

Auswirkungen autokorrelierter Störterme auf OLS-Schätzungen

$$y_t = \beta x_t + u_t \quad \text{Datengenerierender Prozess } (t = 1, \dots, T)$$

$$y_t = \hat{\beta} x_t + \hat{u}_t \quad \text{Schätzgleichung}$$

Schätzung von $\hat{\beta}$ durch Minimierung der Summe der quadrierten Residuen (OLS):

$$RSS = \sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (y_t - \hat{\beta} x_t)^2$$

$$\frac{\partial RSS}{\partial \hat{\beta}} = -2 \sum_{t=1}^T (y_t - \hat{\beta} x_t) x_t = 0$$

$$\sum_{t=1}^T x_t y_t - \hat{\beta} \sum_{t=1}^T x_t^2 = 0 \quad \Rightarrow \quad \boxed{\hat{\beta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} = \frac{S_{xy}}{S_{xx}}}$$

Erwartungswert von $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} = \frac{\sum_{t=1}^T x_t (\beta x_t + u_t)}{\sum_{t=1}^T x_t^2} = \beta + \frac{\sum_{t=1}^T x_t u_t}{\sum_{t=1}^T x_t^2}$$

$$E(\hat{\beta}) = \beta + \frac{1}{\sum_{t=1}^T x_t^2} E\left(\underbrace{\sum_{t=1}^T x_t u_t}_{=0}\right)$$

= 0 falls $\text{cov}(u_t, x_t) = 0$ (d.h. x_t ist echt exogen).

$$\Rightarrow \boxed{E(\hat{\beta}) = \beta}$$

Schätzvarianz von $\hat{\beta}$:

$$\text{var}(\hat{\beta}) = E(\hat{\beta} - \beta)^2 = \frac{1}{\left(\sum x_t^2\right)^2} E\left(\sum x_t u_t\right)^2$$

$$\begin{aligned} E\left(\sum x_t u_t\right)^2 &= E(x_1 u_1 + x_2 u_2 + \dots + x_T u_T)(x_1 u_1 + x_2 u_2 + \dots + x_T u_T) \\ &= (x_1^2 \sigma^2 + x_2^2 \sigma^2 + \dots + x_T^2 \sigma^2) \text{ falls } u_t \sim IN(0, \sigma^2) \text{ und folglich } \boxed{\text{cov}(u_t, u_{t-j}) = 0} \\ &= \sigma^2 \sum x_t^2 \end{aligned}$$

$$\Rightarrow \boxed{\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2}}$$

Falls $u_t = \rho u_{t-1} + \varepsilon_t$ mit $\varepsilon_t \sim IN(0, \sigma_\varepsilon^2)$ und $0 < \rho < 1$ AR(1)-Prozess

sind die Kovarianzen $\text{cov}(u_t, u_{t-j})$ nicht Null sondern:

$$\boxed{\text{cov}(u_t, u_{t-1}) = \rho \sigma^2, \quad \text{cov}(u_t, u_{t-2}) = \rho^2 \sigma^2, \dots} \quad \text{mit } \sigma^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

und folglich:

$$\begin{aligned} E\left(\sum x_t u_t\right)^2 &= E(x_1 u_1 + x_2 u_2 + \dots + x_T u_T)(x_1 u_1 + x_2 u_2 + \dots + x_T u_T) \\ &= \sigma^2 \sum x_t^2 + 2\rho \sigma^2 \sum x_t x_{t-1} + 2\rho^2 \sigma^2 \sum x_t x_{t-2} + \dots \end{aligned}$$

$$\Rightarrow \boxed{\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2} \left(1 + 2\rho \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} + \dots \right)}$$

Der Ausdruck in der Klammer entspricht dem Faktor, um den die Varianz von $\hat{\beta}$ durch die übliche OLS-Formel

$$\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2} \text{ unterschätzt wird.}$$

Dieser Faktor ist insbesondere dann gross (wesentlich grösser als 1), wenn die exogene Variable x selber autokorreliert ist, was im Allgemeinen der Fall ist. Nimmt man für x einen AR(1)-Prozess mit Autokorrelationskoeffizienten η an, so gilt:

$$\boxed{\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2} (1 + 2\rho\eta + 2\rho^2\eta^2 + \dots) = \frac{\sigma^2}{\sum x_t^2} \left(\frac{1 + \rho\eta}{1 - \rho\eta} \right)}$$

