IAW DISCUSSION PAPERS

iaw@iaw.edu | www.iaw.edu

IAW Diskussionspapiere Nr. 53 September 2009 | No. 53 September 2009

IV-Schätzung eines linearen Panelmodells mit stochastisch überlagerten Betriebs- und Unternehmensdaten

Elena Biewen
Gerd Ronning
Martin Rosemann

Institut für Angewandte Wirtschaftsforschung e.V.
Ob dem Himmelreich 1 | 72074 Tübingen | Germany
Tel.: +49 7071 98960 | Fax: +49 7071 989699

ISSN: 1617-5654

IAW-Diskussionspapiere

Das Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen ist ein unabhängiges außeruniversitäres Forschungsinstitut, das am 17. Juli 1957 auf Initiative von Professor Dr. Hans Peter gegründet wurde. Es hat die Aufgabe, Forschungsergebnisse aus dem Gebiet der Wirtschafts- und Sozialwissenschaften auf Fragen der Wirtschaft anzuwenden. Die Tätigkeit des Instituts konzentriert sich auf empirische Wirtschaftsforschung und Politikberatung.

Dieses IAW-Diskussionspapier können Sie auch von unserer IAW-Homepage als pdf-Datei herunterladen:

http://www.iaw.edu/Publikationen/IAW-Diskussionspapiere

ISSN 1617-5654

Weitere Publikationen des IAW:

- IAW-News (erscheinen 4x jährlich)
- IAW-Forschungsberichte

Möchten Sie regelmäßig eine unserer Publikationen erhalten, dann wenden Sie sich bitte an uns:

IAW Tübingen, Ob dem Himmelreich 1, 72074 Tübingen,
Telefon 07071 / 98 96-0
Fax 07071 / 98 96-99
E-Mail: iaw@iaw.edu

Aktuelle Informationen finden Sie auch im Internet unter:

http://www.iaw.edu

IV-Schätzung eines linearen Panelmodells mit stochastisch überlagerten Betriebs-und Unternehmensdaten*

Elena Biewen† Gerd Ronning‡ Martin Rosemann§

1. September 2009

Zusammenfassung

Eines der wichtigsten Verfahren zur Anonymisierung von Betriebs- und Unternehmensdaten ist die stochastische Überlagerung. Ihr Einsatz zur Sicherstellung der faktischen Anonymität der Einheiten eines Datensatzes führt jedoch zu inkonsistenten Schätzungen von linearen Panelmodellen und macht die Verwendung von Korrekturverfahren erforderlich. Dieser Beitrag befasst sich mit der Instrumentvariablen-Schätzung (IV-Schätzung) eines linearen Panelmodells mit Individualeffekten und überprüft die Eignung der IV-Methode zur Korrektur der Verzerrung. Als Instrumente werden (a) eine verzögerte Variable, (b) die Differenz von verzögerten Variablen und (c) eine zusätzlich anonymisierte Variable getestet. Wir kommen zum Ergebnis, dass lediglich das letzte Instrument in konsistenten IV-Schätzern resultiert.

Schlagwörter: Instrumentvariablen-Schätzung, additive und multiplikative stochastische Überlagerungen, Anonymisierung von Paneldaten
JEL: C01, C13, C23

*Der Beitrag entstand im Rahmen des vor kurzem abgeschlossenen Projekts "Wirtschaftsstatistische Paneldaten und faktische Anonymisierung", das vom Bundesministerium für Bildung und Forschung gefördert wurde.

†Institut für Angewandte Wirtschaftsforschung, elena.biewen@iaw.edu.
‡Universität Tübingen, gerd.ronning@uni-tuebingen.de.
§Institut für Angewandte Wirtschaftsforschung, martin.rosemann@iaw.edu.

Inhaltsverzeichnis

1 Einführung

Seit einigen Jahren stellen statistische Ämter Betriebs- und Unternehmensdaten für die wissenschaftliche Nutzung – die so genannten Scientific-Use-Files (SUFs) – zur Verfügung. Ein großer Vorteil solcher Datensätze besteht darin, dass sie von Wissenschaftlern am eigenen Arbeitsplatz genutzt werden können. Da SUFs aber faktisch anonymisierte Datensätze sind, können Schätzungen auf ihrer Basis zur verzerrten Ergebnissen führen, sofern so genannte datenverändernde Anonymisierungsmaßnahmen zum Einsatz kommen.

Eines der Verfahren, das für die Anonymisierung wirtschaftsstatistischer Daten vorgeschlagen wird, ist die stochastische Überlagerung (Ronning et al. 2005). Man unterscheidet zwischen zwei Varianten: additive und multiplikative Überlagerungen, wobei letztere als bessere Methode zur Anonymisierung der Betriebs- und Unternehmensdaten angesehen wird, weil Großunternehmen besser geschützt werden können (vgl. Rosemann (2006)).

Das Problem, das von (Mess-)fehlern für die Schätzung pänelökonometrischer Modelle ausgeht, wurde bereits seit langer Zeit erkannt. Griliches und Hausman (1986) weisen darauf hin, dass die Ergebnisse auf Basis der "Within"-Transformationen häufig nicht zufrieden stellend sind und zu kleine oder inkonsistente Schätzer liefern.[1] Des Weiteren merken die Autoren an, dass das Fehler-in-den-Variablen-Problem in Paneldatenmodellen sogar größer wird. Auf das gleiche Problem wird auch im Beitrag von Biørn (2000) hingewiesen. Die "Within"-Transformation kann das Verhältnis der Messfehlervarianz und zu der Varianz der Originalvariable ("noise-signal ratio") vergrößern. Individualeffekte werden zwar eliminiert und das Problem ihrer Korrelation mit anderen Regressoren gelöst, das Fehler-in-den-Variablen-Problem wird jedoch verschärft (Biørn und Krishnakumar 2008).

In der ökonometrischen Literatur werden mehrere Möglichkeiten zur Korrektur der Verzerrung vorgeschlagen. Eine Idee basiert auf dem Gedanken, aus mehreren inkonsistenten Schätzern einen konsistenten Schätzer herzuleiten. Griliches und Hausman (1986) konstruieren einen konsistenten Schätzer aus der Kombination der FE-Schätzung und der Schätzung in ersten Differenzen ("first differences estimator"). Dieser Ansatz wird in Wansbeek und Konig (1991) verallgemeinert.

[1] Vgl. auch Biewen und Ronning (2008) sowie Ronning (2009), die zeigen, dass der "Within"- oder "Fixed-Effects"-Schätzer (FE-Schätzer) auf Basis der sowohl additiv als auch multiplikativ überlagerten Daten inkonsistent ist.

Häufig wird aber der Einsatz der Instrumentvariablen- bzw. der Verallgemeinerten Momentenmethode (VVM) vorgeschlagen (vgl. Griliches und Hausman (1986), Wansbeek und Konig (1991), Biørn (1992, 1996), Biørn und Klette (1998), Wansbeek (2001), Biørn und Krishnakumar (2008)). In der Regel wird eine Mischung von Niveau- und Differenzvariablen verwendet. Dabei sind zwei Varianten denkbar. (1) Das Panelmodell wird zuerst transformiert und anschließend mit der IV- bzw. VMM-Methode geschätzt. Als Instrumente werden dabei die Niveauwerte der fehlerbehafteten Variable aus anderen Perioden verwendet. (2) Das Panelmodell wird nicht transformiert und mit der IV- bzw. VMM-Methode geschätzt. Als Instrumentvariablen werden jedoch Regressoren aus anderen Perioden in differenzierter Form eingesetzt. Die Differenzbildung sorgt in beiden Varianten dafür, dass die individuenspezifische Effekte eliminiert werden. Biørn und Krishnakumar (2008) untersuchen die beiden Varianten und leiten einen konsistenten Schätzer her, wobei sie als Instrumente verzögerte Variablen und zur Eliminierung der Individualeffekte die Differenz zwischen einer Variable zu unterschiedlichen Zeitpunkten verwenden. Nach unserem Kenntnisstand betrachten alle Beiträge nur den additiven Fehler.

In diesem Beitrag wird der Fall der Anonymisierung mittels additiver und multiplikativer stochastischer Überlagerung untersucht, wobei ein lineares Panelmodell mit Individualeffekten mit der IV-FE-Methode geschätzt wird. Als Instrumente werden getestet: (1) eine verzögerte anonymisierte Variable, (2) die Differenz von verzögerten anonymisierten Variablen und (3) eine zusätzlich anonymisierte Variable, die aus einem zweiten anonymisierten Datensatz stammt. Es wird überprüft, ob die IV-Schätzung mit diesen Instrumentvariablen zu guten Korrekturen der Verzerrung führt.

Der Beitrag ist wie folgt gegliedert. Im zweiten Abschnitt wird das zu schätzende lineare Panelmodell präsentiert. Die Varianten der stochastischen Überlagerung und ihre Auswirkungen auf die Schätzung mit fixen Effekten (FE-Schätzung) eines linearen Panelmodells werden kurz erläutert. Abschnitt 3 beschäftigt sich mit der Auswahl geeigneter Instrumentvariablen. Die Abschnitte 4 bis 8 wenden sich der theoretischen Analyse des zu untersuchenden Problems zu. Nach einer allgemeinen Darstellung der Wahrscheinlichkeitsgrenzwerte des IV-Schätzers (Abschnitt 4) wird die IV-Schätzung eines einfachen linearen Panelmodells bei additiver (Abschnitt 5) und multiplikativer (Abschnitt 6) Überlagerungen detailliert untersucht. In Abschnitt 7 werden die Ergebnisse für ein Modell mit mehreren Regressoren verallgemeinert. Abschnitt 8 stellt die asymptothische Varianz des (konsistenten) IV-Schätzers dar. Anschließend werden die hergeleiteten theoretischen Ergebnisse anhand einer Monte-Carlo-Studie (Abschnitt 9) und unter Verwendung der Daten des Monatsberichts (Wellen 1995-2004) getestet (Abschnitt 10). Probleme, die trotz einer konsistenten Schätzung in Praxis auftreten können, werden in Abschnitt 11 belichtet. Eine Zusammenfassung der im Beitrag vorgestellten Ergebnisse (Abschnitt 12) schließt diese Arbeit ab.

2 Das Modell

Betrachten wir ein einfaches lineares Panelmodell mit Individualeffekten:

$$y_{it} = \alpha_i + x_{it}\beta + \epsilon_{it}, \qquad i = 1, \cdots, N,\ t = 1, \cdots, T\,. \tag{2-1}$$

y_{it} ist die abhängige Variable. Mit α_i wird der Individualeffekt bezeichnet, der in dem vorliegenden Modell als ein stochastischer Effekt aufgefasst wird und damit mit dem Regressor x_{it} korreliert sein kann. β ist der Regressorkoeffizient und ϵ_{it} ist der klassische Störterm (vgl. z.B. Greene (2008)).

Der Regressor ist autokorreliert und folgt einem autokorrelierten Prozess erster Ordnung (AR(1)):

$$x_{it} = \rho_0 + \rho x_{i,t-1} + \tau_{it}\,, \tag{2-2}$$

wobei τ_{it} standardnormalverteilt ist und $|\rho| < 1$ gilt.

Im Folgenden werden Modellvariablen mittels additiver und multiplikativer stochastischer Überlagerung anonymisiert. Zwei Varianten der Überlagerung werden dabei jeweils untersucht: die einfache Überlagerung und die Überlagerung nach dem von Höhne (2008) vorgeschlagenen Verfahren, das auch als Überlagerung mit Faktorstruktur bezeichnet wird.

Additive Überlagerung

Bei der *einfachen additiven Überlagerung* werden die Variablen anonymisiert, indem zur Originalvariable eine stochastisch erzeugte Fehlervariable hinzu addiert wird:

$$x^a_{it} = x_{it} + u_{it}, \qquad y^a_{it} = y_{it} + v_{it}, \tag{2-3}$$

wobei der additive Fehler Erwartungswert Null und Varianz σ^2_u bzw. σ^2_v hat. Der Erwartungswert von Null ist eine erwünschte Eigenschaft, damit die Erwartungswerte der anonymisierten Variablen identisch mit denen der Originalvariablen bleiben.

Im Fall der *additiven Überlagerung mit Faktorstruktur* werden die Fehlervariablen wie folgt erzeugt:

$$x_{it}^a = x_{it} + \underbrace{\delta D_i + \epsilon_{itx}}_{u_{it}^H}, \qquad y_{it}^a = y_{it} + \underbrace{\delta D_i + \epsilon_{ity}}_{v_{it}^H} \tag{2-4}$$

mit

$$D_i = \begin{cases} +1 & \text{mit Wahrscheinlichkeit } 0{,}5 \\ -1 & \text{mit Wahrscheinlichkeit } 0{,}5\,. \end{cases} \tag{2-5}$$

δ bleibt für jede Beobachtungseinheit über die Zeit und für alle Merkmale konstant und es gilt $\epsilon_{itx}, \epsilon_{ity} \sim N(0,\, \sigma_\epsilon^2)$. Die Fehlervariablen u_{it}^H und v_{it}^H haben wie bei der einfachen Überlagerung den Erwartungswert Null.

Multiplikative Überlagerung

Bei der *einfachen multiplikativen Überlagerung* wird eine stochastisch erzeugte Fehlervariable mit dem Wert der Originalvariable multipliziert:[2]

$$x_{it}^a = x_{it} * u_{it}^m\,, \quad y_{it}^a = y_{it} * v_{it}^m\,. \tag{2-6}$$

Zur Sicherstellung der Erwartungstreue haben die Fehlervariablen u_{it} und v_{it} den Erwartungswert Eins. Ihre Varianz ist σ_u^2 bzw. σ_v^2. Des Weiteren wird unterstellt, dass Fehler mit Originalvariablen für alle i, t unkorreliert sind.

Die anonymisierten Variablen, die mittels der *multiplikativen Überlagerung mit Faktorstruktur* erzeugt wurden, sind:

$$\begin{aligned} x_{it}^a &= x_{it} * u_{it}^H = x_{it}\left(1 + \delta D_i + \varepsilon_{itx}\right), \\ y_{it}^a &= y_{it} * v_{it}^H = y_{it}\left(1 + \delta D_i + \varepsilon_{ity}\right) \end{aligned} \tag{2-7}$$

D_i wurde in (2-5) bereits definiert.

Zuerst werden somit alle Werte der i-ten Einheit in eine Richtung mit Hilfe des Grundüberlagerungsfaktors $(1+\delta)$ oder $(1-\delta)$ verschoben. Anschließend wird noch

[2]Der Index m über der multiplikativen Fehlervariable wird in weiteren Ausführungen aus Gründen der Einfachheit unterdrückt.

eine zufällig erzeugte Variable addiert, um eine zusätzliche Veränderung einzelner Werte zu erreichen.[3] Da bei der Überlagerung der Daten der i-ten Querschnittseinheit die gleiche Variable D_i verwendet wird, sind die Fehlervariablen miteinander korreliert. Bei dieser Variante der Überlagerung haben die Fehlervariablen u_{it}^H und v_{it}^H den Erwartungswert Eins, die Varianz $(\delta^2 + \sigma_{\varepsilon_x}^2)$ bzw. $(\delta^2 + \sigma_{\varepsilon_y}^2)$ und sind mit Orignalvariablen unkorreliert. Die zusätzliche Fehlervariable $\varepsilon_{itj}(j = x, y)$ ist normalverteilt $\left(N \sim (0, \sigma_{\varepsilon_j^2})\right)$ und mit allen anderen Variablen unkorreliert.

Das Verfahren von Höhne (2008) entspricht der Überlagerung mit einer Mischungsverteilung und ist insofern vorteilhaft, dass bei gleicher Varianz des Fehlers anonymisierte Variablen weiter von den Originalwerten entfernt sind und der Datenschutz damit höher als bei einfacher Überlagerung ist (vgl. Ronning (2009)).

Schätzung mit anonymisierten Daten

Im Fall der Anonymisierung haben Wissenschaftler/innen i.d.R. keine Informationen über Originalvariablen, können aber die anonymisierten Variablen beobachten. In termini der beobachtbaren Variablen erhält man für die einfache additive Überlagerung

$$\begin{aligned} y_{it} &= \alpha_i + (x_{it}^a - u_{it})\beta + \epsilon_{it} \\ &= \alpha_i + x_{it}^a\beta + \eta_{it} \end{aligned} \tag{2-8}$$

mit $\eta_{it} = \epsilon_{it} - u_{it}\beta$

und für die additive Überlagerung mit Faktorstruktur

$$\begin{aligned} y_{it} &= \alpha_i + (x_{it}^a - \delta D_i - \varepsilon_{it})\beta + \epsilon_{it} \\ &= \alpha_i + x_{it}^a\beta + \eta_{it}^H \end{aligned} \tag{2-9}$$

mit $\eta_{it}^H = \epsilon_{it} - (\delta D_i + \varepsilon_{it})\beta$.

Die Schätzung mit der KQ-Methode führt aus zwei Gründen zu inkonsistenten Parameterschätzern (vgl. auch Hsiao (2005)). Zum einen wird eine mögliche Korrelation zwischen dem Regressor und dem Individualeffekt die Ergebnisse verzerren. Dieses Problem wird mittels des Abziehens der zeitlichen Mittelwerte von den Beobachtungen ("Within"-Transformation) behoben, da dann Individualeffekte aus der Schätz-

[3] Dieses Verfahren wird ausführlich in Höhne (2008) und Ronning (2009) beschrieben.

gleichung ausdifferenziert werden können. Zum anderen ist der Störterm η_{it} bzw. η_{it}^H mit dem anonymisierten Regressor korreliert. Die Konsequenz der Korrelation sind verzerrte und nicht konsistente KQ-Schätzer der Modellparameter. Es lässt sich schnell nachvollziehen, dass auch die Ausführung der "Within"-Transformation die Korrelation zwischen dem Regressor und der Störgröße nicht beseitigt. Dann führt auch die FE-Schätzung zu inkonsistenten Parameterschätzern (vgl. auch Sevestre und Trognon (1996), Arellano (2003)).

Biewen und Ronning (2008), Ronning (2009) zeigen, dass auch im Fall der multiplikativen Überlagerung die FE-Schätzung zu inkonsistenten Ergebnissen führt.

In Literatur wird zur Lösung des Fehler-in-den-Variablen-Problems häufig die Instrumentvariablen-Schätzung als ein Korrekturverfahren vorgeschlagen.

3 Mögliche Instrumente

In der Praxis stellt sich oft die Frage, welche Instrumente bei der IV-Schätzung verwendet werden sollen. Ein gutes Instrument muss folgende Annahmen erfüllen (vgl. z.B. Carroll et al. (2006), S. 129):

A(1): Die Korrelation zwischen dem Instrument z und der fehlerfreien Variable x soll möglichst hoch sein.[4] Die Erfüllung dieser Bedingung zeigt, dass das Instrument tatsächlich relevant ist.

A(2): Das Instrument und der Störterm der abhängigen Variablen dürfen nicht korreliert sein: $C(z, \epsilon) = 0$. Dies ermöglicht es, den vom Fehlerterm unabhängigen Einfluss auf die abhängige Variable zu isolieren.

A(3): Die Fehlervariable soll mit dem Instrument unkorreliert sein: $C(z, u) = 0$.

Paneldaten bieten wegen ihrer Mehrdimensionalität eine größere Auswahl an Instrumentvariablen als reine Querschnitts- bzw. Zeitreihendaten. Wenn Überlagerungsfehler miteinander unkorreliert sind und die Originalvariable x eine Autokorrelation aufweist, könnte jede Beobachtung von x^a zu einem anderen Zeitpunkt als ein Instrument verwendet werden (Sevestre und Trognon 1996). Das würde bedeuten, dass

[4] bzw. im Fall eines einfachen linearen Modells muss $C(z, x) \neq 0$ gelten.

für jede Periode $(T-1)$ Instrumente verfügbar sind. In diesem Fall ist jedoch Vorsicht geboten. Im Kontext der Paneldaten sollte zuerst überprüft werden, ob die verzögerte Variable tatsächlich ein gültiges Instrument ist. Die Korrelation der Individualeffekte mit den Regressoren und die zur Eliminierung des Individualeffekts notwendigen Transformation können zur Verletzung der Validität so eines Instruments führen (Griliches und Hausman 1986).

Als ein weiteres "internes" Instrument bietet sich die Variable an, die als Differenz zwischen zwei Zeitpunkten gebildet wurde ($z_{i(t\theta)} = x^a_{it} - x^a_{i\theta}\,,\ \ t \neq \theta$). Auch in diesem Fall sollte zuerst untersucht werden, ob im zu untersuchenden Modell die Annahmen eines gültigen Instruments nicht verletzt werden.

Im Rahmen der Anonymisierung wäre noch eine weitere Alternative denkbar. Die Daten bereitstellende Institution erzeugt mindestens zwei oder sogar mehrere anonymisierte Versionen eines Datensatzes, die den Datennutzern zur Verfügung gestellt werden. Da die überlagerte Variable aus der Originalvariable (und einem stochastisch erzeugten Fehler) besteht, ist eine Korrelation zwischen den gleichen Variablen aller anonymisierten Datenversionen gegeben. Die Höhe der Korrelation lässt sich dabei über die Wahl der Überlagerungsparameter kontrollieren. Auf diese Weise kann die gleiche Variable eines anderen Datensatzes als Instrument verwendet werden. Die Bereitstellung zusätzlicher anonymisierter Datensätze würde jedoch ein höheres Re-Identifikationsrisiko bedeuten. Aus Gründen des Datenschutzes ist es deswegen zu empfehlen, die weiteren Datensätze stärker zu anonymisieren, jedoch so, dass die Datensicherheit gewährleistet ist und die für die IV-Schätzung notwendige Korrelation gegeben ist.

Die IV-Schätzung könnte problematisch werden, wenn die gewählten Instrumente "schwach" sind. Das bedeutet, dass die Instrumentvariable mit dem unbeobachtbaren fehlerfreien Regressor nur gering korreliert ist. Dieses Problem wird beispielsweise von Bound et al. (1995), Staiger und Stock (1997) ausführlich untersucht. Schwache Instrumentalisierung kann starke Inkonsistenz der IV-Schätzer verursachen, die sogar größer ausfällt als im Fall der KQ-Schätzung. "The use of instruments that jointly explain little of the variation in the endogenous variable can do more harm than good" (Bound et al. 1995, S. 449). Bound et al. (1995) zeigen zwei Probleme auf, die durch eine schwache Instrumentalisierung verursacht werden. In kleinen Stichproben entwickelt sich die Verzerrung der IV-Methode in die gleiche Richtung wie die der KQ-Schätzung. Die Verzerrungen der beiden Methoden sind dann gleich, wenn R^2

zwischen dem Regressor und dem Instrument nahe Null ist. Darüber hinaus führt die geringe Korrelation zwischen x und z dazu, dass die Klein-Stichproben-Verzerrung in großen Stichproben nicht mehr eliminiert wird (vgl. Staiger und Stock (1997)).

