Kapitel 10

Endogenität und Instrumentvariablen

> *"The whole problem with the world is that fools and fanatics are always so certain of themselves, but wiser people are so full of doubts."* (Bertrand Russell)

10.1 Einleitung

Bei Konferenzen ist es gefürchtet, das berühmte *'E'*-Wort, die Frage nach der *Endogenität*. Warum? Probleme mit der Endogenität sind in Sozialwissenschaften omnipräsent, die Konsequenzen können dramatisch sein, und sie erscheint uns häufig schwer fassbar. Wir haben nur wenige Waffen dagegen, und diese sind meist nur bedingt einsetzbar, schwer handhabbar, und obendrein häufig ungenau. Nicht selten sind die gegen die Endogenität vorgeschlagenen Maßnahmen schlimmer als die Krankheit selbst. Aber richtig angewandt versprechen sie die Lösung eines unserer Urprobleme, Kausalitätsaussagen.

Wir haben bereits gezeigt, dass die OLS Schätzer unter den Gauss Markov Annahmen BLUE (*Best Linear Unbiased Estimators*) sind, also erwartungstreu und effizient. Zur Wiederholung fassen wir die Gauss Markov Annahmen noch einmal kurz zusammen

A1 Der Zusammenhang zwischen erklärenden und abhängiger Variable ist in der Grundgesamtheit linear, d.h.

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, \ldots, n$$

und korrekt spezifiziert.

A2 Die $\boldsymbol{X}$ Matrix hat vollen Spaltenrang (keine perfekte Multikollinearität). Dies ist eine notwendige Bedingung für die Identifizierbarkeit der interessierenden Parameter, deshalb können wir darunter etwas breiter eine Identifikationsbedingung verstehen. Zur Erinnerung, etwas salopp gesprochen ist ein Modell eindeutig identifiziert, wenn aus den Daten (bzw. deren gemeinsame Verteilung) die interessierenden Parameter ermittelt werden können (d.h. eine eindeutige Lösung existiert).

A3 Die erklärenden x Variablen und die Störterme ε sind stochastisch unabhängig

$$\mathrm{E}(\varepsilon_i|\boldsymbol{X}) = \mathrm{E}(\varepsilon_i) = 0$$

Falls diese Annahme erfüllt ist sprechen wir von exogenen Regressoren, anderenfalls von *endogenen Regressoren.*

A4 Die Störterme sind

$$\varepsilon_i|\boldsymbol{X} \sim \text{i.i.d.}(0, \sigma^2)$$

das heißt, die Störterme haben einen Erwartungswert Null ($\mathrm{E}(\varepsilon_i) = 0$) und sind weder heteroskedastisch noch autokorreliert.

Die entscheidende Annahme für die Erwartungstreue der OLS Schätzfunktion sind die Annahmen A1 – A3, korrekte Spezifikation, Identifizierbarkeit, und stochastische Unabhängigkeit zwischen Regressoren und Störtermen. Wir erinnern uns, um die Erwartungstreue zu zeigen setzten wir den 'wahren Zusammenhang' (die korrekt spezifizierte PRF) in die Schätzfunktion ein und bildeten den Erwartungswert

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\right] = \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right] \\
&= \boldsymbol{\beta} + \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right]
\end{aligned}$$

Für deterministische Regressoren und korrekt spezifizierte Modelle gilt $\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}\left[\boldsymbol{\varepsilon}\right] = \boldsymbol{0}$, d.h., wenn A1 – A3 erfüllt sind ist die OLS Schätzfunktion erwartungstreu, $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

Diese Annahme deterministischer Regressoren ist in experimentellen Zusammenhängen häufig vernünftig, in Experimenten können die erklärenden x Variablen kontrolliert und von der Forscherin festgelegt werden.

Bei Beobachtungsdaten haben wir in der Regel keine Kontrolle über die erklärenden x Variablen; die Vorstellung ist eher, dass wir aus einer Grundgesamtheit zufällig $(\boldsymbol{x}_i', y_i)$ Paare ziehen(bzw., dass der datengenerierende Prozess insgesamt n i.i.d. verteilte $(\boldsymbol{x}_i', y_i)$ Paare erzeugt). Deshalb ist die Annahme deterministischer x Variablen in den Sozialwissenschaften häufig unrealistisch.

Glücklicherweise bleiben die meisten angenehmen Eigenschaften des OLS Schätzers auch für stochastische Regressoren erhalten, allerdings benötigt man dafür ein leicht erweitertes Annahmenset, bzw. mindestens eine zusätzliche Annahme

A5 Die $(\boldsymbol{x}_i', y_i)$ Paare $(i = 1, \ldots, n)$ sind das Resultat von Zufallsziehungen (*'random sampling'*) aus einer gemeinsamen Verteilung. Wenn die Ziehungen unabhängig sind und aus einer gemeinsamen Verteilung gezogen wird sind die Daten i.i.d., d.h. *'independent and identically distributed'*.

Darüber hinaus sollen große Ausreißer unwahrscheinlich sein, technisch ausgedrückt, die vierten Momente von x und y müssen existieren und dürfen nicht unendlich groß sein

$$0 < \mathrm{E}(x_{ih}^4) < \infty, \qquad 0 < \mathrm{E}(y_i^4) < \infty$$

Im wesentlichen stellt diese Annahme sicher, dass einzelne Beobachtungen keinen 'zu großen' Einfluss haben.

Die Erwartungstreue von Schätzfunktionen kann für stochastische Regressoren für komplexe Schätzfunktionen manchmal nicht bewiesen werden, aber wenn die Annahmen A1 bis A5 erfüllt sind können 'Gesetze der Großen Zahl' angewandt werden, was in günstigen Fällen den Beweis der 'Konsistenz' ermöglicht. Dazu werden wir im Folgenden öfter das *'probability limit'* (plim) benötigen.[1]

Wenn die Annahmen A1 bis A5 erfüllt sind und die Störterme zusätzlich normalverteilt sind, dann ist auch die auf die x Variablen bedingte Stichprobenkennwertverteilung der OLS Schätzer normalverteilt, und deren Varianzen können wie bisher geschätzt werden. Auch alle Konfidenzintervalle und Hypothesentests bleiben gültig.

Wenn die Störterme *nicht* normalverteilt, sondern nur i.i.d. sind, gelten alle Hypothesentests asymptotisch. Das heißt, wenn die Annahmen A1 bis A5 erfüllt sind, sind die OLS Schätzer auch bei stochastischen Regressoren konsistent und asymptotisch normalverteilt.

Etwas technischer kann dies folgendermaßen zusammengefasst werden: Falls die Annahmen A1 bis A5 erfüllt sind und insbesondere

$$\operatorname{plim}\left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right) = \boldsymbol{\Sigma}_{xx}$$

existiert und nicht unendlich groß ist gilt

1. Die OLS Schätzer sind konsistent, d.h. $\operatorname{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, bzw. $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$
2. Die mit $\sqrt{n}$ skalierte Differenz zwischen $\hat{\boldsymbol{\beta}}$ und $\boldsymbol{\beta}$ konvergiert gegen die Normalverteilung
$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(0, \sigma^2 \boldsymbol{\Sigma}_{xx}^{-1}\right)$$
3. Falls die Störterme normalverteilt sind bleiben die üblichen Hypothesentests auch in kleinen Stichproben gültig, anderenfalls gelten sie asymptotisch.

Die entscheidende Annahme für die Konsistenz des OLS Schätzers ist, dass die erklärenden x Variablen nicht mit den Störtermen ε korreliert sind, oder etwas allgemeiner, dass sie stochastisch unabhängig sind ($\mathrm{E}(\varepsilon_i | x_{\cdot 1}, \ldots, x_{\cdot k}) = 0$).

Da diese Annahme die unbeobachtbaren Störterme der Grundgesamtheit betrifft, kann sie empirisch nicht einfach überprüft werden (die Stichprobenresiduen $\hat{\varepsilon}_i$ sind *immer* unkorreliert mit den Regressoren, warum?).

Wir werden gleich sehen, dass endogene Regressoren in den Sozialwissenschaften eher die Regel als die Ausnahme sind. Da in diesen Fällen OLS Schätzer *weder erwartungstreu noch konsistent* sind, spielt diese Annahme in der Ökonometrie eine ganz zentrale Rolle, ja, man kann sogar sagen, dass die Ökonometrie als eigenständige Wissenschaft ganz wesentlich aus der Auseinandersetzung mit diesem Problem entstanden ist.

Um eine intuitive Vorstellung von den Problemen zu bekommen erinnern wir uns, dass durch eine Regression eine zu erklärende Variable y gewissermaßen in zwei

[1] Man kann zeigen, dass der OLS Schätzer auch bei stochastischen Regressoren unverzerrt ist, wenn die Annahmen A1 bis A5 erfüllt sind. Dies gilt nicht für alle Schätzfunktionen.

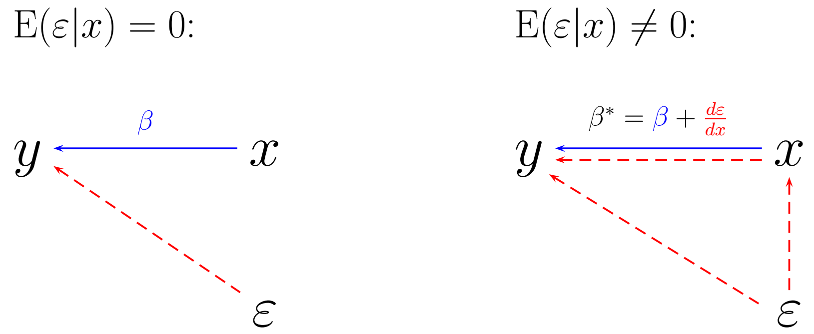

Abbildung 10.1: Wenn $\mathrm{E}(\varepsilon_i|x_1,\ldots,x_n) = 0$ (links) ist der OLS Schätzer unverzerrt, β misst den Effekt von x auf y. Falls $\mathrm{E}(\varepsilon_i|x_1,\ldots,x_n) \neq 0$ (rechts) ist der OLS Schätzer verzerrt; der messbare Parameter β^* misst den direkten Effekt β *und* indirekten Effekt $d\varepsilon/dx$ gemeinsam, der interessierende Parameter β ist in diesem Fall *nicht identifiziert*.

Teile zerlegt wird, in einen durch die Regressoren erklärten (systematischen) Teil, und in einen unerklärten Teil, der im Störterm abgebildet wird. Dies ist im linken Panel von Abbildung 11.1 dargestellt; der zu schätzende Koeffizient β_2 beschreibt gewissermaßen die 'Stärke' des Zusammenhangs zwischen y und x.

Falls aber die erklärende Variable x mit dem Störterm ε korreliert ist, wie im rechten Panel von Abbildung 11.1 dargestellt, wird der eigentlich interessierende Zusammenhang zwischen x und y durch den zusätzlichen Zusammenhang zwischen x und ε gewissermaßen 'verschmutzt'.

Wenn ein Regressor x mit ε korreliert ist können wir uns vorstellen, ε ist eine Funktion von x, das heißt $\varepsilon = \varepsilon(x)$. In diesem Fall misst β_2 nicht mehr den marginalen Effekt von x, da

$$y = \beta_1 + \beta_2 x + \varepsilon(x)$$

In diesem Fall gibt es einen direkten und indirekten Zusammenhang zwischen x und y

$$\frac{dy}{dx} = \beta_2 + \frac{d\varepsilon}{dx} := \beta_2^*$$

Der OLS Schätzer misst den gemeinsamen Einfluss $\beta_2^* := \beta_2 + \frac{d\varepsilon}{dx}$ anstelle des meist interessierenden marginalen Effekts β_2, der OLS Schätzer ist systematisch verzerrt.

Man kann dies auch als Identifikationsproblem verstehen, mit einem endogenen Regressor x können wir nur den Gesamteffekt $\beta_2 + \frac{d\varepsilon}{dx}$ identifizieren, nicht aber den einzelnen interessierenden Effekt β_2.

Dies kann einfach am Beispiel 'fehlender relevanter Variablen' demonstriert werden. Angenommen der datengenerierende Prozess (DGP) könne durch

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

beschrieben werden und wir interessieren uns für β_2. Außerdem nehmen wir an x_3 könne nicht beobachtet werden, aber die fehlende Variable x_3 sein eine lineare Funktion der beobachteten Variable x_2

$$x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$$

Wenn wir dies in den DGP einsetzen erhalten wir

$$y_i = \beta_1 + \beta_2 x_{i2} + \underbrace{[\beta_3(\delta_1 + \delta_2 x_{i2} + \upsilon_i) + \varepsilon_i]}_{\varepsilon^*}$$

mit dem marginalen Effekt

$$\frac{\partial y}{\partial x_2} = \beta_2 + \frac{\partial \varepsilon^*}{\partial x_2} = \beta_2 + \beta_3 \delta_2$$

Wann immer der Regressor x mit dem Störterm ε korreliert ist gibt es *ohne zusätzlicher Information* keine Möglichkeit den Koeffizienten β_2 unverzerrt zu schätzen! Da diese Korrelation zwischen x und ε in der Regel auch mit zunehmender Stichprobengröße nicht verschwinden wird, ist der OLS Schätzer im Fall eines endogenen Regressors auch nicht konsistent.