Simulationsexperiment mit Eviews

Vorgegeben sei ein wahres **Modell mit autokorreliertem Störterm** (datengenerierender Prozess):

$$(1) \quad y_t = \alpha + \beta x_t + u_t \quad \text{mit } u_t = \rho u_{t-1} + \varepsilon_t \quad \text{und } \varepsilon_t \sim IN(0, \sigma^2)$$

$$\alpha = 4, \quad \beta = 2 \quad \sigma = 1.0, \quad \rho = 0.866$$

$$\Rightarrow \text{var}(u) = \frac{\sigma^2}{1 - \rho^2} = \frac{1}{1 - 0.75} = 4$$

$$\text{sdev}(u) = 2$$

Es werden 100 künstliche Stichproben mit je 108 Beobachtungen (Quartalswerte 1970:1 bis 1996:4) generiert. Für jede Stichprobe werden die Parameter α, β und σ mit der Methode der kleinsten Quadrate (OLS) geschätzt. Das Simulationsexperiment dient zur Illustration der Verzerrungen und der Ineffizienz der OLS-Schätzung.

Konstruktion der 100 Stichproben

Zuerst wird die exogene Variable x nach dem folgenden Schema generiert:

$$x_t = \eta x_{t-1} + v_t \quad \text{mit } \eta = 0.707, \quad x_0 = 0, \quad v_t \sim N(0, 1)$$

$$\Rightarrow \text{var}(x) = 2$$

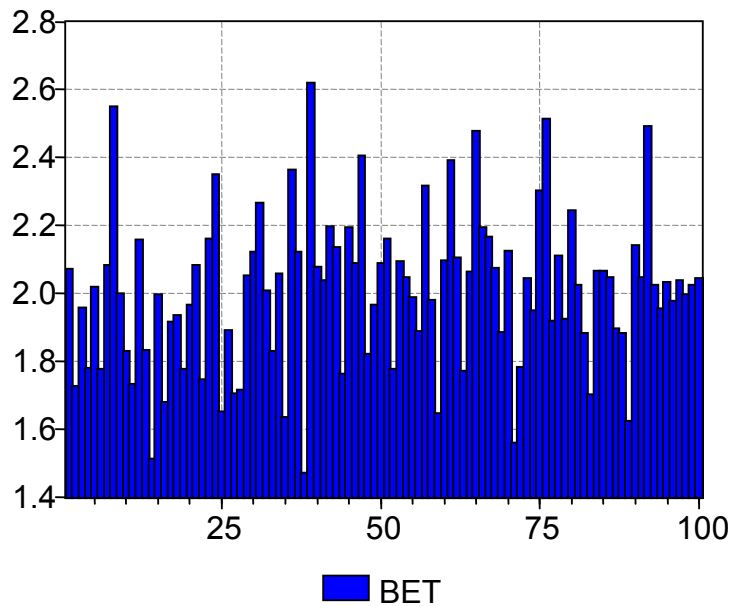
Die Variable x ist in allen 100 Stichproben gleich.

Dann werden mit einem Zufallszahlen-Generator 108 Werte aus $\varepsilon_t \sim IN(0, \sigma^2)$ gezogen und über (1) mit den vorgegebenen Parameterwerten ($\alpha = 4, \beta = 2, \rho = 0.866$) 108 Werte für y bestimmt. Dieser Schritt wird 100 mal wiederholt.

\Rightarrow 100 Stichproben mit je 108 Beobachtungswerten für x und y .

