

Kapitel 9

Autokorrelation

“There is always an easy solution to every human problem — neat, plausible and wrong.”
(H.L. Mencken)

Autokorrelation bedeutet ‘mit sich selbst korreliert’, das heißt, verschiedene Beobachtungen *einer* Variable sind *untereinander* korreliert. Damit ein solches Muster interpretierbar ist, muss die Reihenfolge der Beobachtungen einer logischen Ordnung gehorchen, wie dies zum Beispiel bei Zeitreihen der Fall ist. Weil Autokorrelation in erster Linie ein Zeitreihenproblem ist werden wir in diesem Kapitel anstelle des für Querschnittsdaten üblichen Index i den Index t (für *time*) für die Beobachtungen verwenden, mit $t = 1, 2, \dots, T$, wobei T die Anzahl der Beobachtungen bezeichnet (analog zu n für Querschnittsdaten).

Bei Autokorrelation sind also die Werte einer Variable zum Zeitpunkt t mit Werten dieser Variable in Vorperioden $t - 1, t - 2, t - 3, \dots$ korreliert.

Zum Beispiel sind die Konsumausgaben der Periode t häufig mit den Konsumausgaben der Vorperiode $t - 1$ korreliert. Damit ist eine Annahme des ‘*random sampling*’ verletzt, die Ziehungen sind nicht unabhängig.

Wenn in der Ökonometrie einfach von Autokorrelation (‘*serial correlation*’) gesprochen wird bezieht sich dies fast immer auf eine *Autokorrelation der Störterme* ε_t .

Im einfachsten Fall ist jeder Störterm mit dem Störterm der Vorperiode korreliert ist, das heißt

$$\text{cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$$

Selbstverständlich können auch ‘weiter auseinanderliegende’ Störterme untereinander korreliert sein

$$\text{cov}(\varepsilon_t, \varepsilon_{t-p}) := E[(\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t-p} - \bar{\varepsilon})] = E(\varepsilon_t \varepsilon_{t-p}) \neq 0 \quad \text{für } p = 1, 2, \dots$$

wobei p den *time lag* (Zeitverzögerung) bezeichnet. Wenn $p = 1$ ist impliziert dies, dass jede Beobachtung mit der Beobachtung der Vorperiode ($t - 1$) korreliert ist; wenn z.B. $p = 3$ ist die 4. mit der 1. die 5. mit der 2. Beobachtung usw. korreliert. Man beachte, dass durch die Lag Bildung p Beobachtungen am Anfang der Zeitreihe verloren gehen.

Exkurs: Bildung von zeitverzögerten Variablen (*time lags*) und ersten Differenzen:

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	$\Delta y_t := y_t - y_{t-1}$
1	12	–	–	–	–
2	9	12	–	–	–3
3	14	9	12	–	5
4	16	14	9	12	2
5	20	16	14	9	4



Exkurs:

Wir haben schon früher erwähnt, dass durch eine logarithmische oder Potenz-Transformationen von Variablen manchmal eine *Stabilisierung der Varianz* erreicht werden kann.

Ähnlich kann bei trendbehafteten Zeitreihendaten manchmal durch *Differenzenbildung* eine **Stabilisierung des Mittelwertes** erreicht werden.

Bildung erster Differenzen:

$$\begin{aligned}
 y_t &= \beta_1 + \beta_2 x_{t,2} + \beta_3 x_{t,3} + \cdots + \beta_k x_{t,k} + \varepsilon_t \\
 y_{t-1} &= \beta_1 + \beta_2 x_{t-1,2} + \beta_3 x_{t-1,3} + \cdots + \beta_k x_{t-1,k} + \varepsilon_{t-1} \quad / - \\
 \Delta y_t &= \beta_2 \Delta x_{t,2} + \beta_3 \Delta x_{t,3} + \cdots + \beta_k \Delta x_{t,k} + \Delta \varepsilon_t
 \end{aligned}$$

\Rightarrow Interzept fällt raus!

Erste Differenzen mit Trend

$$\begin{aligned}
 y_t &= \beta_1 + \alpha \text{Trend}_t + \beta_2 x_{t,2} + \beta_3 x_{t,3} + \cdots + \beta_k x_{t,k} + \varepsilon_t \\
 y_{t-1} &= \beta_1 + \alpha \text{Trend}_{t-1} + \beta_2 x_{t-1,2} + \beta_3 x_{t-1,3} + \cdots + \beta_k x_{t-1,k} + \varepsilon_{t-1} \quad / - \\
 \Delta y_t &= \alpha + \beta_2 \Delta x_{t,2} + \beta_3 \Delta x_{t,2} + \cdots + \beta_k \Delta x_{t,k} + \Delta \varepsilon_t
 \end{aligned}$$

\Rightarrow Koeffizient von Trend erscheint als Interzept!
(Beachte: $\text{Trend}_t - \text{Trend}_{t-1} = 1 \quad \forall t$)



Exkurs: Autokorrelation kann häufig schon in einem Residuenplot erkannt werden. Im häufigeren Fall von positiver Autokorrelation sind ‘Cluster’ von positiven und negativen Residuen zu beobachten, wenn eine Beobachtung über (unter) der Regressionsgeraden liegt, liegt die nächste Beobachtung mit hoher Wahrscheinlichkeit ebenfalls wieder über (unter) der Regressionsgerade.

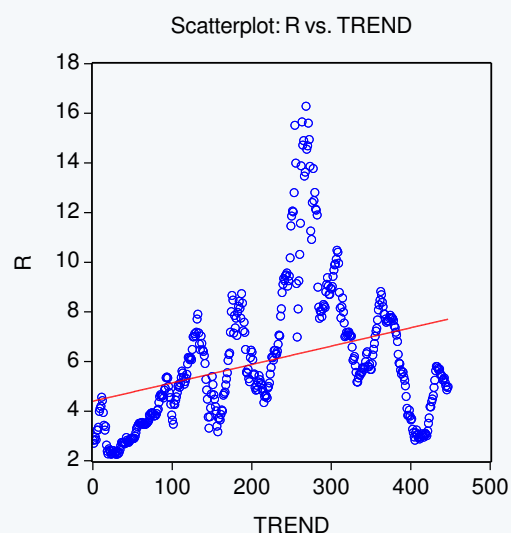
Beispiel: Kurzfristiger Zinssatz (IRS), Österreich 1970 – 2008 (OECD, Economic Outlook)

$$\text{IRS} = 7.5755 - 0.0258 \text{ TREND}$$

$$(23.0764) \quad (-7.0425)$$

$$R^2 = 0.243, \quad \text{DW} = 0.124, \quad T = 156$$

In diesem Fall beobachten wir positive Autokorrelation, wenn der Störterm in Periode t größer Null ist ($\epsilon_t > 0$), ist die Wahrscheinlichkeit für $\epsilon_{t+1} > 0$ größer als 0.5.



Die durch Autokorrelation verursachten Probleme ähneln in mehrerer Hinsicht denen bei Heteroskedastizität. Erinnern wir uns, die Gauss-Markov Annahmen über die Störterme verlangen

$$\epsilon_t \sim \text{i.i.d.}(0, \sigma^2)$$

d.h., dass die Störterme ‘*identical and independently distributed*’ sind. Bei Heteroskedastizität war die Annahme ‘*identical distributed*’ verletzt, d.h. die Varianz σ_i^2 war nicht für alle Beobachtungen gleich sondern in irgendeiner Form von den erklärenden Variablen abhängig. Bei Autokorrelation ist die Annahme ‘*independently distributed*’ verletzt, d.h., die Beobachtungen sind nicht unabhängig von der Reihenfolge der Ziehung.

Selbstverständlich können Heteroskedastizität und Autokorrelation auch gemeinsam auftreten. Heteroskedastische und/oder autokorrelierte Störterme werden manchmal auch als ‘*non-spherical disturbances*’ bezeichnet.

Wir werden später sehen, dass auch die Auswirkungen von Autokorrelation denen von Heteroskedastizität sehr ähnlich sind, d.h. die OLS-Schätzer bleiben zwar *erwartungstreu und konsistent*, sind aber *nicht mehr effizient*. Die mittels OLS geschätzten Standardfehler der Koeffizienten sind darüber hinaus verzerrt, d.h. die Teststatistiken (t- und F-Statistiken) sind bei Vorliegen von Autokorrelation ungültig!

9.1 Autoregressive Prozesse 1. Ordnung, AR(1) Prozesse

Wie schon erwähnt beschäftigen wir uns in diesem Abschnitt ausschließlich mit einer *Autokorrelation der Störterme*.

Im einfachsten Fall ist jeder Störterm mit dem Störterm der Vorperiode korreliert, d.h. wir gehen von folgendem datengenerierenden Prozess aus:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t \quad \text{mit } v_t \sim \text{i.i.d.}(0, \sigma_v^2) \end{aligned}$$

(v ist der griechische Buchstabe Upsilon.) Alle restlichen Gauss-Markov Annahmen seien erfüllt.