10.2 Ursachen von Endogenität

Da endogene Regressoren derart schwerwiegende Probleme aufwerfen stellt sich als nächstes die Frage, inwieweit dieses Problem relevant ist, das heißt, inwieweit wir damit rechnen müssen, mit diesem Problem bei unserer praktischen Arbeit konfrontiert zu werden. Leider stellt sich heraus, dass dieses Problem alltäglich ist, und bei fast allen empirischen Arbeiten im sozialwissenschaftlichen Umfeld berücksichtigt werden muss.

Die wichtigsten Fälle, die zu endogenen Regressoren führen, sind

1. Nichtberücksichtigung zumindest einer relevanten erklärenden Variable, die mit einem Regressor korreliert ist (*'omitted variables'*).
2. Simultane Kausalität, das heißt, wenn zur Beschreibung eines Zusammenhangs mehr als eine Gleichung benötigt wird, und dies zu *feed-back* Mechanismen führt (*'reverse causality'*).
3. Messfehler in der erklärenden x Variable.
4. Autokorrelation mit verzögerten endogenen Variablen.

Im Folgenden wollen wir diese Fälle etwas ausführlicher darstellen.

10.2.1 Nichtberücksichtigung relevanter Variablen (*Omitted Variables*)

Wir haben bereits früher gezeigt, dass die Nichtberücksichtigung relevanter Variablen zu einem Endogenitätsbias führt. Wenn der datengenerierende Prozess durch das 'wahre' Modell

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad \varepsilon_i \sim \text{i.i.d.}(0, \sigma^2)$$

beschrieben beschrieben wird, wir aber fälschlich ein 'kurzes' Modell

$$y_i = \widehat{\beta}_1^* + \widehat{\beta}_2^* x_{i2} + \hat{\varepsilon}_i^*$$

dann ist

$$\mathrm{E}(\widehat{\beta}_2^*) = \beta_2 + \beta_3 * \delta_2$$

wobei δ_2 den linearen Zusammenhang zwischen dem Regressor x_2 und der nicht berücksichtigten Variable x_3 in der Form

$$x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$$

beschreibt.

Wann immer β_3 *und* δ_2 von Null verschieden sind ist der OLS Schätzer $\widehat{\beta}_2^*$ des 'kurzen' Modells verzerrt!

Der Grund ist einfach, wenn wir fälschlich das 'kurze' Modell schätzen ist x_3 im Störterm enthalten, d.h.

$$y_i = \widetilde{\beta}_1 + \widetilde{\beta}_2 x_{i2} + \widetilde{\varepsilon}_i, \quad \text{mit} \quad \widetilde{\varepsilon}_i = \beta_3 x_{i3} + \varepsilon_i$$

Wenn $\mathrm{cov}(x_2, x_3) \neq 0$ und $\beta_3 \neq 0$ führt dies zu einer Korrelation zwischen x_2 und $\widetilde{\varepsilon}$, also Endogenität. Diesen Fall haben wir bereits im Kapitel zur deskriptiven Regressionsanalyse diskutiert.

Man beachte, dass in diesem Fall eine *kausale Interpretation* von $\widetilde{\beta}_2$ als marginaler Effekt falsch wäre, durch die Fehlspezifikation (d.h. Weglassen von x_3) misst $\widehat{\beta}_2$ den direkten *und* indirekten Effekt anstelle des interessierenden marginaler ceteris paribus Effekts von x_2 (d.h. β_2).

Achtung: Falls in einem multiplen Regressionsmodell

$$y_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_{i2} + \widehat{\beta}_3 x_{i3} + \ldots + \widehat{\beta}_k x_{ik} + \hat{\varepsilon}_i$$

ein einziger Regressor x_h endogen ist (d.h. mit dem Störterm korreliert ist), werden in der Regel *alle* Koeffizienten $\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_k$ verzerrt geschätzt! Dies ist wenig verwunderlich, da meist alle Regressoren – und damit auch die Koeffizienten – untereinander korreliert sind.

Hinweis: Während 'fehlende relevante Variablen' ziemlich verheerende Folgen für die Schätzung der Koeffizienten haben kann sind die Auswirkungen häufig weniger dramatisch, wenn das Ziel eine *Prognose* von $\widehat{y}$ ist. Wenn die fehlende Variable eine lineare Funktion der berücksichtigten Variablen ist sind die Vorhersagen sogar unverzerrt (Greene, 2003, 30).

Die Intuition ist einfach: angenommen der DGP sei $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, aber wir schätzen irrtümlich das 'kurze' Modell $y_i = \widehat{\beta}_1^* + \widehat{\beta}_2^* x_{i2} + \hat{\varepsilon}_i^*$, und die fehlende Variable x_3 sei eine lineare Funktion von x_2, d.h. $x_{i3} = \delta_1 + \delta_2 x_{i2} + \upsilon_i$.

Wenn wir in das korrekte 'lange' Model einsetzen erhalten wir

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \\ &= \beta_1 + \beta_2 x_{i2} + \beta_3(\delta_1 + \delta_2 x_{i2} + \upsilon_i) + \varepsilon_i \\ &= \underbrace{\beta_1 + \beta_3 \delta_1}_{\beta_1^*} + \underbrace{(\beta_2 + \beta_3 \delta_2)}_{\beta_2^*} x_{i2} + \underbrace{\beta_3 \upsilon_i + \varepsilon_i}_{\varepsilon_i^*} \\ &= \beta_1^* + \beta_2^* x_{i2} + \varepsilon_i^* \end{aligned}$$

Wenn wir diese fehlspezifizierte Gleichung schätzen erhalten wir $y_i = \widehat{\beta}_1^* + \widehat{\beta}_2^* x_{i2} + \hat{\varepsilon}_i^*$ mit

$$\mathrm{E}(y_i) = \mathrm{E}(\beta_1^* + \beta_2^* x_{i2} + \varepsilon_i^*) = (\beta_1 + \beta_3\delta_1) + (\beta_2 + \beta_3\delta_2)x_{i2}$$

Die Koeffizienten des 'kurzen' Modells haben zwar keine kausale Interpretation, aber für die Vorhersage 'korrigieren' sie gewissermaßen für die fehlende Variable x_3. Natürlich ist die Vorhersage mit dem 'kurzen' Modell nicht effizient, wie man einfach am Störterm $\varepsilon_i^* = \beta_3 \upsilon_i + \varepsilon_i$ erkennt.

10.2.2 Simultane Kausalität

Bisher haben wir stets eine sehr einfache Form von Abhängigkeit zwischen der erklärenden Variable x und der abhängigen Variablen y angenommen, nämlich dass dass die x Variable unmittelbar y beeinflusst.

Was passiert aber, wenn die Variablen auf komplexere Weise verknüpft sind, so dass wir zur Beschreibung mindestens *zwei oder mehrere Gleichungen* benötigen? Dies ist in den Sozialwissenschaften eher die Regel als die Ausnahme, selbst zur Beschreibung des einfachsten Marktes benötigen wir bereits eine Nachfrage- *und* Angebotsfunktion!

Wir werden nun zeigen, dass eine solche gegenseitige Abhängigkeit, zu deren Beschreibung mehr als eine Gleichung benötigt wird, ebenfalls zu einer Korrelation zwischen dem Störterm und den erklärenden x Variablen führt.

Der einfachste Fall einer solchen Korrelation zwischen Regressor und Störterm kann einfach anhand des keynesianischen Eingaben-Ausgaben Modells mit einer stochastischen Konsumfunktion und Einkommensidentität erläutert werden.

Das der Abbildung 11.2 zugrunde liegende datengenerierende Prozess ist

$$\begin{aligned} C_i &= 60 + 0.5Y_i + \varepsilon_i \qquad \text{mit } \varepsilon_i \sim \mathrm{U}(-30, +30) \\ Y_i &= C_i + I_i \qquad \text{mit } I_i \sim \mathrm{U}(30, 80) \end{aligned}$$

wobei U hier für die Gleichverteilung steht, $I_i \sim \mathrm{U}(30, 80)$ bedeutet also, dass die Variable I_i mit gleicher Wahrscheinlichkeit irgendeinen Wert zwischen 30 und 80 annimmt. Die Gleichverteilung wurde nur gewählt um eine übersichtlichere Grafik zu erhalten und spielt ansonsten keine Rolle.

Abbildung 11.2 zeigt diese Funktionen für die Extremwerte von ε_i und I_i (d.h. $C_i = 60 + 0.5Y_i - 30$ und $C_i = 60 + 0.5Y_i + 30$, bzw. $C_i = Y_i - 30$ und $C_i = Y_i - 80$; die strichlierten Linen zeigen die Funktionen für $\varepsilon_i = 0$ und $I_i = 0$).

Die Daten wurden diesem Modell entsprechend vom Computer erzeugt. Eine einfache OLS Schätzung der Konsumfunktion liefert statt der wahren Werte $\widehat{\beta}_1 = 60$ und $\widehat{\beta}_2 = 0.5$ völlig andere Werte.

$$\begin{aligned} \mathrm{C} &= \underset{(5.685)}{-9.175} + \underset{(0.024)^{***}}{0.808}\ \mathrm{Y} \end{aligned}$$

$$R^2 = 0.934, \quad s = 10.818, \quad F\text{-Stat} = 1096.971, \quad n = 80$$

(Standardfehler in Klammern)

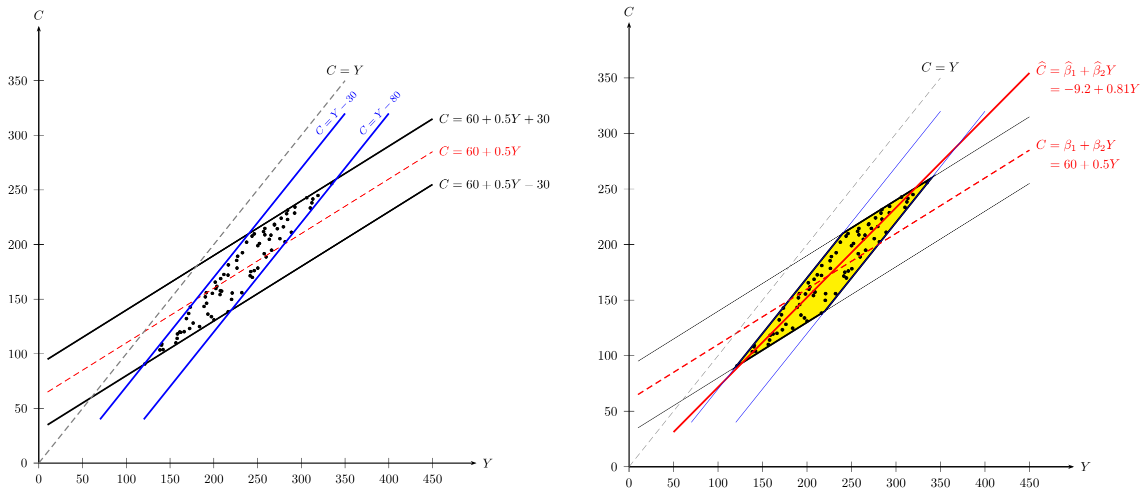

Abbildung 10.2: Ein einfaches Keynesianisches Modell.

Abbildung 11.2 zeigt, was schief gelaufen ist, die Einkommensidentität $Y = C + I$ erzwingt, dass alle Realisationen im gelben Parallelogramm liegen müssen!

Wie Abbildung 11.3 zeigt führt dies zu einer Korrelation zwischen den Störtermen und der erklärenden Variable Y, denn ein in einer Periode zufällig größerer Störterm führt zu erhöhten Konsumausgaben, diese über die Identität zu einem höheren Einkommen, weshalb Einkommen und Störterme korreliert sind.

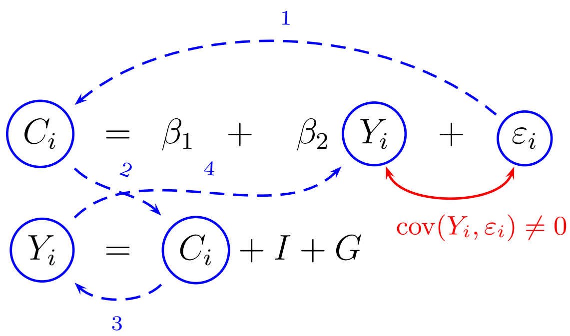

Abbildung 10.3: Simultanität im keynesianischen Eingaben-Ausgaben Modell [local, www]

Wie das rechte Panel von Abbildung 11.2 zeigt, liefert die OLS Gerade zwar eine optimale Beschreibung der realisierten Daten, aber sie liefert keine konsistente Schätzung der Konsumfunktion, weil die Abhängigkeit zwischen Konsum und Einkommen *über die Einkommensidentität* nicht berücksichtigt wird!

Die OLS Koeffizienten $\widehat{\beta}_1$ und $\widehat{\beta}_2$ liefern deshalb keine unverzerrten Schätzer für den autonomen Konsum β_1 und die marginale Konsumneigung β_2! Da dieses Problem unabhängig von der Stichprobengröße existiert wird das Problem mit zunehmender Stichprobengröße auch nicht kleiner, deshalb ist der OLS Schätzer in diesem Fall auch nicht konsistent!