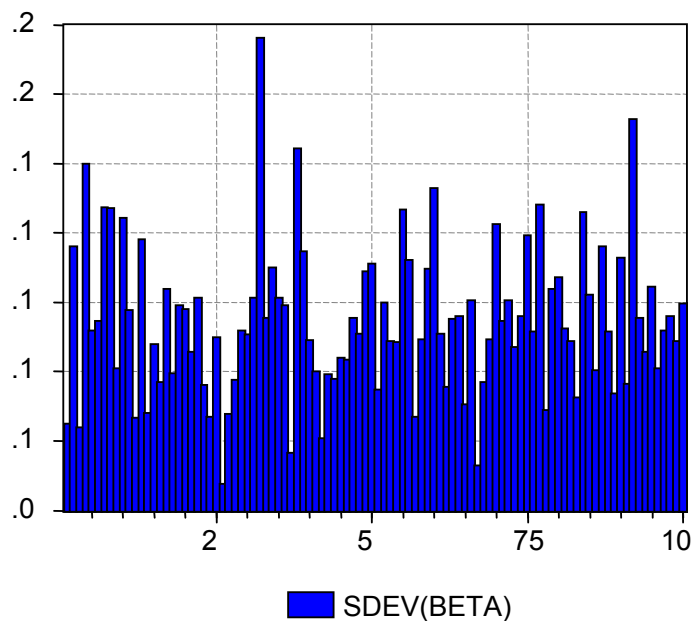
Schätzergebnisse

100 Schätzwerte für β und dazugehörige geschätzte Standardfehler



Geschätzte β 's:

Mean	2.005259
Median	2.025382
Maximum	2.619239
Minimum	1.470410
Std. Dev.	0.227488



Standardfehler:

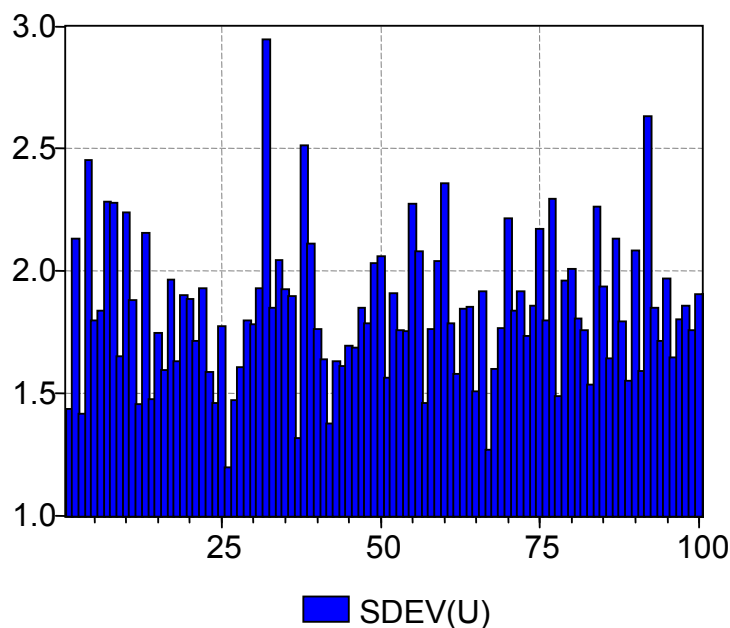
Mean	0.134705
Median	0.131920
Maximum	0.216109
Minimum	0.087612
Std. Dev.	0.021726

Die Unverzerrtheit der OLS-Schätzung für β (trotz autokorreliertem Störterm) zeigt sich darin, dass der Mittelwert der 100 Schätzwerte (2.005) praktisch dem wahren Wert von 2 entspricht.

Dagegen sind die Standardfehler der geschätzten β 's verzerrt. Im Mittel der 100 Schätzungen beträgt der geschätzte Standardfehler 0.1347, während die 100 Schätzwerte von β effektiv eine Standardabweichung von 0.2275 auf-

weisen. Die Autokorrelation des Störterms führt also zu einer beträchtlichen Unterschätzung des Standardfehlers der geschätzten β 's (und folglich zu einer Überschätzung der t-Werte).

Eine weitere Verzerrung zeigt sich hinsichtlich der Störterm-Varianz bzw. des daraus abgeleiteten Standardfehlers der Gleichung. Die im Simulationsexperiment vorgegebene Störterm-Varianz beträgt $\text{var}(u) = 4$ bzw. die Standardabweichung $\text{sdev}(u) = 2$. Die in den 100 Schätzungen ausgewiesenen Standardfehler der Gleichung sind aber etwas kleiner; im Durchschnitt der 100 OLS-Schätzungen ergibt sich ein Wert von 1.8375 (anstatt 2). Die Autokorrelation des Störterms führt also auch zu einer gewissen Unterschätzung der Störterm-Varianz bzw. des Standardfehlers der Gleichung.



