem entstehen würde, dass unberücksichtigte aber relevante Variablen mit den erklärenden Variablen korrelieren.

[37] $LnoM_{it}$ steckt im Vektor der Kontrollvariablen X_{it}.

$Pol3 = 0$ wenn die Partei, welche über alle Wahlen hinweg die meisten Wahlen in einer Gemeinde gewonnen hat, einen durchschnittlichen Stimmvorsprung über dem Median von 13,4% besaß. $Pol3 = 1$ wenn das Gegenteil der Fall ist. Für alle drei Variablen ist $Pol_i = 1$ ein Maß welches für eine Gemeinde mit hohem politischen Wettbewerb steht, die Gemeinde ist folglich politisch umkämpft.[Halling et al., 2010, S.14ff.]

Die Methode des natürlichen Experimentes sieht eine Treatment- und eine Kontrollgruppe vor. In Halling et al. [2010] ist die Treatmentgruppe definiert als jene Banken, deren Eigentümer Gemeinden sind, in denen ein überdurschnittliches Maß an politischem Wettbewerb herrscht, das heißt in welchen Wahlen durchschnittlich „knapp"enden. Die Kontrollgruppe umfasst dagegen Banken, deren Eigentümer Gemeinden mit durchschnittlich klaren Wahlergebnissen sind.

In der Diskussion der Entscheidung über Looting, sind die Autoren zu dem Ergebnis gelangt, dass Looting für Bürgermeister mit niedrigeren Chancen auf eine Wiederwahl tendenziell vorteilhafter ist. Daraus folgt, dass Looting in politisch umkämpften Gemeinden häufiger in Daten zu erkennen sein müsste, als in weniger umkämpften, da amtierende Bürgermeister in politisch umkämpften Gemeinden niedrigere Wiederwahl-Wahrscheinlichkeiten haben. Konkret bedeutet das, dass der exogene Effekt durch das natürliche Experiment einen stärkeren Effekt hatte auf Banken aus politisch umkämpften Gemeinden, als auf Banken aus politisch unumkäpften Gemeinen.

Das Modell. Folgendes Modell wird von Halling et al. [2010, S.19] untersucht:

$$\begin{aligned} RoA_{it} &= a_{LM} LM_{it} + a_E E_t + a_{LM}^E (LM_{it} E_t) \\ &+ a_P Pol_i + a_P^E Pol_i E_t + a_{P,LM} Pol_i LM_{it} + a_{P,LM}^E Pol_i LM_{it} E_t \\ &+ a_x X_{it} + u_i^* + \epsilon_{it} \end{aligned} \tag{58}$$

Wobei u_i^* kein fester Effekt mehr ist, sondern eine Zufallsvariable. Als fester Effekt wird u_i anhand der Daten geschätzt. Als Zufallsvariable wird diese Variable, ähnlich wie der Störterm aus einer einheitlichen Verteilung, über alle Banken gezogen. Wie bereits erwähnt, haben feste Effekte u_i den Vorteil, dass zeitunabhängige, unbeobachtete Effekte für jede Gemeinde i in ihnen stecken. Diese dürfen dann auch mit den erklärenden Variablen für die Gemeinde korrelieren. Bei zufälligen Effekten ist das nicht der Fall. Hier wird eine Kovarianz von null zwischen dem zufälligen Effekt einer Gemeinde i, u_i^*, und allen erklärenden Variablen dieser Gemeinde für alle t angenommen - ansonsten entsteht Endogenität. Die Herleitung des Difference-in-Differences Schätzers ist nicht mehr so einfach wie im Kapitel über Endogenität. Angenommen, wir vernachlässigen den Vektor der Kontrollvariablen und gehen von festen Effekten für ür u_i^* aus, dann würde der DD Schätzer wie folgt aussehen:

$$\hat{\beta}_{DD} = R\bar{o}A_P^E - R\bar{o}A_P - (R\bar{o}A^E - R\bar{o}A) = \hat{a}_P^E + \hat{a}_{P,LM}^E \tag{59}$$