Eine einfache Monte Carlo Simulation soll das Problem wieder veranschaulichen. Das folgende kleine EViews Programm erzeugt tausend Mal ein Modell mit zwei

Gleichungen, einer Konsumfunktion und einer Identität

$$\begin{aligned} \mathrm{C}_i &= 60 + 0.5Y + \varepsilon_i \\ Y_i &= \mathrm{C}_i + Z_i \end{aligned}$$

mit $\varepsilon_i \sim \mathrm{U}(-30, +30)$ und $Z_i \sim \mathrm{U}(30, 80)$, wobei Z alle exogenen Nachfragekategorien enthält (zum Beispiel Investitionen und Staatsausgaben).

Das folgende EViews-Programm löst dieses Modell bei jedem der tausend Durchgänge neu und schätzt jedes Mal *aus der Lösungen* die marginale Konsumneigung. Man beachte, dass die 'wahre' marginale Konsumneigung mit 0.5 vorgegeben wurde.

```
wfcreate u 25
rndseed 123456789
!REP = 1000
vector(!REP) MCP = na
series Cons
series Y
series Z
for !r = 1 to !REP
Z = @runif(30,80)
    ' Modell lösen
    model Keynes
    Keynes.append  Cons = 60 + 0.5*Y + @runif(-30,30)
    Keynes.append  Y = Cons + Z
    Keynes.solve
    delete Keynes
    ' Konsumfunktion mit Modelllösungen schätzen
    equation temp.ls Cons_0 c Y_0
    MCP(!r) = c(2)
next
MCP.distplot hist
```

Abbildung 11.4 zeigt das Ergebnis, offensichtlich ist die Schätzung der marginalen Konsumneigung systematisch verzerrt, obwohl der wahre Wert der marginalen Konsumneigung mit 0.5 vorgegeben wurde, liegt der Mittelwert der empirischen Stichprobenkennwertverteilung des OLS Schätzers bei 0.8 (bei tausend Replikationen).

Da die Sozialwissenschaften in der Regel hochkomplexe *interdependente Systeme* untersuchen tritt dieser Fall sehr häufig auf.

Im Kern tritt diese Problem immer dann auf, wenn zwei Variablen auf *mehr* als eine Art verknüpft sind, das heißt, wenn zur Beschreibung des Verhaltens einer Variable mehr als eine Gleichung benötigt wird. Das klassische Beispiel sind natürlich Angebots- und Nachfragefunktionen oder Makromodelle, aber ähnliche Probleme existieren fast überall in den Sozialwissenschaften.

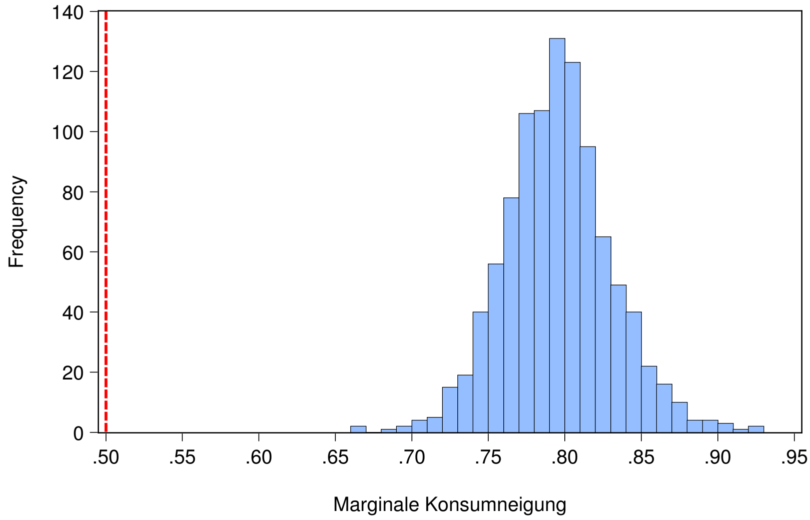

Abbildung 10.4: Monte Carlo Simulation für endogenen Regressor. Der wahre Wert der marginalen Konsumneigung ist 0.5; aufgrund des *feedbacks* über die Einkommensidentität überschätzt OLS diesen Wert systematisch.

10.2.3 Messfehler in den erklärenden (x) Variablen

Eine weitere mögliche Ursache für Endogenität sind Messfehler in mindestens einer x Variablen. Um dies zu zeigen nehmen wir an, dass anstelle des 'wahren' Wertes[2] $\ddot{x}_i$ nur ein fehlerhaft gemessenes

$$\ddot{x}_i^* = \ddot{x}_i + \upsilon_i$$

beobachtet werden kann ($\ddot{x}_i := x_i - \bar{x}$ bezeichne wieder Abweichungen vom Mittelwert), mit $\upsilon_i \sim$ i.i.d.$(0, \sigma^2_{\upsilon_i})$.

Das wahre (und unbeobachtbare) Regressionsmodell sei

$$\begin{aligned} \ddot{y}_i &= \beta\ddot{x}_i + \varepsilon_i \\ &= \beta(\ddot{x}_i^* - \upsilon_i) + \varepsilon_i \\ &= \beta\ddot{x}_i^* + (\varepsilon_i - \beta\upsilon_i) \end{aligned}$$

aber wir können nur das Modell

$$\ddot{y}_i = \beta\ddot{x}_i^* + \varepsilon_i^* = \beta\underbrace{(\ddot{x}_i + \upsilon_i)}_{\ddot{x}_i^*} + \underbrace{(\varepsilon_i - \beta\upsilon_i)}_{\varepsilon_i^*}$$

schätzen.

Wir beschränken uns auf den einfachsten Fall, wenn die Störterme alle angenehmen Eigenschaften aufweisen: $\mathrm{E}(\varepsilon_i) = \mathrm{E}(\upsilon_i) = 0$, $\mathrm{var}(\varepsilon_i) = \mathrm{E}(\varepsilon_i^2) = \sigma^2_\varepsilon$,

[2]Einfachheitshalber beschränken wir uns im folgenden wieder auf den 2-Variablen Fall. Für eine allgemeinere Darstellung in Matrixschreibweise siehe z.B. Johnston 1997, S. 154ff.

$\text{var}(\upsilon_i) = \text{E}(\upsilon_i^2) = \sigma_\upsilon^2$, $\text{cov}(\varepsilon_i, \varepsilon_j) = \text{cov}(\upsilon_i, \upsilon_j) = 0$ für $i \neq j$, und $\text{cov}(\varepsilon_i, \upsilon_j) = 0$ für alle i, j.

Selbst wenn die Störterme all diese angenehmen Eigenschaften aufweisen und mit den wahren Werten $\ddot{x}_i$ unkorreliert sind, so ist der OLS–Schätzer in der Regel dennoch *weder erwartungstreu noch konsistent!* Dies kann einfach gezeigt werden

Der OLS-Schätzer mit dem fehlerhaft gemessenen $\ddot{x}^*$ ist

$$\begin{aligned}
\widehat{\beta}_2 &= \frac{\sum \ddot{x}^* \ddot{y}}{\sum (\ddot{x}^*)^2} \\
&= \frac{\sum \ddot{x}^* (\beta \ddot{x} + \varepsilon)}{\sum (\ddot{x}^*)^2} \\
&= \beta \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2}
\end{aligned}$$

In diesem Fall können wir nicht einfach den Erwartungswert berechnen, da aufgrund des Messfehlers auch der Nenner des Schätzers stochastisch ist.[3] Aber man kann mit Hilfe des *probability limits* immerhin die *Konsistenz* überprüfen.

Wenn die entsprechenden Momente 2. Ordnung existieren und

$$\begin{aligned}
\text{plim} \left(\frac{1}{n} \sum (\ddot{x}^*)^2 \right) &= \sigma_{\ddot{x}}^2 + \sigma_\upsilon^2 \\
\text{plim} \left(\frac{1}{n} \sum (\ddot{x}^*) \ddot{x} \right) &= \sigma_{\ddot{x}}^2 \\
\text{plim} \left(\frac{1}{n} \sum \ddot{x}^* \varepsilon \right) &= 0
\end{aligned}$$

folgt daraus

$$\text{plim}(\widehat{\beta}_2) = \text{plim} \left(\beta \frac{\sum (\ddot{x}^*) \ddot{x}}{\sum (\ddot{x}^*)^2} + \frac{\sum \ddot{x}^* \varepsilon}{\sum (\ddot{x}^*)^2} \right) = \beta \left(\frac{\sigma_{\ddot{x}}^2}{\sigma_{\ddot{x}}^2 + \sigma_\upsilon^2} \right)$$

Da Varianzen immer positiv sind ist der Nenner von $\sigma_{\ddot{x}}^2 / (\sigma_{\ddot{x}}^2 + \sigma_\upsilon^2)$ immer größer als der Zähler ($\sigma_\upsilon^2 > 0$). Deshalb führt der Messfehler im bivariaten Modell dazu, dass der OLS Steigungsparameter $\widehat{\beta}_2$ näher bei Null liegt als der wahre Wert β_2 (auf englisch wird dies *'attenuation bias'* genannt). Die *'attenuation'* gilt allerdings nur im bivariaten Modell, in multiplen Regressionsmodellen kann der Bias komplizierter sein.

Fehlerhaft gemessene Variable treten häufig auf, wenn Proxy-Variablen verwendet werden, da die eigentlich interessierenden Variable unbeobachtbar ist (z.B. das permanente Einkommen in einer Konsumfunktion nach Friedmann, oder die Begabung eines Arbeitnehmers), wenn Fragen in einem Fragebogen zweideutig sind, oder wenn Geräte ungenaue Daten liefern. In allen diesen Fällen ist mit systematisch verzerrten Schätzergebnissen zu rechnen!

Die folgende kleine Monte Carlo Simulation soll dies demonstrieren:

[3] Erinnern Sie sich, $\text{E}(X/Y) \neq \text{E}(X)/\text{E}(Y)$!

```
' Monte Carlo: errors in variables
!n = 300
!REP = 1000

wfcreate u !n
vector(!REP) R
rndseed 1234567
series x = @rnorm
for !r = 1 to !REP
  series u = @rnorm
  series y = 5 + 5*x + u   ' Datengen. Prozess
  series xs = x + @rnorm   ' x wird fehlerhaft gemessen
  equation eq.ls y c xs    ' xs = x + Fehler
  R(!r) = c(2)
next
R.distplot hist
```

Da in diesem einfachen Fall $\beta_2 = 5$, $x_i \sim \mathrm{N}(0,1)$, $\upsilon_i \sim \mathrm{N}(0,1)$ $\varepsilon_i \sim \mathrm{N}(0,1)$ ist

$$\operatorname{plim}(\widehat{\beta}_2) = \beta_2 \left(\frac{\sigma_{\ddot{x}}^2}{\sigma_{\ddot{x}}^2 + \sigma_{\upsilon}^2} \right) = \beta_2 \frac{1}{1+1} = 0.5\beta_2 = 2.5$$

Der Mittelwert dieser tausend Replikationen (gespeichert in Vektor R) ist ungefähr 2.46 und die Standardabweichung 0.11, was denn Erwartungen entspricht. Abbildung 11.5 zeigt das Histogramm

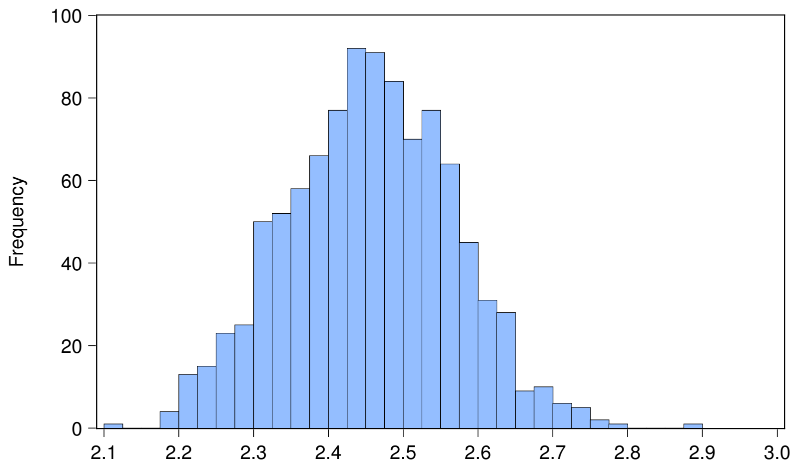

Abbildung 10.5: Monte Carlo Simulation für Messfehler im Regressor. Der wahre Wert ist $\beta_2 = 5$, aufgrund der Messfehler wird dieser Wert massiv unterschätzt.