In diesem Fall folgen die Störterme einem *autoregressiven Prozess 1. Ordnung*, bzw. AR(1), da jeder Störterm nur mit dem Störterm der Vorperiode korreliert ist.

Allgemeiner spricht man von einem autoregressiven Prozess der Ordnung p AR(p), wenn

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + v_t$$

aber wir wollen uns in diesem Abschnitt auf den einfacheren Fall von AR(1) Prozessen $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$ beschränken.

Wenn wir im restlichen Abschnitt einfach von *Autokorrelation* sprechen, so meinen wir damit, dass die *Störterme* ε_t einem autoregressiven Prozess 1. Ordnung folgen. Außerdem wollen wir annehmen, dass die restlichen Gauss-Markov Annahmen (wie z.B. $E(\varepsilon_t) = 0$ und Homoskedastizität $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$) erfüllt seien.

Die Autokorrelation kann positiv oder negativ sein, je nach dem Vorzeichen von ρ (siehe Abbildung 9.1). In makroökonomischen Zeitreihen tritt positive Autokorrelation weitaus häufiger auf als negative Autokorrelation.

9.1.1 Mögliche Ursachen für Autokorrelation

Wir wissen, dass die Vergangenheit häufig Auswirkungen auf die Gegenwart und Zukunft hat, die nicht im systematischen Teil einer Regressionsgleichung erfasst werden kann. Dies kann dazu führen, dass die über die Zeit gesammelten Beobachtungen nicht voneinander unabhängig sind, also zu Autokorrelation führen.

Zum Beispiel wirken sich Ereignisse wie Erdbeben, Börsencrashes oder andere ‘Schocks’ häufig nicht nur in der Periode des Auftretens auf die Wirtschaft aus, sondern auch in den Folgeperioden, aber häufig werden diese Auswirkungen über die Zeit hinweg schwächer. Wenn diese Einflüsse nicht explizit gemessen und in den x -Variablen berücksichtigt werden, sind diese Effekte in den Störtermen enthalten und führen dort zu Autokorrelation. Generell führt *Persistenz* in ökonomischen Zeitreihen häufig zu autokorrelierten Residuen.

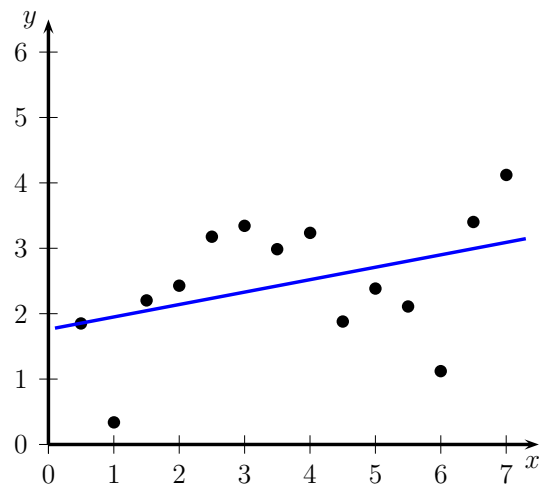
Weitere mögliche Ursachen für Autokorrelation sind

- **Trägheit** (*Inertia*) in der Anpassung: z.B. Konjunkturzyklen.

Positive Autokorrelation: $\rho = +0.8$

$$y_i = 0.5 + 0.7x_i + \varepsilon_i,$$

$$\varepsilon_i = 0.8\varepsilon_{i-1} + v, \quad v \sim N(0, 1)$$



Negative Autokorrelation: $\rho = -0.9$

$$y_i = 0.5 + 0.7x_i + \varepsilon_i$$

$$\varepsilon_i = -0.9\varepsilon_{i-1} + v, \quad v \sim N(0, 1)$$

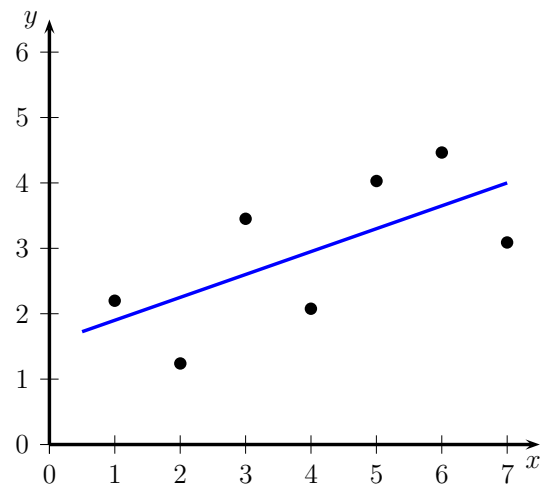


Abbildung 9.1: Autokorrelierte Störterme: Die Störterme sind untereinander korreliert, d.h. $\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0$.

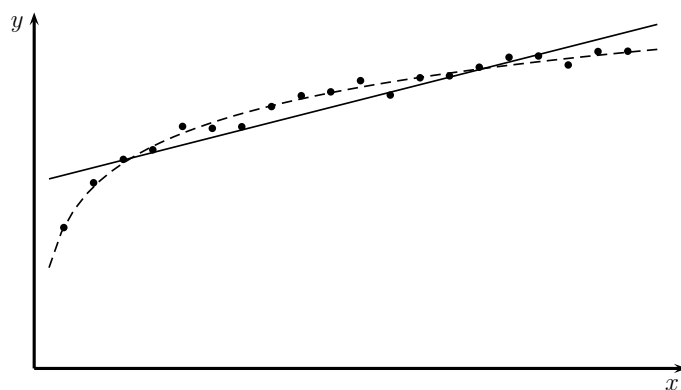


Abbildung 9.2: Eine fehlspezifizierte Funktionsform kann zu Autokorrelation in den Residuen führen.

- **Fehlspezifikation, falsche Funktionsform:** wenn z.B. der wahre Zusammenhang log-linear ist und eine lineare Funktion geschätzt wird (vgl. Abbildung 9.2).
- **Fehlspezifikation, fehlende Variablen:** Der Störterm repräsentiert den Einfluss aller *nicht berücksichtigten* erklärenden Variablen. Wir erwarten, dass der Einfluss dieser Variablen gering ist und dass sie sich in ihrer Wirkung im Durchschnitt gegenseitig aufheben. Wenn sich die ‘ausgelassenen’ Variablen aber sehr ähnlich verhalten kann dies zu Autokorrelation führen.
- **Messfehler** in den abhängigen Variablen.
- **Cobweb-Phänomen** (Schweinezyklus)
- **Daten-Transformationen** (z.B. Glättungs- oder Filterverfahren, Saisonsbereinigungen).

Autokorrelation kann prinzipiell sowohl bei Zeitreihen- als auch Querschnittsdaten auftreten, spielt aber bei Zeitreihenanalysen eine weit größere Rolle. Für Zeitreihendaten stellt Autokorrelation vermutlich das mit Abstand häufigste Problem dar. Bei Querschnittsdaten ist nur dann auf Autokorrelation zu achten, wenn die Reihenfolge der Beobachtungen einer bestimmten logischen Ordnung gehorcht. Zum Beispiel können Daten für regionale Einheiten, wie z.B. Gemeinden oder Bezirke, *räumliche Autokorrelation* (*‘spatial autocorrelation’*) aufweisen.

9.1.2 Stationarität

Da im Fall von Autokorrelation die Annahme der Unabhängigkeit der Störterme verletzt ist, benötigen wir eine zusätzliche Annahme, nämlich dass der Autokorrelationskoeffizient ρ der Beziehung $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ zwischen minus und plus Eins liegt ($-1 < \rho < 1$), die sogenannte *Stationaritätsannahme*.

Wäre der Absolutbetrag von ρ größer als Eins würden die Störterme $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ im Zeitablauf immer größer werden und gewissermaßen ‘explodieren’, was offensichtlich für die meisten Zeitreihen nicht beobachtet wird.

In der Zeitreihenanalyse wird meistens von stochastischen Prozessen ausgegangen. Vereinfacht ausgedrückt ist ein stochastischer Prozess (*stochastic* oder *random process*) eine Folge von Zufallsvariablen in der Zeit, d.h. eine empirische Zeitreihe kann als Realisation *eines* stochastischen Prozesses angesehen werden (analog zu einer Stichprobenziehung aus der Grundgesamtheit bei Querschnittsdaten).

Wiederum vereinfacht ausgedrückt ist ein stochastischer Prozess **stationär**, wenn Mittelwert und Varianz über die Zeit konstant sind, und wenn die Kovarianz zwischen zwei Zeitpunkten nur von der Lag-Länge abhängt, nicht aber von dem Zeitpunkt, zu dem gemessen wird.