$a_{P,LM}^E$ ist folglich ein Difference-in-Differences Effekt. Er misst die Differenz der Auswirkung der EU-Mitgliedschaft auf die Profitabilität von Gemeindekrediten zwischen politisch umkämpften- und politisch unumkämpften Gemeinden. Wie in (59) zu sehen, bestehen beide „Differences” aus den Differenzen der Stichprobenmittelwerte der Treatmentgruppe (i) nach dem EU-Beitritt, $R\bar{o}A_P^E$, und (ii) vor dem EU-Beitritt, $R\bar{o}A_P$, sowie derselben Differenz für die Kontrollgruppe. Die erste Differenz misst die Veränderungen in der Treatmentgruppe, wobei feste Effekte in der Treatmentgruppe durch die Differenz wegfallen, da sie konstant über die Perioden sind. Die zweite Differenz misst Veränderungen in der Kontrollgruppe: Sollte es einen zeitabhängigen Trend geben, wird für diesen, durch das subtrahieren der zweiten Differenz kontrolliert. Wie bereits angeprochen, ist die Treatmentgruppe (Kontrollgruppe) die Menge an Banken, welche sich in politisch umkämpften (unumkäpften) Gemeinden befinden.

Die Ergebnisse. Sollte der Rückgang beim Looting die Erklärung für die gestiegene Profitabilität der Gemeindekredite sein, dann sollten, nach den anfangs getätigten Prognosen, bei Sparkassen in Gemeinden mit hohem politischen Wettbewerb stärkere Anstiege bei der Profitabilität von Gemeindekrediten zu beobachten sein, als bei Banken in Gemeinden mit niedrigem politischen Wettbewerb. Genau diesen Punkt misst $a_{P,LM}^E$. Halling et al. [2010] schätzen (58) mit Generalized Least Squares und bekommen, im Gegensatz zu (57), eine insignifikante Schätzung für a_{LM}^E und im Gegenzug eine signifikant positive Schätzung (5% Type

I Fehler) für $a^E_{P,LM}$ bei allen drei Variablen für politischen Wettbewerb.[Halling et al., 2010, Table 6] Daraus folgt, dass nur die Profitabilität von Gemeindekrediten von Banken aus Gemeinden mit hohem politischen Wettbewerb zunimmt, nicht aber für Banken aus Gemeinden mit niedrigem politischen Wettbewerb. Dazu scheint sich der Grad an politischem Wettbewerb vor allem über den Interaktionterm $Pol_iLM_{it}E_t$ auf die Rentabilität der Sparkassen auszuwirken. Ohne den Interaktionsterm $Pol_iLM_{it}E_t$ ist a^E_P deutlich weniger in der Lage die Variation in der Rentabilität der Sparkassen zu erklären.[Halling et al., 2010, S.20]

Die maßgebliche Bedingung, damit der Schätzer in (59) unverzerrt bleibt, ist, dass neben den Auswirkungen des EU-Beitritts auf Gemeindekredite kein weiterer Faktor besteht, der mit der Treatmentgruppe in der post-EU Beitritt Periode interagiert. Das heißt, β_{DD} sollte in Abwesenheit des Treatments null sein. Es sollte also keinen zeitabhängigen Trend geben, der die Gruppen unproportional beeinflusst. Für den beobachteten Trend, wie z.B. dem Wachstum im pro Kopf-BIP, der beide Gruppen nicht gleichmäßig beeinflusst, kann schon im Kontrollvektor (59) kontrolliert werden. Wichtig ist, dass das Wachstum im pro Kopf-BIP keine endogene Variable ist. Daher ist es kritisch, dass die angesprochenen Änderungen bei der Gemeindesteuer, bzw. dem tertiären Finanzausgleich und der daraus resultierende, signifikante Anstieg am Anteil an Gemeindekrediten exogen erklärt ist. Dieser Effekt darf die Gruppe an Banken aus politisch umkämpften Gemeinden nicht anders beeinflusst haben, als die Gruppe von Banken aus unumkämpften Gemeinden. Es ist zwar unwahrscheinlich, dass dies der Fall ist, da bei der DD-Methode die Differenz der Änderung bei Gemeindesteuern zwischen den beiden Gruppen maßgeblich ist und diese ist insignifikant. [Halling et al., 2010, Table 3] Es könnte aber grundsätzlich schon denkbar sein. Daher führen Halling et al. [2010, S.21ff.] weitere Tests durch, um zu sehen, ob die Schätzungen von (58) robust sind. Sie führen eine ähnliche Regression wie die auf (58) durch, halten aber das Volumen an Gemeindekrediten auf dem Niveau vor dem EU-Beitritt konstant und finden dasselbe Muster wie bei nicht-konstanten Gemeindekrediten. Sie führen außerdem eine Regression durch um zu prüfen, ob die Variable für politischen Wettbewerb einen Effekt auf den Anteil an Gemeindekrediten hat und finden, dass dem nicht so ist. Endogenität durch Änderungen bei der Gemeindesteuer, bzw. dem tertiären Finanzausgleich sind folglich relativ unwahrscheinlich.