Exkurs: Messfehler in der abhängigen Variablen *y* Die vorhergehenden Ausführungen bezogen sich ausschließlich auf Meßfehler in der erklärenden x Variable, Meßfehler in der abhängigen y Variable sind im Vergleich dazu verhältnismäßig unproblematisch. Um dies zu zeigen nehmen wir an, das wahre Regressionsmodell sei durch die Beziehung

$$\ddot{y}_i = \beta \ddot{x}_i + \varepsilon_i$$

gegeben, wobei ε_i ein üblicher Störterm sei. Wenn nun anstelle der 'wahren' abhängigen Variablen $\ddot{y}_i$ nur ein fehlerhaft gemessenes $\ddot{y}_i^*$ beobachtbar ist, wobei $\ddot{y}_i^* = \ddot{y}_i + \upsilon_i$ mit $\operatorname{cov}(\upsilon_i, \varepsilon_i) = 0$, dann berechnen wir die Regression

$$\ddot{y}_i^* = \beta \ddot{x}_i + (\varepsilon_i + \upsilon_i)$$

Eine OLS-Schätzung für β ist weiterhin unverzerrt und effizient. Die Varianz des Störterms $(\varepsilon_i + \upsilon_i)$ ist nun zwar größer, dem wird aber bei der Berechnung von $\hat{\sigma}^2$ aus den beobachteten Störtermen Rechnung getragen. Deshalb bleiben alle statistischen Tests gültig.

Systematische Verzerrungen treten nur auf, wenn die erklärenden x Variablen fehlerhaft gemessen werden.

Wir fassen zusammen: bei endogenen x Variablen liefert der OLS-Schätzer selbst in großen Stichproben systematisch verzerrte Ergebnisse, d.h. OLS Schätzer sind in solchen Fällen weder erwartungstreu noch konsistent! Deshalb sind andere Schätzverfahren erforderlich.

Leider sind keine Schätzer bekannt, die in solchen Fällen BLUE (d.h. auch in kleinen Stichproben unverzerrt und effizient) sind, aber die *Methode Zweistufiger Kleinster Quadrate* (*'two stage least squares'*, 2SLS) und der *Instrumentvariablen Schätzer* (IV-Schätzer) liefern in diesen Fällen zumindest konsistente Ergebnisse (d.h. die Schätzungen werden mit zunehmender Stichprobengröße zunehmend genauer).

10.3 Die Methode der zweistufigen kleinste Quadrate (2SLS)

Erinnern wir uns: die OLS Methode beruht auf einer orthogonalen Zerlegung von y in eine systematische – durch die x Variablen erklärte – Komponente, und in eine stochastische nicht-systematische Komponente, die Residuen $\hat{\varepsilon}$. Der OLS Schätzer stellt sicher, dass die Residuen mit den x Variablen *in der Stichprobe* unkorreliert sind (orthogonal), aber dies muss nicht für die Störterme der Grundgesamtheit ε gelten.

Endogenität im ökonometrischen Sinne ist definiert als eine Abhängigkeit zwischen Störtermen und den x Variablen, d.h. $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) \neq \mathbf{0}$.

Intuitiv können wir uns vorstellen, dass die erklärende x Variable 'zwei Arten von Streuung' enthält, eine *erwünschte Streuung*, die zur Korrelation mit der systematischen Komponente von y führt, und eine *unerwünschte Streuung*, die zur Korrelation mit den Störtermen ε (d.h. der nicht-systematischen Komponente von y) führt.

Diese intuitive Idee der Zerlegung der Streuung von x in eine 'gute' und erwünschte Streuung und in eine unerwünschte problematische Streuung führt unmittelbar zur Idee der Methode der zweistufigen kleinsten Quadrate (*'two stage least squares'* bzw. 2SLS).

Wir beginnen mit dem einfachsten Fall, dem bivariaten Modell und einer einzigen Instrumentvariablen.

Wir gehen von einer Strukturgleichung der Form

$$y_i = \widehat{\beta}_1 + \widehat{\beta}_2 x_i + \varepsilon_i$$

aus. Wenn x endogen ist (d.h. $\text{cov}(x, \varepsilon) \neq 0$ oder allgemeiner $\text{E}(\varepsilon|x) \neq 0$), ist der OLS Schätzer bekanntlich weder erwartungstreu noch konsistent.

Wie schon früher erwähnt besteht das Problem darin, dass der interessierende Effekt von x auf y durch die Korrelation zwischen x und ε gewissermassen 'verschmutzt' wird (siehe Abbildung 11.6 linkes Panel).

Die Idee der 2SLS Schätzung besteht einfach darin, dass man versucht mit Hilfe einer Regression (der 1. Stufe) die Streuung von x in diese zwei Teile zu zerlegen, in eine erwünschte Streuungen, die aus der Korrelation mit y herrührt, und in eine unerwünschte Streuungen, die aus der Korrelation mit ε folgt.

Dies kann nur mit Hilfe zusätzlicher Information gelingen, man benötigt eine (oder mehrere) zusätzliche Variable(n) z, die ganz besondere Eigenschaften haben müssen, sie sollten möglichst hoch mit x korreliert sein, dürfen aber nicht mit ε korreliert sein.

Solche Variablen werden **'Instrumentvariablen'** (z) genannt, und sie müssen die beiden folgenden Anforderungen erfüllen

1. **Relevanz:** $\text{cov}(z, x) \neq 0$, und
2. **Exogenität:** $\text{cov}(z, \varepsilon) = 0$

Mit Hilfe einer solchen Instrumentvariable kann die gewünschte Zerlegung der Streuung des endogenen x in eine 'unproblematische' exogene Streuung und in eine 'problematische' Streuung, die aus der Korrelation mit ε folgt, erfolgen.

Abbildung 11.6 (rechtes Panel) zeigt das Grundprinzip einer 2SLS Schätzung. Die endogene abhängige Variable x wird in zwei Teile zerlegt, einen exogenen Teil $\hat{x}$, der durch die Instrumentvariable erklärt wird, und einen endogenen Teil υ_x. Praktisch erfolgt diese Zerlegung mit Hilfe einer einfachen OLS Regression, der potentiell endogene Regressor x der Gleichung $y = \beta_1 + \beta_2 x + \varepsilon$ wird in einem ersten Schritt mit einer Hilfsregression in zwei Teile zerlegt

$$\begin{aligned} x &= \hat{\alpha}_1 + \hat{\alpha}_2 z + \upsilon_x \\ &= \hat{x} + \upsilon_x \end{aligned}$$

Man kann sich vorstellen, dass die Streuung von x mit Hilfe einer Regression auf die Instrumentvariable in einen unproblematischen *exogenen* Anteil – d.h. die Streuung von $\hat{x}$ – und einen problematischen *endogenen* Anteil – d.h. die Streuung von υ_x – zerlegt wird.

Die Streuung von $\hat{x}$ ist deshalb unproblematisch, da sie ausschließlich durch die per Definition exogene Instrumentvariable erklärt wird. Die problematische Streuung von υ_x wird verworfen, vgl Abbildung 11.6.

Wir werden später sehen, dass – wenn die Annahmen für Instrumentvariablen (*Relevanz* und *Exogenität*) erfüllt sind – Instrumentvariablenschätzer konsistent sind.

Für eine Schätzung mit der Methode der Zweistufigen Kleinsten Quadrate wird – wie der Name schon sagt – zweistufig vorgegangen:

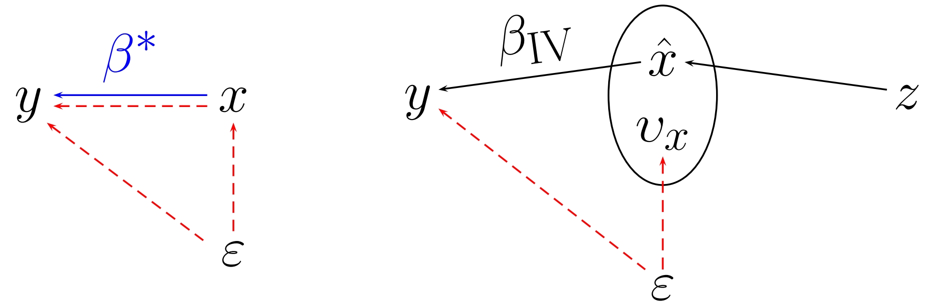

Abbildung 10.6: Falls $\mathrm{E}(\varepsilon|\boldsymbol{X}) \neq 0$ ist der OLS Schätzer β^* verzerrt, da der Einfluss von x auf y durch die Korrelation mit dem Störterm 'verschmutzt' wird (linkes Panel). Bei einer IV Schätzung wird x gewissermaßen in einen exogenen Teil $\hat{x}$ – der durch die Instrumentvariable erklärt wird – und einen endogenen Teil v_x zerlegt. Die IV Schätzer verwenden nur die exogene Streuung $\hat{x}$ (rechtes Panel).

1. Stufe: Zerlege x mit Hilfe einer Regression x in eine Komponente, die durch die exogenen Instrumentvariable z erklärt wird, und in eine problematische 'unerwünschte' Komponente v_i. Dazu wird in einer Hilfsregression die Variable x auf die Instrumentvariable z regressiert

$$x_i = \hat{\alpha}_1 + \hat{\alpha}_2 z_i + v_i$$

und daraus werden die gefitteten Werte $\widehat{x}_i$ berechnet.

Da diese gefitteten Werte $\widehat{x}$ nur von der exogenen Instrumentvariable z abhängen, ist die Streuung von $\widehat{x}_i$ sicher nicht mit dem Störterm der Strukturgleichung ε_i korreliert.

2. Stufe: Deshalb werden auf der 2. Stufe die gefitteten Werte $\widehat{x}$ der 1. Stufe anstelle der 'verschmutzten' x Werte verwendet, um den Effekt von x auf y zu messen

$$y_i = \widehat{\beta}_1 + \widehat{\beta}_2 \widehat{x}_i + \hat{\varepsilon}_i$$

Diese intuitiv sehr einleuchtende Methode der **zweistufigen kleinsten Quadrate** (2SLS oder *'Two Stage Least Squares'*) liefert numerisch die gleichen Ergebnisse wie die allgemeineren Instrumentvariablenschätzer, die wir gleich vorstellen werden.

Wenn man derartig zweistufig vorgeht erhält man zwar die korrekten Koeffizienten, allerdings sind die Standardfehler der 2. Stufe nicht korrekt, da diese die 1. Stufe nicht berücksichtigen! Deshalb sollte man die in allen Ökonometrieprogammen verfügbaren Befehle zur Schätzung von 2SLS verwenden (z.B. `2SLS` für EViews oder `ivregress` für Stata), da diese automatisch die korrekten Standardfehler liefern.[4]

[4] Das Problem besteht darin, dass für die Schätzung von $\hat{\sigma}^2$ die Residuen $\hat{\varepsilon}_i = y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i$ verwendet werden müssen, und nicht die Residuen der 2. Stufe mit den gefitteten $\widehat{x}$, d.h. $\hat{\varepsilon}_i^* = y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} \widehat{x}_i$

Beispiel: welchen Einfluss hat die Anzahl der Polizisten in einer Stadt auf die Kriminalität? (Levitt, 1997)

Dies ist ein typisches Beispiel für Simultanität: Polizisten sollen Kriminalität senken, aber Städte mit höherer Kriminalität heuern mehr Polizisten an. Eine Regression der Anzahl der Polizisten (P) auf die Anzahl der Verbrechen (V) kann das falsche Vorzeichen liefern.

Levitt verwendet Wahljahre (W) als Instrumentvariable. Ist dies ein geeignetes Instrument?

1. Relevanz: Die Ausgaben für Sicherheit nehmen in Wahljahren zu, damit auch die Anzahl der Polizisten.
2. Exogenität: Die Wahlzyklen sind exogen (prädeterminiert).

Zweistufige kleinste Quadrate:

1. Stufe: Zerlege x in eine Komponente, die mit der exogenen Instrumentvariable z erklärt werden kann, und in eine problematische 'unerwünschte' Komponente

$$P_i = \hat{\alpha}_1 + \hat{\alpha}_2 W_i + v_i$$

2. Stufe: Verwende den gefitteten Wert der 1. Stufe um den Effekt auf die Anzahl der Verbrechen zu schätzen

$$V_i = \widehat{\beta}_1 + \widehat{\beta}_2 \widehat{P}_i + \hat{\varepsilon}_i$$

Levitt's Ergebnis: Eine Erhöhung der Anzahl der Polizisten vermindert die Zahl der Gewaltverbrechen (aber hat wenig Einfluss auf die Zahl der Eigentumsdelikte).

Diese Methode kann einfach auf die multiple Regression übertragen werden (die entsprechende Technik folgt später).

10.4 Instrumentvariablen Schätzer (IV)

> "Two things distinguish the discipline of econometrics from the older sister field of statistics. One is the lack of shyness about causality. Causal inference has always been the name of the game in applied econometrics. [...]
>
> The second thing that distinguishes us from most statisticians – and indeed from most other social scientists – is an arsenal of statistical tools that grew out of early econometric research on the problem of how to estimate the parameters in a system of linear simultaneous equations. The most powerful weapon in this arsenal is the method of instrumental variables (IV), the subject of this chapter." (Angrist and Pischke, 2008, p. 113)

Instrumentvariablenschätzer sind *keine* OLS Schätzer, deshalb ist auch das Gauss Markov Theorem nicht anwendbar, deshalb garantiert uns nichts die Unverzerrtheit und Effizienz des 2SLS Schätzers.