Konkret wird ein stochastischer Prozess $\{\varepsilon_t\}$ *schwach stationär* genannt, wenn

- $E(\varepsilon_t)$ unabhängig von t ist;
- wenn $\text{var}(\varepsilon_t)$ ebenfalls unabhängig von t und eine endliche positive Zahl ist ($\text{var}(\varepsilon_t) < \infty$), und

- wenn $\text{cov}(\varepsilon_t, \varepsilon_{t-p})$ nur eine Funktion von der Lag-Länge $t - p$, aber nicht von t oder p ist.

Für einen autoregressiven Prozess 1. Ordnung

$$\begin{aligned}y_t &= \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t\end{aligned}$$

stellt die **Stationaritätsbedingung** $-1 < \rho < 1$ sicher, dass die Auswirkungen verzögerter Störvariablen mit den Verzögerungen (Lags) abnehmen. Wäre dies nicht der Fall, würden die Schwankungen im Zeitablauf ständig zunehmen und das Modell wäre nicht ‘stabil’. Deshalb werden wir im Folgenden stets annehmen, dass die Stationaritätsbedingung erfüllt ist.

Eine Zeitreihe mit Autokorrelation hat ein ‘Gedächtnis’ (*memory*), das heißt, eine Zufallsstörung in der Periode t hat Auswirkungen auf die Zukunft, allerdings werden diese bei stationären Zeitreihen aufgrund von $-1 < \rho < 1$ im Zeitablauf schwächer und verschwinden schließlich zur Gänze.

Die Stationaritätsannahme garantiert, dass

$$\begin{aligned}\text{E}(\varepsilon_t) &= \text{E}(\varepsilon_{t-1}) = \text{E}(\varepsilon_{t-2}) = \dots = 0 \\ \text{var}(\varepsilon_t) &= \text{var}(\varepsilon_{t-1}) = \text{var}(\varepsilon_{t-2}) = \dots = \sigma^2\end{aligned}$$

Intuitiv kann man sich vorstellen, dass die Stationaritätsannahme sicher stellt, dass sich die Zukunft ähnlich verhält wie die Vergangenheit. Wenn dies nicht der Fall wäre, könnten wir aus den vergangenen Realisationen nichts über die Zukunft lernen.

Der Spezialfall $|\rho| = 1$, bzw. $\varepsilon_t = \varepsilon_{t-1} + v_t$, wird ‘*unit roots*’ genannt und hat v.a. in der makroökonomischen Literatur große Bedeutung. Es hat sich nämlich gezeigt, dass für viele makroökonomische Zeitreihen (z.B. das BIP oder der Konsumentenpreisindex) die Nullhypothese $|\rho| = 1$ häufig nicht verworfen werden kann, was weitreichende Konsequenzen für die Schätzung hat.

Eine intuitive Vorstellung von den Problemen vermittelt Abbildung 9.3. Dazu wurden 300 standardnormalverteilte Zufallsvariablen $v_t \sim N(0, 1)$ generiert (v ist der griechische Buchstabe upsilon, nicht zu verwechseln mit den lateinischen Buchstaben u oder v). Die Beobachtung 100 wurde auf 20 gesetzt ($v_{100} = 20$). Mit Hilfe dieser Zufallsvariable erzeugen wir drei Zeitreihen $\{u_t\}$, $\{v_t\}$ und $\{w_t\}$, die alle einem AR(1) Prozesse mit unterschiedlichem ρ folgen

$$\begin{aligned}u_t &= 0.5u_{t-1} + v_t \\ v_t &= 0.9v_{t-1} + v_t \\ w_t &= 1w_{t-1} + v_t\end{aligned}$$

Wie man aus Abbildung 9.3 erkennen kann fällt die erste Zeitreihe u mit $\rho = 0.5$ nach dem einmaligen Schock wieder ziemlich schnell auf ihr ursprüngliches Niveau zurück. Für die Zeitreihe v mit $\rho = 0.9$ geht dies nicht ganz so schnell, aber auch sie kehrt wieder auf ihr ursprüngliches Niveau zurück. Diese Tendenz, zum ursprünglichen Niveau zurückzukehren, wird in der Literatur ‘*mean reversion*’ genannt.

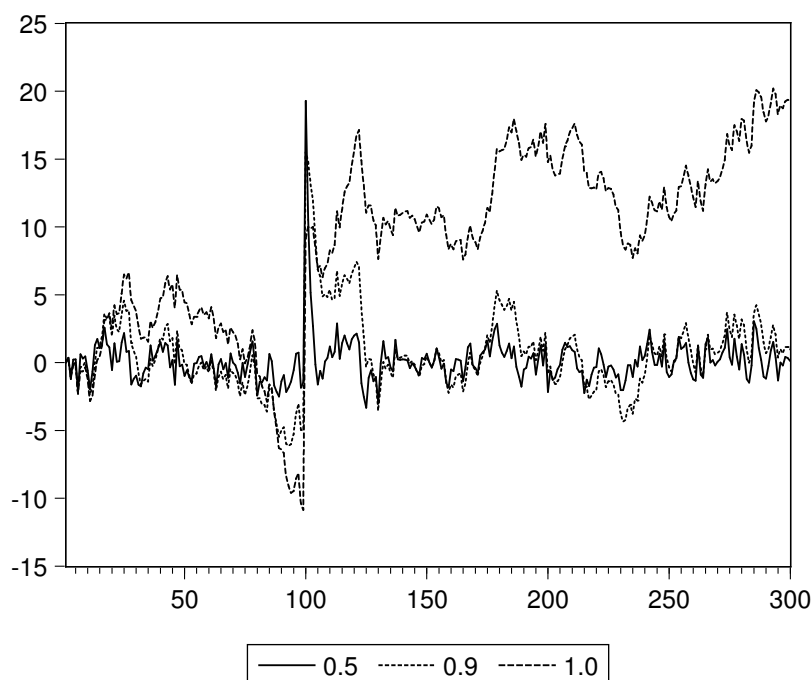


Abbildung 9.3: Drei künstlich erzeugte AR(1) Zeitreihenprozesse $x_t = \rho x_{t-1} + v_t$ mit $\rho = 0.5, 0.9, 1.0$. Bei Beobachtung 100 wurde der Störterm auf 20 gesetzt, d.h. $v_{100} = 20$ ('Schock').

Der AR(1) Prozess w mit $\rho = 1$ hingegen wandert schon vor dem Schock im Zeitpunkt $t = 100$ ziemlich erratisch umher und zeigt nach dem Schock $\varepsilon_{100} = 20$ keine Tendenz zur Rückkehr zum ursprünglichen Niveau. Aufgrund dieser Eigenschaft spricht man bei solchen Variablen auch von einem 'random walk', oder man sagt, die Variable folgt einem 'stochastischen Trend'.

Wenn man zwei solche 'unit roots' Variablen aufeinander regressiert passieren komische Dinge, die übliche statistische Intuition scheint außer Kraft gesetzt. Man kann zeigen, dass in solchen Fällen die Koeffizienten und Standardfehler systematisch verzerrt sind, d.h. die üblichen Teststatistiken sind nicht mehr anwendbar.

Um dies zu demonstrieren erzeugen wir zwei völlig unabhängige Zeitreihen

$$\begin{aligned} y_t &= y_{t-1} + v_t \\ x_t &= x_{t-1} + \nu_t \end{aligned}$$

(ν ist der griechische Buchstabe nu, gesprochen 'nü') mit $\text{cov}(v, \nu) = 0$ sowie $v_i \sim \text{i.i.d.}(0, \sigma_v^2)$, $\nu_i \sim \text{i.i.d.}(0, \sigma_\nu^2)$, Offensichtlich gibt es keine direkte Abhängigkeit zwischen y und x , deshalb würden wir auf den ersten Blick erwarten, dass in der Regression

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 x_t + \varepsilon_t$$

der Koeffizient $\hat{\beta}_2$ nicht signifikant von Null verschieden ist. Wenn man dieses Experiment aber wiederholt durchführt stellt man fest, dass die t-Statistik in sehr vielen

Fällen einen hochsignifikanten Koeffizienten $\widehat{\beta}_2$ anzeigt. Dies ist allerdings eine reine **Scheinkorrelation** (*spurious correlation*), die nur aus der *'unit root'* Eigenschaft $|\rho| = 1$ folgt. Wie man mit diesem Problem umgeht wird in der *Zeitreihenökonomie* ausführlich behandelt.¹

Glücklicherweise sind die Folgen weit weniger dramatisch, wenn die Zeitreihen stationär sind, wenn also $|\rho| < 1$ ist. Mit diesem Fall werden wir uns im Folgenden beschäftigen.

9.1.3 Eigenschaften von AR(1) Prozessen

Für die Berechnung der eigentlich interessierenden Varianz-Kovarianzmatrix der geschätzten Koeffizienten $\text{var}(\widehat{\beta}) = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]$ benötigen wir die Varianz-Kovarianzmatrix der Störterme $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$. Für den Spezialfall *ohne* Heteroskedastizität und Autokorrelation (also mit *'spherical disturbances'*) konnten wir zeigen, dass $\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}$ ist.