Im Unterschied zu klassischen Messfehlern werden im Fall der Anonymisierung die Originaldaten bewusst und überlegt verfremdet. Die Fehlervarianzen können so gewählt werden, dass eine hohe Korrellation zwischen dem Instrument und Regressor gegeben ist.

In der folgenden Analyse wird untersucht, wie gut die IV-Schätzung zur Reduktion der durch die stochastische Überlagerung herbeigeführten Verzerrung beiträgt. Dabei wird davon ausgegangen, dass nur eine Instrumentvariable für den anonymisierten Regressor zur Verfügung steht. Drei oben beschriebenen Instrumente werden betrachtet.

4 Wahrscheinlichkeitsgrenzwert des IV-Schätzers

Im Folgenden wird der Wahrscheinlichkeitsgrenzwert des IV-Schätzers für alle Varianten der stochastischen Überlagerung hergeleitet. Dabei wird angenommen, dass alle Unternehmen identisch verteilt sind.

Additive Überlagerung

Wegen

$$\begin{aligned} y_{it}^a - \overline{y}_i^a &= \big(y_{it} + v_{it}\big) - \frac{1}{T}\sum_{t=1}^{T}(y_{it} + v_{it}) \\ &= (y_{it} - \overline{y}_i) + (v_{it} - \overline{v}_i) \\ &= (x_{it} - \overline{x}_i)\beta + (v_{it} - \overline{v}_i) + (\epsilon_{it} - \overline{\epsilon}_i) \end{aligned} \tag{4-10}$$

resultiert der Wahrscheinlichkeitsgrenzwert des IV-Schätzers bei *der einfachen ad-*

ditiven Überlagerung:

$$
\begin{aligned}
plim\,\widehat{\beta^{IV}} &= \frac{plim\,\frac{1}{NT}\sum_i\sum_t(z_{it}-\overline{z}_i)(y^a_{it}-\overline{y}^a_i)}{plim\,\frac{1}{NT}\sum_i\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)} \\
&= \frac{plim\,\frac{1}{NT}\sum_i\sum_t(z_{it}-\overline{z}_i)\Big((x_{it}-\overline{x}_i)\beta+(v_{it}-\overline{v}_i)+(\epsilon_{it}-\overline{\epsilon}_i)\Big)}{plim\,\frac{1}{NT}\sum_i\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)} \\
&= \frac{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x_{it}-\overline{x}_i)}{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)}\beta \\
&+ \frac{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(v_{it}-\overline{v}_i)}{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)} \\
&= \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta+\frac{E[S_{zv}]}{E[S_{zx^a}]} \qquad (4\text{-}11)
\end{aligned}
$$

$$
\text{mit}\quad S_{zx}=\frac{1}{T}\sum_t(z_t-\overline{z})(x_t-\overline{x})\,,\quad S_{zx^a}=\frac{1}{T}\sum_t(z_t-\overline{z})(x^a_t-\overline{x}^a)
$$

$$
\text{und}\quad S_{zv}=\frac{1}{T}\sum_t(z_t-\overline{z})(v_t-\overline{v})\,. \qquad (4\text{-}12)
$$

Der IV-Schätzer ist damit dann konsistent, wenn der Erwartungswert von S_{zx} demjenigen von S_{zx^a} entspricht und gleichzeitig $E[S_{zv}]$ gleich Null ist. Des Weiteren darf der Erwartungswert $E[S_{zx^a}]$ nicht gleich Null sein (vgl. Annahmen eines gültigen Instruments, S. 8).

Für die *additive Überlagerung mit Faktorstruktur* erhält man wegen

$$
\begin{aligned}
y^a_{it}-\overline{y}^a_i &= \big(y_{it}+\delta D_i+\varepsilon_{ity}\big)-\frac{1}{T}\sum_t^T(y_{it}+\delta D_i+\varepsilon_{ity}) \\
&= (y_{it}-\overline{y}_i)+(\varepsilon_{ity}-\overline{\varepsilon}_{iy}) \\
&= (x_{it}-\overline{x}_i)\beta+(\varepsilon_{ity}-\overline{\varepsilon}_{iy})+(\epsilon_{it}-\overline{\epsilon}_i) \qquad (4\text{-}13)
\end{aligned}
$$

den folgenden Grenzwert:

$$
\begin{aligned}
plim\,\widehat{\beta^{IV}} &= \frac{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x_{it}-\overline{x}_i)\beta}{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)} \\
&+ \frac{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(\varepsilon_{ity}-\overline{\varepsilon}_{iy})}{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x^a_{it}-\overline{x}^a_i)} \\
&= \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta+\frac{E[S_{z\varepsilon_y}]}{E[S_{zx^a}]}\,, \qquad (4\text{-}14)
\end{aligned}
$$

wobei ε_{ity} die zusätzliche Veränderung des Fehlers der abhängigen Variable ist. Da die Differenzbildung zur Eliminierung des δ-Faktors geführt hat, ist hier anstelle des Erwartungswertes der empirischen (Gesamt)Kovarianz zwischen dem Instrument und der Fehlervariable lediglich $E[S_{z\varepsilon_y}]$ von Bedeutung.

Multiplikative Überlagerung

Im Fall der *einfachen multiplikativen Überlagerung* erhält man

$$\begin{aligned}
y_{it}^a - \overline{y}_i^a &= y_{it}v_{it} - \overline{yv}_i \\
&= (\alpha_i + x_{it}\beta + \epsilon_{it})\, v_{it} - \frac{1}{T}\sum_{t=1}^{T}(\alpha_i + x_{it}\beta + \epsilon_{it})\, v_{it} \\
&= \alpha_i(v_{it} - \overline{v}_i) + (x_{it}v_{it} - \overline{xv}_i)\beta + (\epsilon_{it}v_{it} - \overline{\epsilon v}_i) \qquad (4\text{-}15) \\
\text{mit} \quad \overline{xv}_i &= \frac{1}{T}\sum_t x_{it}v_{it}\,, \quad \overline{\epsilon v}_i = \frac{1}{T}\sum_t \epsilon_{it}v_{it}\,.
\end{aligned}$$

Dabei fällt auf, dass durch die "Within"-Transformation die Individualeffekte nicht entfernt werden.

Folglich ist der Wahrscheinlichkeitsgrenzwert des IV-Schätzers

$$\begin{aligned}
plim\,\widehat{\beta^{IV}} &= \frac{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)\Big(\alpha_i(v_{it}-\overline{v}_i) + (x_{it}v_{it}-\overline{xv}_i)\beta + (\epsilon_{it}v_{it}-\overline{\epsilon v}_i)\Big)}{plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)(x_{it}^a-\overline{x}_i^a)} \\
&= \frac{E_{\alpha zv}}{E[S_{zx^a}]} + \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta \qquad (4\text{-}16)
\end{aligned}$$

mit

$$E_{\alpha zv} = plim\,\frac{1}{N}\sum_i\frac{1}{T}\sum_t(z_{it}-\overline{z}_i)\alpha_i(v_{it}-\overline{v}_i)\,.$$

Dabei wurde die Unkorreliertheit zwischen u und v sowie zwischen dem Modellstörterm ϵ und z und v unterstellt.

Bei der Überlagerung mit Faktorstruktur gelangt unter Verwendung von

$$
\begin{aligned}
y_{it}^a - \overline{y}_i^a &= y_{it}(1 + \delta D_i + \varepsilon_{ity}) - \frac{1}{T}\sum_{t=1}^{T} y_{it}(1 + \delta D_i + \varepsilon_{ity}) \\
&= (1 + \delta D_i)(y_{it} - \overline{y}_i) + (y_{it}\varepsilon_{ity} - \overline{y\varepsilon_{y_i}}) && \text{(4-17)} \\
&= (1 + \delta D_i)\left((x_{it} - \overline{x}_i)\beta + (\varepsilon_{it} - \overline{\varepsilon}_i)\right) + (y_{it}\varepsilon_{ity} - \overline{y\varepsilon_{y_i}}) && \text{(4-18)} \\
& && \text{(4-19)}
\end{aligned}
$$

$$
\text{mit} \quad \overline{y\varepsilon_{y_i}} = \frac{1}{T}\sum_{t=1}^{T} y_{it}\varepsilon_{ity} \qquad \text{(4-20)}
$$

zum Wahrscheinlichkeitsgrenzwert

$$
plim\,\widehat{\beta^{IV}} = \frac{plim\,\frac{1}{N}\sum_i \frac{1}{T}\sum_t (z_{it} - \overline{z}_i)(1 + \delta D_i)(x_{it} - \overline{x}_i)\beta}{plim\,\frac{1}{N}\sum_i \frac{1}{T}\sum_t (z_{it} - \overline{z}_i)(x_{it}^a - \overline{x}_i^a)}\,. \qquad \text{(4-21)}
$$

In (4-21) wurde ausgenutzt, dass ε_{ity} mit y_{it} und allen anderen Variablen unkorreliert ist.

Die individuenspezifischen Effekte kommen in der *plim*-Formel nicht vor, da sie aufgrund der additiven Komponenten der Fehlervariablen durch die "Within"-Transformation herausdifferenziert werden.

5 Additive Überlagerung

In diesem Abschnitt wird der Effekt der additiven stochastischen Überlagerung auf die Wahrscheinlichkeitsgrenzwerte des IV-Schätzers untersucht. Zuerst wird die einfache Überlagerung und anschließend die Überlagerung mit Faktorstruktur analysiert. Als Instrumente werden die drei bereits oben vorgestellten Alternativen untersucht.

5.1 Einfache Überlagerung

Verzögerte Variable als Instrument
Als Instrumentvariable wird zunächst die verzögerte anonymisierte Variable verwendet:

$$
z_{it} = x_{it-1}^a = x_{it-1} + u_{it-1}\,, \qquad \text{(5-22)}
$$

wobei der Originalregressor und die Fehlervariable für alle i und t unkorreliert sind und der Erwartungswert des Fehlers gleich Null ist.

Überprüfung der Annahmen eines gültigen Instruments

Zuerst wird überprüft, ob die drei Annahmen eines guten Instruments aus Abschnitt 3 erfüllt sind (vgl. A(1)-A(3), Abschnitt 3).

Für die Kovarianz zwischen dem Regressor und dem Instrument gilt

$$C(x_{it}, z_{is}) = E[x_{it}(x_{is-1} + u_{is-1})] - E[x_{it}]E[x_{is-1} + u_{is-1}]$$

$$= \begin{cases} V(x_{it}), & \text{wenn } t = s-1\,, \\ C(x_{it}, x_{is-1}), & \text{wenn } t \neq s-1\,. \end{cases} \tag{5-23}$$

Die Annahme (A1) ist demnach erfüllt. Wenn $t \neq s-1$ ist, führt eine hohe Autokorrelation des Regressors zu einer höheren Korrelation zwischen x und z.

Die Kovarianz zwischen der Fehlervariable und dem Instrument ist

$$C(z_{it}, u_{is}) = E[u_{it}(x_{is-1} + u_{is-1})] - E[u_{it}]E[x_{is-1} + u_{is-1}]$$

$$= \begin{cases} V(u_{it}), & \text{wenn } t = s-1\,, \\ C(u_{it}, u_{is-1}), & \text{wenn } t \neq s-1\,. \end{cases} \tag{5-24}$$

Für $t = s-1$ wird somit die Annahme A(3) verletzt. In diesem Fall entspricht die Kovarianz zwischen z und u der Varianz der Fehlervariable, die bei der Anonymisierung nicht gleich Null ist. Für $t \neq s-1$ ist das Instrument mit der Fehlervariable dann unkorreliert, sofern der Fehler keine Autokorrelation aufweist. Es bleibt damit noch zu prüfen, ob bei der Herleitung der Wahrscheinlichkeitsgrenzwerte die Situation $t = s-1$ eintritt.

Die Annahme der Unkorreliertheit des Instruments mit dem Modellstörterm ist erfüllt, wenn x_{it-1} und u_{it-1} mit ϵ_{it} unkorreliert sind. Dies wird gemäß der Annahmen eines linearen Modells i.d.R. gegeben sein (vgl. auch Ronning (2009)).

Inkonsistenz des IV-Schätzers

In Abschnitt 4 wurde bereits gezeigt, dass für die Untersuchung der Konsistenz des IV-Schätzers die Erwartungswerte der empirischen Kovarianzen S_{zx} und S_{zx^a} von Bedeutung sind (vgl. Gl. (4-11)). Im Folgenden werden diese Erwartungswerte für die verzögerte Variable als Instrument bestimmt. Bei der Herleitung der Formeln werden einzelne Ergebnisse von Hamilton (1994, S. 186f.) und Ronning (2009) ausgenutzt.

Die *Kovarianz zwischen dem Instrument und dem Originalregressors* ist

$$S_{zx} = \frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}^a - \overline{x_1^a})(x_t - \overline{x_2}) \tag{5-25}$$

mit $\overline{x_1^a} = \frac{1}{T-1}\sum_{t=2}^{T} x_{t-1}^a$ und $\overline{x_2} = \frac{1}{T-1}\sum_{t=2}^{T} x_t$.

Für den Erwartungswert lässt sich dann schreiben

$$\begin{aligned}
E[S_{zx}] &= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}\big((x_{t-1}^a-\mu)-(\overline{x_1^a}-\mu)\big)\left((x_t-\mu)-(\overline{x_2}-\mu)\right)\Bigg] \\
&= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}\left((x_{t-1}-\mu)-(\overline{x_1}-\mu)\right)\left((x_t-\mu)-(\overline{x_2}-\mu)\right)\Bigg] \\
&= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\mu)(x_t-\mu)\Bigg] - E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\mu)(\overline{x_2}-\mu)\Bigg] \\
&\quad - E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(\overline{x_1}-\mu)(x_t-\mu)\Bigg] + E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(\overline{x_1}-\mu)(\overline{x_2}-\mu)\Bigg] .
\end{aligned} \tag{5-26}$$

Für den ersten Summanden von (5-26) erhält man

$$E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\mu)(x_t-\mu)\Bigg] = C(x_t, x_{t-1}) = \gamma_1 . \tag{5-27}$$

Für den zweiten Summanden lässt sich herleiten:

$$\begin{aligned}
&E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\mu)(\overline{x_2}-\mu)\Bigg] = E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\mu)\left(\frac{1}{T-1}\sum_{t'=2}^{T}(x_{t'}-\mu)\right)\Bigg] \\
&= \frac{1}{(T-1)^2}E\Bigg[(x_1-\mu)\Big((x_2-\mu)+(x_3-\mu)+\cdots+(x_T-\mu)\Big) \\
&\qquad\qquad + \cdots + \\
&\qquad\qquad (x_{T-1}-\mu)\Big((x_2-\mu)+(x_3-\mu)+\cdots+(x_T-\mu)\Big)\Bigg] \\
&= \frac{1}{(T-1)^2}\Big[\gamma_1+\gamma_2+\cdots+\gamma_{T-1}+ \quad \cdots \quad +\gamma_{T-3}+\gamma_{T-4}+\cdots+\gamma_0+\gamma_1\Big] \\
&= \frac{1}{(T-1)^2}\Big((T-2)\gamma_0+2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big),
\end{aligned} \tag{5-28}$$

wobei $\gamma_j = E\Big[(x_t - \mu)(x_{t-j} - \mu)\Big]$ ist.

Der dritte Summand von (5-26) ist gegeben durch

$$\begin{aligned}
&E\left[\frac{1}{T-1}\sum_{t=2}^{T}(\overline{x_1}-\mu)(x_t-\mu)\right] = E\left[\frac{1}{T-1}\sum_{t=2}^{T}(x_t-\mu)\left(\frac{1}{T-1}\sum_{t'=2}^{T}(\overline{x}_{t'-1}-\mu)\right)\right] \\
&= \frac{1}{(T-1)^2}E\Big[(x_2-\mu)\Big((x_1-\mu)+(x_2-\mu)+\cdots+(x_{T-1}-\mu)\Big) \\
&\qquad\qquad + \cdots + \\
&\qquad\qquad (x_T-\mu)\Big((x_1-\mu)+(x_2-\mu)+\cdots+(x_{T-1}-\mu)\Big)\Big] \\
&= \frac{1}{(T-1)^2}\Big[\gamma_1+\gamma_0+\cdots+\gamma_{T-3}+ \quad \cdots \quad +\gamma_{T-1}+\gamma_{T-2}+\cdots+\gamma_1\Big] \\
&= \frac{1}{(T-1)^2}\Big((T-2)\gamma_0+2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big). \qquad (5\text{-}29)
\end{aligned}$$

Für den vierten Summanden von (5-26) lässt sich Folgendes herleiten:

$$\begin{aligned}
&E\left[\frac{1}{T-1}\sum_{t=2}^{T}(\overline{x_1}-\mu)(\overline{x_2}-\mu)\right] = E\left[\frac{1}{(T-1)^3}\sum_{t=2}^{T}\sum_{t'=2}^{T}\sum_{s=2}^{T}(x_{t'-1}-\mu)(x_s-\mu)\right] \\
&= \frac{1}{(T-1)^3}\sum_{t=2}^{T}E\Big[(x_1-\mu)\Big((x_2-\mu)+(x_3-\mu)+\cdots+(x_T-\mu)\Big) \\
&\qquad\qquad + (x_2-\mu)\Big((x_2-\mu)+(x_3-\mu)+\cdots+(x_T-\mu)\Big) \\
&\qquad\qquad + \cdots + \\
&\qquad\qquad (x_{T-1}-\mu)\Big((x_2-\mu)+(x_3-\mu)+\cdots+(x_T-\mu)\Big)\Big] \\
&= \frac{1}{(T-1)^2}\Big[\gamma_1+\gamma_2+\cdots+\gamma_{T-1}+ \quad \cdots \quad +\gamma_{T-3}+\gamma_{T-4}+\cdots+\gamma_0+\gamma_1\Big] \\
&= \frac{1}{(T-1)^2}\Big((T-2)\gamma_0+2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big). \qquad (5\text{-}30)
\end{aligned}$$

(5-28), (5-29) und (5-30) sind damit identisch.

Für den Erwartungswert der empirischen Kovarianz zwischen dem Instrument und dem Originalregressor gilt bei der einfachen additiven Überlagerung:

$$\begin{aligned}E[S_{zx}] &= \gamma_1 - \frac{1}{(T-1)^2}\Big((T-2)\gamma_0 + 2(T-2)\gamma_1 + 2(T-3)\gamma_2 + \cdots + 2\gamma_{T-2} + \gamma_{T-1}\Big)\\ &= \gamma_1 - \psi(\gamma, T)\,,\end{aligned} \tag{5-31}$$

mit

$$\psi(\gamma, T) = \frac{1}{(T-1)^2}\Big((T-2)\gamma_0 + 2(T-2)\gamma_1 + 2(T-3)\gamma_2 + \cdots + 2\gamma_{T-2} + \gamma_{T-1}\Big)\,. \tag{5-32}$$

Weiter wird für die Bestimmung des Wahrscheinlichkeitsgrenzwertes des IV-Schätzers der *Erwartungswert der Kovarianz zwischen der Instrumentvariable und dem anonymisierten Regressor* benötigt (vgl. (4-11)). Dabei gelangt man wegen der Unabhängigkeit von x_t und u_s für alle t und s zum folgenden Ergebnis:

$$\begin{aligned}&E\Big[S_{zx^a}\Big] = E\Big[\frac{1}{T-1}\sum_{t=2}^{T}\Big((x_{t-1} - \overline{x_1}) + (u_{t-1} - \overline{u_1})\Big)\Big((x_t - \overline{x_2}) + (u_t - \overline{u_2})\Big)\Big]\\ &= E\Big[\frac{1}{T-1}\sum_{t=2}^{T}\Big((x_{t-1} - \overline{x_1})(x_t - \overline{x_2})\Big)\Big] + E\Big[\frac{1}{T-1}\sum_{t=2}^{T}\Big((u_{t-1} - \overline{u_1})(u_t - \overline{u_2})\Big)\Big]\,.\end{aligned} \tag{5-33}$$

Der erste Summand von (5-33) entspricht $E\big[S_{zx}\big]$ (vgl. (5-26)). Der zweite Ausdruck lässt sich analog zu den Gleichungen (5-26) bis (5-31) herleiten. Dabei werden anstelle von $x_t, x_{t-1}, \mu, \gamma_j$ entsprechend $u_t, u_{t-1}, \mu_u, \gamma_{u_j}$ eingesetzt, wobei $\mu_u = 0$ und $\gamma_{u_j} = E\big[(u_t - \mu_u)(u_{t-j} - \mu_u)\big]$ sind. Wenn die Fehlervariable nicht autokorreliert ist, gelangt man zu (vgl. auch Ronning 2009, Abschnitt 11.9.3):

$$E\Big[\frac{1}{T-1}\sum_{t=2}^{T}\Big((u_{t-1} - \overline{u_1})(u_t - \overline{u_2})\Big)\Big] = -\frac{T-2}{(T-1)^2}\sigma_u^2\,. \tag{5-34}$$

Dann folgt

$$E\big[S_{zx^a}\big] = E\big[S_{zx}\big] - \frac{T-2}{(T-1)^2}\sigma_u^2\,. \tag{5-35}$$

Anschließend sieht der Wahrscheinlichkeitsgrenzwert des IV-Schätzers wie folgt aus:

$$plim\,\widehat{\beta^{IV}} = \frac{E\big[S_{zx}\big]}{E\big[S_{zx}\big] - \frac{T-2}{(T-1)^2}\sigma_u^2}\beta + \frac{E[S_{zv}]}{E[S_{zx^a}]} \neq \beta\,. \tag{5-36}$$

Für $T \geq 3$ ist die Instrumentvariablen-Schätzung nicht mehr konsistent, wenn die verzögerte anonymisierte Variable als Instrument verwendet wird (auch wenn $E[S_{zv}] = 0$ ist). Für die Konsistenz muss $\sigma_u^2 = 0$ erfüllt sein, was aber den Fall ohne Anonymisierung bedeutet. In Gl. (5-24) wurde bereits gezeigt, dass die Annahme der Unkorreliertheit des Instruments und des Fehlers verletzt sein kann. Bei der FE-Schätzung kommt diese Situation vor, da hier wegen der Differenzbildung von zeitlichen Mittelwerten $t = s - 1$ eintritt.

Differenz von verzögerten Variablen als Instrument

Als ein alternatives Instrument könnte die Differenz der verzögerten Variablen verwendet werden,[5] z.B.

$$z_{it} = x^a_{it-1} - x^a_{it-2} = x_{it-1} - x_{it-2} + u_{it-1} - u_{it-2}\,. \tag{5-37}$$

Der Erwartungswert der Fehlervariablen bleibt Null, es gilt weiterhin die Annahme der Unabhängigkeit von x_{it} und u_{it} für alle i und t.