Tatsächlich sind 2SLS Schätzer in kleinen Stichproben meist verzerrt und konvergieren nur mit zunehmender Stichprobengröße zu den wahren Werten, d.h. sie sind konsistent.

Technisch werden 2SLS Schätzer meist mit Hilfe der 'Methode der Momente' hergeleitet, die wir im folgenden kurz vorstellen werden, um sie später anwenden zu können. 'Methode der Momente' Schätzer sind zwar nicht immer unverzerrt und effizient, aber sie sind generell konsistent.

10.4.1 Instrumentvariablen als Methode der Momente Schätzer

Wir haben bisher erst eine Methode zur Schätzung eines unbekannten Parametervektors kennengelernt, nämlich die *Methode der Kleinsten Quadrate* (OLS). Die historisch älteste Methode ist allerdings die *'Methode der Momente'*, die immer konsistente Schätzer liefert, die aber – im Unterschied zu Maximum Likelihood Schätzern – nicht immer asymptotisch effizient sind. Die Grundidee der Methode der Momente Schätzer besteht darin, dass die Momente der Stichprobe (z.B. Mittelwert, Varianz, ...) als Schätzer für die Momente der Grundgesamtheit herangezogen werden.

Wir zeigen dies für das einfache Regressionsmodell. In der Grundgesamtheit bestehe folgender Zusammenhang

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Wir nehmen für die Momente der Grundgesamtheit an, dass

- der Erwartungswert der Störterme der Grundgesamtheit ist Null

$$\mathrm{E}(\varepsilon_i) = 0$$

- die Kovarianz zwischen den Störtermen der Grundgesamtheit ε_i und dem möglicherweise stochastischen x_i ist ebenfalls Null

$$\mathrm{E}(x_i \varepsilon_i) = 0$$

Die entsprechenden Bedingungen *für die Stichprobe* sind

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n}\sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$$

wobei $\hat{\varepsilon}_i$ die Stichprobenresiduen sind.

Durch Einsetzen der Stichprobenresiduen $\hat{\varepsilon}_i = y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i$ in die obigen Bedingungen erhalten wir die beiden Normalgleichungen

$$\begin{aligned} \sum y_i &= n\widehat{\beta}_1 + \widehat{\beta}_2 \sum x_i \\ \sum y_i x_i &= \widehat{\beta}_1 \sum x_i + \widehat{\beta}_2 \sum x_i^2 \end{aligned}$$

wobei $\widehat{\beta}_1$ und $\widehat{\beta}_2$ die Methode der Momente Schätzer für β_1 und β_2 sind.

Dies sind aber exakt die gleichen Normalgleichungen, die wir auch für OLS erhalten haben, deshalb liefert die Methode der Momente in diesem einfachen Fall exakt die gleichen Schätzer wie die Methode der kleinsten Quadrate oder – wie wir in einem späteren Kapitel zeigen werden – auch die Maximum Likelihood Methode.

In diesem Fall einfachen führen also alle drei bekannten Schätzmethoden, d.h. OLS, Maximum Likelihood und die Methode der Momente, zum gleichen Ergebnis, aber dies gilt natürlich nicht immer. In manchen komplexen Situationen bietet die Methode der Momente eine relativ einfache Möglichkeit um zu einem konsistenten Schätzer zu gelangen, z.B. bei der Herleitung des Instrumentenvariablen-Schätzers.

Mit dem Methode der Momente Schätzer haben wir die Bedingungen

$$\mathrm{E}(\varepsilon_i) = 0 \qquad \text{und} \qquad \mathrm{E}(x_i \varepsilon_i) = 0$$

angenommen um einen konsistenten Schätzer zu erhalten, wobei x auch stochastisch sein kann. Wenn aber zum Beispiel eine fehlerhaft gemessene Variable x^* mit ε korreliert ist kann *nicht* angenommen werden, dass $\mathrm{E}(x_i^* \varepsilon_i) = 0$.

Aber angenommen, wir hätten wieder eine Variable z zur Verfügung, die nicht mit den ε korreliert ist, also $\mathrm{E}(z_i, \varepsilon_i) = 0$, die aber mit x_i korreliert ist ($\mathrm{cov}(z_i, x_i) \neq 0$), kann die Methode der Momente herangezogen werden um zumindest einen konsistenten Schätzer zu finden.

Die entsprechenden Stichproben Momente sind in diesem Fall

$$\frac{1}{n} \sum \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n} \sum (z_i \hat{\varepsilon}_i) = 0$$

Eine solche **Instrumentvariable** (z) muss – wie wir bereits im letzten Abschnitt gesehen haben – zwei wichtige Bedingungen erfüllen:

1. die Instrumentvariable z sollte möglichst hoch mit dem x korreliert sein, d.h. sie muss *relevant* sein
$$\mathrm{cov}(z, x) \neq 0$$

2. die Instrumentvariable z darf nicht mit dem Störterm ε der Grundgesamtheit korreliert sein, d.h. sie muss *exogen* sein
$$\mathrm{cov}(z, \varepsilon) = 0$$

Man beachte, dass die zweite Bedingung, $\mathrm{cov}(z, \varepsilon) = 0$, nicht einfach getestet werden kann, da die Störterme der Grundgesamtheit unbekannt sind![5] Man kann auch nicht die Stichprobenresiduen $\hat{\varepsilon}$ für einen solchen Test nützen, da diese per Konstruktion mit x unkorreliert sind.

Hingegen kann die erste Bedingung $\mathrm{cov}(z, x) \neq 0$ sehr einfach getestet werden, indem man eine Regression $x_i = a_0 + a_1 z_i + \nu_i$ rechnet und mit dem üblichen t-Test überprüft, ob a_1 signifikant von Null verschieden ist (bzw. mit einem F-Test im multivariaten Fall).

[5] Ein Test auf Endogenität des Regressors ist der *Hausman Test*, der später vorgestellt wird.

Die Instrumentvariablen Schätzer $\widehat{\beta}_{k,\text{IV}}$ erhalten wir, indem wir die IV-Residuen $\hat{\varepsilon}_i = y_i - \widehat{\beta}_{1,\text{IV}} - \widehat{\beta}_{2,\text{IV}} x_i$ in die Momentbedingungen

$$\frac{1}{n}\sum \hat{\varepsilon}_i = 0 \qquad \text{und} \qquad \frac{1}{n}\sum (z_i \hat{\varepsilon}_i) = 0$$

einsetzen

$$\begin{aligned} \sum \left(y_i - \widehat{\beta}_{1,\text{IV}} - \widehat{\beta}_{2,\text{IV}} x_i\right) &= 0 \\ \sum z_t \left(y_i - \widehat{\beta}_{1,\text{IV}} - \widehat{\beta}_{2,\text{IV}} x_i\right) &= 0 \end{aligned}$$

Daraus können wieder die Normalgleichungen hergeleitet werden

$$\begin{aligned} \sum y_i &= n\widehat{\beta}_{1,\text{IV}} + \widehat{\beta}_{2,\text{IV}} \sum x_i \\ \sum y_i z_i &= \widehat{\beta}_{1,\text{IV}} \sum z_i + \widehat{\beta}_{2,\text{IV}} \sum x_i z_i \end{aligned}$$

Diese Gleichungen können wieder wie üblich gelöst werden, indem wir die erste Gleichung mit $\sum z_i$ und die zweite Gleichung mit n multiplizieren sowie die erste Gleichung von der zweiten Gleichung subtrahieren,

$$\widehat{\beta}_{2,\text{IV}} = \frac{n\sum y_i z_i - \sum y_i \sum z_i}{n\sum x_i z_i - \sum x_i \sum z_i}$$

Dies kann ähnlich wie früher vereinfacht werden und gibt die folgenden Instrumentvariablen Schätzer für den bivariaten Fall

$$\begin{aligned} \widehat{\beta}_{2,\text{IV}} &= \frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (x_i - \bar{x})(z_i - \bar{z})} = \frac{\sum \ddot{y}_i \ddot{z}_i}{\sum \ddot{x}_i \ddot{z}_i} = \frac{\text{cov}(y,z)}{\text{cov}(x,z)} \\ \widehat{\beta}_{1,\text{IV}} &= \bar{y} - \widehat{\beta}_{2,\text{IV}} \bar{x} \end{aligned}$$

wobei $\bar{x} := 1/n \sum x_i$ wieder den Mittelwert von x bezeichnet und $\ddot{x}_i := x_i - \bar{x}$ die Abweichungen vom Mittelwert sind, bzw. $\ddot{z}_i := z_i - \bar{z}$.

Man beachte, dass für den IV Schätzer nicht einfach x durch z 'ersetzt' wird, im Nenner des Schätzers für $\widehat{\beta}_{2,\text{IV}}$ steht nämlich weder var(x) noch var(z), sondern cov(x,z)!

Intuition: Cameron, Trivedi, Microeconometrics using Stata, p. 173

Marginaler Effekt (Kettenregel):

$$\widehat{\beta}_{\text{IV}} = \frac{dy}{dz}\frac{dz}{dx}$$

Warum?

$$\widehat{\beta}_{\text{IV}} = \frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{x}_i \ddot{z}_i} = \frac{\frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2}}{\frac{\sum_i \ddot{x}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2}}$$

Sei

$$\hat{\ddot{y}}_i = \hat{\alpha}\, \ddot{z}_i \qquad \text{und} \qquad \hat{\ddot{x}}_i = \hat{\delta}\, \ddot{z}_i$$

mit

$$\hat{\alpha} = \frac{\sum_i \ddot{y}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2} = \frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}} \qquad \text{und} \qquad \hat{\delta} = \frac{\sum_i \ddot{x}_i \ddot{z}_i}{\sum_i \ddot{z}_i^2} = \frac{d\,\hat{\ddot{x}}}{d\,\ddot{z}}$$

Deshalb

$$\widehat{\beta}_{\mathrm{IV}} = \frac{\hat{\alpha}}{\hat{\delta}} = \frac{\frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}}}{\frac{d\,\hat{\ddot{x}}}{d\,\ddot{z}}} = \frac{d\,\hat{\ddot{y}}}{d\,\ddot{z}}\,\frac{d\,\ddot{z}}{d\,\hat{\ddot{x}}}$$

■

Um die Konsistenz dieses Schätzers zu zeigen verwenden wir wieder die Variablen in Abweichungsform vom Mittelwert und setzen für $\ddot{y}$ den wahren Wert $\beta_2 \ddot{x}_i + \varepsilon_i$ ein. Wir erhalten

$$\widehat{\beta}_{2,\mathrm{IV}} = \beta_2 + \frac{\sum \ddot{z}_i \varepsilon_i}{\sum \ddot{x}_i \ddot{z}_i}$$

Wenn $\ddot{x}_i$ stochastisch ist kann der Erwartungswert nicht einfach ermittelt werden, da wir hier einen Quotienten zweier Zufallsvariablen haben, aber der plim kann einfach berechnet werden

$$\operatorname{plim}(\widehat{\beta}_{2,\mathrm{IV}}) = \beta_2 + \frac{\operatorname{plim}\left[\frac{1}{n}\sum(\ddot{z}_i \varepsilon_i)\right]}{\operatorname{plim}\left[\frac{1}{n}\sum \ddot{x}_i \ddot{z}_i\right]}$$

Der Zähler des letzten Ausdrucks der rechten Seite ist Null, da annahmegemäß $\operatorname{cov}(\ddot{z}_i \varepsilon_i) = 0$. Der Nenner ist der plim der Stichprobenkovarianz zwischen $\ddot{x}_i$ und $\ddot{z}_i$ und ist deshalb ein konsistenter Schätzer für die Kovarianz zwischen $\ddot{x}_i$ und $\ddot{z}_i$ der Grundgesamtheit $\sigma_{\ddot{x}z}$. Wenn $\sigma_{\ddot{x}\ddot{z}} \neq 0$ folgt also

$$\widehat{\beta}_{2,\mathrm{IV}} = \beta_2 + \frac{0}{\sigma_{\ddot{x}\ddot{z}}} = \beta_2$$

Man beachte, dass selbst eine sehr geringe Korrelation zwischen z und ε zu großen Fehlern führen kann, wenn die Korrelation zwischen z und x ebenfalls sehr klein ist. Deshalb sollte man die Korrelation zwischen x und z *immer* mit einer Hilfsregression überprüfen!

Die Konsistenz von $\widehat{\beta}_{1,\mathrm{IV}}$ kann ähnlich gezeigt werden.

Da unterschiedliche Instrumente gewählt werden können, und jedes Instrument zu einem anderen konsistenten Schätzer führt, kann nicht davon ausgegangen werden, dass Instrumentvariablen Schätzer asymptotisch effizient sind.