In diesem Abschnitt wollen wir nun die einzelnen Elemente der

$$\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{pmatrix} \text{var}(\varepsilon_1) & \text{cov}(\varepsilon_1, \varepsilon_2) & \cdots & \text{cov}(\varepsilon_1, \varepsilon_T) \\ \text{cov}(\varepsilon_2, \varepsilon_1) & \text{var}(\varepsilon_2) & \cdots & \text{cov}(\varepsilon_2, \varepsilon_T) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\varepsilon_T, \varepsilon_1) & \text{cov}(\varepsilon_T, \varepsilon_2) & \cdots & \text{var}(\varepsilon_T) \end{pmatrix}$$

Matrix für den Fall von Autokorrelation 1. Ordnung berechnen.

Wir beginnen mit dem Erwartungswert von ε_t . Wenn die Stationaritätsbedingung $-1 < \rho < 1$ erfüllt ist, sind die Erwartungswerte sowie die Varianzen und Kovarianzen im Zeitablauf konstant. Dies impliziert

$$\mathbf{E}(\varepsilon_t) = \mathbf{E}(\varepsilon_{t-1})$$

woraus für $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ folgt

$$\begin{aligned} \mathbf{E}(\varepsilon_t) &= \mathbf{E}(\rho\varepsilon_{t-1} + v_t) \\ &= \rho\mathbf{E}(\varepsilon_{t-1}) + \mathbf{E}(v_t) \\ &= \rho\mathbf{E}(\varepsilon_t) + 0 \\ (1 - \rho)\mathbf{E}(\varepsilon_t) &= 0 \\ \mathbf{E}(\varepsilon_t) &= 0 \end{aligned}$$

wobei ρ ist ein unbekannter Parameter der Grundgesamtheit ist. Das heißt, wenn der Störterm ε_t einem AR(1) Prozess folgt, und $v_t = \varepsilon_t - \rho\varepsilon_{t-1}$ alle Gauss-Markov Annahmen erfüllt, ist der Erwartungswert von ε_t gleich Null.

¹Bekanntere Tests auf diese *'unit root'* Eigenschaft sind z.B. der Dickey-Fuller (DF) Test oder Augmented Dickey-Fuller (ADF) Test.

Damit können wir uns nun der Varianz-Kovarianzmatrix $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$ zuwenden. Wir beginnen mit den Hauptdiagonalelementen:

$$\begin{aligned}\text{var}(\varepsilon_t) := \sigma_{\varepsilon_t}^2 &= E(\varepsilon_t^2) = E[(\rho\varepsilon_{t-1} + v_t)^2] \\ &= E[\rho^2\varepsilon_{t-1}^2 + 2\rho\varepsilon_{t-1}v_t + v_t^2] \\ &= \rho^2 E(\varepsilon_{t-1}^2) + E(v_t^2) \quad (\text{da } \varepsilon_{t-1} \text{ und } v_t \text{ unabh. sind}) \\ &= \rho^2 \text{var}(\varepsilon_t) + \text{var}(v_t) \quad (\text{Stationarität}) \\ &= \rho^2\sigma_\varepsilon^2 + \sigma_v^2 \quad (\text{da } \varepsilon_t \text{ homoskedast.})\end{aligned}$$

$$\text{var}(\varepsilon_t) := \sigma_\varepsilon^2 = \frac{\sigma_v^2}{1 - \rho^2}$$

Man beachte, dass weder σ_v^2 noch ρ einen Subindex t hat, das bedeutet, dass alle Hauptdiagonalelemente der Varianz-Kovarianzmatrix der Störterme ε den gleichen numerischen Wert haben, oder in anderen Worten, dass die Störterme *homoskedastisch* sind!

Wenden wir uns nun den Nebendiagonalwerten der Varianz-Kovarianzmatrix $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$ zu. Die Kovarianzen für den ersten *time-lag* können ähnlich berechnet werden:

$$\begin{aligned}\text{cov}(\varepsilon_t, \varepsilon_{t-1}) &= E(\varepsilon_t\varepsilon_{t-1}) \\ &= E[(\rho\varepsilon_{t-1} + v_t)\varepsilon_{t-1}] \\ &= E[\rho\varepsilon_{t-1}^2 + v_t\varepsilon_{t-1}] \\ &= \rho E(\varepsilon_{t-1}^2) \\ &= \rho \text{var}(\varepsilon_t) \\ \text{cov}(\varepsilon_t, \varepsilon_{t-1}) &= \rho\sigma_\varepsilon^2\end{aligned}$$

Man beachte, dass ρ unter der Stationaritätsannahme als Korrelationskoeffizient zwischen ε_t und ε_{t-1} interpretiert werden kann

$$\text{corr}(\varepsilon_t, \varepsilon_{t-1}) := \frac{\text{cov}(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{\text{var}(\varepsilon_t)\text{var}(\varepsilon_{t-1})}} = \frac{\rho\sigma_\varepsilon^2}{\sigma_\varepsilon^2} = \rho$$

weil aufgrund der Stationaritätsannahme $\text{var}(\varepsilon_t) = \text{var}(\varepsilon_{t-1}) = \dots = \sigma_\varepsilon^2$. Deshalb wird ρ auch *Autokorrelationskoeffizient* genannt. Dies gilt auch allgemeiner

$$\text{corr}(\varepsilon_t, \varepsilon_{t-p}) = \rho^p \sigma_\varepsilon^2 \quad (p > 0)$$

Für die Berechnung der weiteren Nebendiagonal-Elemente der Varianz-Kovarianzmatrix $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$, d.h. $\text{cov}(\varepsilon_t, \varepsilon_{t-2}), \text{cov}(\varepsilon_t, \varepsilon_{t-3}), \dots$, berücksichtigen wir, dass aus

$$\begin{aligned}\varepsilon_t &= \rho\varepsilon_{t-1} + v_t \\ \text{und } \varepsilon_{t-1} &= \rho\varepsilon_{t-2} + v_{t-1}\end{aligned}$$

durch Substitution folgt

$$\begin{aligned}\varepsilon_t &= [\rho(\rho\varepsilon_{t-2} + v_{t-1}) + v_t] \\ &= \rho^2\varepsilon_{t-2} + \rho v_{t-1} + v_t\end{aligned}$$

Also:

$$\begin{aligned} \text{cov}(\varepsilon_t, \varepsilon_{t-2}) &= \text{E} [(\rho^2 \varepsilon_{t-2} + \rho v_{t-1} + v_t) \varepsilon_{t-2}] \\ &= \rho^2 \sigma_\varepsilon^2 \end{aligned}$$

da $\text{E}(v_{t-1} \varepsilon_{t-2}) = \text{E}(v_t \varepsilon_{t-2}) = 0$ und $\text{E}(v_t) = 0$.

Analog dazu folgt durch weitere Substitution

$$\begin{aligned} \text{cov}(\varepsilon_t, \varepsilon_{t-3}) &= \rho^3 \sigma_\varepsilon^2 \\ \text{cov}(\varepsilon_t, \varepsilon_{t-4}) &= \rho^4 \sigma_\varepsilon^2 \\ &\vdots \quad \quad \quad \vdots \end{aligned}$$

Damit haben wir alle Elemente der Varianz-Kovarianzmatrix eines AR(1) Prozesses beisammen, die gesuchte Varianz-Kovarianzmatrix der Störterme ist also

$$\text{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}') = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix} = \sigma_\varepsilon^2 \boldsymbol{\Omega} = \mathbf{V} \quad (9.1)$$

mit

$$\sigma_\varepsilon^2 = \frac{\sigma_v^2}{1 - \rho^2}$$

Man beachte, dass diese Varianz-Kovarianzmatrix eine sehr einfache Form hat, sie enthält nur die unbekannt Parameter σ_v^2 und ρ .

Mit Hilfe dieser Matrix können wir später eine geeignete Schätzmethode für Autokorrelation finden.

9.1.4 Konsequenzen von Autokorrelation

Wir haben schon früher erwähnt, dass Autokorrelation der Störterme nicht die Erwartungstreue der OLS Schätzer für die Koeffizienten $\hat{\boldsymbol{\beta}}$ beeinflusst, da für deterministische \mathbf{X}

$$\text{E}(\hat{\boldsymbol{\beta}}) = \text{E} [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon})] = \boldsymbol{\beta} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \text{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}$$

wenn $\text{E}(\boldsymbol{\varepsilon}) = 0$, bzw. für stochastische \mathbf{X} muss $\text{E}(\boldsymbol{\varepsilon} | \mathbf{X}) = 0$ erfüllt sein.