4.4.4 Wettbewerb und politischer Wettbewerb.

Halling et al. [2010, S.22ff.] unterstreicht, dass durch die EU-Direktiven zur öffentlichen Beschaffung von Bankdienstleistungen, nicht automatisch Gemeindekredite zu marktgerechten Preisen garantiert werden. Dies ist nur der Fall, wenn Wettbewerb unter den Anbietern von Krediten herrscht. Wettbewerb ist wahrscheinlicher in Märkten, die profitabel sind. Ein Anstieg des Wettbewerbs beeinflusst auch die Profitabilität der Banken und hat eventuell Auswirkungen auf den Schätzer für $a^E_{P,LM}$ in (58). Daher wird in einer weiteren Regression der Wettbewerb einer Region in ein Modell, wie das in (58), integriert. Da der Wettbewerb in einer Region nicht ganz einfach beobachtbar und quantifizierbar ist, wird stattdessen das pro Kopf-BIP als Proxy herangezogen. In ihrem Datensatz beobachten Halling et al. [2010] eine starke Korrelation zwischen dem Trend im pro Kopf-BIP und dem politischen Wettbewerb in einer Gemeinde: Gemeinden mit mit hohem politischen Wettbewerb liegen in Regionen mit überdurchschnittlichem Wachstum des pro-Kopf BIP. Das wirft die Frage auf, ob der Effekt des politischen Wettbewerbs, der in $a^E_{P,LM}$ gemessen wurde, nicht eigentlich auf die Unterschiede im pro-Kopf BIP der Gemeinden zurückzuführen ist. Daher wird ein weiterer Dummy $HiGDPC$ gebildet, der eins (null) ergibt für den Fall, dass sich die Bank in einer Region befindet, deren pro Kopf-BIP sich (nicht) über dem Median der pro Kopf-BIPs der Stichprobe befindet.

Halling et al. [2010] substituieren $HiGDPC$ für Pol_i in Gleichung (58), schätzen das Modell und bekommen einen ähnlichen Wert für den Interaktionsterm $a^E_{GDP,LM}$, der auch bei 5% Typ I Fehler signifikant ist. Im nächsten Schritt ist zu sehen, ob dieser Effekt nur wegen der hohen Korrelation zwischen Pol_i und $HiGDPC$ entsteht, oder ob $HiGDPC$, über Pol_i hinaus, zusätzliche Information besitzt, die Profitabilität von Gemeindekrediten zu erklären. Darum kontrollieren Halling et al. [2010, S.24] für Pol_3, indem sie die Daten anhand der beiden Ausprägungen für Pol_3 in zwei Untermengen einteilen. Eine Regression desselben Modells auf beide Untermengen ergibt, dass $HiGDPC$ keinerlei Erklärungskraft für die Difference-in-Differences Effekte auf die Profitabilität von Gemeindekrediten hat. Wird bei dem Regressionsmodell in (58) für $HiGDPC$ kontrolliert, indem die Stichprobe anhand der Ausprägungen für $HiGDPC$ aufgeteilt wird, ist der resultierende

Schätzer $a^E_{P,LM}$ nur für die Untermenge, bei der $HiGDPC = 1$ gilt, signifikant positiv (1% Typ I Fehler). Das bedeutet, in Gemeinden mit niedrigem pro Kopf-BIP ist der Effekt von politischem Wettbewerb auf die Profitabilität von Gemeindekrediten nicht signifikant - die t-Statistik ist nur 0,43: der Schätzer ist also weit davon entfernt, die Null-Hypothese zu verwerfen.

Das kann bedeuten, dass die EU-Richtlinien einen Rückgang an Looting in jenen Gemeinden zur Folge haben, die politisch umkämpft sind und in denen ein hoher Wettbewerb zwischen Anbietern von Krediten vorherrscht, nicht aber in Gemeinden mit niedrigem Wettbewerb zwischen den Anbietern von Krediten.