Überprüfung der Annahmen eines gültigen Instruments

Weiter wird überprüft, ob das verwendete Instrument valide ist (vgl. Abschnitt 3).

Für die Kovarianz zwischen dem Instrument und dem fehlerfreien Regressor erhält man

$$\begin{aligned} C(x_{it}, z_{is}) &= C\Big(x_{it}, ((x_{is-1} - x_{is-2}) + (u_{is-1} - u_{is-2}))\Big) \\ &= C\left(x_{it}, x_{is-1}\right) - C\left(x_{it}, x_{is-2}\right)\,, \end{aligned} \tag{5-38}$$

$$t, s = 1, \cdots, T\,.$$

Da diese Kovarianz für autokorrelierte Regressoren nicht gleich Null sein wird, ist A(1) erfüllt. Allerdings geht die Kovarianz zwischen x_{it} und x_{is-2} in die Gleichung mit einem Minuszeichen ein, weswegen die Autokorrelation zu einer Absenkung der Korrelation zwischen x und z führen kann.

Unter Ausnutzung der üblichen Annahmen gelangt man zur folgenden Kovarianz

[5] $t > 1$

des Instruments und der Fehlervariable

$$C(z_{it}, u_{is}) = C\Big(((x_{it-1} - x_{it-2}) + (u_{it-1} - u_{it-2})), u_{is}\Big)$$
$$= C\,(u_{it-1}, u_{is}) - C\,(u_{it-2}, u_{is})\ , \tag{5-39}$$
$$t, s = 1, \cdots, T\,.$$

Die Annahme A(3) ist damit nur dann erfüllt, wenn die Fehlervariablen nicht autokorreliert sind und gleichzeitig $s \neq t-1$ wie $s \neq t-2$ erfüllt ist.

Das Instrument ist mit dem Störterm der abhängigen Variable unkorreliert, wenn x_{it-1}, x_{it-2}, u_{it-1} und u_{it-2} mit ϵ_{it} unkorreliert sind, was in einem klassischen linearen Modell der Fall ist.

Inkonsistenz des IV-Schätzers

Die *Kovarianz zwischen dem Instrument und dem Originalregressor* lässt sich wie folgt schreiben

$$S_{zx} = \frac{1}{T-2}\sum_{t=3}^{T}(z_t - \overline{z})(x_t - \overline{x})$$

$$= \frac{1}{T-2}\sum_{t=3}^{T}\Big((x_{t-1} + u_{t-1}) - (\overline{x_1} + \overline{u_1}) - (x_{t-2} + u_{t-2}) + (\overline{x_2} + \overline{u_2})\Big)\,(x_t - \overline{x}) \tag{5-40}$$

$$\text{mit}\quad \overline{x} = \frac{1}{T-2}\sum_{t=3}^{T} x_t\,,\ \ \overline{x_1} = \frac{1}{T-2}\sum_{t=3}^{T} x_{t-1}\,,\ \ \overline{x_2} = \frac{1}{T-2}\sum_{t=3}^{T} x_{t-2} \tag{5-41}$$

und analog für u.

Der Erwartungswert kann wie folgt hergeleitet werden:

$$
\begin{aligned}
E[S_{zx}] = E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}\bigg(&(x_{t-1}-\mu) - (\overline{x_1}-\mu) - (x_{t-2}-\mu) + (\overline{x_2}-\mu) \\
&+ (u_{t-1}-\mu_u) - (\overline{u_1}-\mu_u) - (u_{t-2}-\mu_u) + (\overline{u_2}-\mu_u)\bigg)\bigg((x_t-\mu) - (\overline{x}-\mu)\bigg)\Bigg] \\
= E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}&\bigg((x_{t-1}-\mu) - (\overline{x_1}-\mu) - (x_{t-2}-\mu) + (\overline{x_2}-\mu)\bigg)\times \\
&\times\bigg((x_t-\mu) - (\overline{x}-\mu)\bigg)\Bigg]\,.
\end{aligned} \tag{5-42}
$$

Gleichung (5-42) lässt sich in 8 Terme zerlegen:

$$
\begin{aligned}
E[S_{zx}] = {}& E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\mu)(x_t-\mu)\right] - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\mu)(\overline{x}-\mu)\right] \\
&- E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_1}-\mu)(x_t-\mu)\right] + E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_1}-\mu)(\overline{x}-\mu)\right] \\
&- E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}-\mu)(x_t-\mu)\right] + E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}-\mu)(\overline{x}-\mu)\right] \\
&+ E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_2}-\mu)(x_t-\mu)\right] - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_2}-\mu)(\overline{x}-\mu)\right]
\end{aligned} \tag{5-43}
$$

Im Folgenden sind die einzelnen Summanden von Gleichung (5-43) zu bestimmen.

Für den ersten und fünften Summanden erhält man

$$
E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\mu)(x_t-\mu)\right] = C(x_t, x_{t-1}) = \gamma_1\,, \tag{5-44}
$$

$$
E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}-\mu)(x_t-\mu)\right] = C(x_t, x_{t-2}) = \gamma_2\,. \tag{5-45}
$$

Für den zweiten Term lässt sich schreiben

$$
\begin{aligned}
E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\mu)(\overline{x}-\mu)\Bigg] &= E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\mu)\frac{1}{T-2}\sum_{s=3}^{T}(x_s-\mu)\Bigg]\\
= \frac{1}{(T-2)^2}E\Bigg[&(x_2-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\\
&+(x_3-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\\
&\quad+\cdots+\\
&+(x_{T-1}-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\Bigg]\\
= \frac{1}{(T-2)^2}\Big(&\gamma_1+\gamma_2+\cdots+\gamma_{T-2}+\quad\cdots\quad+\gamma_{T-4}+\gamma_{T-5}+\cdots+\gamma_0+\gamma_1\Big]\\
= \frac{1}{(T-2)^2}\Big[&(T-3)\gamma_0+2(T-3)\gamma_1+2(T-4)\gamma_2+\cdots+2\cdot\gamma_{T-3}+1\cdot\gamma_{T-2}\Big].
\end{aligned}
\tag{5-46}
$$

Für den dritten Summanden gilt

$$
E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_1}-\mu)(x_t-\mu)\Bigg] = E\Bigg[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{s-1}-\mu)(x_t-\mu)\Bigg]. \tag{5-47}
$$

Da der dritte Summand der Gleichung (5-46) entspricht, wird er nicht weiter hergeleitet.

Das Gleiche gilt für den vierten Summanden:

$$
\begin{aligned}
E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_1}-\mu)(\overline{x}-\mu)\Bigg] &= E\Bigg[\frac{1}{(T-2)^3}\sum_{t=3}\sum_{t'=3}\sum_{s=3}(x_{t'-1}-\mu)(x_s-\mu)\Bigg]\\
= \frac{1}{(T-2)^3}\sum_{t=3}E\Bigg[&\sum_{t'=3}\sum_{s=3}(x_{t'-1}-\mu)(x_s-\mu)\Bigg]\\
= \frac{1}{(T-2)}E\Bigg[&(x_2-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\\
&+(x_3-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\\
&\quad+\cdots+\\
&+(x_{T-1}-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\Bigg].
\end{aligned}
\tag{5-48}
$$

Gleichungen (5-46), (5-47) und (5-48) sind damit identisch.

Des Weiteren kann man für den sechsten Summanden schreiben:

$$
\begin{aligned}
E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}-\mu)(\overline{x}-\mu)\right] &= E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{t-2}-\mu)(x_s-\mu)\right] \\
= \frac{1}{(T-2)^2}E\Bigg[(x_1-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big) \\
&+ (x_2-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big) \\
&\quad + \cdots + \\
&+ (x_{T-2}-\mu)\Big((x_3-\mu)+(x_4-\mu)+\cdots+(x_T-\mu)\Big)\Bigg] \\
= \frac{1}{(T-2)^2}\Big(\gamma_2+\gamma_3+\cdots+\gamma_{T-1}+ \quad \cdots \quad +\gamma_{T-5}+\gamma_{T-6}+\cdots+\gamma_0+\gamma_1+\gamma_2\Big) \\
= \frac{1}{(T-2)^2}\Big((T-4)\gamma_0+2(T-4)\gamma_1+2(T-4)\gamma_2+2(T-5)\gamma_3 \\
&+ 2(T-6)\gamma_4+\cdots+2\cdot 2\gamma_{T-4}+3\gamma_{T-3}+2\gamma_{T-2}+\gamma_{T-1}\Big)\,. \qquad (5\text{-}49)
\end{aligned}
$$

Anschließend resultiert für zwei letzten Summanden

$$
E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_2}-\mu)(x_t-\mu)\right] = E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{s-2}-\mu)(x_t-\mu)\right] \qquad (5\text{-}50)
$$

und

$$
\begin{aligned}
E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{x_2}-\mu)(\overline{x}-\mu)\right] &= E\left[\frac{1}{(T-2)^3}\sum_{t=3}^{T}\sum_{t'=3}^{T}\sum_{s=3}^{T}(x_{t'-2}-\mu)(x_s-\mu)\right] \\
&= \frac{1}{(T-2)^2}E\left[\sum_{t'=3}^{T}\sum_{s=3}^{T}(x_{t'-2}-\mu)(x_s-\mu)\right] \qquad (5\text{-}51)
\end{aligned}
$$

das gleiche Ergebnis wie für (5-49).

Die oben hergeleiteten Einzelergebnisse führen zum Erwartungswert

$$
\begin{aligned}
E[S_{zx}] &= \gamma_1 - \gamma_2 - \frac{1}{(T-2)^2}\Big[(T-3)\gamma_0 + 2(T-3)\gamma_1 + \cdots + \gamma_{T-2}\Big] \\
&+ \frac{1}{(T-2)^2}\Big[(T-4)\gamma_0 + 2(T-4)\gamma_1 + 2(T-4)\gamma_2 + \cdots + \gamma_{T-1}\Big] \\
&= -\frac{1}{(T-2)^2}\gamma_0 - \psi(\gamma, T)
\end{aligned}
\tag{5-52}
$$

mit

$$
\begin{aligned}
\psi(\gamma, T) = \frac{1}{(T-2)^2}\Big[&2(T-3)\gamma_1 - (T-2)^2\gamma_1 + \cdots + \gamma_{T-2} \\
&- 2(T-4)\gamma_1 - 2(T-4)\gamma_2 + (T-2)^2\gamma_2 - \cdots - \gamma_{T-1}\Big] .
\end{aligned}
\tag{5-53}
$$

Im nächsten Schritt wird der *Erwartungswert der empirischen Kovarianz der Instrumentvariable und des anonymisierten Regressors* bestimmt:

$$
\begin{aligned}
E[S_{zx^a}] &= E\left[\frac{1}{T-2}\sum_{t=3}^{T}(z_t - \overline{z})(x_t^a - \overline{x}^a)\right] \\
&= E\left[\frac{1}{T-2}\sum_{t=3}^{T}\left(x_{t-1}^a - x_{t-2}^a - \overline{x}_1^a + \overline{x}_2^a\right)(x_t^a - \overline{x}^a)\right] \\
&= E\Bigg[\frac{1}{T-2}\sum_{t=3}^{T}\Big((x_{t-1} - \overline{x_1}) - (x_{t-2} - \overline{x_2}) + (u_{t-1} - \overline{u_1}) - (u_{t-2} - \overline{u_2})\Big) \\
&\times \Big((x_t - \overline{x}) + (u_t - \overline{u})\Big)\Bigg] \\
&= E[S_{zx}] + E[S_{uu}] ,
\end{aligned}
\tag{5-54}
$$

wobei der Erwartungswert von S_{uu} an dieser Stelle wie folgt definiert ist:

$$
E[S_{uu}] = E\left[\frac{1}{T-2}\sum_{t=3}^{T}(u_{t-1} - \overline{u_1})(u_t - \overline{u}) - (u_{t-2} - \overline{u_2})(u_t - \overline{u})\right] . \tag{5-55}
$$

Der IV-Schätzer wird dann konsistent sein, wenn der Ausdruck (5-55) gleich Null ist. Um das zu überprüfen, wird dieser Erwartunsgwert weiter umgeformt:

$$
\begin{aligned}
E[S_{uu}] = & \; E\left[\frac{1}{T-2}\sum_{t=3}^{T}(u_{t-1}-\mu)(u_t-\mu)\right] - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(u_{t-1}-\mu)(\overline{u}-\mu)\right] \\
& - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{u_1}-\mu)(u_t-\mu)\right] + E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{u_1}-\mu)(\overline{u}-\mu)\right] \\
& - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(u_{t-2}-\mu)(u_t-\mu)\right] + E\left[\frac{1}{T-2}\sum_{t=3}^{T}(u_{t-2}-\mu)(\overline{u}-\mu)\right] \\
& + E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{u_2}-\mu)(u_t-\mu)\right] - E\left[\frac{1}{T-2}\sum_{t=3}^{T}(\overline{u_2}-\mu)(\overline{u}-\mu)\right]
\end{aligned}
\tag{5-56}
$$

Gleichung (5-56) wird hier nicht im Einzelnen weiter umgeformt. Die Summanden lassen sich auf die gleiche Weise wie für die Gleichung (5-43) bestimmen.

Wenn die Fehlervariablen nicht autokorreliert sind, gelangt man dann zu

$$E[S_{uu}] = -\frac{1}{(T-2)^2}\sigma_u^2\,, \tag{5-57}$$

wobei σ_u^2 die Varianz der Fehlervariable ist.

Anschließend erhält man für den Wahrscheinlichkeitsgrenzwert des IV-Schätzers

$$plim\,\widehat{\beta^{IV}} = \frac{E\big[S_{zx}\big]}{E\big[S_{zx}\big] - \frac{1}{(T-2)^2}\sigma_u^2}\beta + \frac{E[S_{zv}]}{E[S_{zx^a}]} \neq \beta\,. \tag{5-58}$$

Der IV-Schätzer ist damit nicht konsistent (auch wenn $E[S_{zv}] = 0$ erfüllt ist). Die Inkonsistenz resultiert daraus, dass die Varianz des Fehlers von Null verschieden ist.

Zusätzliche anonymisierte Variable als Instrument

Im Folgenden wird die Variante der IV-Schätzung mit einem Instrument analysiert, das aus einem weiteren anonymisierten Datensatz erzeugt wird. Die Instrumentvariable ist dabei wie folgt definiert:

$$z_{it} = x_{it} + \nu_{it}\,. \tag{5-59}$$

Die Fehlervariable ν_t wird unabhängig vom Regressor x und von der Fehlervariable des ursprünglichen anonymisierten Datensatzes (u) generiert, hat den Erwartungswert Null und die Varianz σ_ν^2.

Überprüfung der Annahmen eines gültigen Instruments

Das vorgeschlage Instrument erfüllt die Annahmen eines validen Instruments. Es ist mit dem fehlerfreien Regressor korreliert, da z den Originalregressor enthält:

$$C(z_{it}, x_{is}) = C(x_{it} + \nu_{it}, x_{is}) = C(x_{it}, x_{is})\,, \tag{5-60}$$

$$t, s = 1, \cdots, T\,.$$

Am Beispiel eines AR(1)-Prozesses und für $t \neq s$ wird deutlich, dass mit steigender (positiver) Autokorrelation des Regressors die Korrelation zwischen z und x größer wird:

$$Corr(x_{it}, x_{is}) = \frac{\gamma_0 \rho^{|t-s|}}{\gamma_0} = \rho^{|t-s|}\,. \tag{5-61}$$

γ_0 steht dabei für die Varianz des nicht anonymisierten Regressors.

Da der zweite anonymisierte Datensatz unabhängig von dem ersten überlagert wird, ist das Instrument mit der Fehlervariable unkorreliert:

$$C(z_{it}, u_{is}) = C(x_{it} + \nu_{it}, u_{is}) = C(x_{it}, u_{is}) + C(\nu_{it}, u_{is}) = 0\,, \tag{5-62}$$

$$t, s = 1, \cdots, T\,.$$

Die Instrumentvariable ist mit dem Modellstörterm ϵ unkorreliert, wenn sowohl der Originalregressor als auch der Fehler mit der Störgröße keine Korrelation aufweisen. Diese Bedingung ist in einem klassischen linearen Modell normalerweise erfüllt.

Konsistenz des IV-Schätzers

Der Erwartungswert der *Kovarianz zwischen dem Instrument und dem Originalre-*

gressor lässt sich wie folgt umformen:

$$
\begin{aligned}
E[S_{zx}] &= E\left[\frac{1}{T}\sum_t (z_t - \overline{z})(x_t - \overline{x})\right] \\
&= E\left[\frac{1}{T}\sum_t \Big((z_t - \mu_z) - (\overline{z} - \mu_z)\Big)\Big((x_t - \mu_x) - (\overline{x} - \mu_x)\Big)\right] \\
&= E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(x_t - \mu_x)\right] - E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(\overline{x} - \mu_x)\right] \\
&+ E\left[\frac{1}{T}\sum_t (\overline{z} - \mu_z)(\overline{x} - \mu_x)\right] - E\left[\frac{1}{T}\sum_t (\overline{z} - \mu_z)(x_t - \mu_x)\right] .
\end{aligned} \tag{5-63}
$$

Im Folgenden werden die vier Summanden in (5-63) genauer untersucht.

Für den ersten Summanden aus (5-63) erhält man:

$$
E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(x_t - \mu_x)\right] = C(z_t, x_t)\,. \tag{5-64}
$$

Setzen wir die Instrumentenvariable (5-59) in (5-64) ein:

$$
C(z_t, x_t) = C(x_t + \nu_t, x_t) = V(x_t) = \gamma_0\,. \tag{5-65}
$$

Der zweite Summand aus Gleichung (5-63) lässt sich umschreiben als:

$$
\begin{aligned}
E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(\overline{x} - \mu_x)\right] &= E\left[\frac{1}{T^2}\sum_t \sum_{t'} (z_t - \mu_z)(x_{t'} - \mu_x)\right] \\
&= \frac{1}{T^2}\sum_t \sum_{t'} C(z_t, x_{t'})\,.
\end{aligned} \tag{5-66}
$$

Für diese Kovarianz im Fall des Instruments (5-59) gilt:

$$
C(z_t, x_{t'}) = C(x_t + \nu_t, x_{t'}) = C(x_t, x_{t'})\,. \tag{5-67}
$$

Einsetzen von (5-67) in (5-66) führt zu

$$
\frac{1}{T^2}\sum_t \sum_{t'} C(z_t, x_{t'}) = \frac{1}{T^2}\sum_t \sum_{t'} C(x_t, x_{t'})
$$

$$
\begin{aligned}
&= \frac{1}{T^2} E[(x_1-\mu_x)^2 + (x_1-\mu_x)(x_2-\mu_x) + \cdots + (x_1-\mu_x)(x_T-\mu_x) \\
&\qquad + (x_2-\mu_x)(x_1-\mu_x) + (x_2-\mu_x)^2 + \cdots + (x_2-\mu_x)(x_T-\mu_x) \\
&\qquad + \cdots + \\
&\qquad + (x_T-\mu_x)(x_1-\mu_x) + (x_T-\mu_x)(x_2-\mu_x) + \cdots + (x_T-\mu_x)^2] \\
&= \frac{1}{T^2}(\gamma_0 + \gamma_1 + \cdots + \gamma_{T-1} + \gamma_1 + \gamma_0 + \cdots + \gamma_{T-2} \\
&\qquad + \gamma_2 + \gamma_1 + \gamma_0 + \cdots + \gamma_{T-3} + \quad \cdots \quad + \gamma_{T-1} + \gamma_{T-2} + \cdots + \gamma_0) \\
&= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + 2(T-3)\gamma_3 + 2(T-4)\gamma_4\right. \\
&\qquad \left. + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) \, .
\end{aligned}
\tag{5-68}
$$

Für den dritten Summanden aus (5-63) erhält man:

$$
\begin{aligned}
E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(\overline{x}-\mu_x)\right] &= E\left[\frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''}(z_{t'}-\mu_z)(x_{t''}-\mu_x)\right] \\
&= \frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(z_{t'}, x_{t''}) \, .
\end{aligned}
\tag{5-69}
$$

Für die Kovarianz lässt sich herleiten:

$$
C(z_{t'}, x_{t''}) = C(x_{t'} + \nu_{t'}, x_{t''}) = C(x_{t'}, x_{t''}) \, . \tag{5-70}
$$

Damit erhält man

$$
\begin{aligned}
&\frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(z_{t'}, x_{t'}) = \frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(x_{t'}, x_{t''}) = \frac{1}{T^2}\sum_{t'}\sum_{t''} C(x_{t'}, x_{t''}) \\
&= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) \, .
\end{aligned}
\tag{5-71}
$$

Umformung des vierten Summanden liefert:

$$
\begin{aligned}
E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(x_t-\mu_x)\right] &= E\left[\frac{1}{T^2}\sum_t\sum_{t'}(z_{t'}-\mu_z)(x_t-\mu_x)\right] \\
&= \frac{1}{T^2}\sum_t\sum_{t'} C(z_{t'}, x_t) \, .
\end{aligned}
\tag{5-72}
$$

Weiter gelangt man zu:

$$
C(z_{t'}, x_t) = C(x_{t'} + \nu_{t'}, x_t) = C(x_{t'}, x_t) \tag{5-73}
$$

und

$$\frac{1}{T^2}\sum_t\sum_{t'} C(z_{t'}, x_t) = \frac{1}{T^2}\sum_t\sum_{t'} C(x_{t'}, x_t)$$

$$= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad (5\text{-}74)$$

Wegen Identität von (5-68), (5-71) und (5-74) lässt sich für den Erwartungswert der Kovarianz zwischen z und x wie folgt schreiben schreiben:

$$E(S_{zx}) = \gamma_0 - \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right)$$

$$= \gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad (5\text{-}75)$$

Im Folgenden wird der *Erwartungswert der Kovarianz zwischen dem Instrument und dem anonymisierten Regressor* hergeleitet. Es werden dabei die oben dargestellten Ergebnisse ausgenutzt.

$$\begin{aligned} E[S_{zx^a}] &= E\left[\frac{1}{T}\sum_t (z_t-\overline{z})(x_t^a-\overline{x^a})\right] \\ &= E\left[\frac{1}{T}\sum_t (z_t-\mu_z)(x_t^a-\mu_x)\right] - E\left[\frac{1}{T}\sum_t (z_t-\mu_z)(\overline{x^a}-\mu_x)\right] \\ &+ E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(\overline{x^a}-\mu_x)\right] - E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(x_t^a-\mu_x)\right] . \end{aligned} \qquad (5\text{-}76)$$

Für den ersten Summanden erhält man:

$$E\left[\frac{1}{T}\sum_t (z_t-\mu_z)(x_t^a-\mu_x)\right] = \frac{1}{T}\sum_t C(z_t, x_t^a) = \frac{1}{T}\sum_t C(x_t+\nu_t, x_t+u_t) = \gamma_0 . \qquad (5\text{-}77)$$

Dabei wurde ausgenutzt, dass ν_t, u_t und x_t miteinander unkorreliert sind.