Allerdings würden wir bei Anwendung der einfachen OLS Methode die ‘falsche’ Varianz-Kovarianzmatrix der Koeffizienten $\widehat{\text{var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}_\varepsilon^2 (\mathbf{X}' \mathbf{X})^{-1}$ verwenden anstatt der ‘richtigen’ Varianz-Kovarianzmatrix der Koeffizienten

$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma_\varepsilon^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1}$$

vgl. Gleichung (9.1), deshalb ist der OLS-Schätzer *nicht effizient*, und die *Standardfehler der Koeffizienten sind verzerrt!*

Wir fassen zusammen:

- Der OLS-Schätzer ist bei Vorliegen von Autokorrelation in den Residuen zwar weiterhin **unverzerrt und konsistent**, aber **nicht mehr effizient**. Zudem hat eine einmalige Störung langfristige Auswirkungen, allerdings mit abnehmenden Gewichten.
- Schlimmer, die **Standardfehler der geschätzten Koeffizienten sind in der Regel verzerrt**, d.h. die Teststatistiken sind nicht länger gültig.

Im Falle der (häufigeren) positiven Autokorrelation werden die Standardfehler der Koeffizienten systematisch *unterschätzt* (d.h. sie sind nach unten verzerrt), wodurch die Präzision der geschätzten Parameter *überschätzt* wird. Dies führt dazu, dass die Nullhypothese zu häufig verworfen wird, wenn sie tatsächlich akzeptiert werden sollte. Außerdem führt dies häufig zu einem sehr großen Bestimmtheitsmaß und zeichnet somit ein zu optimistisches Bild von der Schätzung.

- Wie wir im nächsten Abschnitt zeigen werden, führt Autokorrelation in Regressionen mit einer verzögerten abhängigen Variablen als Regressor (z.B. $y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + \varepsilon_t$) zu Endogenität (d.h. zu einer Korrelation zwischen dem Regressor y_{t-1} und dem Störterm ε_t), deshalb liefert die OLS-Schätzung in diesem Fall verzerrte und nicht konsistente Ergebnisse.

Eine intuitive Idee von den Auswirkungen (positiver) Autokorrelation vermittelt Abbildung 9.4. Je nachdem, ob das erste Residuum ε_1 positiv oder negativ ist, wird die Steigung über- oder unterschätzt. Da aber das erste Residuum mit gleicher Wahrscheinlichkeit positiv oder negativ ist, ist die Schätzung weiterhin unverzerrt. Allerdings minimiert OLS die Quadratsumme der Residuen *ohne Berücksichtigung der Autokorrelation*, deshalb gibt OLS einen ‘zu guten Fit’. Aus diesem Grund ist das Bestimmtheitsmaß R^2 bei positiver Autokorrelation in der Regel verhältnismäßig groß.

9.1.5 Autokorrelation mit verzögerten endogenen Variablen

Besondere Vorsicht ist geboten, wenn auf der rechten Seite der Regressionsgleichung eine verzögerte abhängige Variable vorkommt und der Störterm dieser Regression autokorreliert ist. Wenn z.B.

$$y_t = \beta_1 + \beta_2 y_{t-1} + \varepsilon_t$$

mit $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$

mit $v_t \sim \text{i.i.d.}(0, \sigma_{v_t}^2)$ führt dies zu einer Korrelation zwischen dem Störterm und dem Regressor, also zu *Endogenität*.

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \varepsilon_t \\ &= \beta_1 + \beta_2 y_{t-1} + \underbrace{\rho \varepsilon_{t-1} + v_t}_{\varepsilon_t} \\ &= \beta_1 + \beta_2 y_{t-1} + \underbrace{\rho (y_{t-1} - \beta_1 - \beta_2 y_{t-2}) + v_t}_{\varepsilon_t} \end{aligned}$$

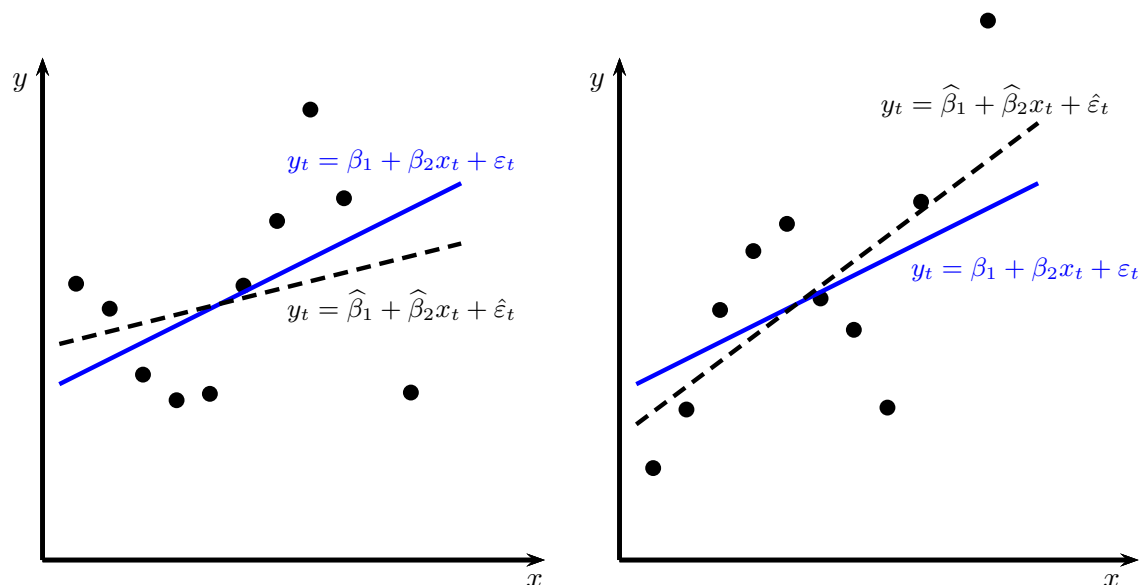


Abbildung 9.4: Positive Autokorrelation, wahrer (durchgezogene Linie) und geschätzter (strichlierte Linie) Zusammenhang; OLS ist erwartungstreu, aber nicht effizient, im linken Fall wird die Steigung unterschätzt, weil der erste Störterm ε_1 positiv ist, im rechten Fall wird die Steigung überschätzt, weil der erste Störterm ε_1 negativ ist.

durch einsetzen von $\varepsilon_{t-1} = y_{t-1} - \beta_1 - \beta_2 y_{t-2}$. Da y_{t-1} sowohl als erklärende Variable als auch im Störterm vorkommt sind diese korreliert!

Während die OLS Schätzer für die Koeffizienten bei Autokorrelation ohne verzögerten endogenen Variablen erwartungstreu sind, führt *Autokorrelation gemeinsam mit verzögerten endogenen Variablen* zu Endogenität, und in diesem Fall OLS Schätzer für die Koeffizienten $\hat{\beta}$ weder erwartungstreu noch konsistent!

In solchen Fällen werden häufig weitere lags von y als Regressoren verwendet, bis die Autokorrelation im Störterm vernachlässigbar ist.

Wenn genügend Beobachtungen zur Verfügung stehen werden in der Praxis oft so viele Zeitverzögerungen (y_{t-p}) verwendet, bis eine geeignete Teststatistik auf weißes Rauschen in den Residuen schließen lässt (vgl. Wooldridge, 2005, S. 378f). Dieses Verfahren liefert oft erstaunlich gute Ergebnisse, da die verzögerten endogenen Variablen den Einfluss unbeobachtbarer Variablen erfassen können, die sich im Zeitablauf nur langsam ändern.

Dies liefert allerdings ein dynamisches System und hat Auswirkungen auf die Interpretation der Koeffizienten. Näheres dazu erfahren Sie in Veranstaltungen zur Zeitreihenökonometrie.

9.2 Tests auf Autokorrelation

Das Problem bei den Tests auf Autokorrelation besteht darin, dass die Störterme der Grundgesamtheit ε_t nicht beobachtbar sind, wir können nur den Residuenvektor

der Stichprobe $\hat{\boldsymbol{\varepsilon}}$ beobachten. Wir wissen bereits, dass $\hat{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$ mit $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Deshalb gilt selbst unter den Gauss-Markov Annahmen $E(\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}') = E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{M}') = \mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\mathbf{M} = \sigma_{\varepsilon}^2\mathbf{M}$.

Da aber die Nebendiagonal-Elemente von \mathbf{M} nicht gleich Null sind hängen die OLS-Residuen von der Matrix \mathbf{X} ab und eignen sich deshalb nicht direkt für einen Test auf Autokorrelation.

9.2.1 Durbin–Watson Statistik

Der Durbin–Watson Test war zumindest früher der gebräuchlichste Test auf Autokorrelation. Im Unterschied zu den später folgenden Tests gilt dieser Test auch in kleinen Stichproben, er ist also nicht nur asymptotisch gültig.

James Durbin and Geoffrey Watson (1950) konnten – aufbauend auf einer Arbeit des Mathematikers John von Neumann – zeigen, dass auf Grundlage der geschätzten OLS-Residuen $\hat{\varepsilon}_t$ die Null-Hypothese $\rho = 0$ getestet werden kann.

