GML

Ökonometrie: SURE

All models are wrong but some are useful, turn data into assets

Prof. Dr. Frank Lehrbass

Copyright

© FOM Hochschule für Oekonomie & Management
gemeinnützige Gesellschaft mbH (FOM), Leimkugelstraße 6, 45141 Essen

Dieses Werk ist urheberrechtlich geschützt und nur für den persönlichen Gebrauch im Rahmen der Veranstaltungen der FOM bestimmt.

Die durch die Urheberschaft begründeten Rechte (u. a. Vervielfältigung, Verbreitung, Übersetzung, Nachdruck) bleiben dem Urheber vorbehalten.

Das Werk oder Teile daraus dürfen nicht ohne schriftliche Genehmigung des Urhebers / der FOM reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Dies schließt auch den Upload in soziale Medien oder andere digitale Plattformen ein.

Quellen

- Wir verwenden die genannten Bücher/Artikel und geben die im Text Quellen wie folgt an:

- Assenmacher, 2002, Einführung in die Ökonometrie, Oldenbourg
- Zellner, 1962, An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, in: Journal of the American Statistical Association, 57, S. 348–368, 1962
- Verbeek, 2017, Seite xx
- Schröder, 2002, dito
- Danielsson, 2011, dito
- Gehrke, 2022, dito

Paneldaten

- Bisher: Datenerhebung einmalig, d. h. an einem einzigen Zeitpunkt, an jedem Subjekt.
- Beispiel: Messung des Stimmrechtsanteils des größten Aktionärs und der Kapitalmarktbewertung zu einem Zeitpunkt. In diesem Fall wird der Zusammenhang zwischen diesem Stimmrechtsanteil und der Kapitalmarktbewertung durch eine lineare Regression geprüft.
- Jetzt: Mehrmalige Datenerhebung an den gleichen Subjekten zu verschiedenen Zeitpunkten.
- Daten dieser Art heißen Paneldaten, oder auch longitudinale Daten bzw. Kohorten.

Wichtige Variablen

- Subjekt: Untersuchungseinheiten, an denen Messungen (zu verschiedenen Zeitpunkten) vorgenommen werden. Die Subjekte werden in einer eigenen Variablen erfasst.
- Zeit: Zeitpunkte, zu denen die einzelnen Messungen an den Subjekten durchgeführt werden. Die Zeitpunkte werden ebenfalls in einer eigenen Variablen erfasst.

Wichtige Vorteile von Paneldaten

- Variationen über die Zeit können erfasst und analysiert werden.
- Die Kombination von Zeitreihen mit Querschnittsdaten beinhaltet mehr Informationen als reine Querschnittsdaten.
- Paneldaten ermöglichen Analysen zur Dynamik von Veränderungsprozessen.

Paneldaten

Arten von Paneldaten

- Balanced Panel: Für jedes Subjekt ist die gleiche Anzahl an Beobachtungen vorhanden, d. h. für jedes Subjekt liegen zu jedem Zeitpunkt Daten vor.
- Unbalanced Panel: Die Subjekte weisen ungleiche Anzahlen an Beobachtungen auf, d. h. für jedes Subjekt liegen nicht zu jedem Zeitpunkt Daten vor.
- Short Panel: Die Zahl der Subjekte ist größer als die Zahl der Messzeitpunkte.
- Long Panel: Die Zahl der Subjekte ist kleiner als die Zahl der Messzeitpunkte.

Modelle für Paneldaten

- Pooled Regression
- Fixed Effects Model (FEM)
- Random Effects Model (REM)
- Seemingly Unrelated Regressions Model (SURE)

- Die ersten drei Modelle kann man sich durch das Buch von Gehrke (2022) erarbeiten und/oder durch diese Publikation https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3453055