Der zweite Summand aus (5-76) ergibt sich zu:

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_{t}(z_t-\mu_z)(\overline{x^a}-\mu_x)\right]=\frac{1}{T^2}\sum_{t}\sum_{t'}C(z_t,x^a_{t'})\\
&=\frac{1}{T^2}\sum_{t}\sum_{t'}C(x_t+\nu_t,x_{t'}+u_{t'})=\frac{1}{T^2}\sum_{t}\sum_{t'}C(x_t,x_{t'})\\
&=\frac{1}{T^2}\left(T\cdot\gamma_0+2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 1\gamma_{T-1}\right)\,. \qquad (5\text{-}78)
\end{aligned}
$$

Der dritte Summand kann umgeschrieben werden zu:

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_{t}(\overline{z}-\mu_z)(\overline{x^a}-\mu_x)\right]=\frac{1}{T^3}\sum_{t}\sum_{t'}\sum_{t''}C(z_{t'},x^a_{t''})=\frac{1}{T^2}\sum_{t'}\sum_{t''}C(x_{t'},x_{t''})\\
&=\frac{1}{T^2}\left(T\cdot\gamma_0+2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 1\gamma_{T-1}\right)\,. \qquad (5\text{-}79)
\end{aligned}
$$

Für den vierten Summanden aus (5-76) ergibt sich:

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_{t}(\overline{z}-\mu_z)(x^a_t-\mu_x)\right]=\frac{1}{T^2}\sum_{t}\sum_{t'}C(z_{t'},x^a_t)=\frac{1}{T^2}\sum_{t}\sum_{t'}C(x_{t'},x_t)\\
&=\frac{1}{T^2}\left(T\cdot\gamma_0+2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 1\gamma_{T-1}\right)\,. \qquad (5\text{-}80)
\end{aligned}
$$

Wiederum sind der zweite, dritte und vierte Summand ((5-78), (5-79) und (5-80)) identisch.

Zusammenfassend erhält man

$$
E[S_{zx^a}]=\gamma_0\left(1-\frac{1}{T}\right)-\frac{1}{T^2}\left(2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 2\gamma_{T-2}+2\cdot 1\gamma_{T-1}\right). \qquad (5\text{-}81)
$$

(5-81) entspricht somit (5-75).

Für den Wahrscheinlichkeitsgrenzwert des IV-Schätzers lässt sich damit schreiben:

$$
plim\,\widehat{\beta^{IV}}=\frac{E[S_{zx}]}{E[S_{zx^a}]}\beta+\frac{E[S_{zv}]}{E[S_{zx^a}]}
$$

$$
=\frac{\gamma_0\left(1-\frac{1}{T}\right)-\frac{1}{T^2}\left(2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 2\gamma_{T-2}+2\cdot 1\gamma_{T-1}\right)}{\gamma_0\left(1-\frac{1}{T}\right)-\frac{1}{T^2}\left(2(T-1)\gamma_1+2(T-2)\gamma_2+\cdots+2\cdot 2\gamma_{T-2}+2\cdot 1\gamma_{T-1}\right)}\beta\,. \qquad (5\text{-}82)
$$

Da der Zähler und der Nenner in (5-82) gleich sind und $E[S_{zv}]=0$ annahmegemäß erfüllt ist, ist der IV-Schätzer konsistent:

$$plim\,\widehat{\beta^{IV}} = \beta\,.$$

Für den AR(1)-Prozess kann man wegen $\gamma_j = \gamma_0 \cdot \rho^j,\ j = 1, ..., T-1$ (vgl. Ronning (2009)) schreiben:[6]

$$plim\,\widehat{\beta^{IV}} = \beta\left[1 - \frac{\frac{\gamma_0}{T^2}\left(2(T-1)\rho + 2(T-2)\rho^2 + \cdots + 2\cdot 2\rho^{T-2} + 2\cdot 1\rho^{T-1}\right)}{\frac{\gamma_0}{T^2}\left(2(T-1)\rho + 2(T-2)\rho^2 + \cdots + 2\cdot 2\rho^{T-2} + 2\cdot 1\rho^{T-1}\right)}\right]. \tag{5-83}$$

γ_0 bezeichnet hier die Varianz des Originalregressors, ρ ist der Autokorrelationsparameter.

5.2 Überlagerung mit Faktorstruktur

Bei der additiven Überlagerung mit Faktorstruktur werden ebenfalls die drei alternativen Instrumente getestet:

Verzögerte Variable:

$$z_{it} = x^a_{it} = x_{it-1} + \delta D_i + \varepsilon_{it-1}\,, \tag{5-84}$$

Differenz von verzögerten Variablen:

$$z_{it} = x^a_{it-1} - x^a_{it-2} = (x_{it-1} - x_{it-2}) + (\varepsilon_{it-1} - \varepsilon_{it-2}) \tag{5-85}$$

und *zusätzlich anonymisierte Variable:*

$$z_{it} = x^{a,neu}_{it} = x_{it} + \delta^z D^z_i + \varepsilon^z_{it}\,. \tag{5-86}$$

Bei der letzteren Variante wird die Fehlervariable unabhängig vom Fehler in x^a_{it} generiert. Dies wird durch den Index z verdeutlicht.

Überprüfung der Annahmen eines gültigen Instruments

Zuerst wird wieder überprüft, ob und unter welchen Bedingungen die Annahmen eines validen Instruments im Fall der additiven Überlagerung mit Faktorstruktur erfüllt sind (vgl. (A1)-(A3), Abschnitt 3).

[6] Der Ausdruck in der Klammer entspricht Eins und kann natürlich gekürzt werden. Da er jedoch in späteren Ausführungen eine Rolle spielen wird, wird er hier in dieser Form dargestellt.

- Korrelation zwischen Regressor und Instrument

Erstens muss der fehlerfreie Regressor mit dem Instrument möglichst hoch korrelieren, d.h. $C(z_{it}, x_{it}) \neq 0$.

Für die *verzögerte anonymisierte Variable* als Instrument ergibt sich

$$
\begin{aligned}
C(z_{it}, x_{is}) &= C\Big((x_{it-1} + \delta D_i + \varepsilon_{it-1}), x_{is}\Big) \\
&= \begin{cases} V(x_{is})\,, & \text{wenn } s = t-1\,, \\ C(x_{it-1}, x_{is})\,, & \text{wenn } s \neq s-1\,, \end{cases} \qquad \text{(5-87)} \\
& \qquad t, s = 1, \cdots, T\,.
\end{aligned}
$$

Erwartungsgemäß ist die Annahme (A1) erfüllt. Für den AR(1)-Prozess und $s \neq t-1$ lässt sich weiter schreiben:

$$C(x_{it-1}, x_{is}) = \gamma_0 \rho^{|t-1-s|}\,. \qquad \text{(5-88)}$$

Mit steigender positiver Autokorrelation des Regressors verbessert sich der Zusammenhang zwischen x und z.

Für die*Differenz von verzögerten Variablen* erhält man

$$
\begin{aligned}
C(x_{it}, z_{is}) &= C\Big(x_{it}, ((x_{is-1} - x_{is-2}) + (\varepsilon_{is-1} - \varepsilon_{is-2}))\Big) \\
&= C\left(x_{it}, x_{is-1}\right) - C\left(x_{it}, x_{is-2}\right)\,, \qquad \text{(5-89)} \\
& \qquad t, s = 1, \cdots, T\,.
\end{aligned}
$$

Dabei resultiert das gleiche Ergebnis wie bei der einfachen additiven Überlagerung (vgl. Gl. (5-38)). Die bestehende Korrelation zwischen x und z wird mit steigendem $C\left(x_{it}, x_{is-2}\right)$ abgeschwächt.

Für die *zusätzlich anonymisierte Variable* als Instrument gilt

$$C(z_{it}, x_{is}) = C\Big((x_{it} + \delta^z D_i^z + \varepsilon_{it}^z), x_{is}\Big)$$

$$= \begin{cases} V(x_{is})\,, \quad \text{wenn } s = t\,, \\ C(x_{it}, x_{is}) = \gamma_0 \rho^{|t-s|}\,, \quad \text{wenn } s \neq t\,. \end{cases} \tag{5-90}$$

Die Bedingung ist damit erfüllt. Dabei ist zu beachten, dass die Fehler des ersten Regressors $(\delta D_i + \varepsilon_{it})$ und des Instruments $(\delta^z D_i^z + \varepsilon_{it}^z)$ voneinander stochastisch unabhängig generiert werden.

- Korrelation zwischen Regressor und Störterm

Zweitens soll die Annahme der Unkorreliertheit des Instruments mit dem Störterm der abhängigen Variable gelten.

Im Fall der *verzögerten Variable* ist

$$C(z_{it}, \epsilon_{is}) = C\Big((x_{it-1} + \delta D_i + \varepsilon_{it-1}), \epsilon_{is}\Big) = 0\,, \tag{5-91}$$

$$t, s = 1, \cdots, T$$

dann erfüllt, wenn der Regressor und die Fehlervariable bzw. ihre einzelnen Komponenten mit dem Modellstörterm zu allen Zeitpunkten unkorreliert sind.

Die gleichen Bedingungen müssen auch für die *Differenzvariable* und die *zusätzlich anonymisierte Variable* als Instrumente erfüllt sein:

$$C(z_{it}, \epsilon_{is}) = C\Big(((x_{it} - x_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})), \epsilon_{is}\Big) = 0\,, \tag{5-92}$$

$$C(z_{it}, \epsilon_{is}) = C\Big((x_{it} + \delta^z D_i^z + \varepsilon_{it}^z), \epsilon_{is}\Big) = 0\,. \tag{5-93}$$

- Korrelation zwischen Regressor und Fehlervariable

Drittens wird unterstellt, dass das Instrument und die Fehlervariable keine Korrelation aufweisen. Dabei ist zu berücksichtigen, dass diese Annahme für alle Zeitpunkte t und s gelten soll.

Für die *verzögerte Variable* lässt sich schreiben:

$$C\Big(z_{it}, (\delta D_i + \varepsilon_{is})\Big) = E\Big[(x_{it-1} + \delta D_i + \varepsilon_{it-1})(\delta D_i + \varepsilon_{is})\Big]$$

$$= \begin{cases} \delta^2\,, \quad \text{wenn } s \neq t-1\,, \\ \delta^2 + \sigma_\varepsilon^2\,, \quad \text{wenn } s = t-1\,. \end{cases} \tag{5-94}$$

Dabei wurden die Annahmen ausgenutzt, die bei der Generierung der Fehlervariablen unterstellt werden, d.h. $C(x_{it}, \varepsilon_{is}) = 0$ (für alle t und s) und $C(\varepsilon_{it}, \varepsilon_{is}) = 0$ $(t \neq s)$.

Insbesondere wegen der gemeinsamen Variablen D_i sind die Fehlervariablen miteinander korreliert. Die Forderung $C(z, \text{Fehler}) = 0$ ist damit verletzt.

Bei der *Differenz der verzögerten Variable* als Instrument wird der δ-Faktor zwar herausdifferenziert. Wegen

$$\begin{aligned} C\Bigg(z_{it}, (\delta D_i + \varepsilon_{is})\Bigg) &= C\Bigg(((x_{it-1} - x_{it-2}) + (\varepsilon_{it-1} - \varepsilon_{it-2})), (\delta D_i + \varepsilon_{is})\Bigg) \\ = C\Bigg((\varepsilon_{it-1} - \varepsilon_{it-2}), \varepsilon_{is}\Bigg) &= C(\varepsilon_{it-1}, \varepsilon_{is}) - C(\varepsilon_{it-2}, \varepsilon_{is})\,, \qquad \text{(5-95)} \end{aligned}$$

$$t, s = 1, \cdots, T$$

ist jedoch die Annahme verletzt, sobald die Situation mit gleichen Zeitpunkten ($t = s - 1$ bzw. $t = s - 1$) eintritt.

Im Fall der *zusätzlichen anonymisierten Variable* werden die Fehlervariablen unabhängig voneinander erzeugt, was zur Erfüllung der Annahme führt:

$$C\Bigg(z_{it}, (\delta D_i + \varepsilon_{is})\Bigg) = E\Bigg[(x_{it} + \delta^z D_i^z + \varepsilon_{it}^z)(\delta D_i + \varepsilon_{is})\Bigg] = 0\,. \qquad \text{(5-96)}$$

Konsistenz des IV-Schätzers (zusätzlich anonymisierte Variable als Instrument)

Da die *verzögerte anonymisierte Variable* eine der Bedingungen für ein valides Instrument vorab nicht erfüllt, führt die IV-Schätzung zu inkonsistenten Schätzern. Aber auch im Fall der *Differenz der verzögerten anonymisierten Variablen* resultieren verzerrte Schätzer, da die "Within"- Transformation zur Verletzung einer für die Konsistenz notwendigen Annahme beiträgt (vgl. Gl. (5-95)). Aus diesem Grund wird im Folgenden lediglich der Wahrscheinlichkeitsgrenzwert des IV-Schätzers für das Instrument *zusätzliche anonymisierte Variable* hergeleitet. Dabei wird die Zerlegung in Gleichung (5-63) ausgenutzt. Des Weiteren wird bei der Herleitung auf einige Ergebnisse in Abschnitt 5.1 zurückgegriffen.

Wegen

$$C(z_t, x_t) = E\left[(x_t + \delta^z D_i^z + \varepsilon_t^z)x_t\right] = V(x_t) = \gamma_0 \tag{5-97}$$

resultiert, dass der erste Summand in Gleichung (5-63) ebenfalls wie bei der einfachen Überlagerung der Varianz des Regressors entspricht.

Für den zweiten Summanden ergibt sich (vgl. Gl. (5-66)):

$$\begin{aligned}
&E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(\overline{x} - \mu_x)\right] = \frac{1}{T^2}\sum_t^T \sum_{t'}^T C\left(z_t, x_{t'}\right) \\
&= \frac{1}{T^2}\sum_t^T \sum_{t'}^T (x_t + \delta^z D_i^z + \varepsilon_t^z, x_{t'}) = \frac{1}{T^2}\sum_t \sum_{t'} C(x_t, x_{t'}) \\
&= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\cdot\gamma_{T-2}^H + \gamma_{T-1}^H\right) .
\end{aligned} \tag{5-98}$$

Wegen Gleichung (5-71) erhält man für den dritten Summanden:

$$\begin{aligned}
&E\left[\frac{1}{T}\sum_t (\overline{z} - \mu_z)(\overline{x} - \mu_x)\right] = \frac{1}{T^2}\sum_{t'} \sum_{t''} C(x_{t'}, x_{t''}) \\
&= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\cdot\gamma_{T-2}^H + \gamma_{T-1}^H\right) .
\end{aligned} \tag{5-99}$$

Der vierte Summand entspricht wegen Gleichung (5-72) und

$$C(z_{t'}, x_t) = C(x_{t'}, x_t) \tag{5-100}$$

dem in Formel (5-74) hergeleiteten Ergebnis.

Die Einzelergebnisse kann man anschließend zusammenfassen:

$$E[S_{zx}] = \gamma_0 - \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1^H + \cdots + 2\cdot 2\gamma_{T-2}^H + 2\gamma_{T-1}^H\right) . \tag{5-101}$$

Für die Herleitung von $E[S_{zx^a}]$ gilt weiterhin die Gleichung (5-76). Die *Kovarianz zwischen dem Instrument und dem anonymisierten Regressor* ist

$$C(z_t, x_t^a) = C\left(x_t + \delta^z D_i^z + \varepsilon_t^z, x_t + \delta D_i + \varepsilon_t\right) = V(x_t) = \gamma_0\,, \tag{5-102}$$

da die einzelnen Komponenten der Fehlervariable miteinander und mit x unkorreliert sind.

Der erste Summand enspricht dann γ_0. Wegen der Kovarianz

$$C(z_t, x_s^a) = C\left(x_t, x_s\right) \ , \tag{5-103}$$

erhält man für den zweiten Summanden:

$$E\left[\frac{1}{T}\sum_t (z_t - \mu_z)(\overline{x^a} - \mu_x)\right] = \frac{1}{T^2}\sum_t\sum_{t'} C(z_t, x_{t'}^a) = \frac{1}{T^2}\sum_t\sum_{t'} C(x_t, x_{t'})$$
$$= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\cdot 1\gamma_{T-1}^H\right) \ , \tag{5-104}$$

für den dritten Summanden:

$$E\left[\frac{1}{T}\sum_t (\overline{z} - \mu_z)(\overline{x^a} - \mu_x)\right] = \frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(z_{t'}, x_{t''}^a) = \frac{1}{T^2}\sum_{t'}\sum_{t''} C(x_{t'}, x_{t''})$$
$$= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\gamma_{T-1}\right) \tag{5-105}$$

und anschließend für den vierten Summanden:

$$E\left[\frac{1}{T}\sum_t (\overline{z} - \mu_z)(x_t^a - \mu_x)\right] = \frac{1}{T^2}\sum_t\sum_{t'} C(z_{t'}, x_t^a) = \frac{1}{T^2}\sum_t\sum_{t'} C(x_{t'}, x_t)$$
$$= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\gamma_{T-1}^H\right) \ . \tag{5-106}$$

Damit resultiert

$$E[S_{zx^a}] = \gamma_0 - \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1^H + \cdots + 2\cdot 2\gamma_{T-2}^H + 2\gamma_{T-1}^H\right) \ . \tag{5-107}$$

Zusammenfassend lässt sich an dieser Stelle festhalten, dass aufgrund der Gleichheit von $E[S_{zx^a}]$ und $E[S_{zx}]$ und aufgrund von $E[S_{z\varepsilon_y}] = 0$ der IV-Schätzer mit zusätzlicher anonymisierter Variable bei additiver Überlagerung mit Faktorstruktur konsistent ist (vgl. Gl. (4-14)):

$$plim\,\widehat{\beta^{IV}} = \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta + \frac{E[S_{z\varepsilon_y}]}{E[S_{zx^a}]} = \frac{E[S_{zx}]}{E[S_{zx}]}\beta = \beta \, . \tag{5-108}$$

6 Multiplikative Überlagerung

In diesem Abschnitt wird die multiplikative stochastische Überlagerung untersucht. Die Vorgehensweise des Kapitels zur additiven Überlagerung wird beibehalten. Zuerst werden die Wahrscheinlichkeitsgrenzwerte des IV-Schätzers für die einfache und anschließend für die Überlagerung mit Faktorstruktur bestimmt.

6.1 Einfache Überlagerung

Verzögerte Variable als Instrument

Die folgende Instrumentvariable wird verwendet:

$$z_{it} = x^a_{it-1} = x_{it-1} u_{it-1} \,. \tag{6-109}$$

Die Fehlervariable hat den Erwartungswert 1 und die Varianz σ_u^2 und ist vom Originalregressor stochastisch unabhängig.

Überprüfung der Annahmen eines gültigen Instruments

Der obigen Vorgehensweise folgend werden zuerst die Annahmen angeschaut, die die IV-Schätzung voraussetzt ((A1)-(A3), Abschnitt 3).

Es lässt sich schnell nachvollziehen, dass der Regressor und das Instrument miteinander korreliert sind. Dabei gilt

$$C(z_{it}, x_{is}) = C(x_{it-1} u_{it-1}, x_{is}) = C(x_{it-1}, x_{is}) \,, \tag{6-110}$$

$$t, s = 1, \cdots, T \,.$$

Für die Kovarianz zwischen dem Regressor und der Instrumentvariable ergibt sich im Vergleich zur additiven Überlagerung ein veränderter Ausdruck:

$$C(z_{it}, u_{is}) = C\left(x_{it-1} u_{it-1}, u_{is}\right) = E[x_{it-1}] C\left(u_{it-1}, u_{is}\right) \tag{6-111}$$

$$t, s = 1, \cdots, T \,.$$

Diese Kovarianz ist nur dann gleich Null, wenn die Fehler nicht autokorreliert sind und $s \neq t-1$ gilt.

Unter Gültigkeit der Annahmen eines linearen Modells und der für die Fehlervariablen unterstellten Bedingungen ist auch $C(z_{it}, \epsilon_{is}) = 0$ gegeben.

Inkonsistenz des IV-Schätzers

Für den *Erwartungswert der empirischen Kovarianz zwischen dem Originalregressor* *und der Instrumentvariable* erhält man

$$\begin{aligned}
E[S_{zx}] &= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x^a_{t-1}-\overline{x}^a_1)(x_t-\overline{x_2})\Bigg] \\
&= \Bigg[\frac{1}{T-1}\sum_{t=2}^{T}\left(x_{t-1}u_{t-1}-\frac{1}{T-1}\sum_{t'=2}^{T}x_{t'-1}u_{t'-1}\right)\left(x_t-\frac{1}{T-1}\sum_{s=2}^{T}x_s\right)\Bigg] \\
&= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}x_{t-1}x_t u_{t-1}\Bigg]-E\Bigg[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{s=2}^{T}x_{t-1}x_s u_{t-1}\Bigg] \\
&\quad - E\Bigg[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{t'=2}^{T}x_{t'-1}x_t u_{t'-1}\Bigg]+E\Bigg[\frac{1}{(T-1)^3}\sum_{t=2}^{T}\sum_{t'=2}^{T}\sum_{s=2}^{T}x_{t'-1}x_s u_{t'-1}\Bigg] \\
&= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}\left(x_{t-1}-\frac{1}{T-1}\sum_{t'=2}^{T}x_{t'-1}\right)\left(x_t-\frac{1}{T-1}\sum_{s=2}^{T}x_{s-1}\right)\Bigg] \\
&= E\Bigg[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}-\overline{x}_1)(x_t-\overline{x_2})\Bigg]\,. \qquad (6\text{-}112)
\end{aligned}$$

Gleichung (6-112) entspricht dem Erwartungswert $E[S_{zx}]$ im Fall der einfachen additiven Überlagerung (vgl. Gl. (5-26)). Aus diesem Grund gilt das in Gleichung (5-31) aufgezeigte Ergebnis ebenfalls für $E[S_{zx}]$ bei multiplikativer Überlagerung.