Die Durbin–Watson (DW) Teststatistik ist

$$DW = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\varepsilon}_t^2}$$

Diese Teststatistik kann Werte zwischen 0 und 4 annehmen, wobei Werte nahe bei Null auf positive Autokorrelation und Werte nahe bei 4 auf negative Autokorrelation hindeuten. *Im Idealfall sollten die Werte der DW Statistik möglichst nahe bei 2 liegen.* Dies ist erkennbar, wenn man obige Formel ausmultipliziert:

$$DW = \frac{\sum \hat{\varepsilon}_t^2 + \sum \hat{\varepsilon}_{t-1}^2 - 2 \sum (\hat{\varepsilon}_t \hat{\varepsilon}_{t-1})}{\sum \hat{\varepsilon}_t^2}$$

Da sich $\sum \hat{\varepsilon}_t^2$ und $\sum \hat{\varepsilon}_{t-1}^2$ nur durch eine Beobachtung unterscheiden werden sie in großen Stichproben annähernd gleich sein. In diesem Fall gilt näherungsweise

$$DW \approx \frac{2 \sum \hat{\varepsilon}_t^2 - 2 \sum (\hat{\varepsilon}_t \hat{\varepsilon}_{t-1})}{\sum \hat{\varepsilon}_t^2} = 2 \left(1 - \frac{\sum (\hat{\varepsilon}_t \hat{\varepsilon}_{t-1})}{\sum \hat{\varepsilon}_t^2} \right)$$

Der Korrelationskoeffizient zwischen $\hat{\varepsilon}_t$ und $\hat{\varepsilon}_{t-1}$ ist $\hat{\rho} = \sum (\hat{\varepsilon}_t \hat{\varepsilon}_{t-1}) / \sum \hat{\varepsilon}_t^2$, deshalb gilt *ungefähr*

$$DW \approx 2(1 - \hat{\rho})$$

Daraus folgt, wenn $\hat{\rho} = -1$ ist die $DW \approx +4$, wenn $\hat{\rho} = +1$ ist die $DW \approx 0$, also gilt für $-1 \leq \rho \leq +1$, dass $0 \leq DW \leq 4$.

Wenn der Korrelationskoeffizient $\hat{\rho}$ gleich Null ist, hat die Durbin-Watson Statistik den Wert 2.

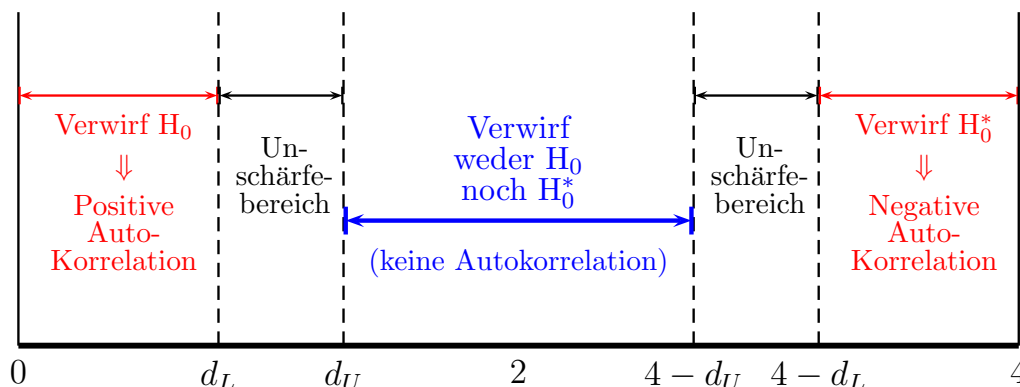
Allerdings hängen die geschätzten OLS-Residuen von den Werten der \mathbf{X} Matrix ab, wir erinnern uns, $E(\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}') = \sigma_{\varepsilon}^2\mathbf{M}$, deshalb ist die Verteilung der DW Statistik etwas komplizierter.

Durbin und Watson konnten zeigen, dass sich für die Verteilung der DW Statistik Grenzen angeben lassen, die nur von der Anzahl der x -Variablen und der Anzahl der Beobachtungen (T) abhängen, nicht aber von den konkreten Werten der x -Variablen.

Deshalb finden sich in den Tabellen für die kritischen Werte der DW Statistik eine Untergrenze d_L und eine Obergrenze d_U . Liegt der berechnete Wert der DW Statistik zwischen diesen beiden Werten liefert der DW Test keine interpretierbare Aussage. Liegt der berechnete Wert der DW Statistik aber unter dem kritischen Wert der Untergrenze d_L muss die Nullhypothese $\rho = 0$ (d.h. keine Autokorrelation) zugunsten der Hypothese positiver Autokorrelation verworfen werden.

Konkret ist die DW Statistik in Bezug auf *Autokorrelation 1. Ordnung* folgendermaßen zu interpretieren:

$0 <$	DW	$< d_L$	Verwirf Nullhypothese $\rho = 0$, → positive Autokorrelation
$d_L <$	DW	$< d_U$	keine Aussage möglich
$d_U <$	DW	< 2	Akzeptiere Nullhypothese $\rho = 0$
$2 <$	DW	$< 4 - d_U$	Akzeptiere Nullhypothese $\rho = 0$
$4 - d_U <$	DW	$< 4 - d_L$	keine Aussage möglich
$4 - d_L <$	DW	< 4	Verwirf Nullhypothese $\rho = 0$, → negative Autokorrelation



H_0 : keine positive Autokorrelation

H_0^* : keine negative Autokorrelation

Abbildung 9.5: Die Durbin-Watson Statistik

Beispiel: Angenommen wir möchten eine lineare Kostenfunktion

$$COST = \hat{\beta}_1 + \hat{\beta}_2 OUTPUT + \hat{\varepsilon}$$

schätzen und erhalten folgendes Ergebnis

$$COST = 166.47 + 19.93 OUTPUT + \hat{\varepsilon}_i$$

(8.75) (6.50)

$$R^2 = 0.84 \quad DW = 0.71 \quad T = 25$$

Tabelle 9.1: Durbin-Watson Statistik: Untere (d_L) und obere (d_U) Schranken der kritischen Werte des Durbin-Watson Tests; 5% Signifikanzniveau ($\alpha = 0.05$). T ist die Anzahl der Beobachtungen, und k^s die Anzahl der erklärenden Variablen ohne Interzept!

	$k^s = 1$		$k^s = 2$		$k^s = 3$		$k^s = 4$		$k^s = 5$	
T	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
10	0.88	1.32	0.70	1.64	0.52	2.02	0.38	2.41	0.24	2.82
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Da der empirische Wert der DW Statistik kleiner ist als der kritische Wert $d_L = 1.29$ muss die Nullhypothese $\rho = 0$ (keine Autokorrelation) zugunsten der Hypothese *positive Autokorrelation* verworfen werden (für $\alpha = 0.05$).

Wir haben bereits erwähnt, dass positive Autokorrelation häufig die Folge einer Fehlspezifikation ist, z.B. falsche Funktionsform oder fehlende relevante x -Variablen. Deshalb liefert die Durbin Watson Statistik häufig auch Hinweise auf eine Spezifikationsfehler allgemeiner Art. In diesem Fall wäre offensichtlich eine kubische Funktionsform

$$\text{COST} = \hat{\beta}_1 + \hat{\beta}_2 \text{OUTPUT} + \hat{\beta}_3 \text{OUTPUT}^2 + \hat{\beta}_4 \text{OUTPUT}^3 + \hat{\varepsilon}$$

geeigneter gewesen, wie die Abbildung des Residuenplots (Abb. 9.6) zeigt

Achtung: Die Durbin-Watson Statistik ist nur gültig, wenn die Regression ein Interzept enthält und wenn alle x -Variablen strikt exogen sind!

Der Durbin-Watson Test darf auch **nicht** verwendet werden, wenn eine verzögerte abhängige Variable (z.B. y_{t-1}) im Schätzansatz vorkommt. In diesem Fall ist entweder *Durbin's h Test* oder ein allgemeinerer Lagrange-Multiplier Test zu verwenden.

9.2.2 Durbin's h Test

Wenn $y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_t + \varepsilon_t$, dann ist

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T[\text{var}(\hat{\beta})]}} = \left(1 - \frac{DW}{2}\right) \sqrt{\frac{T}{1 - T[\text{var}(\hat{\beta})]}}$$

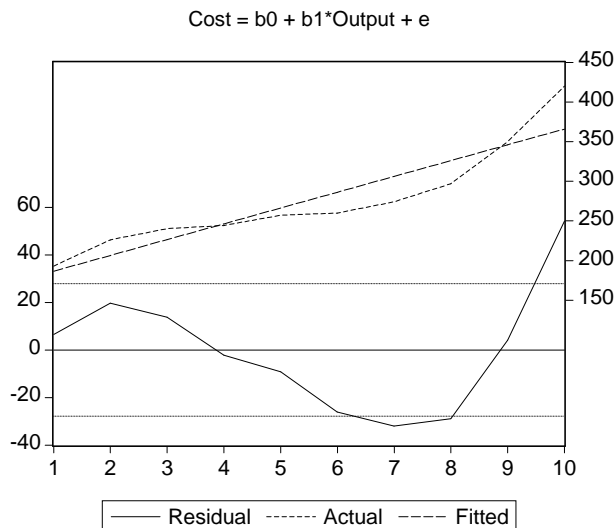


Abbildung 9.6: Gefittete Werte und Residuen

asymptotisch standard-normalverteilt ($\hat{\beta}$ ist die Schätzung für β). Durbin's h Test wird in der ökonometrischen Praxis kaum noch angewandt, meist wird in solchen Fällen der allgemeinere Breusch-Godfrey Test auf Autokorrelation vorgezogen.