Panelmodelle

Die einfachste Form, Paneldaten zu modellieren ist, für alle Beobachtungen direkt eine einfache lineare Regressionsgleichung aufzustellen:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it}$$

mit:	y_{it}	Messwert für die Zielgröße für Subjekt i zum Zeitpunkt t
	x_{kit}	Messwerte für die unabhängigen Variablen X_{kit} für Subjekt i zum Zeitpunkt t
	β_0	Konstante
	β_k	Regressionskoeffizienten $(k=1,\ldots,K)$
	i	Index für die Subjekte ($i=1,\ldots,I$)
	t	Index für die Zeitpunkte ($t=1,\ldots,T$)
	u_{it}	Residualgröße für Subjekt i zum Zeitpunkt t, uiv (unabhängig identisch verteilt) mit E(u)=0 und Var(u)=σ_u^2

Paneldaten

- Bei der Pooled Regression wird davon ausgegangen, dass es keine subjektspezifischen Unterschiede gibt. Für alle Subjekte gelten die gleichen Regressionskoeffizienten.
- Subjektspezifische Einflüsse werden ausschließlich in u_{it} erfasst und als Störung begriffen.
- Die Pooled Regression entspricht damit der normalen multiplen Regression.

Problem

- Der Fehlerterm korreliert mit einem oder mehreren Prädiktoren.
- Die zugehörigen Regressionskoeffizienten werden verzerrt und inkonsistent geschätzt.

Lässt man den Achsenabschnitt individuell werden, bewegt man sich zum FEM und REM. Wir springen sofort Richtung SURE, d.h. wir schätzen je Individuum eine Gleichung:

$$y_{it} = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{kit} + u_{it}$$

Was hat sich geändert?

Fallstudie

Event Study in Long Panel

Ausführen Schritt 1 in GML_05_01 sure.R

Betrachtung zweier Firmen in <u>kontrollierter</u> Umgebung

Erzeuge basierend auf dem CAPM die Aktienrenditen zweier Firmen

Firma a hat ein Beta von 1, Firma b von 2, das idiosynkratische Risiko habe eine Standardabweichung von 3% bzw 2%. Beide Störterme sind mit 85% korreliert, weil beide Firmen derselben Branche angehören.

Die Renditen des Marktportfolios seien zwecks Vereinfachung durch 60 Monatsrenditen des DAX gegeben.

```
EE <- mvrnorm(n,means,COV)
epsa = EE[,1]
epsb = EE[,2]X=coredata(indexReturn)
Ya = 1 * X + epsa
Yb = 2 * X + epsb
```

Was macht jeder einzelne Schritt?

Event Study in Long Panel

Betrachtung zweier Firmen in <u>kontrollierter</u> Umgebung
Nun werden zwei Extrarenditen ergänzt

Ya = 1 * X + epsa & mit „ExtraRendite 10% am Tag 30“
Yb = 2 * X + epsb & mit „ExtraRendite 2% am Tag 40“

```
#now event of xtra return of 10%/2% on day 30/40
Ya[30]=Ya[30]+0.1
Yb[40]=Yb[40]+0.02
Da = rep(0,60)
Db=Da
Da[30]=1
Db[40]=1
```

Was macht jeder einzelne Schritt?

Event Study in Long Panel

Ausführen Schritt 2 in GML_05_01 sure.R

Was tun wir hier 2x separat?

```
> Da[1:30]
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
> eq1 <- Ya~X+Da
> eq2 <- Yb~X+Db
> #two separate OLS
> summary(lm(eq1))

Call:
lm(formula = eq1)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.057746 -0.017792  0.000531  0.017210  0.062473 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.002862   0.003534  -0.810  0.42150    
X            1.021027   0.052647  19.394  < 2e-16 ***
Da           0.090559   0.026590   3.406  0.00122 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02628 on 57 degrees of freedom
Multiple R-squared:  0.8755,    Adjusted R-squared:  0.8711 
F-statistic: 200.4 on 2 and 57 DF,  p-value: < 2.2e-16
```

```
> summary(lm(eq2))

Call:
lm(formula = eq2)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.030982 -0.009791 -0.000948  0.008082  0.051850 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.002523   0.002314  -1.090    0.280    
X            1.987918   0.034902  56.958   <2e-16 ***
Db           0.013267   0.017628   0.753    0.455    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01724 on 57 degrees of freedom
Multiple R-squared:  0.9833,    Adjusted R-squared:  0.9827 
F-statistic:  1675 on 2 and 57 DF,  p-value: < 2.2e-16
```