Weiter ist zu bestimmen:

$$
\begin{aligned}
E[S_{zx^a}] &= E\left[\frac{1}{T-1}\sum_{t=2}^{T}(x^a_{t-1}-\overline{x}^a_1)(x^a_t-\overline{x}^a_2)\right]\\
&= E\left[\frac{1}{T-1}\sum_{t=2}^{T}\Big((x^a_{t-1}-\mu)-(\overline{x}^a_1-\mu)\Big)\Big((x^a_t-\mu)-(\overline{x}^a_2-\mu)\Big)\right]\\
&= E\left[\frac{1}{T-1}\sum_{t=2}^{T}\Big((x^a_{t-1}-\mu)(x^a_t-\mu)-(x^a_{t-1}-\mu)(\overline{x}^a_2-\mu)\right.\\
&\left. -(x^a_t-\mu)(\overline{x}^a_1-\mu)+(\overline{x}^a_1-\mu)(\overline{x}^a_2-\mu)\Big)\right]\\
&= E\left[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_tu_t-\mu)\right]\\
&- E\left[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{s=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_su_s-\mu)\right]\\
&- E\left[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{t'=2}^{T}(x_tu_t-\mu)(x_{t'-1}u_{t'-1}-\mu)\right]\\
&+ E\left[\frac{1}{(T-1)^3}\sum_{t=2}^{T}\sum_{t'=2}^{T}\sum_{s=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_su_s-\mu)\right]. \qquad (6\text{-}113)
\end{aligned}
$$

Die einzelnen Summanden lassen sich analog zum Fall der additiven Überlagerung herleiten.

Für den ersten Summanden von (6-113) resultiert:

$$
\begin{aligned}
&E\left[\frac{1}{T-1}\sum_{t=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_tu_t-\mu)\right] = C(x^a_t, x^a_{t-1})\\
&= E[x_tx_{t-1}]E[u_tu_{t-1}] - E[x_t]E[x_{t-1}] = C(x_t, x_{t-1}) = \gamma_1\,, \qquad (6\text{-}114)
\end{aligned}
$$

wenn die Fehlervariable u nicht autokorreliert ist ($C(u_t, u_{t-1}) = 0$).

Der zweite Summand von (6-113) ist

$$
\begin{aligned}
&E\left[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{s=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_s u_s-\mu)\right] = \frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{s=2}^{T}C(x^a_{t-1},x^a_s)\\
&= \frac{1}{(T-1)^2}\sum_{t=2}^{T-1}V(x_t u_t) + \frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{s=2,t-1\neq s}^{T}C(x_{t-1},x_s)\\
&= \frac{(T-2)}{(T-1)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&+\frac{1}{(T-1)^2}\Big(2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big)\,.
\end{aligned}
\tag{6-115}
$$

Der dritte Summand von (6-113) ist gegeben als

$$
\begin{aligned}
&E\left[\frac{1}{(T-1)^2}\sum_{t=2}^{T}\sum_{t'=2}^{T}(x_t u_t-\mu)(x_{t'-1}u_{t'-1}-\mu)\right]\\
&= \frac{1}{(T-1)^2}E\Bigg[(x_2u_2-\mu)\Big((x_1u_1-\mu)+(x_2u_2-\mu)+\cdots+(x_{T-1}u_{T-1}-\mu)\Big)\\
&\qquad +\cdots+\\
&\qquad (x_Tu_T-\mu)\Big((x_1u_1-\mu)+(x_2u_2-\mu)+\cdots+(x_{T-1}u_{T-1}-\mu)\Big)\Bigg]\\
&= \frac{1}{(T-1)^2}\Big[\gamma_1+V(x_tu_t)+\cdots+\gamma_{T-3}+\ \gamma_2+\gamma_1+V(x_tu_t)+\cdots+\gamma_{T-4}\\
&\qquad +\cdots+\gamma_{T-1}+\gamma_{T-2}+\cdots+\gamma_1\Big]\\
&= \frac{(T-2)}{(T-1)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&+\frac{1}{(T-1)^2}\Big(2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big)\,.
\end{aligned}
\tag{6-116}
$$

Anschließend lässt sich für den vierten Summanden herleiten:

$$
\begin{aligned}
&E\Big[\frac{1}{(T-1)^3}\sum_{t=2}^{T}\sum_{t'=2}^{T}\sum_{s=2}^{T}(x_{t-1}u_{t-1}-\mu)(x_s u_s-\mu)\Big]\\
&=\frac{1}{(T-1)^2}E\Big[(x_1u_1-\mu)\Big((x_2u_2-\mu)+\cdots+(x_Tu_T-\mu)\Big)\\
&\qquad\qquad +\cdots+\\
&\qquad\qquad (x_{T-1}u_{T-1}-\mu)\Big((x_2u_3-\mu)+\cdots+(x_Tu_T-\mu)\Big)\Big]\\
&=\frac{(T-2)}{(T-1)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&+\frac{1}{(T-1)^2}\Big(2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big)\,. \qquad (6\text{-}117)
\end{aligned}
$$

Diese Einzelergebnisse führen zum Erwartungswert

$$
\begin{aligned}
E[S_{zx^a}] &= \gamma_1-\frac{(T-2)}{(T-1)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&-\frac{1}{(T-1)^2}\Big(2(T-2)\gamma_1+2(T-3)\gamma_2+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big)\\
&=E[S_{zx}]-\frac{(T-2)}{(T-1)^2}\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\,. \qquad (6\text{-}118)
\end{aligned}
$$

Ferner ergibt sich für den Wahrscheinlichkeitswert des IV-Schätzers im Fall der einfachen multiplikativen Überlagerung (vgl. Gl. (4-16)):

$$
\begin{aligned}
plim\,\widehat{\beta^{IV}} =& \frac{E_{\alpha z v}}{E[S_{zx}]-\frac{(T-2)}{(T-1)^2}\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)}+\\
&\frac{E\big[S_{zx}\big]}{E[S_{zx}]-\frac{(T-2)}{(T-1)^2}\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)}\beta\neq\beta\,. \qquad (6\text{-}119)
\end{aligned}
$$

Der IV-Schätzer ist damit bei Verwendung der verzögerten anonymisierten Variable inkonsistent, wobei die Inkonsistenz insbesondere aufgrund von $\sigma_u^2 \neq 0$ resultiert. Da Individualeffekte in die *plim*-Formel enthalten sind und sie in der Regel mit Originalregressoren korrelieren, wird die Verzerrung darurch noch vergrößert.

Differenz von verzögerten Variablen

Die Instrumentvariable ist wie folgt definiert:

$$z_{it} = x^a_{it-1} - x^a_{it-2} = x_{it-1}u_{it-1} - x_{it-2}u_{it-2}\,, \tag{6-120}$$

dabei gelten die üblichen Annahmen für die multiplikative Fehlervariable.

Überprüfung der Annahmen eines gültigen Instruments

Die Kovarianz zwischen z und x ist

$$C(z_{it}, x_{is}) = C\Big((x_{it-1}u_{it-1} - x_{it-2}u_{it-2}), x_{is}\Big) = C(x_{it-1}, x_{is}) - C(x_{it-2}, x_{is})\,, \tag{6-121}$$

$$t, s = 1, \cdots, T\,.$$

Dies entspricht dem Ergebnis der additiven Überlagerung.

Ist das Ausgangsmodell richtig spezifiziert und der Regressor mit der Fehlervariable zu allen Zeitpunkten unkorreliert, so wird $C(z_{it}, \epsilon_{is}) = 0$ normalerweise erfüllt sein.

Das Problem könnte aber von der Kovarianz zwischen z und der Fehlervariable ausgehen. Hier erhält man

$$\begin{aligned} C(z_{it}, u_{is}) &= C\Big((x_{it-1}u_{it-1} - x_{it-2}u_{it-2}), u_{is}\Big) \\ &= E[x_{t-1}]C(u_{it-1}, u_{is}) - E[x_{t-2}]C(u_{it-2}, u_{is})\,, \end{aligned} \tag{6-122}$$

$$t, s = 1, \cdots, T\,.$$

Es lässt sich keine eindeutige Aussage bezüglich der Validität der Annahme (A3) machen. Entscheidend ist, ob die Fehler keine Autokorrelation aufweisen und $s \neq t-1$ wie $s \neq t-2$ gegeben ist.

Inkonsistenz des IV-Schätzers

Folgend der obigen Vorgehensweise wird hier zunächst der *Erwartungswert der empirischen Kovarianz zwischen dem Instrument und dem Originalregressor* bestimmt.

Die Kovarianz ist

$$
\begin{aligned}
S_{zx} &= \frac{1}{T-2}\sum_{t=3}^{T}(z_t-\overline{z})(x_t-\overline{x})\\
&= \frac{1}{T-2}\sum_{t=3}^{T}\left((x_{t-1}u_{t-1}-x_{t-2}u_{t-2})-\frac{1}{T-2}\sum_{s=3}^{T}(x_{s-1}u_{s-1}-x_{s-2}u_{s-2})\right)(x_t-\overline{x})\\
&= \frac{1}{T-2}\sum_{t=3}^{T}\left(x_{t-1}u_{t-1}-\frac{1}{T-2}\sum_{s=3}^{T}x_{s-1}u_{s-1}\right)\left(x_t-\frac{1}{T-2}\sum_{s'=3}^{T}x_{s'}\right)\\
&- \frac{1}{T-2}\sum_{t=3}^{T}\left(x_{t-2}u_{t-2}-\frac{1}{T-2}\sum_{s=3}^{T}x_{s-2}u_{s-2}\right)\left(x_t-\frac{1}{T-2}\sum_{s'=3}^{T}x_{s'}\right).
\end{aligned}
\tag{6-123}
$$

Nach einigen Umformungen lässt sich für den Erwartungswert schreiben

$$
\begin{aligned}
E[S_{zx}] &= E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}-\overline{x}_1)(x_t-\overline{x})\right]-E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}-\overline{x}_2)(x_t-\overline{x})\right]\\
&= E\left[\frac{1}{T-2}\sum_{t=3}^{T}\left((x_{t-1}-\mu)-(\overline{x_1}-\mu)-(x_{t-2}-\mu)+(\overline{x_2}-\mu)\right)\times\right.\\
&\quad\left.\times\left((x_t-\mu)-(\overline{x}-\mu)\right)\right]
\end{aligned}
\tag{6-124}
$$

$$
\text{mit}\quad \overline{x}=\frac{1}{T-2}\sum_{t=3}^{T}x_t\,,\ \ \overline{x_1}=\frac{1}{T-2}\sum_{t=3}^{T}x_{t-1}\,,\ \ \overline{x_2}=\frac{1}{T-2}\sum_{t=3}^{T}x_{t-2}\,.
$$

Die Gleichung (6-124) ist damit mit der Gleichung (5-42) identisch. Dann gilt das Ergebnis (vgl. Gl. (5-52)):

$$
E[S_{zx}] = -\frac{1}{(T-2)^2}\gamma_0-\psi(\gamma,T) \tag{6-125}
$$

mit

$$
\begin{aligned}
\psi(\gamma,T) = \frac{1}{(T-2)^2}\Big[&2(T-3)\gamma_1-(T-2)^2\gamma_1+\cdots+\gamma_{T-2}\\
&-2(T-4)\gamma_1-2(T-4)\gamma_2+(T-2)^2\gamma_2-\cdots-\gamma_{T-1}\Big].
\end{aligned}
$$

Weiter wird der *Erwartungswert der empirischen Kovarianz zwischem dem Instrument und* *dem anonymisierten Regressor* hergeleitet:

$$
\begin{aligned}
E[S_{zx^a}] &= E\left[\frac{1}{T-2}\sum_{t=3}^{T}(z_t-\overline{z})(x_t^a-\overline{x}^a)\right]\\
&= E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}^a-x_{t-2}^a-\overline{x}_1^a+\overline{x}_2^a)(x_t^a-\overline{x}^a)\right]\\
&= E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}u_{t-1}-\mu)(x_tu_t-\mu)\right]\\
&- E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{t-1}u_{t-1}-\mu)(x_su_s-\mu)\right]\\
&- E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{t'=3}^{T}(x_tu_t-\mu)(x_{t'-1}u_{t'-1}-\mu)\right]\\
&+ E\left[\frac{1}{(T-2)^3}\sum_{t=3}^{T}\sum_{t'=3}^{T}\sum_{s=3}^{T}(x_{t'-1}u_{t'-1}-\mu)(x_su_s-\mu)\right]\\
&\\
&- E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}u_{t-2}-\mu)(x_tu_t-\mu)\right]\\
&+ E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{t-2}u_{t-2}-\mu)(x_su_s-\mu)\right]\\
&+ E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{t'=3}^{T}(x_tu_t-\mu)(x_{t'-2}u_{t'-2}-\mu)\right]\\
&- E\left[\frac{1}{(T-2)^3}\sum_{t=3}^{T}\sum_{t'=3}^{T}\sum_{s=3}^{T}(x_{t'-2}u_{t'-2}-\mu)(x_su_s-\mu)\right]
\end{aligned}
\tag{6-126}
$$

Folglich werden die acht Summanden bestimmt. Für den ersten und fünften Summanden resultiert im Fall der nicht autokorrelierten Fehlervariablen

$$E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-1}u_{t-1}-\mu)(x_tu_t-\mu)\right]=C(x_{t-1}u_{t-1},x_tu_t)=C(x_{t-1},x_t)=\gamma_1\,, \tag{6-127}$$

$$E\left[\frac{1}{T-2}\sum_{t=3}^{T}(x_{t-2}u_{t-2}-\mu)(x_tu_t-\mu)\right]=C(x_t,x_{t-2})=\gamma_2\,. \tag{6-128}$$

Der zweite Summand lässt sich wie folgt umformen

$$E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{t-1}u_{t-1}-\mu)(x_su_s-\mu)\right]$$

$$=\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3,\,t-1\neq s}^{T}C(x_{t-1},x_s)+\frac{1}{(T-2)^2}\sum_{t=3}^{T-1}V(x_tu_t)$$

$$=\frac{(T-3)}{(T-2)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)+ \tag{6-129}$$

$$+\frac{1}{(T-2)^2}\Big[2(T-3)\gamma_1+2(T-4)\gamma_2+\cdots+2\gamma_{T-3}+\gamma_{T-2}\Big] \tag{6-130}$$

Das Ergebnis in Gleichung (6-130) gilt auch für den dritten und vierten Summanden.

Weiter kann man für den sechsten Summanden schreiben:

$$E\left[\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3}^{T}(x_{t-2}u_{t-2}-\mu)(x_su_s-\mu)\right]$$

$$=\frac{1}{(T-2)^2}\sum_{t=3}^{T}\sum_{s=3,\,t-2\neq s}^{T}C(x_{t-2},x_s)+\frac{1}{(T-2)^2}\sum_{t=3}^{T-2}V(x_tu_t)$$

$$=\frac{(T-4)}{(T-2)^2}\left(\gamma_0+\sigma_u^2\left(\gamma_0+\left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)+ \tag{6-131}$$

$$+\frac{1}{(T-2)^2}\Big[2(T-4)\gamma_1+2(T-4)\gamma_2+2(T-5)\gamma_3+\cdots+2\gamma_{T-2}+\gamma_{T-1}\Big]. \tag{6-132}$$

Das gleiche Resultat wie in Gleichung (6-132) ergibt sich auch für die letzten beiden Summanden.

Der Erwartungswert der empirischen Kovarianz ist dann

$$\begin{aligned}
E[S_{zx^a}] &= \gamma_1 - \frac{(T-3)}{(T-2)^2}\left(\gamma_0 + \sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&- \frac{1}{(T-2)^2}\Big[2(T-3)\gamma_1 + 2(T-4)\gamma_2 + \cdots + 2\gamma_{T-3} + \gamma_{T-2}\Big]\\
&- \gamma_2 + \frac{(T-4)}{(T-2)^2}\left(\gamma_0 + \sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right)\right)\\
&+ \frac{1}{(T-2)^2}\Big[2(T-4)\gamma_1 + 2(T-4)\gamma_2 + 2(T-5)\gamma_3 + \cdots + 2\gamma_{T-2} + \gamma_{T-1}\Big]\\
&= -\frac{1}{(T-2)^2}\gamma_0 - \frac{1}{(T-2)^2}\sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right)\\
&- \frac{1}{(T-2)^2}\Big(2(T-3)\gamma_1 - (T-2)^2\gamma_1 + \cdots + 2\gamma_{T-3} + \gamma_{T-2}\\
&- 2(T-4)\gamma_1 - 2(T-4)\gamma_2 + (T-2)^2\gamma_2 - \cdots - \gamma_{T-1}\Big)\\
&= E[S_{zx}] - \frac{1}{(T-2)^2}\sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right).
\end{aligned} \tag{6-133}$$

Für den Wahrscheinlichkeitsgrenzwert des IV-Schätzers resultiert

$$plim\,\widehat{\beta^{IV}} = \frac{E_{\alpha z v}}{E[S_{zx}] - \frac{1}{(T-2)^2}\sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right)} + \frac{E\big[S_{zx}\big]}{E[S_{zx}] - \frac{1}{(T-2)^2}\sigma_u^2\left(\gamma_0 + \left(\frac{\rho_0}{1-\rho}\right)^2\right)}\beta \neq \beta\,. \tag{6-134}$$

Damit lassen sich die gleichen Folgerungen wie im Fall der einfachen additiven Überlagerung ziehen. Da die Fehlervarianz bei der Anonymisierung nicht gleich Null ist, kürzt sich der zweite Term in der *plim*-Formel (6-134) nicht zu β. Die zweite Quelle der Verzerrung ist auf die Korrelation zwischen α und x (und somit auch z) zurückzuführen, die in der Praxis in der Regel nicht Null ist.

Zusätzlich anonymisierte Variable als Instrument

Als Instrumentvariable wird

$$z_{it} = x_{it}\nu_{it} \tag{6-135}$$

betrachtet.

ν_{it} ist die Überlagerungsvariable, mit der die Originaldaten erneut überlagert werden. ν_{it} hat den Erwartungswert Eins und die Varianz σ_ν^2 und ist mit dem Fehler aus dem ersten anonymisierten Datensatz u_{it} für alle i, t nicht korreliert. Des Weiteren wird der Fall betrachtet, in dem ν_{it} nicht autokorreliert ist, d.h. $C(\nu_{it}, \nu_{is}) = 0$ $(t \neq s)$.

Überprüfung der Annahmen eines gültigen Instruments

Da

$$C(z_{it}, x_{is}) = C(x_{it}\nu_{it}, x_{is}) = C(x_{it}, x_{is})\,, \tag{6-136}$$

$$C(z_{it}, u_{is}) = E[x_{it}]E[\nu_{it}]E[u_{is}] - E[x_{it}]E[\nu_{it}]E[u_{is}] = 0 \tag{6-137}$$

und

$$C(z_{it}, \epsilon_{is}) = C(x_{it}\nu_{it}, \epsilon_{is}) = E[x_{it}]E[\nu_{it}]E[\epsilon_{is}] = 0\,. \tag{6-138}$$

$$\tag{6-139}$$

$$t, s = 1, \cdots, T$$

gilt, sind alle Annahmen eines validen Instruments bei dieser Variante der Instrumentvariable erfüllt.

Konsistenz des IV-Schätzers

Bei der Herleitung der Wahrscheinlichkeitsgrenzwerte des IV-Schätzers werden einzelne Ergebnisse aus vorherigen Abschnitten (insbesondere Abschnitt 5.1) verwendet. Der *Erwartungswert* $E[S_{zx}]$ setzt sich aus vier Summanden zusammen, die in (5-63) bereits hergeleitet worden sind.

Im Fall der einfachen multiplikativen Überlagerung erhält man für den ersten Summanden den folgenden Ausdruck:

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(z_t - \mu_z)(x_t - \mu_x)\right] = C(z_t, x_t) = E[x_t^2\nu_t] - E[x_t\nu_t]E[x_t] = V(x_t) = \gamma_0\,. \tag{6-140}$$

Für den zweiten Summanden resultiert:

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t}(z_t - \mu_z)(\overline{x} - \mu_x)\right] &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T} C(z_t, x_{t'}) = \frac{1}{T^2}\sum_{t}\sum_{t'} C(x_t, x_{t'}) \\ &= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right)\,. \end{aligned} \tag{6-141}$$

Der dritte Summand in (5-63) ist

$$E\left[\frac{1}{T}\sum_{t}(\overline{z} - \mu_z)(\overline{x} - \mu_x)\right] = \frac{1}{T^3}\sum_{t}\sum_{t'}\sum_{t''} C(z_{t'}, x_{t''})\,.$$

Wegen

$$C(z_{t'},x_{t''}) = C(x_{t'}\eta_{t'},x_{t''}) = E[x_{t'}x_{t''}] - E[x_{t'}]E[x_{t''}] = C(x_{t'},x_{t''})$$

gelangt man zu

$$\begin{aligned}
&\frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(z_{t'},x_{t''}) = \frac{1}{T^2}\sum_{t'}\sum_{t''} C(x_{t'},x_{t''}) \\
&= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad \text{(6-142)}
\end{aligned}$$

Auf einem ähnlichen Weg lässt sich auch der vierte Summand bestimmen:

$$\begin{aligned}
&E\left[\frac{1}{T}\sum_t(\overline{z}-\mu_z)(x_t-\mu_x)\right] \\
&= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad \text{(6-143)}
\end{aligned}$$

Folglich gilt für den Erwartungswert der empirischen Kovarianz von z und x:

$$E[S_{zx}] = \gamma_0 - \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad \text{(6-144)}$$

Für den *Erwartungswert der empirischen Kovarianz zwischen dem Instrument und dem anonymisierten Regressor* kann man unter Verwendung der Ergebnisse aus Abschnitt 5.1 schreiben:

Für den ersten Summanden in Gleichung (5-76) gilt:

$$\begin{aligned}
&E\left[\frac{1}{T}\sum_t(z_t-\mu_z)(x_t^a-\mu_x)\right] = \frac{1}{T}\sum_t C(z_t,x_t^a) \\
&= \frac{1}{T}\sum_t\Big(E[x_t\eta_t x_t u_t] - E[x_t\eta_t]E[x_t u_t]\Big) = \frac{1}{T}\sum_t\Big(E[x_t^2]E[\eta_t u_t] - E[x_t]E[x_t]\Big) \\
&= \frac{1}{T}\sum_t\gamma_0 = \gamma_0 . \qquad \text{(6-145)}
\end{aligned}$$

Der zweite Summand lässt sich umformen zu

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_t (z_t-\mu_z)(\overline{x^a}-\mu_x)\right] = \frac{1}{T^2}\sum_t\sum_{t'} C(z_t, x^a_{t'}) \\
&= \frac{1}{T^2}\sum_t\sum_{t'} E[x_t\eta_t x_{t'}u_{t'}] - E[x_t\eta_t]E[x_{t'}u_{t'}] \\
&= \frac{1}{T^2}\sum_t\sum_{t'} E[x_t x_{t'}]E[\eta_t u_{t'}] - E[x_t]E[x_{t'}] = \frac{1}{T^2}\sum_t\sum_{t'} C(x_t, x_{t'}) \\
&= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 1\gamma_{T-1}\right) . \qquad (6\text{-}146)
\end{aligned}
$$

Für den dritten Summanden erhält man:

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(\overline{x^a}-\mu_x)\right] = \frac{1}{T^3}\sum_t\sum_{t'}\sum_{t''} C(z_{t'}, x^a_{t''}) = \frac{1}{T^2}\sum_{t'}\sum_{t''} C(x_{t'}, x_{t''}) \\
&= \frac{1}{T^2}\left(T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 1\gamma_{T-1}\right) . \qquad (6\text{-}147)
\end{aligned}
$$

Der vierte Summand in (5-76) lässt sich wie folgt schreiben:

$$
\begin{aligned}
&E\left[\frac{1}{T}\sum_t (\overline{z}-\mu_z)(x^a_t-\mu_x)\right] = \frac{1}{T^2}\sum_t\sum_{t'} C(z_{t'}, x^a_t) = \frac{1}{T^2}\sum_t\sum_{t'} C(x_{t'}, x_t) \\
&= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 1\gamma_{T-1}\right) . \qquad (6\text{-}148)
\end{aligned}
$$

Für den gesamten Erwartungswert lässt sich damit schreiben:

$$
E[S_{zx^a}] = \gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right) . \qquad (6\text{-}149)
$$

Da die beiden Erwartungswerte $E[S_{zx}]$ und $E[S_{zx^a}]$ identisch sind, resultiert auch im Fall der einfachen multiplikativen Überlagerung ein konsistenter IV-Schätzer:

$$
\begin{aligned}
plim\, \widehat{\beta^{IV}} &= \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta \\
&= \frac{\gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right)}{\gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \cdots + 2\cdot 2\gamma_{T-2} + 2\cdot 1\gamma_{T-1}\right)}\beta \qquad (6\text{-}150) \\
&= \beta\, .
\end{aligned}
$$

6.2 Überlagerung mit Faktorstruktur

Die Instrumentvariablen im Fall der multiplikativen Überlagerung mit Faktorstruktur sind gegeben als

Verzögerte Variable:

$$z_{it} = x_{it}^a = x_{it-1}\left(1 + \delta D_i + \varepsilon_{it-1}\right) , \tag{6-151}$$

Differenz von verzögerten Variablen:

$$z_{it} = x_{it-1}^a - x_{it-2}^a = x_{it-1}(1 + \delta D_i + \varepsilon_{it-1}) - x_{it-2}(1 + \delta D_i + \varepsilon_{it-2}) \tag{6-152}$$

und *zusätzlich anonymisierte Variable:*

$$z_{it} = x_{it}^a = x_{it}\left(1 + \delta^z D_i^z + \varepsilon_{it}^z\right) . \tag{6-153}$$

Bei der zusätzlich anonymisierten Variable werden die Fehlervariable verschiedener Datensätze stochastisch unabhängig voneinander erzeugt.