In der Schätzung Nr. 24 ergibt sich beispielsweise:

$$Y = \text{ALFA} + \text{BETA} * X$$

	Coefficient	Std. Error	t-Statistic	Prob.
ALFA	4.011372	0.148934	26.93391	0.0000
BETA	2.350018	0.106991	21.96465	0.0000

R-squared	0.819865	Mean dependent var	5.100430
Adjusted R-squared	0.818165	S.D. dependent var	3.422620
S.E. of regression	1.459477	Akaike info criterion	3.612378
Sum squared resid	225.7877	Schwarz criterion	3.662047
Log likelihood	-193.0684	Durbin-Watson stat	0.482271

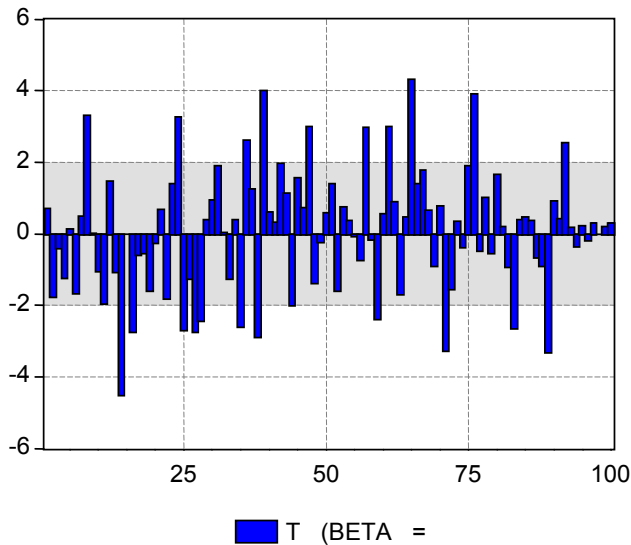
- 1) Die DW-Statistik von 0.482 zeigt das Autokorrelationsproblem an.
- 2) Der geschätzte Standardfehler von β beträgt 0.10699, während die 100 geschätzten β 's effektiv mit einer Standardabweichung von 0.2275 um den wahren Wert von 2 schwanken.
- 3) Der Standardfehler der Gleichung wird auf 1.4595 geschätzt. Die wahre Standardabweichung des Störterms u beträgt aber 2.
- 4) Die richtige Null-Hypothese $\beta = 2$ wird in einem t-Test abgelehnt:

$$t = \frac{\hat{\beta} - 2}{SE(\hat{\beta})} = \frac{2.35 - 2}{0.10699} = 3.27 \gg 2$$

Dafür verantwortlich ist die Unterschätzung des Standardfehlers.

Zu 4)

In wie vielen der 100 Schätzungen wird die richtige Null-Hypothese $\beta = 2$ abgelehnt? Es ergeben sich die folgenden 100 berechneten t-Werte:



Auf einem Signifikanzniveau von 5% ($t = 2$) sollte die richtige Null-Hypothese mit 5% Wahrscheinlichkeit abgelehnt werden.

In den 100 Schätzungen wird sie aber – wegen der Unterschätzung des Standardfehlers – in 21 Fällen abgelehnt.

Zu 2 und 3)

Im vorliegenden Fall mit $\rho = 0.866$ (Autokorrelation des Störterms u) und $\eta = 0.707$ (Autokorrelation von x) lässt sich das Ausmass der Unterschätzung des Standardfehlers von β und des Standardfehlers der Gleichung analytisch bestimmen (Maddala, Kapitel 6.5):

Der Standardfehler von β ist approximativ um den Faktor

$$\sqrt{\frac{1+\eta\rho}{1-\eta\rho}} = \sqrt{4.158} = 2.039 \text{ unterschätzt.}$$

Der Standardfehler der Gleichung ist approximativ um den Faktor

$$1 / \sqrt{1 - \frac{1+\eta\rho}{1-\eta\rho} \frac{1}{T}} = 1 / \sqrt{0.9615} = 1.020 \text{ unterschätzt (} T = 108 \text{).}$$

Diese Unterschätzung vermindert sich mit zunehmendem Stichprobenumfang T .

Schlussfolgerung

In einem Modell mit autokorreliertem Störterm (**ohne zeitverzögerte endogene Variable auf der rechten Gleichungsseite**) führt die Methode der kleinsten Quadrate (OLS):

- zu einer Unterschätzung der geschätzten Standardfehler der Parameter und somit zu einer Überschätzung der t-Werte.
- zu einer gewissen Unterschätzung des Standardfehlers der Gleichung.