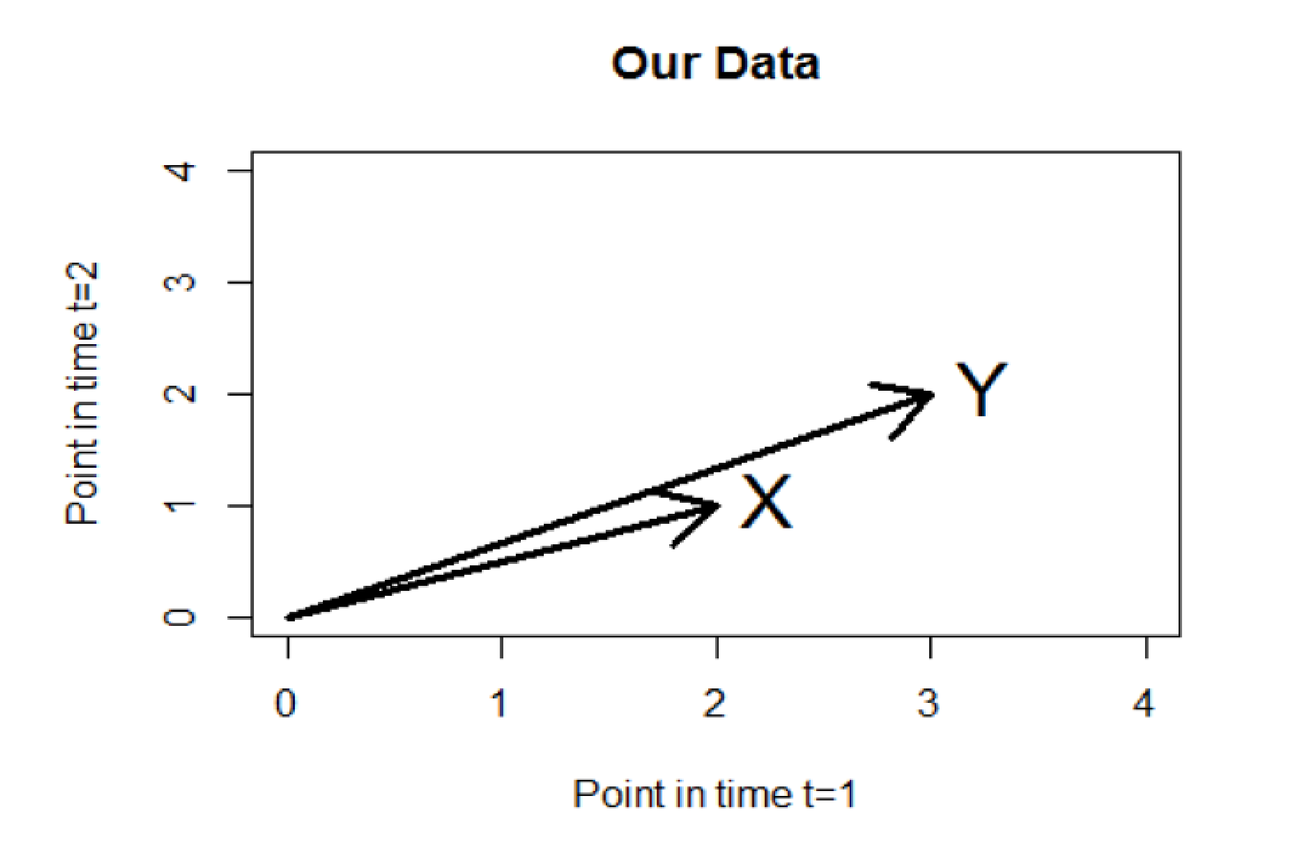
Wurden die Extrarenditen erkannt?