5 Schlussfolgerung

In dieser Arbeit wurden Methoden zur Verhinderung von Endogenität anhand von ausgewählten, wissenschaftlichen Artikeln aus dem Bereich *Politik und Banken* untersucht. Um ein tieferes Verständnis für die Problematik der Endogenität zu erhalten, liegt der Schwerpunkt in den Kapiteln zwei und drei auf diesem Thema. Das zweite Kapitel diente dazu, dem Leser einige Eigenschaften von OLS-Schätzern näherzubringen und Grundlagen kurz anzusprechen. Kriterien für unverzerrte und konsistente Schätzer wurden definiert. Das dritte Kapitel erklärte wie Endogenität entsteht, was die Auswirkungen sind und mit welchen Methoden sie verhindert werden kann - drei Methoden wurden dabei näher vorgestellt. In Kapitel vier wurde wissenschaftliche Literatur im Bereich *Politik und Banken*, genauer im Bereich der Kreditvergabe von staatseigenen Banken, (i) inhaltlich analysiert und (ii) auf die Verwendung der, in Kapitel drei vorgestellten Methoden zur Verhinderung von Endogenität hin untersucht. Zur Verwendung der Methoden zur Verhinderung von Endogenität ist zu sagen, dass in vier der acht Studien zumindest einmal ein natürliches Experiment (Halling et al. [2010] und Cole [2009]) oder eine Instrumentvariable (Cole [2007] und Dinc [2005]) eingesetzt wurde. In allen Studien setzen die Autoren feste Effekte[38] ein, um die Endogenität durch nicht berücksichtigte, relevante Variablen zu verhindern und für die Heterogenität zwischen Banken zu kontrollieren. Inhaltlich handeln die beiden näher analysierten

[38]Einzig Halling et al. [2010] benützen auch zufällige Effekte.

Arbeiten (Halling et al. [2010] und Cole [2007]) von dem Zusammenhang zwischen der Wahrscheinlichkeit wiedergewählt zu werden und dem Ausmaß an Looting von staatlichen Banken, bzw. von dem Effekt den der Wahlzyklus auf das Verhalten von staatlichen Banken bei der Vergabe von Agrarkrediten hat. In beiden Artikeln finden die Autoren in den Daten Hinweise auf eine Ausnutzung der staatlichen Banken durch regierende Politiker, die die Wahrscheinlichkeit ihrer Wiederwahl erhöhen wollen. Halling et al. [2010] finden auf Gemeindeebene einen negativen Zusammenhang zwischen der Wahrscheinlichkeit, mit der ein regierender Bürgermeister wiedergewählt wird und dem Ausmaß, in welchem dieser bei einer Bank, auf deren Entscheidungen er Einfluss hat, Looting betreibt. In Coles Artikel [2007] wird ein Zusammenhang zwischen dem Wahlzyklus in indischen Wahlkreisen und der Vergabe von Agrarkrediten von staatlichen Banken hergestellt. In Wahljahren steigt das Volumen an neu-gewährten Agrarkrediten signifikant an. Zudem sind die Agrarkredite vor allem in politisch umkämpfte Wahlkreise gerichtet. Micco und Dinc finden in ihren Studien ähnliche Ergebnisse an Wahljahren. Cole [2009] ist eine ergänzende Studie aus einem ähnlichen Datensatz. Diese Studien haben also wichtige Implikationen auf den Entwurf von Gesetzen, welche es zum Ziel haben, Politiker daran zu hindern, Banken für eigene Motive auszunützen. In Cole [2007] wird die Effizienz dieser politisch motivierten Zunahme in Agrarkrediten untersucht. Trotz der Kredite gibt es keinen signifikanten Anstieg der Investitionen im Agrarsektor. Diese Art von Umverteilung ist folglich kostspielig und sollte mit effektiven Maßnahmen unterbunden werden. Halling et al. [2010] finden Indizien, dass Regeln, welche es Politikern schwerer machen Looting zu betreiben, nur in Gemeinden mit überdurchschnittlichem Wettbewerb unter Finanzdienstleistern einen signifikanten Effekt haben.

Literatur

Joshua Angrist. Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review*, 80: 313–336, 1990.