In großen Stichproben gilt allerdings, dass

$$\widehat{\beta}_{2,\mathrm{IV}} \overset{a}{\sim} \mathrm{N}\left(\beta_2, \operatorname{var}(\widehat{\beta}_{2,\mathrm{IV}})\right)$$

mit der geschätzten Varianz

$$\widehat{\operatorname{var}}\left(\widehat{\beta}_{2,\mathrm{IV}}\right) = \frac{\hat{\sigma}^2 \sum (z_i - \bar{z})^2}{\left[\sum (x_i - \bar{x})(z_i - \bar{z})\right]^2} = \frac{\hat{\sigma}^2 \sum \ddot{z}_i^2}{\left[\sum \ddot{x}_i \ddot{z}_i\right]^2}$$

Einen asymptotischen Schätzer $\hat{\sigma}^2$ für σ^2 erhalten wir aus

$$\hat{\sigma}^2 = \frac{1}{n}\sum \hat{\varepsilon}_i^2$$

wobei $\hat{\varepsilon}_i = y_i - \widehat{\beta}_{1,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{IV}} x_i$.

Daraus wird offensichtlich, dass die Varianz von $\widehat{\beta}_{2,\mathrm{IV}}$ umso kleiner ist, je größer die Korrelation zwischen x_i und z_i ist.

Instrumentvariablen dürfen nicht mit Proxy-Variablen verwechselt werden. Proxy Variablen sollten eine fehlende Variable gewissermaßen 'ersetzen', kommen also in der Regressionsgleichung direkt vor, während Instrumentvariablen nur 'indirekt' vorkommen (d.h. nicht als Regressor aufscheinen).

Die Instrumentvariablen z sollten möglichst hoch mit dem endogenen Regressor x korreliert sein, und *nur* über x auf y einwirken (d.h. mit ε unkorreliert sein).

Proxy-Variablen sollten möglichst hoch mit den fehlenden Variablen korreliert sein, während Instrumentvariablen nicht mit den fehlenden Variablen korreliert sein dürfen.

Im Falle einer Lohngleichung Wage $= \widehat{\beta}_1 + \widehat{\beta}_2$Education $+ \widehat{\beta}_3$Ability $+ \hat{\varepsilon}$ könnte als Proxy-Variable für die unbeobachtbaren Fähigkeiten z.B. der gemessene Intelligenzquotient verwendet werden.

Eine Instrumentvariable sollte hingegen mit den unbeobachtbaren Fähigkeiten und dem Störterm unkorreliert sein, aber möglichst hoch mit der Variablen 'Education' korrelieren. Arbeitsmarktökonomen haben z.B. die Entfernung zwischen Heimat- und Studienort (Card 1995, *Aspects of Labour Market Behavior*) oder die Anzahl der Geschwister als Instrumentvariable für 'Education' vorgeschlagen, da Jugendliche aus kinderreichen Familien typischerweise eine kürzere Schulbildung erhalten, aber (hoffentlich) nicht über weniger Fähigkeiten verfügen (vgl. Wooldridge 2000, Chapter 15.1).[6]

Beispiel: Mit Hilfe einer Monte Carlo Simulation kann die Wirksamkeit der IV-Methode wieder einfach demonstriert werden. Wir haben früher gezeigt, dass eine einfache OLS Schätzung der Konsumfunktion im Modell

$$\begin{aligned} \mathrm{C}_i &= 60 + 0.5Y + \varepsilon_i \\ Y_i &= \mathrm{C}_i + Z_i \end{aligned}$$

aufgrund des *feed-backs* mit der Identität zu systematisch verzerrten Ergebnissen führt. In diesem einfachen Fall bietet sich allerdings ein fast ideales Instrument an, die exogene Variable Z. Da es sich um ein simultanes Gleichungssystem handelt hängen alle endogenen Variablen von der exogenen Variable Z ab, sie sind also korreliert (Relevanz).

[6] Allerdings könnten Kinder aus kinderreichen Familien über mehr soziale Fähigkeiten verfügen. Würde dies die Qualität der Instrumentvariable 'Anzahl der Geschwister' beeinflussen?

Hingegen hat Z keinen direkten Einfluss auf den Konsum Cons, sondern wirkt nur indirekt über das Einkommen Y auf Cons. In anderen Worten, Z kommt in der Konsumfunktion nicht vor und ist mit dem Störterm ε unkorreliert (Exogenität).

Deshalb bietet sich in diesem einfachen Modell Z als ideales Instrument an. Wir können dies wieder mit Hilfe der Monte Carlo Simulation überprüfen. Das folgende Programm entspricht der früheren Simulation, aber anstatt einer früheren Stichprobengröße $n = 25$ wählen wir hier eine größere Stichprobe mit $n = 250$ um zu zeigen, dass die Verzerrung des OLS Schätzers auch bei größeren Stichproben auftritt, und wir führen zusätzlich 2SLS (IV) Schätzungen durch.

```
wfcreate u 250
rndseed 123456789
!REP = 1000
vector(!REP) MCP_OLS = na
vector(!REP) MCP_IV = na
series Cons
series Y
series Z
for !r = 1 to !REP
Z = @runif(30,80)
    ' Modell lösen
    model Keynes
    Keynes.append  Cons = 60 + 0.5*Y + @runif(-30,30)
    Keynes.append  Y = Cons + Z
    Keynes.solve
    delete Keynes
    ' Konsumfunktion mit Modelllösungen schätzen
    equation eqOLS.ls Cons_0 c Y_0
    MCP_OLS(!r) = c(2)
    equation eqIV.tsls Cons_0 c Y_0 @ Z
    MCP_IV(!r) = c(2)
next
MCP_OLS.distplot hist
MCP_IV.distplot hist
```

Abbildung 11.7 zeigt das Ergebnis, offensichtlich ist die OLS Schätzung der marginalen Konsumneigung systematisch verzerrt, die IV-Schätzung (2SLS) liegt dagegen im Mittelwert deutlich näher am wahren Wert 0.5. Man beachte, dass der Standardfehler des IV Schätzers deutlich größer ist.

Übung: Angenommen wir haben Daten für den Lernerfolg (Y) und Lernzeit (LZ) einer Gruppe Studierende gesammelt und möchten den Zusammenhang $Y = \widehat{\beta}_1 + \widehat{\beta}_2\mathrm{LZ} + \hat{\varepsilon}$ schätzen. Da fällt Ihnen ein, dass möglicherweise ein *'omitted variable'* Problem auftritt, denn Sie haben keine Daten für die 'Begabung' der Studierenden.

- Ein Kollege empfiehlt Ihnen, den Intelligenzquotienten (IQ) der Studierenden zu messen und als Instrument zu verwenden.

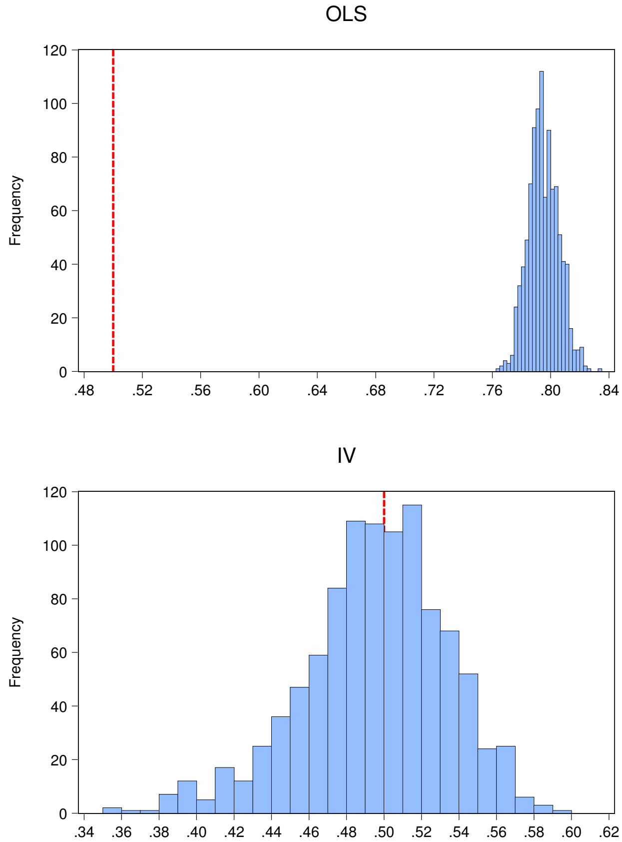

Abbildung 10.7: Monte Carlo Simulation für endogenen Regressor. OLS und IV Schätzungen der marginalen Konsumneigung (der wahre Wert ist 0.5, $n = 250$).

Hinweis: schlechte Idee! (überlegen Sie sich, ob Auswirkungen des IQ im Störterm auftauchen könnten; $\rightarrow$ Proxyvariablen)

- Eine Kollegin empfiehlt Ihnen das Wetter am Vorabend der Prüfung (W) als Instrument zu verwenden (angenommen, die Studierenden schreiben die Prüfung an sehr unterschiedlichen Tagen).
  Hinweis: schon weit besser! (in einer Regression $Y = \widehat{\beta}_1 + \widehat{\beta}_2\text{LZ} + \widehat{\beta}_3 W + \hat{\varepsilon}$ würden wir erwarten, dass $\widehat{\beta}_3$ *nicht* signifikant von Null verschieden ist, deshalb sollte W nicht mit dem Störterm korreliert sein)

Schwache Instrumente (*'weak instruments'*)

Wenn die Annahmen für Instrumentvariablen nicht erfüllt sind können IV-Schätzer extrem verzerrte Ergebnisse liefern. Aber selbst wenn die Annahmen im strengen Sinne erfüllt sind, aber die Instrumente nur schwach mit den Regressoren korreliert sind, können IV Schätzer sehr schlechte Resultate liefern, selbst wenn die Stichprobe groß ist.

Im einfachsten bivariaten Fall ist der IV-Schätzer für den Steigungskoeffizienten

$$\widehat{\beta}_{2,\text{IV}} = \frac{\widehat{\text{cov}}(y_i, z_i)}{\widehat{\text{cov}}(x_i, z_i)} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})(z_i - \overline{z})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(z_i - \overline{z})}$$

Wenn der Nenner $\widehat{\text{cov}}(x_i, z_i)$ dieses Schätzers in der Grundgesamtheit ungleich Null ist, ist dieser Schätzer in (sehr) großen Stichproben annähernd normalverteilt. Wenn der Nenner exakt Null ist, ist der Schätzer in großen Stichproben Cauchy verteilt, die weder Erwartungswert noch Varianz besitzt.

Von *'weak instruments'* spricht man, wenn die Kovarianz sehr kleine ist, also nahe bei Null liegt. Es zeigte sich nämlich, dass in solchen Fällen der 2SLS Schätzer auch in ziemlich großen Stichproben noch verzerrt ist und die Konfidenzintervalle ein zu optimistisches Bild zeichnen (siehe z.B. Bound et al., 1995).

Eine der Lehren aus dieser Literatur ist, dass die Signifikanz der Instrumente auf der 1. Stufe ein Indikator für die *Relevanz* der Instrumente ist. Deshalb sollte die Signifikanz Signifikanz der Instrumente auf der 1. Stufe immer überprüft werden, und in Publikationen auch angegeben werden!

Beginnend mit Staiger and Stock (1997) entwickelte sich eine umfangreiche Literatur zu der Problematik 'schwacher Instrumente', die noch keineswegs abgeschlossen ist.

Stock & Watson's Faustregel: Die F-Statistik auf gemeinsame Signifikanz aller Instrumentvariablen in der 1. Stufe sollte größer als 10 sein, bzw. im bivariaten Fall sollte die t-Statistik größer als 3.3 sein!

Ein kleiner Wert der F-Statistik ist ein Indikator für *schwache* Instrumente (sie erklären wenig Streuung in y_2)! Im Fall schwacher Instrumente liefern IV Schätzungen häufig schlechtere Ergebnisse als OLS Schätzungen!

Dies ist allerdings nur eine sehr grobe Faustregel, das Problem schwacher Instrumente war die letzten Jahre eines der aktivsten Forschungsgebiete mit sehr fruchtbaren Ergebnissen. Für eine Übersicht siehe z.B. Murray (2006).[7]

Identifikation

Im multivariaten Fall muss *mindestens* für jeden *endogenen* Regressor eine Instrumentvariable zur Verfügung stehen.

Allgemeiner, die Regressionskoeffizienten sind

- *exakt identifiziert*, wenn die Anzahl der Instrumente *gleich* der Anzahl der endogenen Regressoren ist.
- *überidentifiziert*, wenn die Anzahl der Instrumente *größer* ist als die Anzahl der endogenen Regressoren.
- *unteridentifiziert*, wenn die Anzahl der Instrumente *kleiner* ist als die Anzahl der endogenen Regressoren.

Unteridentifizierte Systeme können nicht geschätzt werden!

[7] `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=843185`

10.4.2 Tests auf Endogenität und überidentifizierende Restriktionen

Für die Anwendung der IV-Methode sind zwei Fragen wichtig

- ist eine erklärende Variable der Strukturgleichung tatsächlich endogen? Diese Frage kann mit Hilfe eines *Hausman Tests* getestet werden. Sollte dieser Test Hinweise auf Endogenität liefern kann z.B. ein IV-Schätzer verwendet werden.
- wenn auf einen IV-Schätzer zurückgegriffen werden muss, sind die verwendeten Instrumente tatsächlich exogen? Ein Test dieser Hypothese ist nur in überidentifizierten Gleichungen möglich (*Test auf überidentifizierende Restriktionen*).