Ein einfaches Beispiel* (Coding GML_05_02 projection.R)

Man beobachte an 2 Zeitpunkten t=1,2 folgende Ausprägungen für X und Y:

Zeit t = 1: X=2 und Y=3

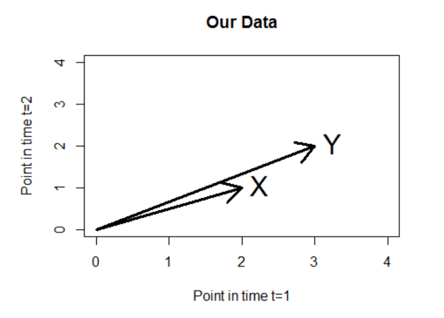
Zeit t = 2: X=1 und Y=2



*inspiriert durch Irizarry & Love, Data Analysis for the Life Sciences with R, Chapman Hall, 2017

Vertiefung Algebra

Ein einfaches Beispiel



- Wir wollen nun Y durch X erklären (also Y auf X zurückführen=regressieren). Man könnte versucht sein, noch einen zweiten Regressor hinzuzunehmen – etwa den Vektor (1,1). Aber in diesem Fall würde man durch eine Linearkombination zwei Unbekannte haben, um zwei Y Koordinaten zu treffen. Das Modell würde also ein Zweigleichungssystem mit 2 Unbekannten sein.

- Nach Ockham* wollen wir ja gerade mit wenig viel erklären. Hier sind Punkte in der Ebene (= 2-dimensional) zu erklären. Eine Herausforderung ist es, wenn man sich auf einen Unterraum beschränkt. Der eindimensionale Raum ist die Gerade, welche durch den X Vektor erzeugt wird via Stauchung und Streckung, d.h. ß*X

Wir versuchen also Y bestmöglich durch b*X zu approximieren!

Haben Sie ein Gefühl für die Grösse von ß?
Wie kann ich durch Streckung von X möglichst nah an Y ran?
Was für eine Regression haben Sie damit vor Augen?

* nach Wilhelm von Ockham (1288–1347) siehe: https://de.wikipedia.org/wiki/Ockhams_Rasiermesser

Vertiefung Algebra

Ein einfaches Beispiel

- Wir wollen, dass $\hat{Y}$ = ß * X bestmöglich! Konkret wollen wir, dass die Residuen senkrecht zu X sind, weil dies die **kürzeste Distanz von ß*X und Y** darstellt!
- Senkrecht bedeutet, das das Skalarprodukt vom Residuum Y-$\hat{Y}$ mit X Null ist, also ($\hat{Y}$-ß X) X = 0. Äquivalent mit YX – ß XX = 0. Woraus folgt ß = YX/XX. Das wird weiter unten nochmal sauberer geschrieben …
- Dies ist im R Coding umgesetzt:

```
sample_XY = crossprod(X,Y)#scalar product x1y1+x2y2
sample_XY

##      [,1]
## [1,]    8

sample_XX = crossprod(X,X)

b_hat = sample_XY / sample_XX
```