Orley Ashenfelter. Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 60(1):47–57, 1978.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quaterly Journal of Economics*, 119(1):249–275, February 2004.

Shawn Cole. Fixing market failures or fixing elections? agricultural credit in india. April 2007.

Shawn Cole. Financial development, bank ownership, and growth: Or, does quantity imply quality? *The Review of Economics and Statistics*, 91(1):33–51, February 2009.

Russel Davidson and James G. MacKinnon. *Econometric Theory and Methods*, volume 1. Oxford University Press, 2003.

Serdar Dinc. Politicians and banks: Political influences on government owned banks in emerging countries. *Journal of Financial Economics*, 77:453–479, 2005.

Abteilung für Finanz-und Handelspolitik. Grundsätzliches und WKÖ-Forderungen zum Finanzausgleich. Juni 2007.

William H. Greene. *Econometric Analysis*. Prentice Hall, 5th edition, 2002.

Michael Halling, Pegaret Pichler, and Alex Stomper. The politics of related lending. November 2010.

Bruce E. Hansen. Econometrics. 2011.

James J. Heckman and Jeffrey A. Smith. The pre-programme earnings dip and the determinants of participation in a social programme. implications for simple programme evaluation strategies. *The Economic Journal*, (109):313–348, 1999.

Jack Johnston and John DiNardo. *Econometric Methods.* McGraw-Hill, 4th edition, 1997.

Asim Khwaja and Atif Mian. Do lenders favor politically connected firms? *Quarterly Journal of Economics*, 120:1371–1411, 2005.

Rafael LaPorta, Florencio Lopez-De-Silanes, and Andrei Shleifer. Government ownership of banks. *The Journal of Finance*, LVII(1):265–301, February 2002.

Rafael LaPorta, Florencio Lopez-De-Silanes, and Guillermo Zamarripa. Related lending. *Quarterly Journal of Economics*, (118):231–267, 2003.

Bruce D. Meyer. Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics*, 13(2):151–161, April 1995.

Alejandro Micco, Ugo Panizza, and Monica Yañez. Bank ownership and performance does politics matter? *Central Bank of Chile Working Papers*, (356), Decembre 2005.

Willam Nordhaus. The political business cycle. *Review of Economic Studies*, 42: 169–190, 1975.

Paola Sapienza. The effects of government ownership on bank lending. *Journal of Financial Economics*, 72:357–384, 2004.

Carl P. Simon and Lawrence Blume. *Mathematics for Economists.* W. W. Norton, 1994.

Herbert Stocker. Einführung in die angewandte Ökonometrie. 2011.

Gilbert Strang. *Linear Algebra and its applications*, volume 3. Brooks / Cole Thomson Learning, 1988.

Jeffery M. Wooldridge. *Introductory Econometrics - A Modern Approach.* South-Western College Pub, 2nd edition, 2002a.

Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, 2002b.

Anhang

Zusammenfassung

In dieser Arbeit werden Methoden zur Verhinderung von Endogenität anhand von ausgewählten, wissenschaftlichen Artikeln aus dem Bereich *Politik und Banken* untersucht. Im zweiten Kapitel werden Grundlagen sowie Eigenschaften von OLS-Schätzern definiert. Außerdem werden Kriterien für unverzerrte und konsistente Schätzer definiert. Das dritte Kapitel erklärt wie Endogenität entsteht, was die Auswirkungen sind und mit welchen Methoden sie verhindert werden kann - drei Methoden werden dabei näher vorgestellt. In Kapitel vier wird wissenschaftliche Literatur im Bereich *Politik und Banken* (i) inhaltlich analysiert und (ii) auf die Verwendung der, in Kapitel drei vorgestellten Methoden zur Verhinderung von Endogenität hin, untersucht. Vier von acht Studien benützen ein natürliches Experiment (Halling et al. [2010] und Cole [2009]) oder eine Instrumentvariable (Cole [2007] und Dinc [2005]). In allen Studien setzen die Autoren feste Effekte ein. Inhaltlich werden zwei Arbeiten (Halling et al. [2010] und Cole [2007]) näher analysiert. In beiden Artikeln finden die Autoren in den Daten Hinweise, die darauf hindeuten, dass regierende Politiker staatliche Banken ausnützen, um die Wahrscheinlichkeit ihrer Wiederwahl zu erhöhen.