Überprüfung der Annahmen eines gültigen Instruments

Für die *verzögerte anonymisierte Variable* erhält man

$$C(z_{it}, x_{is}) = C\Big(x_{it-1}(1 + \delta D_i + \varepsilon_{it-1}), x_{is}\Big) = C(x_{it-1}, x_{is})\,, \tag{6-154}$$

$$C(z_{it}, \epsilon_{is}) = 0\,, \tag{6-155}$$

$$t, s = 1, \cdots, T$$

und

$$C\Big(z_{it}, (1 + \delta D_i + \varepsilon_{is})\Big) = C\Big(x_{it-1}(1 + \delta D_i + \varepsilon_{it-1}), (1 + \delta D_i + \varepsilon_{is})\Big)$$
$$= E[x_{it-1}]E[(1 + \delta D_i + \varepsilon_{it-1})(1 + \delta D_i + \varepsilon_{is})] - E[x_{it-1}]$$

$$= \begin{cases} E[x_{it-1}](\delta^2 + \sigma_\varepsilon^2)\,, & \text{wenn } s = t-1\,, \\ E[x_{it-1}]\delta^2\,, & \text{wenn } s \neq t-1\,. \end{cases} \tag{6-156}$$

Damit ist die letzte Annahme (Unkorreliertheit des Instruments mit der Fehlervariable) verletzt, wenn als Instrument die verzögerte anonymisierte Variable verwendet wird.

Folglich lässt sich für die *Differenz von verzögerten Variablen* zeigen, dass diese Annahme auch hier nicht erfüllt ist:

$$C\Big(z_{it}, (1 + \delta D_i + \varepsilon_{is})\Big) = E[x_{it-1}]\Big(\delta^2 + C(\varepsilon_{it-1}, \varepsilon_{is})\Big) - E[x_{it-1}]\Big(\delta^2 + C(\varepsilon_{it-2}, \varepsilon_{is})\Big)\,, \tag{6-157}$$

$$t, s = 1, \cdots, T\,,$$

wobei gemäß den Annahmen bezüglich der Fehlervariable $C(\varepsilon_{it}, \varepsilon_{is}) = 0$ $(t \neq s)$ gilt.

Lediglich bei der *zusätzlich anonymisierten Variable* sind alle Bedingungen eines validen Instruments gegeben:

$$C(z_{it}, x_{is}) = C\Big(x_{it}(1 + \delta^z D_i^z + \varepsilon_{it}^z), x_{is}\Big) = C(x_{it}, x_{is})\,, \tag{6-158}$$

$$C(z_{it}, \epsilon_{is}) = C\Big(x_{it}(1 + \delta^z D_i^z + \varepsilon_{it}^z), \epsilon_{is}\Big) = 0\,, \tag{6-159}$$

$$C(z_{it}, (1 + \delta D_i + \varepsilon_{is})) = C\Big(x_{it}(1 + \delta^z D_i^z + \varepsilon_{it}^z), (1 + \delta D_i + \varepsilon_{is})\Big) = 0\,,$$

$$t, s = 1, \cdots, T\,.$$

Konsistenz des IV-Schätzers (zusätzlich anonymisierte Variable als Instrument)

Die folgenden Ausführungen beschränken sich nur auf das Instrument *zusätzliche anonymisierte Variable*, da die beiden anderen Instrumente die Ausgangsannahmen einer konsistenten Schätzung nicht erfüllen.

Der Wahrscheinlichkeitsgrenzwert für die multiplikative Überlagerung mit Faktorstruktur wurde bereits in Gleichung (4-21) hergeleitet. Da im Fall der zusätzlichen Variable als Instrument die Zufallsvariable D_i mit z und x unkorreliert ist und $\delta E[D_i] = 0$ gilt, lässt sich (4-21) zu

$$plim\,\widehat{\beta^{IV}} = \frac{plim\,\frac{1}{N}\sum_i \frac{1}{T}\sum_t (z_{it} - \overline{z}_i)(x_{it} - \overline{x}_i)\beta}{plim\,\frac{1}{N}\sum_i \frac{1}{T}\sum_t (z_{it} - \overline{z}_i)(x_{it}^a - \overline{x}_i^a)} = \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta \tag{6-160}$$

vereinfachen.

Bei der Herleitung der Erwartungswerte der empirischen Kovarianzen in (6-160) können Ergebnisse aus Abschnitt 6.1 verwendet werden. Die relevanten Erwartungswerte sind in Gleichungen (5-63) und (5-76) definiert.

Für die Bestimmung von $E[S_{zx}]$ und $E[S_{zx^a}]$ werden folgende Kovarianzen benötigt (vgl. Abschnitt 6.1):

$$C(z_t, x_s) = C(x_t + \delta^z D_i^z + \varepsilon_{tx}^z, x_s) = C\left(x_t, x_s\right)\,, \tag{6-161}$$

$$C(z_t, x_s^a) = C(x_t + \delta^z D_i^z + \varepsilon_{tx}^z, x_s + \delta D_i + \varepsilon_{sx}) = C\left(x_t, x_s\right)\,, \tag{6-162}$$

wobei man für $s = t$

$$C(z_t, x_t) = V(x_t) = \gamma_0\,, \tag{6-163}$$

$$C(z_t, x_t^a) = V(x_t) = \gamma_0 \tag{6-164}$$

erhält.

Da diese Kovarianzen denjenigen bei der einfachen multiplikativen Überlagerung entsprechen, lässt sich der Herleitungsweg (Gl. (6-140)-(6-149)) direkt auf die Überlagerung mit Faktorstruktur übertragen.

Anschließend folgt:

$$plim\,\widehat{\beta^{IV}} = \frac{E[S_{zx}]}{E[S_{zx^a}]}\beta$$

$$= \frac{\gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\cdot 2\gamma_{T-2}^H + 2\cdot 1\gamma_{T-1}^H\right)}{\gamma_0\left(1-\frac{1}{T}\right) - \frac{1}{T^2}\left(2(T-1)\gamma_1^H + 2(T-2)\gamma_2^H + \cdots + 2\cdot 2\gamma_{T-2}^H + 2\cdot 1\gamma_{T-1}^H\right)}\beta \tag{6-165}$$

Der IV-Schätzer mit zusätzlicher anonymisierter Variable als Instrument ist auch bei multiplikativer Überlagerung mit Faktorstruktur konsistent.

7 Multiples Modell

7.1 Additive Überlagerung

Die oben dargestellten Ausführungen werden auf ein Modell mit mehreren Regressoren ausgeweitet. Dabei wird davon ausgegangen, dass das Instrument alle Annahmen A(1)-A(3) (Abschnitt 3) erfüllt.

Für ein lineares Panelmodell im Fall der *einfachen additiven Überlagerung* lässt sich schreiben:

$$\boldsymbol{Y} = \boldsymbol{D\alpha} + (\boldsymbol{X}^a - \boldsymbol{U})\,\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{D\alpha} + \boldsymbol{X}^a\boldsymbol{\beta} + \boldsymbol{\omega} \tag{7-166}$$

mit $\boldsymbol{\omega} = \boldsymbol{\epsilon} - \boldsymbol{U\beta}$.

Für den Wahrscheinlichkeitsgrenzwert des IV-Schätzers resultiert:

$$
\begin{aligned}
plim\,\widehat{\boldsymbol{\beta}}^{IV} &= plim\left(\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1} plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Y}^a \\
&= \left(plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1} plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{D}\boldsymbol{\alpha}+\boldsymbol{X}^a\boldsymbol{\beta}+\boldsymbol{\omega}+\boldsymbol{V}\right) \\
&= \left(plim\,\frac{1}{NT}\left(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{U}\right)\right)^{-1} plim\,\frac{1}{NT}\left(\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{X}+\boldsymbol{U}\right)\boldsymbol{\beta}+\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega}+\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{V}\right) .
\end{aligned}
\tag{7-167}
$$

$\boldsymbol{V}$ ist der Vektor der Überlagerungsvariablen von $\boldsymbol{Y}$.

Bei Unkorreliertheit der Instrumentvariable mit den Fehlervariablen und dem Störterm zu allen Zeitpunkten gilt

$$plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega} = plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{\epsilon}-\boldsymbol{U}\boldsymbol{\beta}\right) = \boldsymbol{0}-\boldsymbol{0}\boldsymbol{\beta} = \boldsymbol{0}\,, \tag{7-168}$$

$$plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{U} = plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{V} = \boldsymbol{0} \tag{7-169}$$

und

$$plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X} = \boldsymbol{M}_{zx}\,. \tag{7-170}$$

Der Grenzwert des IV-Schätzers ist dann

$$plim\,\widehat{\boldsymbol{\beta}}^{IV} = \boldsymbol{M}_{zx}^{-1}\boldsymbol{M}_{zx}\boldsymbol{\beta} = \boldsymbol{\beta}\,. \tag{7-171}$$

Der IV-Schätzer ist bei Gültigkeit von (7-168) und (7-169) konsistent.

Das Panelmodell im Fall der *additiven Überlagerung mit Faktorstruktur* ist[7]

$$\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{\alpha} + \left(\boldsymbol{X}^a - \delta\boldsymbol{D}_i\otimes\boldsymbol{\iota}_{TK} - \boldsymbol{\varepsilon}_x\right)\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}^a\boldsymbol{\beta} + \boldsymbol{\omega}^F \tag{7-174}$$

mit $\boldsymbol{\omega}^F = \boldsymbol{\epsilon} - \left(\delta\boldsymbol{D}_i\otimes\boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x\right)\boldsymbol{\beta}$. $\otimes$ bezeichnet das Kronecker-Produkt.

[7] Zusammengefasst für alle Variablen erhält man in diesem Fall das Fehlermodell

$$\boldsymbol{X}^a = \boldsymbol{X} + \boldsymbol{W}_x = \boldsymbol{X} + \delta\boldsymbol{D}_i\otimes\boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x\,, \tag{7-172}$$

$$\boldsymbol{Y}^a = \boldsymbol{Y} + \boldsymbol{W}_y = \boldsymbol{Y} + \delta\boldsymbol{D}_i\otimes\boldsymbol{\iota}_{T} + \boldsymbol{\varepsilon}_y\,. \tag{7-173}$$

δ ist ein fester Parameter. $\boldsymbol{D}_i$ ist dabei ein $(N\times 1)$-Vektor, der für jede Beobachtung den Wert 1 oder -1 erhält.

Für den Grenzwert ergibt sich

$$
\begin{aligned}
&plim\,\widehat{\boldsymbol{\beta}}^{IV} = plim\left(\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1} plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Y}^a \\
&= \left(plim\,\frac{1}{NT}\Big(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{Z}'\boldsymbol{Q}(\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x)\Big)\right)^{-1} \\
&\times plim\,\frac{1}{NT}\Big(\boldsymbol{Z}'\boldsymbol{Q}\Big(\boldsymbol{X} + (\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x)\Big)\boldsymbol{\beta} + \boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega} + \boldsymbol{Z}'\boldsymbol{Q}(\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_y)\Big)\,.
\end{aligned}
\tag{7-175}
$$

Wenn das Instrument mit den einzelnen Komponenten der Fehlervariable (D_i, ε_{itx} und ε_{ity}) zu allen Perioden unkorreliert ist, gilt

$$
plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}(\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK}) = plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}(\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_T) = \boldsymbol{0}\,, \tag{7-176}
$$

sowie

$$
plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\varepsilon}_x = plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\varepsilon}_y = \boldsymbol{0}\,. \tag{7-177}
$$

Die Unkorreliertheit der Instrumentvariablen mit dem Modellstörterm führt weiter zu

$$
plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega}^F = plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{\epsilon} - (\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x)\right) = \boldsymbol{0}\,. \tag{7-178}
$$

Damit gelangt man bei additiver Überlagerung mit Faktorstruktur zu

$$
plim\,\widehat{\boldsymbol{\beta}}^{IV} = \boldsymbol{M}_{zx}^{-1}\boldsymbol{M}_{zx}\boldsymbol{\beta} = \boldsymbol{\beta}\,. \tag{7-179}
$$

Der IV-Schätzer ist damit konsistent, wobei die entscheidende Annahme der Unkorreliertheit der Instrumentvariable mit der Fehlervariable gültig sein muss. Das bedeutet, dass lediglich die IV-Schätzung mit der zusätzlich anonymisierten Variable zu Konsistenz führt. Nur in diesem Fall ist die Fehlervariable, die bei der Generierung der Instrumentvariable verwendet wird ($\boldsymbol{Z} = \delta^z\boldsymbol{D}_i^z \otimes \boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_x^z$), nicht mit dem Fehler in $\boldsymbol{X}^a$ korreliert.

7.2 Multiplikative Überlagerung

Für die *einfache multiplikative Überlagerung* resultiert der naive IV-Schätzer der Form

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{IV} &= \left(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Y}^a \\
&= \left(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\Big((\boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \odot \boldsymbol{V}\Big) \\
&= \left(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{D}\boldsymbol{\alpha} \odot \boldsymbol{V} + \boldsymbol{X}\boldsymbol{\beta} \odot \boldsymbol{V} + \boldsymbol{\epsilon} \odot \boldsymbol{V}\right) \\
&= \left(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X} \odot \boldsymbol{U}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\left(\boldsymbol{X}\boldsymbol{\beta} \odot \boldsymbol{V} + \boldsymbol{\epsilon} \odot \boldsymbol{V}\right)\,.
\end{aligned}
\tag{7-180}
$$

Für die Erwartungswerte der Terme von (7-180) erhält man

$$E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X} \odot \boldsymbol{U}\big] = E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}\big] \odot E\big[\boldsymbol{U}\big] = E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}\big] = \boldsymbol{M}_{zx}\,, \tag{7-181}$$

$$E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta} \odot \boldsymbol{V}\big] = E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}\boldsymbol{\beta}\big] \odot E\big[\boldsymbol{V}\big] = E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}\big]\boldsymbol{\beta} = \boldsymbol{M}_{zx}\boldsymbol{\beta}\,, \tag{7-182}$$

$$E\big[\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\epsilon} \odot \boldsymbol{V}\big] = \boldsymbol{0}\,, \tag{7-183}$$

wobei ausgenutzt wurde, dass die Instrumentvariable mit den Fehlervariablen des Regressors und der abhängigen Variable und dem Modellstörterm unkorreliert ist. Unter diesen Annahmen resultiert für den Wahrscheinlichkeitsgrenzwert des IV-Schätzers:

$$plim\,\widehat{\boldsymbol{\beta}}^{IV} = \boldsymbol{M}_{zx}^{-1}\boldsymbol{M}_{zx}\boldsymbol{\beta} = \boldsymbol{\beta}\,. \tag{7-184}$$

Setzt man anstelle von $\boldsymbol{U}$ und $\boldsymbol{V}$ die Fehlervariablen

$$\boldsymbol{X}^a = \boldsymbol{X} \odot \boldsymbol{W}_x = \boldsymbol{X} \odot (1 + \delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x)\,, \tag{7-185}$$

$$\boldsymbol{Y}^a = \boldsymbol{Y} \odot \boldsymbol{W}_y = \boldsymbol{Y} \odot (1 + \delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{T} + \boldsymbol{\varepsilon}_y) \tag{7-186}$$

ein, wobei $\odot$ für das Hadamard-Produkt steht, so lässt sich schnell zeigen, dass der IV-Schätzer bei Gültigkeit der Annahmen (A1) bis (A3) (Abschnitt 3) auch im Fall der *multiplikativen Überlagerung mit Faktorstruktur* konsistent ist.

8 Varianz des IV-Schätzers

Weiter wird die asymptotische Varianz des konsistenten IV-Schätzers hergeleitet.

Der IV-Schätzer lässt sich auch wie folgt darstellen:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{IV} &= \big(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\big)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Y}^a \\ &= \big(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\big)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\Big(\boldsymbol{D}\boldsymbol{\alpha} + \boldsymbol{X}^a\boldsymbol{\beta} + \boldsymbol{\Omega}\Big) \\ &= \boldsymbol{\beta} + \big(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\big)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\Omega}\,. \end{aligned} \tag{8-187}$$

Bei einfacher Überlagerung gilt

$$\boldsymbol{\Omega} = \boldsymbol{\epsilon} + \boldsymbol{V} - \boldsymbol{U}\boldsymbol{\beta}$$

und bei der Überlagerung mit Faktorstruktur wird für den Störterm

$$\boldsymbol{\Omega} = \boldsymbol{\epsilon} + (\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_T + \boldsymbol{\varepsilon}_y) - (\delta\boldsymbol{D}_i \otimes \boldsymbol{\iota}_{TK} + \boldsymbol{\varepsilon}_x)\boldsymbol{\beta}$$

verwendet.

Weiter gilt

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}^{IV}-\beta\right)=\left(\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1}\frac{1}{\sqrt{NT}}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega}. \tag{8-188}$$

Für die asymptotische Varianz-Kovarianzmatrix lässt sich herleiten

$$\begin{aligned}Asy.Var.\left(\widehat{\boldsymbol{\beta}}^{IV}\right)&=plim\left[\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}^{IV}-\boldsymbol{\beta}\right)\right]\left[\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}^{IV}-\boldsymbol{\beta}\right)\right]'\\&=plim\left(\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a\right)^{-1}plim\,\frac{1}{NT}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{Q}'\boldsymbol{Z}\,plim\left(\frac{1}{NT}\boldsymbol{X}^{a\prime}\boldsymbol{Q}\boldsymbol{Z}\right)^{-1}.\end{aligned} \tag{8-189}$$

Ferner erhält man

$$plim\,\frac{1}{N}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a=plim\,\frac{1}{N}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}=\boldsymbol{M}_{zx}\,, \tag{8-190}$$

$$plim\,\frac{1}{N}\boldsymbol{X}^{a\prime}\boldsymbol{Q}\boldsymbol{Z}=plim\,\frac{1}{N}\boldsymbol{X}'\boldsymbol{Q}\boldsymbol{Z}=\boldsymbol{M}_{xz} \tag{8-191}$$

und

$$plim\,\frac{1}{N}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{\omega}\boldsymbol{\omega}'\boldsymbol{Q}'\boldsymbol{Z}=\sigma^2\boldsymbol{M}_{zz}\,. \tag{8-192}$$

σ^2 ist dabei die Varianz von Ω.

Daraus resultiert die asymptotische Varianz-Kovarianz-Matrix

$$Asy.Var.\left(\widehat{\boldsymbol{\beta}}_{FE}^{IV}\right)=\sigma^2\boldsymbol{M}_{zx}^{-1}\boldsymbol{M}_{zz}\boldsymbol{M}_{xz}^{-1}\,. \tag{8-193}$$

Diese Varianz lässt sich schätzen mit

$$Est.Asy.Var.\left(\widehat{\boldsymbol{\beta}}^{IV}\right)=\widehat{\sigma}^2(\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{X}^a)^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Z}(\boldsymbol{X}^{a\prime}\boldsymbol{Q}\boldsymbol{Z})^{-1}\,, \tag{8-194}$$

wobei $\widehat{\sigma}^2=\frac{1}{NT}\boldsymbol{e}'\boldsymbol{Q}\boldsymbol{e}$ ist und $\boldsymbol{e}$ die Residuen bezeichnet.

9 Simulationsstudie

9.1 Simulationsdesign

Mit einer Simulationsstudie wird die Leistungsfähigkeit der IV-Schätzung bei stochastisch überlagerten Daten untersucht. Dazu wird das einfache lineare Panelmodell simuliert.

Der Störterm ϵ_{it} ist normalverteilt mit Erwartungswert Null und Varianz $\sigma_\epsilon^2=0{,}25$. α_i

ist ein Individualeffekt, der mit x_{it} korreliert ist. Die Korrelation zwischen α_i und x_{it} wird nach der Formel von Biørn (1996) mit $\lambda = 1$ erzeugt.

Um die Darstellung einfach zu halten, wird nur ein Regressor verwendet, der einem AR(1)-Prozess (vgl. (2-2)) mit $\rho_0 = 4{,}35$ und $\tau_{it} \sim N(0,\, 1)$ folgt. ρ wird variiert und nimmt folgende Werte an: $\rho = 0{,}1$, 0,5 und 0,9.

Es werden $N = 1.000$ Querschnittseinheiten simuliert, die Anzahl der Wellen des Panels ist $T = 4$ und 10. Die Anzahl der Monte-Carlo-Replikationen beträgt 500, in jeder Zeitreihe werden 200 erste Beobachtungen herausgenommen, um den Effekt des Startwertes zu eliminieren. Der wahre Parameter β sei $-2{,}5$.