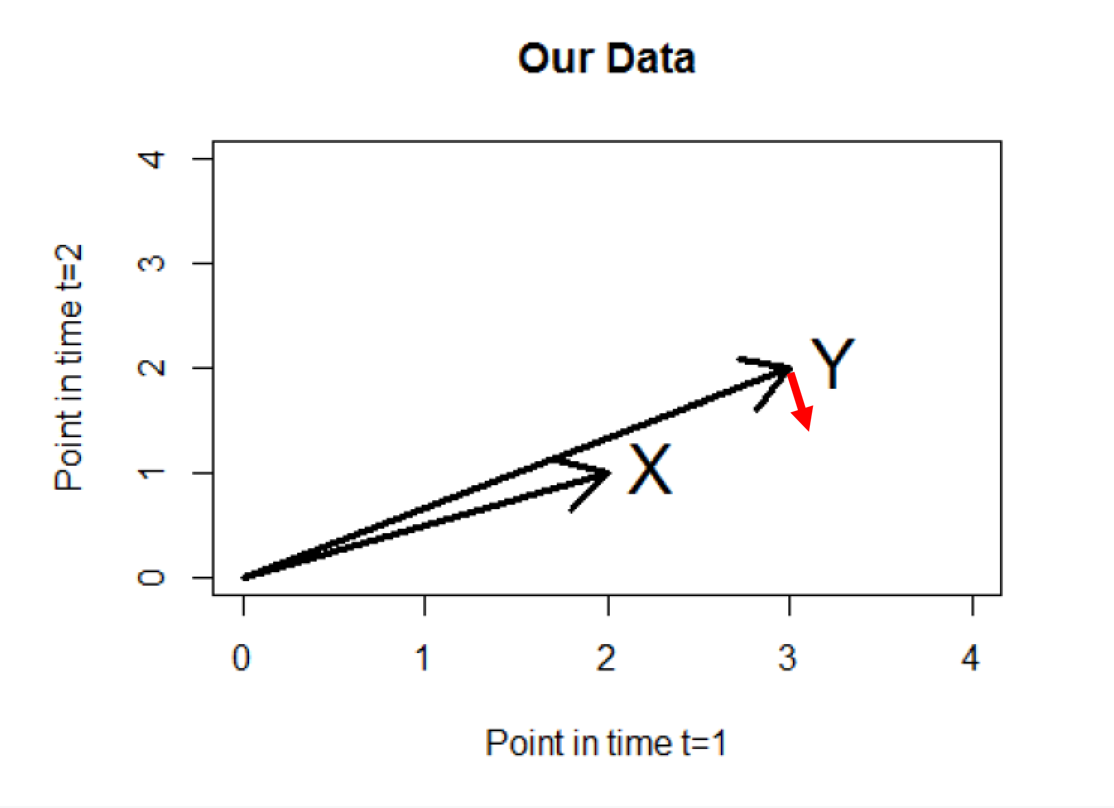
Das erste Skalarprodukt ist (2,1)*t(3,2) =2*3+1*2= 8, das zweite 5, **ß = 1,6**

Vertiefung Algebra

Ein einfaches Beispiel

- Wir wollen, dass $\hat{Y}$ = ß * X bestmöglich!
- Geometrisch ermitteln wir die kürzeste Distanz von Y zu ß*X, indem wir von ‚Ys Spitze das Lot auf die Gerade c*X fällen. Da wo es auftrifft, können wir ß ablesen! (Details, Assenmacher, S 91)

Wichtig: Der rechte Winkel steht auf X Vektor! Nicht auf Abzisse (x-Achse)!



Vertiefung Algebra

Ein einfaches Beispiel

Das erste Skalarprodukt ist (2,1)*t(3,2) =2*3+1*2= 8, das zweite 5, **ß = 1,6**

Und was kommt bei „lm“ raus?

```
reg = lm(Y~X-1)
summary(reg)

##
## Call:
## lm(formula = Y ~ X - 1)
##
## Residuals:
##    1    2
## -0.2  0.4
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## X      1.6        0.2       8   0.0792 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4472 on 1 degrees of freedom
## Multiple R-squared:  0.9846, Adjusted R-squared:  0.9692
## F-statistic:    64 on 1 and 1 DF,  p-value: 0.07917
```

Wie hätte man eine Regression mit zusätzlichem Regressor (1,1) aufgerufen?

Was wäre rausgekommen?

Ein einfaches Beispiel

```
#do dumb stuff
reg = lm(Y~X)
summary(reg)

##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
## ALL 2 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1         NA      NA       NA
## X                  1         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 1 and 0 DF,  p-value: NA

#end
```

Wie überprüfe ich den perfect fit? Was ergibt X+(1,1)?