Die Parameterschätzung selbst ist **unverzerrt (A)**, aber **nicht effizient (B)**.

Zu A: Zeitverzögerte endogene Variablen als Erklärende auf der rechten Gleichungsseite => Verzerrte Parameterschätzungen

Erscheint die endogene Variable zeitverzögert auf der rechten Gleichungsseite – wie z.B. im "partial adjustment"-Modell - so sind auch die Parameterschätzungen **verzerrt**.

Grund:

$$y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t \quad \leftarrow \quad u_t = \rho u_{t-1} + \varepsilon_t$$

$$y_{t-1} = \alpha + \beta_1 x_{t-1} + \beta_2 y_{t-2} + u_{t-1}$$

Da y_{t-1} gemäss der zweiten Gleichung von u_{t-1} abhängig ist, ergibt sich bei autokorreliertem u_t in der ersten Gleichung eine Abhängigkeit zwischen y_{t-1} und u_t . Somit kann y_{t-1} nicht als exogen betrachtet werden. Bei positiver Autokorrelation des Störterms ($0 < \rho < 1$) ist y_{t-1} positiv mit u_t korreliert, was zu einer Überschätzung von β_2 führt. Ein Teil der positiven Autokorrelation des Störterms wird fälschlicherweise von y_{t-1} aufgefangen. Dies führt weiter dazu, dass die berechneten Residuen \hat{u}_t im Vergleich zum Störterm u_t eine zu geringe Autokorrelation anzeigen. Deshalb ist die **DW-Statistik nicht mehr aussagekräftig**; sie tendiert dazu, kein Autokorrelationsproblem anzuzeigen, auch wenn effektiv ein solches vorliegt. Man muss zum sogenannten **Durbin-h-Test** übergehen (vgl. Übung).

Zu B: Effizienz der Schätzung

Im Modell

$$(1) \quad y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + u_t$$

mit autokorreliertem Störterm

$$(2) \quad u_t = \rho u_{t-1} + \varepsilon_t \quad \text{mit} \quad -1 < \rho < 1 \quad \text{und} \quad \varepsilon_t \sim IN(0, \sigma^2)$$

kann durch eine einfache Gleichungstransformation (nach Cochrane-Orcutt) eine unverzerrte und effiziente Schätzung gemacht werden.

Schiebt man (1) um eine Periode verzögert und multipliziert mit ρ , so ergibt sich:

$$(3) \quad \rho y_{t-1} = \rho \alpha + \rho \beta_1 x_{1,t-1} + \rho \beta_2 x_{2,t-1} + \dots + \rho u_{t-1}$$

Subtrahiert man (3) von (1), so erhält man:

$$(4) \quad y_t - \rho y_{t-1} = (1 - \rho)\alpha + \beta_1(x_{1t} - \rho x_{1,t-1}) + \beta_2(x_{2t} - \rho x_{2,t-1}) + \dots + u_t - \rho u_{t-1}$$

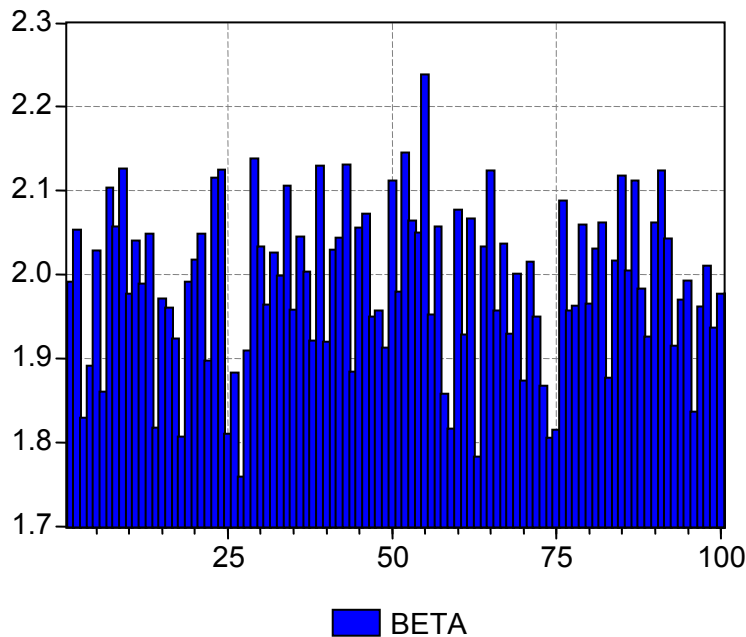
Gemäss Annahme (2) gilt nun $u_t - \rho u_{t-1} = \varepsilon_t$, d.h. Gleichung (4) hat einen Störterm mit der gewünschten Eigenschaft $\varepsilon_t \sim IN(0, \sigma^2)$. Man kann (4) auch wie folgt schreiben:

$$(4') \quad y_t - \rho y_{t-1} = (1 - \rho)\alpha + \beta_1(x_{1t} - \rho x_{1,t-1}) + \beta_2(x_{2t} - \rho x_{2,t-1}) + \dots + \varepsilon_t$$

$$(4'') \quad y_t = (1 - \rho)\alpha + \rho y_{t-1} + \beta_1(x_{1t} - \rho x_{1,t-1}) + \beta_2(x_{2t} - \rho x_{2,t-1}) + \dots + \varepsilon_t$$

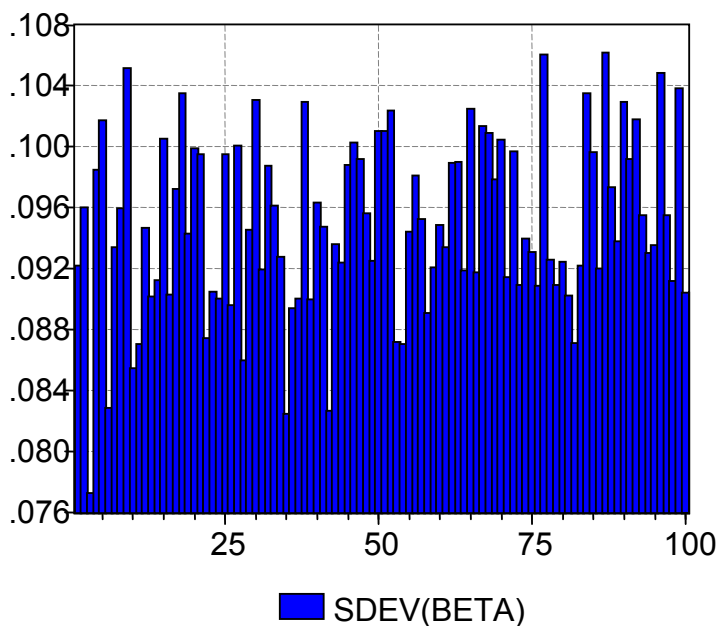
Wiederholt man die 100 Schätzungen nach diesem Verfahren (unter sonst gleichen Annahmen), so ergeben sich die folgenden Resultate:

Die 100 Schätzungen für β :



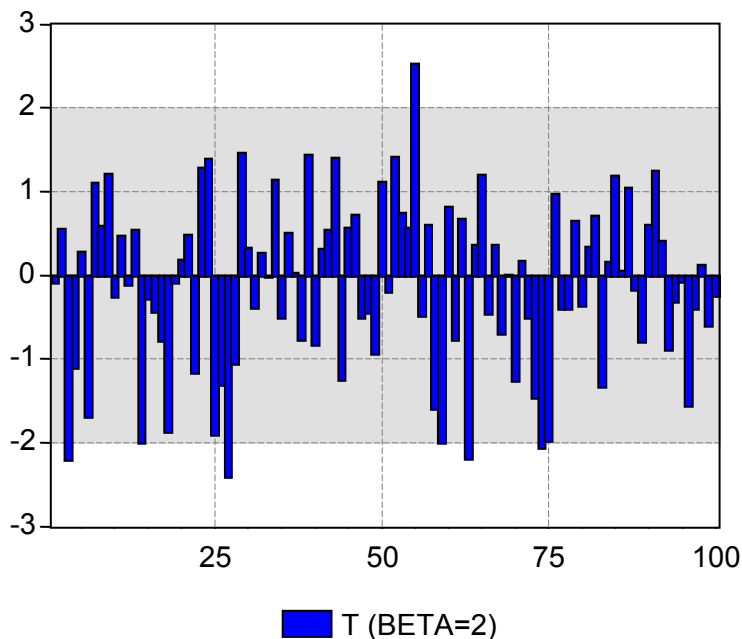
Mean 1.988590
 Median 1.992198
 Maximum 2.238542
 Minimum 1.758983
 Std. Dev. 0.096735