Es werden die Varianten der Überlagerung betrachtet, die in den theoretischen Abschnitten beschrieben wurden: additive Überlagerung (einfache und mit Faktorstruktur) sowie multiplikative Überlagerung (einfache und mit Faktorstruktur).

Es werden drei Schätzungen durchgeführt: eine FE-Schätzung mit Originaldaten, eine FE-Schätzung mit anonymisierten Daten (naive Schätzung) sowie eine Instrumentvariablen-Schätzung. Im Fall der IV-Schätzung werden drei Instrumentvariablen getestet: (1) die verzögerte anonymisierte Variable $z_{it} = x^a_{it-1}$ (in Tabelle "verz.Var."), (2) die Variable, die aus der Differenz der verzögerten anonymisierten Variablen gebildet wurde, $z_{it} = x^a_{it} - x^a_{it-1}$ ("Dif.Var.") und (3) die Variable, die aus einem anderen anonymisierten Datensatz entnommen wurde ("anon.Var."). Im letzten Fall wird der neue Datensatz mit dem Ziel des Datenschutzes stärker überlagert. Da die Ergebnisse der Original- und der naiven Schätzungen sehr ähnlich sind, werden sie in der Tabelle nur einmal berichtet. Die Werte in Klammern sind die Standardabweichungen der Schätzer über alle Monte-Carlo-Wiederholungen.

9.2 Simulationsergebnisse

9.2.1 Additive Überlagerung

In der vorliegenden Studie wird sowohl der Regressor als auch die abhängige Variable additiv stochastisch überlagert, dabei wird sowohl die einfache als auch die Überlagerung mit Faktorstruktur getestet.

Die additiven Fehler sind normalverteilt mit $u_{it}, v_{it} \sim N(0,\, 0{,}5^2)$, sie sind miteinander unkorreliert. Im Fall der *Überlagerung mit Faktorstruktur* werden die Fehlervariablen nach Gleichung (2-4) erzeugt, wobei $\delta_i = 0{,}5$ und $\epsilon_{itx}, \epsilon_{ity} \sim N(0,\, 0{,}05^2)$ ist. Die Fehlervariablen von X und Y sind hier miteinander korreliert.

Weiter werden im Fall der IV-Schätzung mit 'anon.Var.' bei einfacher Überlagerung die Varianz $0{,}82^2$ und bei der Überlagerung mit Faktorstruktur die Parameter $\delta = 0{,}8$ und σ_{ε_x} bzw. σ_{ε_y} gleich 0,2 verwendet, was einer Gesamtvarianz des Fehlers von in etwa $0{,}82^2$ entspricht.

Tabelle 9/1 stellt die Ergebnisse für unterschiedliche (kleine, mittlere und hohe) Werte der Autokorrelation und zwei kleine T dar. Den theoretischen Ausführungen folgend zeigt die Simulation, dass die IV-Schätzungen mit *verzögerter Variable* und mit *Differenz der verzögerten Variable* als Instrumente zu verzerrten Ergebnissen führen. Die Verzerrung ist insbesondere bei der einfachen Überlagerung sehr groß. Dabei liegen die naiven Schätzer in den meisten Fällen näher bei dem wahren $\beta = -2{,}5$ als die IV-Schätzer. Die IV-Schätzer, die besonders stark vom wahren Parameter abweichen, besitzen besonders hohe Standardabweichungen. Bei der Überlagerung mit Faktorstruktur fallen die Verzerrungen wesentlich geringer aus als bei der einfachen Variante, was durch die Konstruktion der Überlagerungsvariable zu erklären ist. Für $T = 10$ bessern sich die IV-Schätzer mit zunehmender Autokorrelation, wenn 'verz.Var.' als Instrument verwendet wird. Die Verwendung dieser Variante der IV-Schätzung wird trotzdem nicht empfohlen, da die Verzerrung bei kleinem und mittleren ρ noch bestehen bleibt.

Die Verwendung der *zusätzlich anonymisierten Variable* als Instrument führt hingegen zu unverzerrten Ergebnissen. Bei der Überlagerung mit Faktorstruktur bleiben die IV-Schätzer konstant für alle Werte der Autokorrelation. Bei einfacher Überlagerung lässt sich eine kleine Verzerrung mit steigendem ρ zwar beobachten, die Schätzer bleiben aber für alle ρ-Werte sehr nahe beim wahren β.

Tabelle 9/1: Additive Überlagerung und IV-Schätzung: Simulation

	$T = 4$			$T = 10$		
	$\rho = 0,1$	$\rho = 0,5$	$\rho = 0,9$	$\rho = 0,1$	$\rho = 0,5$	$\rho = 0,9$
Original		−2,4999			−2,4999	
		(0,0045)			(0,0027)	
Einfache Überl.:						
Naiv	−1,9777	−1,9334	−1,9192	−1,9913	−2,0323	−2,1581
	(0,0211)	(0,0222)	(0,0239)	(0,0120)	(0,0120)	(0,0119)
IV: verz.Var.	−1,8746	−0,5383	−8,5320	−1,0569	−2,7243	−2,5773
	(0,0864)	(0,4290)	(8,8065)	(0,4454)	(0,0609)	(0,0229)
IV: Dif.Var.	−1,9490	−1,7612	−1,5502	−1,9559	−1,8070	−1,6530
	(0,0298)	(0,0344)	(0,0391)	(0,0178)	(0,0199)	(0,0233)
IV: anon.Var.	−2,5009	−2,5002	−2,4995	−2,4999	−2,4998	−2,4997
	(0,0340)	(0,0360)	(0,0362)	(0,0196)	(0,0182)	(0,0136)
Überl. mit Faktorstr.:						
Naiv	−2,4935	−2,4926	−2,4923	−2,4935	−2,4942	−2,4960
	(0,0052)	(0,0055)	(0,0056)	(0,0031)	(0,0029)	(0,0023)
IV: verz.Var.	−2,4931	−2,4662	−2,5140,	−2,3878	−2,5025	−2,5008
	(0,0232)	(1,2718)	(0,0396)	(0,9485)	(0,0096)	(0,0043)
IV: Dif.Var.	−2,4933	−2,4899	−2,4850	−2,4929	−2,4903	−2,4871
	(0,0078)	(0,0096)	(0,0117)	(0,0044)	(0,0053)	(0,0062)
IV: anon.Var.	−2,5001	−2,5001	−2,4999	−2,4998	−2,4998	−2,4998
	(0,0053)	(0,0056)	(0,0056)	(0,0031)	(0,0030)	(0,0024)

9.2.2 Multiplikative Überlagerung

Die Variablen werden mittels der einfachen multiplikativen Überlagerung (vgl. Gl. (2-6)) bzw. mittels der Überlagerung mit Faktorstruktur (Gl. (2-7)) anonymisiert. Die Fehlervariablen bei der *einfachen Überlagerung* sind normalverteilt mit Erwartungswert Null und Varianz $0{,}114^2$, wobei die Fehler voneinander unkorreliert sind. Bei der *Überlagerung mit Faktorstruktur* werden Parameter $\delta = 0{,}11$ und $\sigma_{\varepsilon_j} = 0{,}03$ $(j = x, y)$ verwendet.

Die Instrumentvariable 'anon.Var.' wird bei einfacher Überlagerung mit Varianz 0,206 erzeugt. Bei der Überlagerung mit Faktorstruktur wird sie mit den Parametern $\delta = 0{,}2$ und $\sigma_{\varepsilon_j} = 0{,}05$ $(j = x, y)$ generiert.

Tabelle 9/2 stellt die Simulationsergebnisse dar. Mit steigendem ρ wächst die Verzerrung der naiven FE-Schätzung, wobei sie bei der Überlagerung mit Faktorstruktur für alle Werte der Autokorrelation viel geringer ausfällt. Im Fall von kleinem und mittlerem ρ ($\rho = 0{,}1$ und 0,5) führt die IV-Schätzung nur dann zu zufriedenstellenden Ergebnissen, wenn als Instrument der Regressor aus einem anderen anonymisierten Datensatz verwendet wird. Auch für sehr kleines T erhält man IV-Schätzer, die nahe beim wahren Wert liegen. Sowohl die verzögerte als auch die Differenzvariable als Instrumente führen zu verzerrten Schätzergebnissen. Dieses Ergebnis entspricht den theoretischen Ausführungen. Weiter fällt auf, dass das dritte Instrument 'anon.Var.' bei einer hohen Autokorrelation ($\rho = 0{,}9$) Probleme aufweist. Bei der einfachen Überlagerung ($\rho = 0{,}9$ und $T = 4$) verändert der IV-Schätzer das Vorzeichen und weist eine Standardabweichung von 66,63 auf. Auch andere Simulationsstudien, über die hier nicht weiter berichtet wird, zeigten, dass die unzureichenden IV-Korrekuren bei der IV-Schätzung mit 'anon.Var.' mit sehr hohen Standardabweichungen einhergehen. Der Grund liegt darin, dass in einigen Monte-Carlo-Wiederholungen manchmal sehr starke Ausreißer (stark vom wahren Parameter abweichende Schätzwerte) resultieren, die dementsprechend den in der Tabelle ausgewiesenen durchschnittlichen Schätzer stark verzerren.

Tabelle 9/2: Multiplikative Überlagerung und IV-Schätzung: Simulation

	$T=4$			$T=10$		
	$\rho=0,1$	$\rho=0,5$	$\rho=0,9$	$\rho=0,1$	$\rho=0,5$	$\rho=0,9$
Original		−2,4999			−2,4999	
		(0,0045)			(0,0026)	
Einfache Überl.:						
Naiv	−1,8775	−1,1539	−0,0824	−1,8918	−1,3059	−0,1521
	(0,0324)	(0,0430)	(0,0473)	(0,0179)	(0,0238)	(0,0281)
IV: verz.Var.	−1,7634	−0,1674	0,0456	−0,9448	−3,7091	1,3672
	(0,1366)	(0,2805)	(0,1734)	(0,5129)	(0,3322)	(0,4983)
IV: Dif.Var.	−1,8458	−0,9376	−0,0411	−1,8521	−0,9912	−0,0489
	(0,0482)	(0,0666)	(0,0686)	(0,0283)	(0,0381)	(0,0388)
IV: anon.Var.	−2,5017	−2,5020	3,4578	−2,5001	−2,5011	−2,8006
	(0,0560)	(0,1581)	(66,6270)	(0,0329)	(0,0739)	(1,4403)
Überl. mit Faktorstr.:						
Naiv	−2,4450	−2,3159	−0,8269	−2,4458	−2,3525	−1,2135
	(0,0110)	(0,0187)	(0,0467)	(0,0066)	(0,0100)	(0,0274)
IV: verz.Var.	−2,4345	−1,2798	0,7846	−2,1640	−2,5601	−3,1477
	(0,0510)	(5,1383)	(0,3839)	(0,7830)	(0,0332)	(0,1477)
IV: Dif.Var.	−2,4419	−2,2468	−0,4942	−2,4409	−2,2626	−0,5636
	(0,0166)	(0,0317)	(0,0681)	(0,0090)	(0,0169)	(0,0382)
IV: anon.Var.	−2,4994	−2,4994	−2,5132	−2,5003	−2,5004	−2,5024
	(0,0119)	(0,0214)	(0,2468)	(0,0063)	(0,0107)	(0,0791)

9.3 Problemfälle

Die Simulationsstudie ergab, dass die (konsistente) IV-Schätzung mit zusätzlicher anonymisierter Variable in einigen Fällen problematisch sein kann, d.h. sie liefert Schätzer, die sehr stark vom wahren Schätzwert abweichen. Dieses Problem ensteht dann, wenn der Nenner des IV-Schätzers nahe bei Null liegt (vgl. auch Carroll et al. (2006), S. 133). In diesem Fall "explodiert" der gesamte Ausdruck, und es resultieren sehr hohe Schätzwerte.

Der Nenner des IV-Schätzer entspricht approximativ dem Nenner des Wahrscheinlichkeitsgrenzwertes. Aus diesem Grund wird weiter der Nenner von (6-150) bzw. (6-165) genauer untersucht.[8] Wenn man für den Regressor einen autokorrelierten Prozess erster Ordnung (AR(1)) unterstellt, lässt sich der Nenner wie folgt darstellen:

$$\begin{aligned}&\gamma_0\left(1-\frac{1}{T}\right)-\frac{1}{T^2}\gamma_0\left(2(T-1)\rho+2(T-2)\rho^2+\ldots+2\rho^{T-1}\right)\\ =&\gamma_0\left(\left(1-\frac{1}{T}\right)-\frac{1}{T^2}\left(2(T-1)\rho+2(T-2)\rho^2+\ldots+2\rho^{T-1}\right)\right)\,. \end{aligned}\tag{9-195}$$

Mithilfe einer Simulationsstudie wird überprüft, bei welchen Werten von ρ und T der Ausdruck (9-195) in Richtung Null geht. Da γ_0 sehr unterschiedlich sein kann, wird nur der Term in der Klammer von (9-195) programmiert.

Die Ergebnisse werden in Abbildung 9/1 präsentiert. In der ersten Grafik sind die Kurven für unterschiedliche Werte *positiver* Autokorrelation und in der zweiten Grafik für *negative* Autokorrelation zu sehen. Die Kurvenverläufe sind in allen Fällen sehr ähnlich. Für kleines T nähert sich die Kurve dem Wert Null. Mit steigender Wellenanzahl entfernt sie sich von 0.

Im Fall der positiven Autokorrelation verschiebt sich die in der Grafik abgebildete Funktion mit größer werdendem ρ nach unten (Richtung Nullwert). Die Werte für $\rho = 0{,}9$ und kleines T liegen besonders kritisch nahe bei Null. Je kleiner die Autokorrelation ist, desto besser werden die Ergebnisse.

Im Fall der negativen Autokorrelation verhält es sich genau umgekehrt. Am weitesten vom Nullpunkt entfernt ist die Kurve für $\rho = -0{,}9$. Die näher zu Null liegende

[8] Die folgenden Ergebnisse gelten auch für die additive Überlagerung (vgl. Gleichungen (5-82) und (5-108)).

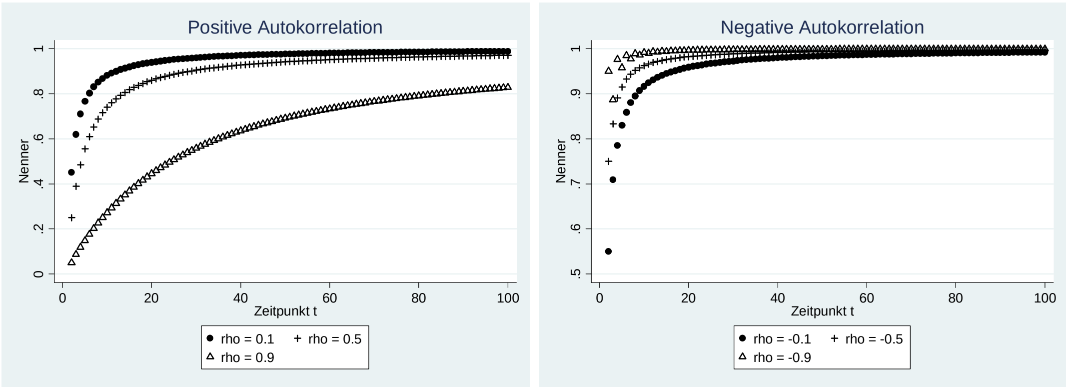

Abbildung 9/1: Verzerrung bei IV-Schätzung

Kurve ergibt sich für $\rho = -0{,}1$. Allerdings ist anzumerken, dass die Situation bei negativer Autokorrelation nicht so problematisch erscheint, da der kleinste Wert weiter von Null entfernt liegt als bei positivem ρ.

10 Empirisches Beispiel

10.1 Untersuchungsmodell

Im Folgenden werden die Auswirkungen der stochastischen Überlagerung unter Verwendung einer empirischen Datengrundlage untersucht. Dabei wird der Monatsbericht für Betriebe im Verarbeitenden Gewerbe und Bergbau (als nicht balanciertes Panel für Wellen 1995 bis 2004) verwendet. Diese Statistik ist eine Vollerhebung und umfasst alle Industriebetriebe, die mindestens 20 Personen beschäftigen.[9] Die Analyse basiert auf dem Modell von Wagner (2007), der auf Basis der Originaldaten des Monatsberichts den Zusammenhang zwischen der Exporttätigkeit und der Produktivität bei deutschen Betrieben untersucht.

Das empirische Untersuchungsmodel lautet:

$$ln\,AP_{it} = \alpha_i + \beta\,Export_{it} + \boldsymbol{Kontroll}'_{it}\boldsymbol{\gamma} + \eta_{it}\,, \qquad (10\text{-}196)$$

$$i = 1, \ldots, N, \quad t = 1, \ldots, T\,.$$

[9]2007 wurde die Abschneidegrenze auf 50 Beschäftigte angehoben.

ln AP bezeichnet die logarithmierte Arbeitsproduktivität. Die Variable *Export* ist eine Exportvariable. Es werden drei Spezifikationen der Exportvariable getestet: Modell 1 mit Exportdummy (in Tabellen als *export* bezeichnet), Modell 2 mit Exportanteil am Gesamtumsatz (*expant*) und Modell 3 mit *expant* und quadriertem Exportanteil (*expantsq*). Weitere Kontrollvariablen sind *pers* (Beschäftigtenanzahl), *perssq* (quadrierte Beschäftigtenanzahl) und *hc* (Humankapitalintensität).

Anschließend werden alle stetigen Variablen mittels der stochastischen Überlagerung anonymisiert.

10.2 Ergebnisse

10.2.1 Additive Überlagerung

Tabellen 10/3 und 10/4 präsentieren die Schätzungen für die additive stochastische Überlagerung. Es werden die FE-Schätzung mit Original- und anonymisierten Daten berechnet und anschließend die IV-Schätzung mit drei Varianten der Instrumentvariablen ausgeführt: die verzögerte Variable ("verz. Var."), die Differenz der verzögerten Variablen ("Dif.Var.") und die zusätzlich anonymisierte Variable ("anon. Var."). Die Fehlervariablen besitzen gleiche Überlagerungsparameter bei allen Variablen, d.h. $\sigma_u = \sigma_v = 0{,}5$ (einfache Überlagerung) und $\delta = 0{,}5$ und $\sigma_{\varepsilon_x} = \sigma_{\varepsilon_y} = 0{,}05$ (Überlagerung mit Faktorstruktur). Das Instrument 'anon.Var' wird mit Varianz $0{,}82^2$ bei einfacher Überlagerung und Parametern $\delta = 0{,}8$ und $\sigma_{\varepsilon_x} = \sigma_{\varepsilon_y} = 0{,}2$ bei der Überlagerung mit Faktorstruktur generiert.

Da die Modelle für Ost- und Westdeutschland (mit Ausreißer und ohne Ausreißer) zu ähnlichen Ergebnissen führten, werden hier nur Resultate für das Modell für Westdeutschland ohne Ausreißer vorgestellt.

Im Fall der einfachen Überlagerung sind alle drei IV-Schätzungen in den meisten Fällen verzerrt. Bei der Überlagerung mit Faktorstruktur sind allerdings nur die Schätzungen mit Instrumenten "verz.Var" und "Dif.Var." stark verzerrt. Die Verwendung von "anon.Var." bei der einfachen Überlagerung führt zu den Schätzergebnissen, die von der Originalschätzung abweichen. Dieses widerspricht für die zusätzliche anonymisierte Variable als Instrument den theoretischen Resultaten sowie der Simulationsstudie. Die Erklärung hierfür wird später (vgl. Abschnitt 10.3) vorgestellt.

Tabelle 10/3: Additive einfache Überlagerung und IV-Schätzung: Empirische Studie

	Original		Anonymisiert		IV verz.Var.		IV Dif.Var.		IV anon.Var.	
	$\widehat{\beta_{FE}}$	t-Wert	$\widehat{\beta_{FE}^{a}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert
Modell 1:										
export	0,070	33,29	0,077	14,53	0,071	11,73	0,074	12,69	0,066	12,39
pers	−0,040	−8,83	−0,034	−3,07	−0,091	−3,52	0,008	0,33	−0,012	−1,10
perssq	0,099	8,36	0,047	1,58	0,151	2,72	−0,045	−0,69	0,040	1,37
hc	0,167	141,18	0,165	55,90	0,293	7,40	0,131	10,73	0,166	56,41
Modell 2:										
expant	0,003	48,43	0,003	21,01	0,002	4,33	0,003	12,04	0,003	18,93
pers	−0,046	−10,23	−0,041	−3,67	−0,093	−3,62	0,006	0,26	−0,019	−1,68
perssq	0,106	9,01	0,055	1,86	0,151	2,73	−0,043	−0,66	0,048	1,64
hc	0,165	139,57	0,163	55,11	0,000	7,07	0,132	10,78	0,164	55,67
Modell 3:										
expant	0,003	23,06	0,003	10,50	0,004	4,26	0,003	4,48	0,002	7,49
expantsq	−0,025	−1,61	−0,044	−1,14	−0,330	−2,48	0,090	1,26	0,031	0,76
pers	−0,046	−10,29	−0,042	−3,71	−0,097	−3,78	0,007	0,29	−0,019	−1,64
perssq	0,107	9,06	0,056	1,89	0,160	2,89	−0,044	−0,68	0,048	1,61
hc	0,165	139,56	0,163	55,11	0,282	7,06	0,132	10,77	0,164	55,67

Tabelle 10/4: Additive Überlagerung mit Faktorstruktur und IV-Schätzung: Empirische Studie

	Original		Anonymisiert		IV verz.Var.		IV Dif.Var.		IV anon.Var.	
	$\widehat{\beta_{FE}}$	t-Wert	$\widehat{\beta_{FE}^a}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert
Modell 1:										
export	0,070	33,29	0,070	32,46	0,066	26,45	0,071	29,23	0,070	32,46
pers	−0,040	−8,83	−0,040	−8,78	−0,101	−9,41	−0,003	−0,31	−0,040	−8,78
perssq	0,099	8,36	0,101	8,30	0,200	8,64	0,019	0,73	0,101	8,30
hc	0,167	141,18	0,166	137,31	0,304	18,37	0,132	26,37	0,166	137,31
Modell 2:										
expant	0,003	48,43	0,003	47,29	0,002	10,26	0,003	26,56	0,003	47,32
pers	−0,046	−10,23	−0,047	−10,14	−0,100	−9,36	−0,008	−0,85	−0,047	−10,15
perssq	0,106	9,01	0,108	8,95	0,194	8,43	0,028	1,07	0,108	8,95
hc	0,165	139,57	0,164	135,72	0,295	17,75	0,133	26,57	0,164	135,72
Modell 3:										
export	0,003	23,06	0,003	22,07	0,003	7,73	0,003	10,70	0,003	22,16
expantsq	−0,025	−1,61	−0,017	−1,08	−0,182	−3,35	0,047	1,57	−0,019	−1,19
pers	−0,046	−10,29	−0,047	−10,18	−0,102	−9,57	−0,008	−0,81	−0,047	−10,19
perssq	0,107	9,06	0,108	8,97	0,199	8,64	0,028	1,05	0,109	8,98
hc	0,165	139,56	0,164	135,72	0,294	17,73	0,133	26,56	0,164	135,71

10.2.2 Multiplikative Überlagerung

Im Folgenden wird der Einfluss der multiplikativen stochastischen Überlagerung auf die IV-Schätzung illustriert. Die Fehlervariablen haben jetzt den Erwartungswert Eins. Bei einfacher Überlagerung sind sie miteinander unkorreliert und haben die Varianz $0{,}114^2$. Die Fehler der Überlagerung mit Faktorstruktur erhalten die üblichen Parameter $\delta = 0{,}11, \sigma_\varepsilon = 0{,}03$. Das Instrument 'anon.Var.' wird mit Varianz 0,13 (einfache Überlagerung) und $\delta = 0{,}2, \sigma_\varepsilon = 0{,}3$ generiert.