Vertiefung Algebra

Ein einfaches Beispiel

- Im R Code werden weitere Details von lm von Hand nachgefahren.
- Hier wollen wir noch eine zentrale Erkenntnis vermitteln (Assenmacher, S 102):
- Wir wissen ß = YX/XX = XY/XX wegen Symmetrie im Zähler. Das wird nun sauberer geschrieben, wobei X'=: Transponierte Matrix(X) und Superindex -1 = Inverse

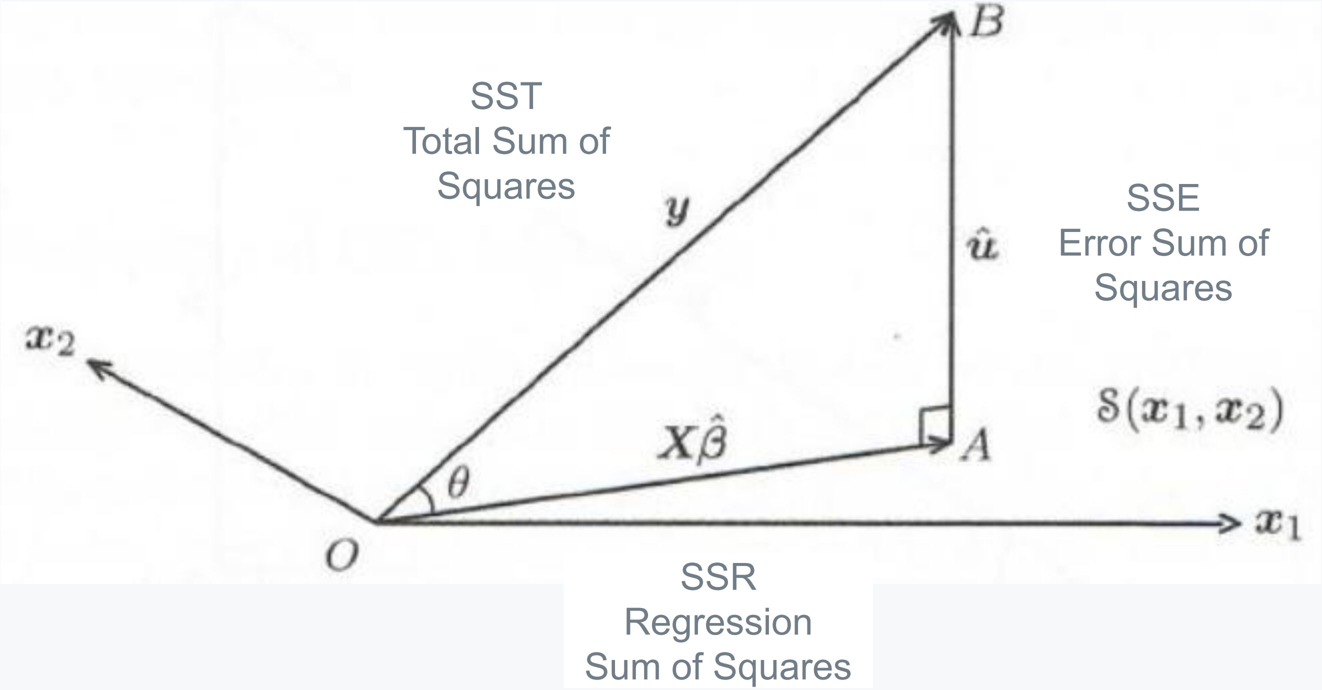
$$ß = (X'X)^{-1}X'Y = [(X'X)^{-1}X'] Y = [.] Y$$

- Der Schätzer für ß ergibt sich also als ein 1x2 Zeilenvektor [.] multipliziert mit einem 2x1 Spaltenvektor Y. **Es handelt sich bei ß also um eine Linearkombination von normalen ZV Y (Y=ßX+Störterm).**
- Deshalb wundert uns nicht, dass ß eine ZV ist und wir ihre Verteilung berechnen können (Details bei Davidson & MacKinnon, 2009, 124).
- Der Standardfehler von ß wird im R Code berechnet. **Wir verstehen nun, warum der ZGWS hilfreich ist. Für n gg unendlich ist ß normalverteilt, solange die Störterme iid (uiv)! Die Verteilung muss also nicht normal sein!!!**
- **Und wir verstehen das L in BLUE!**