9.2.3 Breusch-Godfrey Serial Correlation LM Test

Wenn die DW Statistik z.B. aufgrund verzögerter endogener Variablen, Endogenität oder fehlendem Interzept nicht angewandt werden kann bietet sich ein asymptotischer *Lagrange Multiplier* (LM) Test an, der im allgemeinen robuster ist und darüber hinaus auch allgemeiner anwendbar ist als Durbin's h Test. Dieser Test auf Autokorrelation der Ordnung p ist auch mit verzögerten abhängigen Variablen und für Instrumentvariablenschätzer anwendbar!

Allerdings ist dieser Test nur asymptotisch gültig, in kleinen Stichproben kann er verzerrte Ergebnisse liefern.

Konkret wird die Teststatistik folgendermaßen berechnet: für das Modell

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t$$

wird die Teststatistik mittels der folgenden Hilfsregression berechnet:

$$\hat{\varepsilon}_t = \hat{\beta}_1 + \hat{\beta}_2 x_{t2} + \dots + \hat{\beta}_k x_{tk} + \hat{\alpha}_1 \hat{\varepsilon}_{t-1} + \dots + \hat{\alpha}_p \hat{\varepsilon}_{t-p} + v_t$$

wobei $\hat{\varepsilon}_t$ die geschätzten Residuen für die Störterme ε_t sind.

Falls keine Autokorrelation vorliegt würden wir erwarten, dass die Koeffizienten der verzögerten Residuen ($\hat{\alpha}_1, \dots, \hat{\alpha}_p$) nicht gemeinsam signifikant von Null verschieden sind.

Einige Programme, wie z.B. EViews, geben eine F-Statistik für die gemeinsame Signifikanz dieser Koeffizienten α_h aus. Allerdings ist die exakte Verteilung dieser

F-Statistik unbekannt, weil die Residuen keine unabhängigen, deterministischen Variablen sind.

Eine zumindest asymptotisch gültige Teststatistik ist die sogenannte ‘Obs*R-squared’ Statistik $TR_{\hat{\varepsilon}}^2$ (d.h. Anzahl der Beobachtungen mal Bestimmtheitsmaß aus der Hilfsregression) aus.

Diese ‘Obs*R-squared’ Statistik ist die eigentliche Breusch-Godfrey LM Test Statistik. Unter ziemlich allgemeinen Bedingungen ist diese Statistik asymptotisch $\chi^2(p)$ verteilt.

Die Nullhypothese besagt, dass in den Residuen keine Autokorrelation bis zur angegebenen Ordnung vorliegt.

In *EViews* ist dieser LM-Test, oder genauer der *Breusch-Godfrey Serial Correlation Lagrange Multiplier (LM) Test* für ARMA Störterme p-ter Ordnung, mit dem Befehl `eqname.auto(order, options)` verfügbar.

In R ist dieser Test nach Laden des AER packages mit dem Befehl `bgtest(eqname)` verfügbar. Der entsprechende (postestimation) Befehl für *Stata* ist `estat bgodfrey, lags(1)`.

Als nächstes stellt sich die Frage, was zu tun ist, wenn die Tests auf autokorrelierte Störterme hinweisen.

9.3 Maßnahmen bei Autokorrelation

Ähnlich wie bei der Heteroskedastizität können auch bei Autokorrelation durch eine geeignete Transformation der Daten Bedingungen hergestellt werden, unter denen eine OLS-Schätzung BLUE ist. Dazu gehen wir folgendermaßen vor:

Da das Modell annahmegemäß in jeder Periode gelten soll, können wir die um eine Periode verzögerte Gleichung mit ρ (dem unbekanntem Autokorrelationskoeffizienten der Grundgesamtheit) multiplizieren und von der ursprünglichen Gleichung subtrahieren:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t \\ \rho y_{t-1} &= \rho \beta_1 + \rho \beta_2 x_{t-1,2} + \dots + \rho \beta_k x_{t-1,k} + \rho \varepsilon_{t-1} \quad /- \end{aligned}$$

daraus folgt

$$\begin{aligned} \underbrace{y_t - \rho y_{t-1}}_{y_t^*} &= (1 - \rho)\beta_1 + \beta_2 \underbrace{(x_{t2} - \rho x_{t-1,2})}_{x_{t2}^*} + \dots + \\ &\quad + \beta_k \underbrace{(x_{tk} - \rho x_{t-1,k})}_{x_{tk}^*} + \underbrace{(\varepsilon_t - \rho \varepsilon_{t-1})}_{\varepsilon_t^* = v_t} \end{aligned}$$

oder kürzer

$$y_t^* = (1 - \rho)\beta_1 + \beta_2 x_{t2}^* + \dots + \beta_k x_{tk}^* + \varepsilon_t^*$$

Wir erinnern uns, dass wir für den Störterm angenommen haben $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, bzw. $v_t = \varepsilon_t - \rho \varepsilon_{t-1}$

Der neue Störterm

$$\varepsilon_t^* := \varepsilon_t - \rho\varepsilon_{t-1} = v_t$$

erfüllt deshalb annahmegemäß alle Gauss-Markov Annahmen, deshalb wäre die Schätzung dieses transformierten Modells BLUE, d.h. unverzerrt und effizient.

Würden wir den unbekannt Parameter ρ kennen könnten wir einfach das transformierte Modell

$$y_t^* = (1 - \rho)\beta_1 + \beta_2 x_{t2}^* + \dots + \beta_k x_{tk}^* + \varepsilon_t^* \quad \text{für } t = 2, \dots, T$$

schätzen. Man nennt diese Transformation auch eine "Quasi-Differenzenbildung", da von jeder Beobachtung der mit ρ multiplizierte Wert der Vorperiode subtrahiert wird. Diese Transformation wird nach ihren Entdeckern "**Cochrane-Orcutt**" **Transformation** genannt.

Allerdings verlieren wir durch die Quasi-Differenzen (*lag*) - Bildung die erste Beobachtung. **Prais & Winsten** haben deshalb eine spezielle Transformation der *ersten* Beobachtung vorgeschlagen, die diesen Nachteil behebt.

Dazu wird nur die *erste* Beobachtung

$$y_1 = \beta_1 + \beta_2 x_{11} + \varepsilon_1$$

mit $\sqrt{1 - \rho^2}$ multipliziert.

$$\underbrace{\sqrt{1 - \rho^2} y_1}_{y_1^*} = \beta_1 \underbrace{\sqrt{1 - \rho^2}}_{x_{10}^*} + \beta_2 \underbrace{\sqrt{1 - \rho^2} x_{11}}_{x_{11}^*} + \underbrace{\sqrt{1 - \rho^2} \varepsilon_1}_{\varepsilon_1^*}$$

das heißt

$$y_1^* = \beta_1 x_{10}^* + \beta_2 x_{11}^* + \varepsilon_1^*$$

Diese Transformation der ersten Beobachtung liefert das gewünschte Ergebnis, da ε_1^* die gleichen Eigenschaften wie v_1 hat, d.h. Erwartungswert

$$E(\varepsilon_1^*) = \sqrt{1 - \rho^2} E(\varepsilon_1) = 0$$

und Varianz

$$\text{var}(\varepsilon_1^*) = (1 - \rho^2) \text{var}(\varepsilon_1) = (1 - \rho^2) \frac{\sigma_v^2}{1 - \rho^2} = \sigma_v^2$$

Das komplette Modell ist also

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{v}$$

mit

$$E(\mathbf{v}) = 0 \quad \text{und} \quad \text{var}(\mathbf{v}) = E(\mathbf{v}\mathbf{v}') = \sigma_v^2 \mathbf{I}_T$$

wobei

$$\mathbf{y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_T^* \end{bmatrix} = \begin{bmatrix} \sqrt{1 - \rho^2} y_1 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} \sqrt{1 - \rho^2} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}$$

$$\mathbf{X}^* = \begin{bmatrix} x_{11}^* & x_{12}^* & \cdots & x_{1k}^* \\ x_{21}^* & x_{22}^* & \cdots & x_{2k}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{T1}^* & x_{T2}^* & \cdots & x_{Tk}^* \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_{12} & \cdots & \sqrt{1-\rho^2}x_{1k} \\ 1-\rho & x_{22}-\rho x_{12} & \cdots & x_{2k}-\rho x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1-\rho & x_{T2}-\rho x_{T-1,2} & \cdots & x_{Tk}-\rho x_{T-1,k} \end{bmatrix}$$