Tabellen 10/5 und 10/6 zeigen Ergebnisse für Westdeutschland ohne Ausreißer. Kurz zusammengefasst lässt sich an dieser Stelle Folgendes festhalten. Die IV-Schätzungen mit der verzögerten Variable sowie der Differenzvariable als Instrument führen zu starken Verzerrungen. Auffällig ist aber auch, dass die Schätzer bei dem Instrument 'anon.Var.' bei einigen Variablen ebenfalls stark verzerrt sind, insbesondere bei *pers*, *perssq*, *expantsq*. Bei diesen Einflussgrößen verändert sich in einigen Fällen die Richtung der Verzerrung und die Signifikanz der IV-Schätzer.

Tabelle 10/5: Multiplikative einfache Überlagerung und IV-Schätzung: Empirische Studie

	Original		Anonymisiert		IV verz.Var.		IV Dif.Var.		IV anon.Var.	
	$\widehat{\beta_{FE}}$	t-Wert	$\widehat{\beta_{FE}^{a}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert
Modell 1:										
export	0,070	33,29	0,070	31,78	0,064	26,38	0,066	27,81	0,069	31,45
pers	−0,040	−8,83	−0,013	−4,49	−0,087	−6,31	0,010	1,75	−0,019	−1,29
perssq	0,099	8,36	0,025	3,50	0,158	5,03	0,010	0,71	0,021	0,62
hc	0,167	141,18	0,158	132,13	0,267	23,66	0,147	21,93	0,170	111,53
Modell 2:										
expant	0,003	48,43	0,003	44,59	0,002	14,83	0,002	22,27	0,003	32,49
pers	−0,046	−10,23	−0,015	−5,28	−0,089	−6,52	0,009	1,58	−0,021	−1,40
perssq	0,106	9,01	0,024	3,45	0,160	5,09	0,009	0,64	0,016	0,49
hc	0,165	139,57	0,156	130,82	0,259	22,82	0,150	22,36	0,168	110,08
Modell 3:										
expant	0,003	23,06	0,002	21,04	0,005	8,37	0,002	9,83	0,003	6,84
expantsq	−0,025	−1,61	0,065	5,35	−0,338	−4,86	0,143	7,23	0,003	0,06
pers	−0,046	−10,29	−0,015	−5,20	−0,095	−6,89	0,009	1,60	−0,021	−1,40
perssq	0,107	9,06	0,024	3,42	0,171	5,42	0,009	0,65	0,016	0,49
hc	0,165	139,56	0,156	130,79	0,259	22,73	0,149	22,34	0,168	110,03

Tabelle 10/6: Multiplikative Überlagerung mit Faktorstruktur und IV-Schätzung: Empirische Studie

	Original		Anonymisiert		IV verz.Var.		IV Dif.Var.		IV anon.Var.	
	$\widehat{\beta_{FE}}$	t-Wert	$\widehat{\beta^a_{FE}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert	$\widehat{\beta^{IV}}$	t-Wert
Modell 1:										
export	0,070	33,29	0,070	32,80	0,063	27,11	0,066	28,36	0,070	32,62
pers	−0,040	−8,83	−0,031	−7,54	−0,061	−8,95	0,005	0,46	−0,024	−1,91
perssq	0,099	8,36	0,077	7,21	0,108	7,45	0,070	1,99	0,049	1,75
hc	0,167	141,18	0,165	139,53	0,269	24,81	0,158	22,73	0,163	116,77
Modell 2:										
expant	0,003	48,43	0,003	48,04	0,002	16,41	0,003	24,67	0,003	35,86
pers	−0,046	−10,23	−0,037	−8,93	−0,064	−9,48	0,000	0,03	−0,029	−2,31
perssq	0,106	9,01	0,083	7,84	0,111	7,71	0,073	2,08	0,054	1,92
hc	0,165	139,57	0,163	137,99	0,261	24,02	0,160	23,13	0,161	115,44
Modell 3:										
expant	0,003	23,06	0,003	23,56	0,004	11,81	0,003	10,39	0,003	8,02
expantsq	−0,025	−1,61	−0,028	−1,83	−0,208	−5,26	0,047	1,65	0,003	0,07
pers	−0,046	−10,29	−0,037	−8,99	−0,067	−9,81	0,001	0,05	−0,028	−2,31
perssq	0,107	9,06	0,084	7,89	0,115	8,00	0,072	2,07	0,054	1,91
hc	0,165	139,56	0,163	137,98	0,261	24,00	0,160	23,12	0,161	115,41

10.3 Problemfälle

Wie bei den Simulationen führte die empirische Untersuchung ebenfalls in einigen Fällen zu unplausiblen Ergebnissen. Aus diesem Grund wird noch eine empirische Studie (vgl. das Untersuchungsmodell in Abschnitt 10.1) durchgeführt und das Verhalten der IV-Schätzer in einzelnen Schätzungen näher angeschaut. Das Modell für Westdeutschland ohne Ausreißer wird mit der IV-Methode mit der zusätzlichen anonymisierten Variable als Instrument geschätzt, wobei ein balancierter Datensatz verwendet und die Anonymisierung 2.500 mal durchgeführt werden. Die Schätzer, die in Tabelle 10/7 ausgewiesen werden, sind die Durchschnitte der Schätzer über alle 2.500 Wiederholungen. Das Instrument wird mit den Überlagerungsparametern $\sigma_u = 0{,}539$ (einfache multiplikative Überlagerung) und $\delta = 0{,}5$ und $\sigma_\epsilon = 0{,}2$ (multiplikative Überlagerung mit Faktorstruktur) verwendet.

Tabelle 10/7: IV-Schätzung mit zusätzlich anonymisierter Variable: Empirische Studie, Westdeutschland ohne Ausreißer

		Naiv		IV	
	Original	Einfach	Faktorstr.	Einfach	Faktorstr.
expant	0,025	0,022	0,024	0,025	0,025
pers	−0,602	−0,244	−0,546	−1,384	−0,406
perssq	1,281	0,429	1,147	3,134	0,713
hc	1,666	1,589	1,661	1,662	1,667

Tabelle 10/7 zeigt die Ergebnisse. Zwar gelingt die Korrektur bei zwei Variablen *expant* (Exportanteil) und *hc* (Humankapitalintensität) sehr gut, jedoch schneidet bei *pers* (Beschäftigtenanzahl) und *perssq* (Beschäftigtenanzahl im Quadrat) die naive Schätzung wesentlich besser ab. Wenn man sich die einzelnen Schätzungen anschaut, so fällt auf, dass die Schätzwerte sehr stark streuen. Tabelle 10/8 zeigt die minimalen und maximalen Werte, die die IV-Schätzung ergab.

Sowohl bei der einfachen als auch bei der Überlagerung mit Faktorstruktur resultieren in einzelnen Monte-Carlo-Läufen IV-Schätzer, die extrem stark vom wahren Wert abweichen. Dies deutet darauf hin, dass der Nenner des Schätzers in diesen Fällen kritisch nahe bei Null liegt.

Tabelle 10/8: IV-Schätzung mit zusätzlicher anonymisierter Variable: Ausreißer

	Einfache Überlagerung		Überl. mit Faktorstruktur	
	Minimum	Maximum	Minimum	Maximum
expant	−0,217	1,850	−0,252	0,072
pers	−2450,459	329,722	−70,836	468,408
perssq	−614,619	5275,269	−1272,714	169,113
hc	−8,054	2,548	1,217	4,541

11 Zusammenfassung

Im vorliegenden Beitrag wurde die Instrumentvariablen-Schätzung als Verfahren zur Korrektur der durch die stochastische Überlagerung verursachten Verzerrung von panelökonometrischen Schätzungen analysiert. Drei mögliche Instrumentvariablen wurden betrachtet: verzögerte anonymisierte Variable, Differenz von verzögerten anonymisierten Variablen und (zusätzlich) anonymisierte Variable, die aus einem zweiten anonymisierten Datensatz stammt. Die Instrumente "verzögerte Variable" und "Differenzvariable" weisen bei allen Varianten der Überlagerung Verzerrungen auf. Das aus dem zusätzlich anonymisierten Datensatz erzeugte Instrument führt generell zu einem konsistenten Schätzer. Allerdings resultieren bei höheren Werten der Autokorrelation sowohl in Simulationen als auch in empirischen Studien manchmal starke Abweichungen der Instrumentvariablen-Schätzer von Originalwerten. Es konnte festgestellt werden, dass solche Probleme eher numerischer Natur sind. Der Grund liegt darin, dass in einigen Fällen der Nenner des konsistenten IV-Schätzers sehr nahe bei Null liegt, was dann zu sehr variablen Schätzwerten führt. Diese Probleme treten vor allem bei kleiner Wellenanzahl und hoher Autokorrelation auf.

Literatur

Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press.

Biewen, E. und G. Ronning (2008): Estimation of Linear Models with Anonymised Panel Data, in: *AStA Advances in Statistical Analysis*, 92 (4), S. 423–438.

Biørn, E. (1992): The Bias of Some Estimators for Panel Data Models with Measurement Errors, in: *Empirical Economics*, 17, S. 51–66.

Biørn, E. (1996): Panel Data with Measurement Errors, in: Mátyás, L. und P. Sevestre, (Hrsg.), *The Econometrics of Panel Data. Handbook of the Theory with Applications*, Kapitel 10, Kluwer, Dordrecht, S. 236–279.

Biørn, E. (2000): Panel Data with Measurement Errors: Instrumental Variables and GMM Procedures Combining Levels and Differences, in: *Econometric Reviews*, 19(4), S. 391–424.

Biørn, E. und T. Klette (1998): Panel Data with Errors-in-Variables: Essential and Redundant Orthogonality Conditions in GMM-Estimation, in: *Economics Letters*, 59, S. 275–282.

Biørn, E. und J. Krishnakumar (2008): Measurement Errors and Simultaneity, in: *The Econometrics of Panel Data*, Kapitel 10, Springer, S. 323–368.

Bound, J., D. Jaeger und R. Baker (1995): Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak, in: *Journal of the American Statistical Association*, 90(430), S. 443–450.

Carroll, R., D. Ruppert, L. Stefanski und C. Crainiceanu (2006): *Measurement Error in Nonlinear Models: A Modern Perspective*, Monographs on Statistics and Applied Probability 105, Chapman and Hall, London.

Greene, W. (2008): *Econometric Analysis*, Prentice Hall: Upper Saddle River, 5 Aufl.

Griliches, Z. und J. Hausman (1986): Errors in Variables in Panel Data, in: *Journal of Econometrics*, 31, S. 93–118.

Hamilton, J. (1994): *Time Series Analysis*, Princeton: Princeton University Press.

Höhne, J. (2008): Anonymisierungsverfahren für Paneldaten, in: *AStA Wirtschafts- und Sozialstatistisches Archiv*, 2 (3), S. 259–276.

Hsiao, C. (2005): *Analysis of Panel Data*, 2nd Edition, Cambridge University Press.

Ronning, G. (2009): Stochastische Überlagerung mit Hilfe einer Mischungsverteilung, in: *Institut für Angewandte Wirtschaftsforschung: IAW-Diskussionspapiere, Nr. 48, Tübingen.*

Ronning, G., R. Sturm, J. Höhne, R. Lenz, M. Rosemann, M. Scheffler und D. Vorgrimler (2005): *Handbuch der Anonymisierung wirtschaftsstatistischer Mikrodaten*, Band 4/2005 von *Statistik und Wissenschaft*, Statistisches Bundesamt.

Rosemann, M. (2006): *Auswirkungen datenverändernder Anonymisierungsverfahren auf die Analyse von Mikrodaten*, Institut für Angewandte Wirtschaftsforschung: IAW-Forschungsberichte, Nr. 66, Tübingen.

Sevestre, P. und A. Trognon (1996): Linear Models with Random Regressors, in: *Chapter 6 in L. Mátyás and P. Sevestre (eds), The Econometrics of Panel Data. A Handbook of the Theory with Applications*, Kluwer, Dordrecht.

Staiger, D. und J. Stock (1997): Instrumental Variables Regression with Weak Instruments, in: *Econometrica*, 65(3), S. 557–586.

Wansbeek, T. (2001): GMM Estimation in Panel Data Models with Measurement Error, in: *Journal of Econometrics*, 104, S. 259–268.

Wansbeek, T. und R. Konig (1991): Measurement Error and Panel Data, in: *Statistica Nederlandia*, 45, S. 85–92.

IAW-Diskussionspapiere

Bisher erschienen:

Nr. 1 (September 2001)

Das Einstiegsgeld – eine zielgruppenorientierte negative Einkommensteuer: Konzeption, Umsetzung und eine erste Zwischenbilanz nach 15 Monaten in Baden-Württemberg

Sabine Dann / Andrea Kirchmann / Alexander Spermann / Jürgen Volkert

Nr. 2 (Dezember 2001)

Die Einkommensteuerreform 1990 als natürliches Experiment. Methodische und konzeptionelle Aspekte zur Schätzung der Elastizität des zu versteuernden Einkommens

Peter Gottfried / Hannes Schellhorn

Nr. 3 (Januar 2001)

Gut betreut in den Arbeitsmarkt? Eine mikroökonomische Evaluation der Mannheimer Arbeitsvermittlungsagentur

Jürgen Jerger / Christian Pohnke / Alexander Spermann

Nr. 4 (Dezember 2001)

Das IAW-Einkommenspanel und das Mikrosimulationsmodell SIMST

Peter Gottfried / Hannes Schellhorn

Nr. 5 (April 2002)

A Microeconometric Characterisation of Household Consumption Using Quantile Regression

Niels Schulze / Gerd Ronning

Nr. 6 (April 2002)

Determinanten des Überlebens von Neugründungen in der baden-württembergischen Industrie – eine empirische Survivalanalyse mit amtlichen Betriebsdaten

Harald Strotmann

Nr. 7 (November 2002)

Die Baulandausweisungsumlage als ökonomisches Steuerungsinstrument einer nachhaltigkeitsorientierten Flächenpolitik

Raimund Krumm

Nr. 8 (März 2003)

Making Work Pay: U.S. American Models for a German Context?

Laura Chadwick, Jürgen Volkert

Nr. 9 (Juni 2003)

Erste Ergebnisse von vergleichenden Untersuchungen mit anonymisierten und nicht anonymisierten Einzeldaten am Beispiel der Kostenstrukturerhebung und der Umsatzsteuerstatistik

Martin Rosemann

IAW-Diskussionspapiere

Nr. 10 (August 2003)
Randomized Response and the Binary Probit Model
Gerd Ronning

Nr. 11 (August 2003)
Creating Firms for a New Century: Determinants of Firm Creation around 1900
Joerg Baten

Nr. 12 (September 2003)
Das fiskalische BLAU-Konzept zur Begrenzung des Siedlungsflächenwachstums
Raimund Krumm

Nr. 13 (Dezember 2003)
Generelle Nichtdiskontierung als Bedingung für eine nachhaltige Entwicklung?
Stefan Bayer

Nr. 14 (Februar 2003)
Die Elastizität des zu versteuernden Einkommens. Messung und erste Ergebnisse zur empirischen Evidenz für die Bundesrepublik Deutschland.
Peter Gottfried / Hannes Schellhorn

Nr. 15 (Februar 2004)
Empirical Evidence on the Effects of Marginal Tax Rates on Income – The German Case
Peter Gottfried / Hannes Schellhorn

Nr. 16 (Juli 2004)
Shadow Economies around the World: What do we really know?
Friedrich Schneider

Nr. 17 (August 2004)
Firm Foundations in the Knowledge Intensive Business Service Sector. Results from a Comparative Empirical Study in Three German Regions
Andreas Koch / Thomas Stahlecker

Nr. 18 (Januar 2005)
The impact of functional integration and spatial proximity on the post-entry performance of knowledge intensive business service firms
Andreas Koch / Harald Strotmann

Nr. 19 (März 2005)
Legislative Malapportionment and the Politicization of Germany's Intergovernmental Transfer System
Hans Pitlik / Friedrich Schneider / Harald Strotmann

Nr. 20 (April 2005)
Implementation ökonomischer Steuerungsansätze in die Raumplanung
Raimund Krumm

Nr. 21 (Juli 2005)
Determinants of Innovative Activity in Newly Founded Knowledge Intensive Business Service Firms
Andreas Koch / Harald Strotmann

Nr. 22 (Dezember 2005)
Impact of Opening Clauses on Bargained Wages
Wolf Dieter Heinbach

Nr. 23 (Januar 2006)
Hat die Einführung von Gewinnbeteiligungsmodellen kurzfristige positive Produktivitätswirkungen? – Ergebnisse eines Propensity-Score-Matching-Ansatzes
Harald Strotmann

Nr. 24 (März 2006)
Who Goes East? The Impact of Enlargement on the Pattern of German FDI
Claudia M. Buch / Jörn Kleinert

Nr. 25 (Mai 2006)
Estimation of the Probit Model from Anonymized Micro Data
Gerd Ronning / Martin Rosemann

Nr. 26 (Oktober 2006)
Bargained Wages in Decentralized Wage-Setting Regimes
Wolf Dieter Heinbach

Nr. 27 (Januar 2007)
A Capability Approach for Official German Poverty and Wealth Reports: Conceptual Background and First Empirical Results
Christian Arndt / Jürgen Volkert

Nr. 28 (Februar 2007)
Typisierung der Tarifvertragslandschaft – Eine Clusteranalyse der tarifvertraglichen Öffnungsklauseln
Wolf Dieter Heinbach / Stefanie Schröpfer

Nr. 29 (März 2007)
International Bank Portfolios: Short- and Long-Run Responses to the Business Cycles
Sven Blank / Claudia M. Buch

Nr. 30 (April 2007)
Stochastische Überlagerungen mit Hilfe der Mischungsverteilung
Gerd Ronning

Nr. 31 (Mai 2007)
Openness and Growth: The Long Shadow of the Berlin Wall
Claudia M. Buch / Farid Toubal

IAW-Diskussionspapiere

Nr. 32 (Mai 2007)
International Banking and the Allocation of Risk
Claudia M. Buch / Gayle DeLong / Katja Neugebauer

Nr. 33 (Juli 2007)
Multinational Firms and New Protectionisms
Claudia M. Buch / Jörn Kleinert

Nr. 34 (November 2007)
Within-Schätzung bei anonymisierten Paneldaten
Elena Biewen

Nr. 35 (Dezember 2007)
What a Difference Trade Makes – Export Activity and the
Flexibility of Collective Bargaining Agreements
Wolf Dieter Heinbach / Stefanie Schröpfer

Nr. 36 (Dezember 2007)
To Bind or Not to Bind Collectively? Decomposition of
Bargained Wage Differences Using Counterfactual Distributions
Wolf Dieter Heinbach / Markus Spindler

Nr. 37 (Dezember 2007)
Neue Ansätze zur flächenschutzpolitischen Reform des Kommunalen
Finanzausgleichs
Raimund Krumm

Nr. 38 (Januar 2008)
Banking Globalization: International Consolidation and Mergers in Banking
Claudia M. Buch / Gayle L. DeLong

Nr. 39 (Januar 2008)
Multiplicative Measurement Error and the Simulation Extrapolation Method
Elena Biewen / Sandra Nolte / Martin Rosemann

Nr. 40 (Juni 2008)
Das Konzept des „Regionalen Gewerbeflächenpools" aus ökonomischer Sicht
Raimund Krumm

Nr. 41 (Juli 2008)
Openness and Income Disparities: Does Trade Explain the 'Mezzogiorno' Effect?
Claudia M. Buch / Paola Monti

Nr. 42 (August 2008)
Flächenschutzpolitische Implikationen eines Regionalen Gewerbeflächenpools
Raimund Krumm

Nr. 43 (September 2008)
Mikroökonomische Determinanten und Effekte von FDI am Beispiel Baden-Württemberg
Christian Arndt / Anselm Mattes

Nr. 44 (September 2008)
The Impact of Macroeconomic Factors on Risks in the Banking Sector: A Cross-Country Empirical Assessment
Olga Bohachova

Nr. 45 (Oktober 2008)
Effects of Dismissal Protection Legislation on Individual Employment Stability in Germany
Bernhard Boockmann / Daniel Gutknecht / Susanne Steffes

Nr. 46 (November 2008)
Trade's Impact on the Labor Share: Evidence from German and Italian Regions
Claudia M. Buch / Paola Monti / Farid Toubal

Nr. 47 (März 2009)
Network and Border Effects: Where Do Foreign Multinationals Locate in Germany?
Julia Spies

Nr. 48 (März 2009)
Stochastische Überlagerung mit Hilfe der Mischungsverteilung
(Stand: 18. März 2009 – Version 49)
Gerd Ronning

Nr. 49 (April 2009)
Außenwirtschaftliche Verbindungen der deutschen Bundesländer zur Republik Österreich
Anselm Mattes /Julia Spies

Nr. 50 (Juli 2009)
New Firms – Different Jobs? An Inquiry into the Quality of Employment in Start-ups and Incumbents
(Stand: 28. Juli 2009 – Version 1.3)
Andreas Koch / Jochen Späth

Nr. 51 (Juli 2009)
Poverty and Wealth Reporting of the German Government: Approach, Lessons and Critique
Christian Arndt / Jürgen Volkert

Nr. 52 (August/September 2009)
Barriers to Internationalization: Firm-Level Evidence from Germany
Christian Arndt / Claudia M. Buch / Anselm Mattes

IAW-Diskussionspapiere

Nr. 53 (September 2009)
IV-Schätzung eines linearen Panelmodells mit stochastisch überlagerten Betriebs- und Unternehmensdaten
Elena Biewen / Gerd Ronning / Martin Rosemann