Der Effizienzgewinn zeigt sich in einer wesentlich kleineren Standardabweichung der geschätzten β 's von 0.0967 (gegenüber 0.2275 in der untransformierten OLS-Schätzung). Die Standardabweichung von β wird zudem in den 100 Schätzungen ziemlich richtig geschätzt; sie beträgt im Mittel 0.0949:



Mean 0.094942
 Median 0.094462
 Maximum 0.106184
 Minimum 0.077259
 Std. Dev. 0.005788

Entsprechend wird die richtige Null-Hypothese $\beta = 2$ auf dem 5%-Signifikanzniveau ($t = 2$) in ungefähr 5 der 100 Fälle fälschlicherweise abgelehnt:



Das für dieses Simulationsexperiment verwendete EViews-Programm ist auf der folgenden Seite abgedruckt. Sie finden den Programmcode auf der Internetseite im Anhang zu Ökonometrie I als Simulationsprogramm 3. Der Zufallszahlengenerator ist mit "rndseed(type=kn4) auf die Version 4 zurückgesetzt.

Die verzerrten OLS-Schätzungen erhält man, indem man a) wählt und b) "auskommentiert":

```
' a) OLS (verzerrt):
equation test!j.ls y = alfa(!j)+beta(!j)*x
ncoef = 2

' b) Transformation nach Cochrane/Orcutt (korrekt)
' equation test!j.ls y = (1-rho(!j))*alfa(!j)+beta(!j)*(x-rho(!j)*x(-1))+rho(!j)*y(-1)
' ncoef = 3
```

Die unverzerrten Cochrane/Orcutt-Schätzungen erhält man, indem man b) wählt und a) "auskommentiert":

```
' a) OLS (verzerrt):
' equation test!j.ls y = alfa(!j)+beta(!j)*x
' ncoef = 2

' b) Transformation nach Cochrane/Orcutt (korrekt)
equation test!j.ls y = (1-rho(!j))*alfa(!j)+beta(!j)*(x-rho(!j)*x(-1))+rho(!j)*y(-1)
ncoef = 3
```

Programmcode EViews

```
' SimAutoCorr.prg:
' Regression mit künstlich generierten Daten
' Schätzung einer Gleichung  $y = \alpha + \beta x + u$ 
' mit autokorreliertem Störterm u

' Öffnen eines Workfiles und Festlegen des Stichprobenumfangs
workfile simpreg Q 1969:4 1996:4

' Generieren einer erklärenden Variablen x mit Erwartungswert 0
' und Standardabweichung 2
rdseed(type=kn4) 312
genr v = nrnd
genr x = 0
smpl 1970:1 1996:4
genr x = 0.707*x(-1)+v

FOR lj = 1 to 100

' Generieren von y mit einem "wahren Modell" mit den Parametern
'  $\alpha=4$ ,  $\beta=2$ ,  $\sigma=1$  und  $\rho = 0.866$ 
scalar xseed = 571+lj
rdseed(type=kn4) xseed
genr e = nrnd ' sigma=1
smpl 1969:4 1996:4
genr u = 0
smpl 1970:1 1996:4
genr u = 0.866*u(-1)+e ' rho=0.866
genr y = 4 + 2*x + u ' alpha=4, beta=2
genr ulj = u

scalar ncoef
coef(100) alfa
coef(100) beta
coef(100) rho
coef(100) tbeta2
coef(100) sdbeta
coef(100) seofeq
coef(100) rquadrat
coef(100) dw

' Schätzung der Gleichung: a) oder b) wählen

' a) OLS (verzerrt):
' equation test!j.ls y = alfa(!j)+beta(!j)*x
' ncoef = 2

' b) Transformation nach Cochrane/Orcutt (korrekt)
equation test!j.ls y = (1-rho(!j))*alfa(!j)+beta(!j)*(x-rho(!j)*x(-1))+rho(!j)*y(-1)
ncoef = 3

genr res!j = resid
sdbeta(!j) = sqr(@covariance(ncoef,ncoef))
tbeta2(!j) = (beta(!j)-2)/sdbeta(!j)
seofeq (!j) = @se
rquadrat(!j) = @r2
dw(!j) = @dw
```

NEXT