Wenn das ρ der Grundgesamtheit bekannt ist, ist die Schätzung dieses Modells BLU (*best linear unbiased*). Der Schätzer $\hat{\beta}$ für β ist ein GLS - Schätzer (*Generalized Least Squares Estimator*)

$$\hat{\beta} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^*$$

mit der Varianz-Kovarianz Matrix

$$\text{var}(\hat{\beta}) = \sigma_v^2(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}$$

Ein Schätzer $\hat{\sigma}_v^2$ für die Varianz der Störterme σ_v^2 kann aus dem transformierten Modell geschätzt werden

$$\hat{\sigma}_v^2 = \frac{(\mathbf{y}^* - \mathbf{X}^*\hat{\beta})'(\mathbf{y}^* - \mathbf{X}^*\hat{\beta})}{T - k} = \frac{\hat{\varepsilon}^{*'}\hat{\varepsilon}^*}{T - k}$$

Das Problem dabei ist, dass für diese Transformation eine Schätzung für ρ benötigt wird, da ρ ein unbekannter Parameter der Grundgesamtheit ist. Für diese Schätzung von ρ wurden verschiedene Verfahren vorgeschlagen, für die allerdings *nur asymptotische Eigenschaften bekannt* sind.

9.3.1 Schätzung des Autokorrelationskoeffizienten

Die gebräuchlichsten Verfahren zur Schätzung von ρ sind:

Cochrane–Orcutt: Die Cochrane–Orcutt Prozedur ist ein *iteratives Verfahren* zur Schätzung von ρ . Für den bivariaten Fall $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ kann die Prozedur einfach veranschaulicht werden:

Man beginnt mit einem beliebigen Startwert für $\hat{\rho}$, transformiert mit diesem Startwert die Variablen wie in Gleichung (9.2) (Quasi-Differenzenbildung) und schätzt die Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ mit OLS.

$$y_t - \hat{\rho}y_{t-1} = (1 - \hat{\rho})\hat{\beta}_1 + \hat{\beta}_2(x_t - \hat{\rho}x_{t-1}) + \hat{\varepsilon}_t \quad (9.2)$$

$$(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t) = \hat{\rho}(y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 x_{t-1}) + \hat{\varepsilon}_t^* \quad (9.3)$$

Die zweite Gleichung (9.3) ist eine einfache Umformung der ersten Gleichung. Die Schätzungen für $\hat{\beta}_1$ und $\hat{\beta}_2$ aus der Schätzung der ersten Gleichung werden nun verwendet, um in einem zweiten Schritt die zwei transformierten Datenreihen $y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t$ und $y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 x_{t-1}$ zu berechnen. Mit Hilfe dieser transformierten Variablen kann aus der zweiten Gleichung mittels OLS ein neues – genaueres – $\hat{\rho}$ geschätzt werden.

Dann beginnt man mit dieser neuen Schätzung für $\hat{\rho}$ von vorne, man berechnet aus Gleichung (9.2) verbesserte Schätzungen für $\hat{\beta}_1$ und $\hat{\beta}_2$ und verwendet diese, um aus Gleichung (9.3) ein neues besseres $\hat{\rho}$ zu berechnen.

Dieses Verfahren wird wiederholt, bis $\hat{\beta}_1, \hat{\beta}_2$ und $\hat{\rho}$ konvergieren, bzw. bis die *Durbin-Watson* Statistik auf weißes Rauschen der Residuen schließen lässt.

Achtung: Dieses Verfahren darf nicht angewandt werden, wenn in der Gleichung verzögerte endogene Variablen (z.B. y_{t-1}) vorkommen!

Ein weiterer Nachteil dieses Verfahrens ist, dass es zu einem lokalen anstatt globalen Maximum führen kann. Dies wird bei den folgenden Methoden vermieden.

Hildreth-Lu Die Hildreth-Lu Prozedur benutzt eine “grid-search” und ähnelt deshalb einer Maximum-Likelihood Schätzung.

Maximum-Likelihood Maximum-Likelihood Schätzungen erfordern nicht-lineare Schätzverfahren und sind deshalb rechenintensiver; diese Verfahren werden in Fortgeschrittenenveranstaltungen diskutiert.

Nicht-lineare Schätzverfahren Durch geeignete Substitution erhält man eine Gleichung, die *nicht-linear in den Parametern* ist, z.B. für Autokorrelation 1. Ordnung

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t \end{aligned}$$

Einsetzen der zweiten Gleichung in die erste gibt:

$$y_t = \beta_1 + \beta_2 x_t + \rho \varepsilon_{t-1} + v_t \quad (9.4)$$

Da $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ in jeder Periode gilt kann dies umgeschrieben werden zu $\varepsilon_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{t-1}$. Multiplizieren dieser Gleichung mit ρ gibt

$$\rho \varepsilon_{t-1} = \rho y_{t-1} - \rho \beta_1 - \rho \beta_2 x_{t-1}$$

Wenn wir dies in (9.4) einsetzen folgt

$$y_t = \beta_1(1 - \rho) + \rho y_{t-1} + \beta_2 x_t - \beta_2 \rho x_{t-1} + v_t$$

Diese Gleichung ist zwar linear in den Variablen, aber *nicht linear* in den Parametern β_1, β_2 und ρ ! Deshalb kann diese Gleichung nicht mittels OLS geschätzt werden, aber so gut wie alle ökonometrischen Programmpakete können numerisch $\sum_t v_t^2$ minimieren und derart konsistente Schätzer für die Parameter berechnen.

9.3.2 HAC Standardfehler

Ähnlich wie bei Heteroskedastizität gibt es auch für Autokorrelation robuste Standardfehler, sogenannte ‘*heteroskedasticity and autocorrelation consistent*’ (HAC) Standardfehler.

Auch diese haben die Sandwich Form. Die Varianz- Kovarianzmatrix der Koeffizienten ist

$$\text{var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

mit

$$\boldsymbol{\Omega} = \text{E}(\hat{\varepsilon}\hat{\varepsilon}') = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(\varepsilon_i\varepsilon_j)x_ix_j$$

Diese Matrix $\boldsymbol{\Omega}$ wird geschätzt durch

$$\hat{\boldsymbol{\Omega}} = \sum_{i=1}^n \sum_{j=1}^n w(i-j) \text{cov}(\hat{\varepsilon}_i\hat{\varepsilon}_j)x_ix_j$$

wobei $w()$ eine Gewichtungsfunktion ist (für eine ausführlichere Diskussion siehe z.B. Davidson and MacKinnon (2003, 362)).

Die bekannteste dieser HAC Schätzer wurden von Newey and West (1987) vorgeschlagen (zur Berechnung siehe z.B. Wooldridge, 2005, S. 432ff).

In Stata sind diese mit dem Befehl `newey` verfügbar, in R im sind sie im `sandwich` Paket von A. Zeileis verfügbar.

Die geschätzten Koeffizienten werden davon nicht berührt, aber die Standardfehler der Koeffizienten werden damit *konsistent* geschätzt, weshalb auch Hypothesentests asymptotisch gültig bleiben.

Achtung: Von einer unreflektierten Korrektur von Autokorrelation ist abzuraten. Wir haben bereits gesehen, dass aus Autokorrelation erster Ordnung eine nicht-lineare Gleichung in den Lags von y und x resultiert.

$$y_t = \rho y_{t-1} + \beta_1(1 - \rho) + \beta_2(x_t - \rho x_{t-1}) + v_t$$

Ökonometrikerinnen würden es im allgemeinen bevorzugen diese nicht-lineare Gleichung mit geeigneten Methoden zu schätzen und auf die aus der Autokorrelation folgenden Restriktionen zu testen!

Autokorrelation ist sehr häufig ein Indikator für (dynamische) Fehlspezifikation. Deshalb sollte man auf jeden Fall versuchen eine geeignetere Spezifikation zu finden, bevor man sich auf eines der herkömmlichen Verfahren zur Korrektur der Autokorrelation verlässt!

Dieser Punkt wurde u.a. bereits von Mizon (1995) betont – der Artikel trägt den vielsagenden Titel “*A simple message for autocorrelation correctors: Don’t*”.

Literaturverzeichnis

Davidson, R. and MacKinnon, J. G. (2003), *Econometric Theory and Methods*, Oxford University Press, USA.

Mizon, G. E. (1995), 'A simple message for autocorrelation correctors: Don't', *Journal of Econometrics* **69**(1), 267–288.

URL: <http://www.sciencedirect.com/science/article/pii/030440769401671L>

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), pp. 703–708.

Wooldridge, J. (2005), *Introductory Econometrics: A Modern Approach*, 3 edn, South-Western College Pub.