Hausman Test (Test auf Endogenität)

Da die Auswirkungen einer Korrelation zwischen den erklärenden Variablen und den Residuen der Grundgesamtheit sehr weitreichend sind (d.h. der Schätzer wäre systematisch verzerrt) wäre ein Test auf diese Korrelation sehr hilfreich. Allerdings können dafür nicht die aus der Stichprobe geschätzten Residuen $\hat{\varepsilon}_i$ herangezogen werden, da diese per Konstruktion immer mit den x unkorreliert sind (diese Orthogonalität ist eine Bedingung erster Ordnung!).

Hausman (1978) hat dafür einen sehr einfachen asymptotischen Test vorgeschlagen.

Die entsprechende Null- und Alternativhypothese ist

$$\mathrm{H}_0: \ \operatorname{plim} \frac{1}{n} \sum (x_i - \bar{x})\varepsilon_i = 0 \qquad \mathrm{H}_1: \ \operatorname{plim} \frac{1}{n} \sum (x_i - \bar{x})\varepsilon_i \neq 0$$

Die Grundidee des Hausman Tests besteht in einem Vergleich zweier Schätzer, wovon einer sowohl unter der Null-Hypothese als auch unter der Alternativhypothese konsistent ist, während der zweite nur unter Nullhypothese konsistente Schätzresultate liefert. Ein großer Unterschied zwischen diesen beiden Schätzungen wird als Evidenz zugunsten der Alternativhypothese gedeutet.

Hausman hat gezeigt, dass unter Gültigkeit der Nullhypothese die folgende Statistik asymptotisch χ^2 verteilt ist mit einem Freiheitsgrad

$$H = \frac{\left(\widehat{\beta}_{2,\mathrm{IV}} - \widehat{\beta}_{2,\mathrm{OLS}}\right)^2}{\widehat{\operatorname{var}}\left(\widehat{\beta}_{2,\mathrm{IV}}\right) - \widehat{\operatorname{var}}\left(\widehat{\beta}_{2,\mathrm{OLS}}\right)} \overset{a}{\sim} \chi^2_1$$

wobei $\widehat{\beta}_{2,\mathrm{IV}}$ der Instrumentvariablen und $\widehat{\beta}_{2,\mathrm{OLS}}$ der OLS-Schätzer ist.

Für die Berechnung der Varianzen sollten in beiden Fällen die gleiche Schätzung für $\hat{\sigma}^2$ herangezogen werden!

Wir verwerfen die Nullhypothese und schließen auf eine kontemporäre Korrelation zwischen x_i und ε_i wenn $H > \chi^2_c$ (für ein 5% Signifikanzniveau ist der kritische Wert $\chi^2_c = 3.84$).

Hausman-Tests werden vor allem verwendet für die Erkennung von

- Simulanitäts-Bias
- Messfehlern
- Omitted Variables

oder allgemeiner, in Fällen in denen die Störterm ε mit erklärenden Variablen x korreliert ist.

Eine einfache Version dieses Tests mittels Hilfsregression wurde von Davidson and MacKinnon (1989) vorgeschlagen. Die Idee beruht einfach darauf, dass eine z.B. möglicherweise fehlerhaft gemessene Variable in einer ersten Stufe auf ein Instrument regressiert wird, und die Residuen dieser Hilfsregression in der ursprünglichen Regression als zusätzlicher Regressor berücksichtigt wird. Wenn der Koeffizient dieser Residuen signifikant von Null verschieden ist wird dies als Hinweis dafür interpretiert, dass die eine Korrelation zwischen x und ε (möglicherweise hervorgerufen durch Messfehler, fehlende Variablen, Endogenität) einen Einfluss auf das Schätzergebnis hat. Da es ein asymptotischer Test ist wird die Normalverteilung angenommen (bzw. wenn mehrere Variablen getestet werden die χ^2 Verteilung), aber es hat sich eingebürgert die meist einfacher im Regressionsoutput verfügbare t-Statistik (bzw. F-Statistik) heranzuziehen. Aufgrund der asymptotischen Gültigkeit macht dies keinen großen Unterschied (für nähere Hinweise zum Wu-Hausman Test siehe z.B. Johnston, DiNardo 1997, S. 257ff).

Angenommen, das Modell in Strukturform sei

$$y_1 = \beta_1 + \beta_2 y_2 + \beta_4 x_1 + \beta_5 x_2 + \varepsilon$$

und wir vermuten, dass y_2 endogen ist, d.h. mit ε korreliert ist. Wenn wir für y_2 über ein Instrument z verfügen gehen wir 2-stufig vor

1. Man schätzt eine *reduzierte Form* für y_2

   $$y_2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \hat{\alpha}_3 z + \hat{\nu}$$

   und berechnet daraus die Residuen $\hat{\nu}$.

2. Man schätzt die ursprüngliche Strukturform und verwendet die Stichprobenresiduen der reduzierten Form als *zusätzlichen* Regressor.

   $$y_1 = \widehat{\beta}_1 + \widehat{\beta}_2 y_2 + \widehat{\beta}_4 x_1 + \widehat{\beta}_5 x_2 + \widehat{\beta}_6 \hat{\nu} + \tilde{\varepsilon}$$

   Wenn $\widehat{\beta}_6$ signifikant von Null verschieden ist schließen wir daraus, dass y_2 tatsächlich endogen ist, OLS also verzerrte Ergebnisse liefert.

Dieser Test kann einfach für mehrere endogene Variablen verallgemeinert werden. Man berechnet für jede endogene Variable die Residuen der reduzierten Form, verwendet diese als zusätzliche Regressoren in der Strukturform, und testet deren gemeinsame Signifikanz mit Hilfe eines einfachen F-Tests.

Beispiel: Dies soll anhand der Abbildung 11.2 zugrunde liegenden Daten veranschaulicht werden.

Die einfache OLS Schätzung lieferte die verzerrten Schätzer

$$\begin{array}{rclcl} \text{C} & = & -9.175 & + & 0.808\ \text{Y} \\ & & (5.685) & & (0.024)^{***} \end{array}$$

$$R^2 = 0.934, \quad s = 10.818, \quad F\text{-Stat} = 1096.971, \quad n = 80$$
(Standardfehler in Klammern)

Wir berechnen aus der Hilfsregression

$$\begin{array}{rclcl} \text{Y} & = & 105.804 & + & 2.305\ \text{INV} \\ & & (16.049)^{***} & & (0.293)^{***} \end{array}$$

$$R^2 = 0.442, \quad s = 37.49, \quad F\text{-Stat} = 61.889, \quad \text{DW} = 2.073, \quad n = 80$$
(Standardfehler in Klammern)

die Residuen Res dieser Hilfsregression, und anschließend die eigentlich interessierende Regression

$$\begin{array}{rclclcl} \text{CONS} & = & 45.901 & + & 0.566\ \text{Y} & + & 0.434\ \text{Res} \\ & & (0.000)^{***} & & (0.000)^{***} & & (0.000)^{***} \end{array}$$

$$R^2 = 0.999, \quad s = 0, \quad F\text{-Stat} = 4.77\text{e}{+}13, \quad n = 80$$
(Standardfehler in Klammern)

Der t-Wert von Res ist riesengroß und der p-Wert entsprechend klein, es kann also kaum ein Zweifel an der Endogenität von Y bestehen!

Die Anwendung der 2SLS Methode mit den Investitionen als Instrument liefert eine deutlich bessere Schätzung der wahren Werte $\beta_1 = 60$ und $\beta_2 = 0.5$

$$\begin{array}{rclcl} \text{C} & = & 45.901 & + & 0.566\ \text{Y} \\ & & (12.687)^{***} & & (0.055)^{***} \end{array}$$

$$R^2 = 0.85, \quad s = 16.264, \quad F\text{-Stat} = 105.402, \quad n = 80$$
(Standardfehler in Klammern)

Das folgende EViews Programm produzierte diese Ergebnisse:

```
wfopen http://www.uibk.ac.at/econometrics/data/iv_keynesgraph.xls
equation eq_ols.ls Cons c Y
equation eq_2sls.tsls Cons c Y @ Inv
'Hausman:
equation eq1.ls Y c Inv
eq1.makeresids r
equation eq_h.ls Cons c Y r
show eq_h
```

Test auf überidentifizierende Restriktionen

Instrumente müssen *relevant* und *exogen* sein. Während ein F-Test auf die gemeinsame Signifikanz aller Instrumente in der 1. Stufe Hinweise auf die Relevanz der Instrumente liefert kann die Korrelation zwischen den Instrumenten und den unbeobachtbaren Störtermen der Grundgesamtheit (d.h. die Exogenität) nicht unmittelbar getestet werden.

Allerdings gibt es einen Test für überidentifizierte Systeme. Angenommen, die Strukturgleichung sei

$$y_{i1} = \beta_1 + \beta_2 y_{i2} + \beta_3 x_i + \varepsilon_i$$

und wir verfügen über zwei potentielle Instrumentvariablen z_1 und z_2.

Wir könnten diese Gleichung mit nur einer Instrumentvariable 2-stufig schätzen, z.B. mit z_1, daraus die Residuen $\hat{\varepsilon}_1$ berechnen und überprüfen, ob diese mit der nicht verwendeten Instrumentvariable z_2 korreliert sind. Wenn $\text{corr}(\hat{\varepsilon}_1, z_2) \neq 0$ wäre offensichtlich z_1 *oder* z_2 ein schlechtes Instrument! Umgekehrt könnte auch mit z_2 als Instrument begonnen werden und die Korrelation der Residuen dieser 2SLS Schätzung mit z_1 überprüft werden.

Ein einfacher Test auf überidentifizierende Restriktionen kann wieder mit Hilfe einer Hilfsregression durchgeführt werden:

1. Schätze die überidentifizierte Strukturgleichung mit 2SLS und berechne daraus die Stichproben-Residuen $\hat{\varepsilon}$, z.B.

   $$y_{i1} = \widehat{\beta}_1 + \widehat{\beta}_2 \widehat{y}_{i2} + \widehat{\beta}_3 x_i + \hat{\varepsilon}_i \qquad \rightarrow \quad \text{Stichprobenresiduen } \hat{\varepsilon}$$

2. Schätze eine Hilfsregression von $\hat{\varepsilon}$ auf alle exogenen Variablen und Instrumente und bestimme daraus das Bestimmtheitsmaß R^2_*, z.B.

   $$\hat{\varepsilon}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_{i1} + \hat{\alpha}_2 z_{i2} + \hat{\alpha}_3 x_i + \upsilon_i \qquad \rightarrow \qquad R^2_*$$

3. Unter der Nullhypothese, dass *alle* Instrumentvariablen exogen sind, ist

   $$nR^2_* \overset{a}{\sim} \chi^2_q$$

   wobei n die Anzahl der Beobachtungen ist. Die Anzahl der Freiheitsgrade q ist die Anzahl der Instrumentvariablen außerhalb des Modells abzüglich der Anzahl der endogenen erklärenden Variablen, d.h. der Grad der ‘Überidentifikation’.

Wenn der Wert von nR^2_* größer ist als der entsprechende kritische Wert der χ^2-Verteilung muss die Nullhypothese, dass alle Instrumente exogen sind, verworfen werden.

Dieser Test ist ein Spezialfall des *Sargan-Tests* und ist nur für überidentifizierte Gleichungen anwendbar. Er gibt allerdings keine Hinweise darauf, *welche* Instrumente möglicherweise endogen sind!

10.4.3 IV-Schätzer für den multivariaten Fall bei exakter Identifikation

Für den den multivariaten Fall haben wir bereits gezeigt

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

woraus folgt

$$\text{plim}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \text{plim}\left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \cdot \text{plim}\left(\frac{1}{n}(\boldsymbol{X}'\boldsymbol{\varepsilon})\right)$$

Wenn wir wieder annehmen, dass $\text{plim}(\boldsymbol{X}'\boldsymbol{X}/n) = \boldsymbol{\Sigma}_{XX}$ eine positiv definite Matrix mit vollem Rang ist und $\text{plim}(\boldsymbol{X}'\boldsymbol{\varepsilon}/n) = \boldsymbol{\Sigma}_{X\boldsymbol{\varepsilon}} \neq 0$, dann ist

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{XX}^{-1} \cdot \boldsymbol{\Sigma}_{X\varepsilon}$$

kein konsistenter Schätzer, da einer oder mehrere der Regressoren mit dem Störterm $\boldsymbol{\varepsilon}$ der Grundgesamtheit korreliert ist.

Das Modell ist also $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit $\text{var}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$, aber mit $\text{plim}\boldsymbol{X}'\boldsymbol{\varepsilon}/n \neq 0$!