Übergang zu SURE

Idee

Mathematisch ist die lineare Regression die Minimierung der Euklidschen Norm des Residuums. Grafisch:



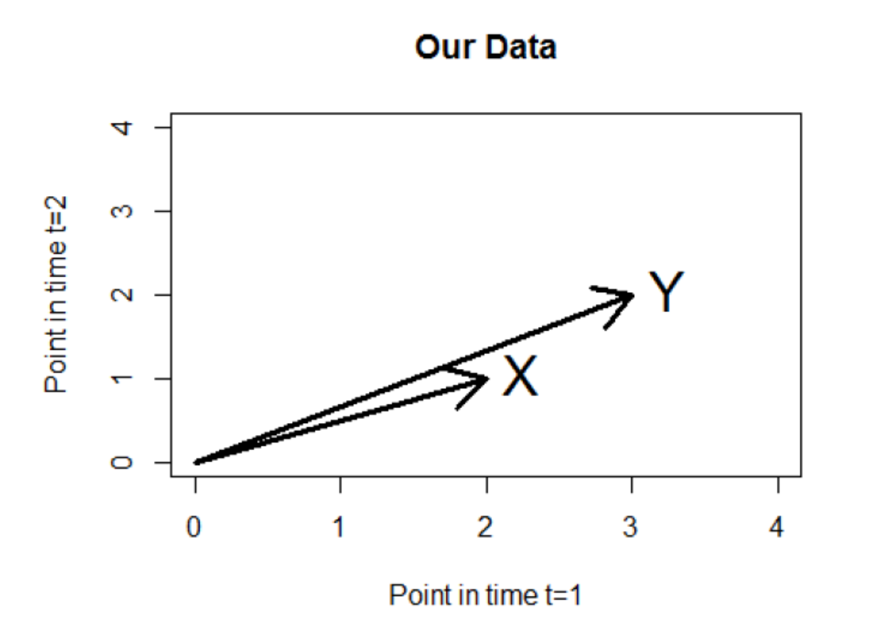
Quelle: Grafik aus Davidson & MacKinnon, 2009, 56

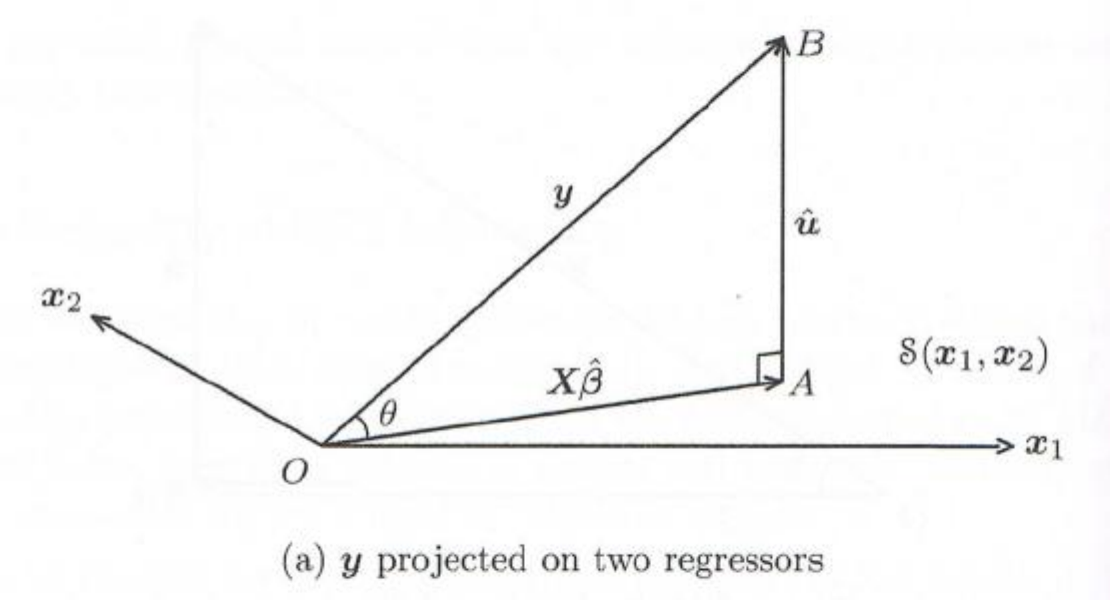
Denken Sie an den Pythagoras! Was erkennen wir zusätzlich ausser dem Schritt hin zum R3?