Abstract

This paper examines methods for the prevention of endogeneity based on selected scientific articles from the field of *politics and finance*. The second chapter defines basic principles regression analysis as well as properties of OLS estimators. In addition, criteria for unbiased and consistent estimators are defined. The third chapter explains how endogeneity arises, what the implications are and by which methods it can be prevented - three methods are presented here in detail. Chapter four analyzes scientific literature from the field of *politics und banking* (i) by content and (ii) by the use of the methods for the prevention of endogeneity presented in chapter three. Four out of eight studies use a natural experiment (Halling et al. [2010] and Cole [2009]) or an instrumental variable (Cole [2007] and Dinc [2005]). In all studies the authors make use of fixed effects. Countentwise, two papers (Halling et al. [2010] and Cole [2007]) are analyzed in detail. In both articles, the authors find evidence in the data, which suggests that politicians in power exploit state banks in order to increase the likelihood of their reelection.

Curriculum Vitae Simon Baumgartner

Kontaktdaten Usterberg 26
6365 Kirchberg in Tirol
Tel: 0664 / 19 281 75
E-Mail: simon.baumgartner@univie.ac.at

Studienadresse Phorusgasse 3/4
1040 Wien
Österreich

Persönliche Daten * 14.02.1986 in München
Staatsangehörigkeit: deutsch

Ausbildung

09/08 – 02/11	Universität Wien, Wien, Österreich Schwerpunkte: - Investmentanalyse - Risikomanagement - Strategiemanagement Angestrebter Abschluss: Magister Voraussichtliche Note: 1,1 (A)
07/10 – 12/10	Institut für Höhere Studien (IHS), Wien, Österreich Economics
09/05 – 06/08	Fachhochschule Kufstein, Kufstein, Österreich Internationale Wirtschaft und Management Schwerpunkte: - Mikroökonomie - Finanzwirtschaft - Rechnungswesen Abschluss: Bachelor of Arts Abschlussnote: „mit gutem Erfolg" 1,9 (B) Top 10% des Jahrgangs
06/07 – 05/08	Korea University Business School, Seoul, Korea Auslandsstudium
09/01 – 06/05	Oscar Paret Gymnasium, Freiberg am Neckar Schwerpunkte: Mathematik, Physik, Deutsch, Englisch, Gemeinschaftskunde Abschluss: Abitur Abschlussnote: 2,4

Praxiserfahrung

05/10 – 07/10	Traitorr srl, Civitanova (Marche), Italien Bereich Controlling - Rechnungsprüfung - Unterstützung bei der Fallbearbeitung - Übersetzung von Verträgen deutsch- und englischsprachiger Kunden

08/09 – 05/10	Universität Wien, Lehrstuhl für Finanzwirtschaft Studienassistent - Unterstützung bei Forschung - Organisation von Kursen - Eigenständiges Halten von Vorlesungen
03/07 – 07/07	Panalpina, Montgomery (Alabama), U.S.A. Bereiche Controlling, Corporate Finance - Durchführung von Projekten zur Effizienzsteigerung und Kostenverringerung - Internes Rechnungswesen - Durchführung einer SWOT-Analyse
07/06 – 09/06	Ferdinand Porsche AG Bereiche Vertrieb - Projektmanagement - Roll-out eines elektronischen Werkstattsystems

---

Fremdsprachen

English:	verhandlungssicher
Italienisch:	Grundkenntnisse
Chinesisch:	Grundkenntnisse
Koreanisch:	Grundkenntnisse

DV-Kenntnisse

Betriebssysteme:	Windows (95/XP/NT/Vista, 7), Mac OS X
Anwendungen:	Sehr gute Kenntnisse: Word, Excel, Powerpoint, Grundkenntnisse: Access, LaTeX, Mathematica

---

Außercurriculare Aktivitäten / Workshops

- Leiter des Sport Arbeitskreises während der Oberstufe
- ehrenamtliche Nachhilfe während des Studiums
- Gründungsmitglied einer Studentenorganisation in Südkorea

Hobbies

Reisen, Wintersport, Basketball, Ökonomie, Politik

Stipendien

06/07-02/08	Stipendium für Austauschstudenten, Korea University
02/08-05/08	Global Merit Scholarship, Korea University
07/10-11/10	Master of Science Economics Scholarship, IHS
08/10	Leistungsstipendium, Universität Wien