Die Methode der Instrumenten-Variablen liefert zumindest konsistente Schätzer, wann immer der Störterm $\boldsymbol{\varepsilon}$ mit den erklärenden Variablen $\boldsymbol{X}$ korreliert ist. Wir fassen die Instrumente in einer $n \times k$ Matrix $\boldsymbol{Z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k)$ zusammen, wobei für jede erklärende Variable ein Instrument mit n Beobachtungen existieren muss.[8] Diese Matrix $\boldsymbol{Z}$ darf nicht singulär sein und muss folgende Eigenschaften erfüllen:

$$\begin{aligned}
\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\right) &= 0 \\
\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right) &= \boldsymbol{\Sigma}_{ZX} \quad \text{existiert und ist nicht singulär.} \\
\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{y}\right) &= \boldsymbol{\Sigma}_{Zy} \quad \text{existiert und ist nicht singulär}
\end{aligned}$$

Die erste Bedingung garantiert, dass die Korrelation zwischen Störterm und Instrumenten–Variable *'in Wahrscheinlichkeit gegen Null konvergiert'*. Die zweite Annahme garantiert eine positive Korrelation zwischen den X und den Z. Die dritte Gleichung ist lediglich eine Definition und unproblematisch.

Wir prämultiplizieren die ursprüngliche Gleichung $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit der transponierten Instrument-Matrix $\boldsymbol{Z}$ und erhalten

$$\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}'\boldsymbol{\varepsilon}$$

anschließend dividieren wir beide Seiten durch n und bilden das *probability limit*

$$\begin{aligned}
\text{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{y}}{n} &= \text{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\boldsymbol{\beta} + \text{plim}\,\frac{\boldsymbol{Z}'\boldsymbol{\varepsilon}}{n} \\
\boldsymbol{\Sigma}_{Zy} &= \boldsymbol{\Sigma}_{ZX}\boldsymbol{\beta} + \boldsymbol{0}
\end{aligned}$$

[8] Allerdings kann (und muss) eine Variable x, die nicht mit dem Störterm korreliert ist, als ihr eigenes Instrument verwendet werden!

da $\text{plim}(\boldsymbol{Z}'\boldsymbol{\varepsilon}/n) = 0$ können wir nach $\boldsymbol{\beta}$ lösen

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{ZX}^{-1}\boldsymbol{\Sigma}_{Zy}$$

Auf diese Gleichung können wir die Methode der Momente anwenden und die Momente der Grundgesamtheit $\boldsymbol{\Sigma}_{yX}$ und $\boldsymbol{\Sigma}_{Zy}$ durch die aus der Stichprobe berechneten Momente $\boldsymbol{y}'\boldsymbol{X}/n$ und $\boldsymbol{Z}'\boldsymbol{X}/n$ ersetzen. Dies gibt uns den Instrumentvariablen-Schätzer

$$\hat{\boldsymbol{\beta}}^{\text{IV}} = \left(\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{Z}'\boldsymbol{y}}{n} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

Wie man einfach sehen kann resultiert der OLS–Schätzer als Spezialfall, wenn $\boldsymbol{Z} = \boldsymbol{X}$, d.h. wenn jede x Variable als Instrument für sich selbst dient. Dies ist erlaubt, wenn die x exogen sind. Wenn hingegen die Instrumente überhaupt nicht mit den ursprünglichen Variablen korreliert sind, d.h. wenn $\boldsymbol{Z}'\boldsymbol{X} = \boldsymbol{0}$, bricht dieser Ansatz zusammen. Aber auch wenn die Korrelation zwischen den x und z Variablen gering ist liefert dieser Ansatz sehr schlechte Resultate ($\boldsymbol{Z}'\boldsymbol{X}$)!

Um die Konsistenz zu prüfen gehen wir wie üblich vor. Wir wissen, dass

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\text{IV}} &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{X})\boldsymbol{\beta} + (\boldsymbol{Z}'\boldsymbol{X})^{-1}(Z'\boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})
\end{aligned}$$

Daraus folgt

$$\begin{aligned}
\text{plim}(\hat{\boldsymbol{\beta}}^{\text{IV}}) &= \boldsymbol{\beta} + \text{plim}\left[(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})\right] \\
&= \boldsymbol{\beta} + \text{plim}\left[\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\right)\right] \\
&= \boldsymbol{\beta} + \boldsymbol{\Sigma}_{ZX}^{-1}\boldsymbol{0} = \boldsymbol{\beta}
\end{aligned}$$

womit gezeigt wurde, dass der Instrumentvariablen Schätzer tatsächlich konsistent ist.

Als nächstes müssen wir die asymptotische Varianz– Kovarianzmatrix $\boldsymbol{V}$ für $\hat{\boldsymbol{\beta}}^{\text{IV}}$ ermitteln. Unter Berücksichtigung von $\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{\varepsilon})$ erhalten wir

$$(\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}^{\text{IV}} - \boldsymbol{\beta})' = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{Z})^{-1}$$

und

$$\text{var}(\hat{\boldsymbol{\beta}}^{\text{IV}}) = \left[\frac{1}{n}\,\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\right]\left[\text{plim}\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{Z}\right)\right]\left[\text{plim}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)^{-1}\right]$$

woraus unter Bezugnahme auf die obigen Annahmen über die Instrumente folgt

$$\text{var}(\hat{\boldsymbol{\beta}}^{\text{IV}}) = \frac{1}{n}\sigma^2\boldsymbol{\Sigma}_{ZX}^{-1}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\Sigma}_{ZX}^{-1}$$

Praktisch erfolgt die Schätzung einer konsistenten Varianz–Kovarianzmatrix durch

$$\hat{\sigma}^2(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{X}'\boldsymbol{Z})^{-1}$$

wobei $\hat{\sigma}^2$ (ein konsistenter Schätzer für σ^2) folgendermaßen ermittelt wird:

$$\hat{\sigma}^2 = \frac{1}{n-k}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{IV}}\right)'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\mathrm{IV}}\right) = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-k}$$

wobei es keine große Rolle spielt, ob wir durch n oder $n-k$ dividieren, da es sich um einen asymptotischen Test handelt.

Man beachte, dass all dies nur anwendbar ist, wenn für jede potentiell endogene x Variable genau eine Instrumentvariable zur Verfügung steht, weil sonst $\boldsymbol{Z}'\boldsymbol{X}$ nicht definiert ist! Der GIVE Schätzer ist auch für überidentifizierte Systeme anwendbar.

10.4.4 Der Verallgemeinerte Instrumentvariablenschätzer (GIVE)

GIVE (*'Generalized Instrumental Variable Estimator'*) ist ein Instrumentvariablenschätzer für den Fall, dass die Anzahl der Instrumente größer ist als die Zahl der Regressoren.

Bisher haben wir stets angenommen, dass wir exakt gleich viele Instrumente wie Regressoren haben, d.h. dass die Instrumenten Matrix $\boldsymbol{Z}$ und die Regressor Matrix $\boldsymbol{X}$ beide die gleiche Dimension ($n \times k$) haben, wobei exogene X als Instrumente für sich selbst verwendet werden.

Wenn die Anzahl der Instrumente l größer ist als die Anzahl der X-Variablen k kann man die ursprüngliche Gleichung $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit $\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ vormultiplizieren

$$\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}$$

Wenn der letzte Term in Wahrscheinlichkeit gegen Null konvergiert und wir mit der Matrix $[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}']^{-1}$ vormultiplizieren erhalten wir nach Anwendung der Methode der Momente einen allgemeineren Instrumentvariablen Schätzer

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = \left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$$

Wenn wir $\boldsymbol{P}_Z := \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ definieren können wir dies einfacher anschreiben

$$\hat{\boldsymbol{\beta}}^{\mathrm{IV}} = [\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y}$$

Die entsprechende Varianz-Kovarianzmatrix der Parameter ist

$$\mathrm{var}(\hat{\boldsymbol{\beta}}^{\mathrm{IV}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X})^{-1}$$

wobei ein Schätzer $\hat{\sigma}^2$ für σ^2 wie üblich aus den Residuen berechnet werden kann (zeigen Sie, dass $\boldsymbol{P}_Z$ symmetrisch und idempotent ist).

Die Intuition dahinter wird klarer, wenn man beachtet, dass $\boldsymbol{P}_Z := \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ eine Projektionsmatrix ist!

Wenn die die Matrix $\boldsymbol{X}$ mit der Projektionsmatrix $\boldsymbol{P}_Z$ vormultipliziert wird, werden die $\boldsymbol{X}$ in den Spaltenraum der Instrumente $\boldsymbol{Z}$ projiziert, d.h.

$$\widehat{\boldsymbol{X}} = \boldsymbol{P}_Z\boldsymbol{X}$$

Daraus folgt eine alternative Möglichkeit den den IV Schätzer herzuleiten. Man kann eine einfache Datentransformation vornehmen, indem man das Modell $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ mit der Projektionsmatrix $\boldsymbol{P}_Z$ vormultipliziert

$$\underbrace{\boldsymbol{P}_Z\boldsymbol{y}}_{\widehat{\boldsymbol{y}}} = \underbrace{\boldsymbol{P}_Z\boldsymbol{X}}_{\widehat{\boldsymbol{X}}}\boldsymbol{\beta} + \underbrace{\boldsymbol{P}_Z\boldsymbol{\varepsilon}}_{\widehat{\varepsilon}}$$

und für die derart transformierten Daten $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{X}}\boldsymbol{\beta} + \widehat{\varepsilon}$ den OLS Schätzer anwendet. Der OLS Schätzer auf die derart transformierten Daten ist der IV Schätzer

$$\hat{\boldsymbol{\beta}}^{\text{IV}} = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\widehat{\mathbf{y}}$$

weil

$$\begin{aligned}
[(\boldsymbol{P}_Z\boldsymbol{X})'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}(\boldsymbol{P}_Z\boldsymbol{X})'\boldsymbol{P}_Z\boldsymbol{y} &= \\
[\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{y} &= \\
[\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y} &= \\
\left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} &= \hat{\boldsymbol{\beta}}^{\text{IV}}
\end{aligned}$$

($\boldsymbol{P}_Z$ ist symmetrisch und idempotent, d.h. $\boldsymbol{P}_Z = \boldsymbol{P}_Z'$ und $\boldsymbol{P}_Z\boldsymbol{P}_Z = \boldsymbol{P}_Z$)
Die geschätzte Varianz-Kovarianzmatrix des Koeffizientenvektors $\hat{\boldsymbol{\beta}}$ ist

$$\begin{aligned}
\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) &= \hat{\sigma}^2(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1} \\
&= \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X})^{-1} \\
&= \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X})^{-1}
\end{aligned}$$

Achtung 1: Die IV Residuen, die u.a. zur Berechnung von $\hat{\sigma}^2$ benötigt werden, werden als

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{IV}}$$

berechnet, und nicht wie manchmal irrtümlich angenommen aus $\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{\beta}}^{\text{IV}}$ oder $\boldsymbol{y} - \widehat{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^{\text{IV}}$!

Achtung 2: Die meisten Programme geben für Instrumentvariablenschätzer ein Bestimmtheitsmaß R^2 aus, aber dieses R^2 darf *nicht* wie beim OLS-Schätzer als Anteil der durch die Regressoren erklärten Variation von y interpretiert werden! Außerdem kann das R^2 von Instrumentvariablenschätzungen negativ sein!

Instrumentvariablenschätzer und die Methode 2-stufigen Kleinsten Quadrate (2SLS)

Da $\boldsymbol{P}_Z$ eine Projektionsmatrix ist kann der vorhin entwickelte IV Schätzer auch als eine zweistufige Anwendung des OLS-Schätzers begriffen werden

- **1. Stufe:** Regressiere jede der Variablen in der X-Matrix auf die Instrumente $\boldsymbol{Z}$ und forme daraus die Matrix der gefitteten Werte $\widehat{\boldsymbol{X}}$

  $$\widehat{\boldsymbol{X}} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X} = \boldsymbol{P}_Z\boldsymbol{X}$$

  mit $\boldsymbol{P}_Z = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$, wobei die Projektionsmatrix $\boldsymbol{P}_Z$ symmetrisch und idempotent ist.

- **2. Stufe:** Regressiere $\boldsymbol{y}$ auf die gefitteten $\widehat{\boldsymbol{X}}$ um um einen Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ zu erhalten

  $$\begin{aligned}
  \hat{\boldsymbol{\beta}}_{\text{2SLS}} &= (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}(\widehat{\boldsymbol{X}}'\boldsymbol{y}) \\
  &= \left[\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{P}_Z\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z'\boldsymbol{y} \\
  &= \left[\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{P}_Z\boldsymbol{y} \\
  &= \left[\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y} \\
  &= \hat{\boldsymbol{\beta}}^{\text{IV}}
  \end{aligned}$$

Deshalb ist der 2SLS Schätzer ein Spezialfall des IV Schätzers.

Literaturverzeichnis

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

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

Davidson, R. and MacKinnon, J. G. (1989), 'Testing for consistency using artificial regressions', *Econometric Theory* **5**(3), 363–384.
URL: *http://www.jstor.org/stable/3532374*

Greene, W. (2003), *Solutions Manual: Econometric Analysis*, 5th edn, Pearson Education.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica* **46**(6), 1251–1271.
URL: *http://www.jstor.org/stable/1913827*

Levitt, S. D. (1997), 'Using electoral cycles in police hiring to estimate the effect of police on crime', *The American Economic Review* **87**(3), 270–290.
URL: *http://www.jstor.org/stable/2951346*

Murray, M. P. (2006), 'The Bad, the Weak, and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation', *Journal of Economic Perspectives* **20**(4), 111–132.

Staiger, D. and Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.