Übergang zu SURE

Idee

Mathematisch ist die lineare Regression die Minimierung der Euklidschen Norm des Residuums. Grafisch:





(a) y projected on two regressors

Quelle: Grafik aus Davidson & MacKinnon, 2009, 56

Wir approximieren ein höherdimensionales Y durch Streckung/Stauchung eines X aus einem Raum mit weniger Dimension. Dies ist verallgemeinerbar. R2 durch R1, R3 durch R2, und R-Stipo# durch R-k-Regressoren …

Übergang zu SURE

Idee

Wenn wir die zu schätzenden Modelle untereinanderschreiben, ergibt sich ein langer Residuumsvektor. Dessen Norm können wir wie bisher minimieren. D.h. wir können das System in einem Schritt lösen. Die Störterme dürfen nun auch korreliert sein. Das erfordert lediglich eine Transformation. Damit sind wir bei GLS.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_E \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & X_E \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_E \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_E \end{bmatrix}$$

Quelle: In Anlehnung an *Zellner, A.*, Estimating Seemingly Unrelated Regressions, 1962, S. 349

Anders als die Grafik suggeriert, dürfen Regressoren auch in jeder der Gleichungen auftauchen, so wie die DAX Rendite, aber es darf auch X geben, die nur bei einem Subjekt auftauchen wie die Dummies.

Fallstudie mit SURE

Ergebnis

Ausführen Schritt 2 in GML_05_01 sure.R

```
> coeftest(fitOLS)

t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)
eq1_(Intercept) -0.0028616  0.0035343 -0.8097  0.419824
eq1_X            1.0210271  0.0526470 19.3938 < 2.2e-16 ***
eq1_Da           0.0905595  0.0265902  3.4058  0.000912 ***
eq2_(Intercept) -0.0025229  0.0023140 -1.0903  0.277893
eq2_X            1.9879178  0.0349017 56.9577 < 2.2e-16 ***
eq2_Db           0.0132669  0.0176276  0.7526  0.453229
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(fitSUR)

t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)
eq1_(Intercept) -0.0029718  0.0035211 -0.8440  0.400443
eq1_X            1.0197683  0.0525315 19.4125 < 2.2e-16 ***
eq1_Da           0.0984919  0.0149879  6.5714 1.556e-09 ***
eq2_(Intercept) -0.0026712  0.0023085 -1.1571  0.249632
eq2_X            1.9835196  0.0345785 57.3628 < 2.2e-16 ***
eq2_Db           0.0267810  0.0099361  2.6953  0.008096 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wurden die Extrarenditen erkannt?

Fallstudie mit SURE

Rückblick

An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias
Author(s): Arnold Zellner
Source: *Journal of the American Statistical Association*, Vol. 57, No. 298 (Jun., 1962), pp. 348-368

In this paper a method of estimating the parameters of a set of regression equations is reported which involves application of Aitken's generalized least-squares [1] to the whole system of equations. Under conditions generally encountered in practice, it is found that the regression coefficient estimators so obtained are at least asymptotically more efficient than those obtained by an equation-by-equation application of least squares. This gain in efficiency can be quite large if "independent" variables in different equations are not highly correlated and if disturbance terms in different equations are highly correlated. Further,

Warum habe ich im Beispiel eine Korrelation von 85% gewählt??
Wie unabhängig waren die Dummies?

Quellen

ARIMA & GARCH

Nachbereitung

Assenmacher (2002), Kap. 12 und Coding inkl. Diagnostik und Test auf Korr. Bei Resis

Zur Abrundung FEM/REM

Ch. 10 in Verbeek (2017)

Kap. 3 in Gehrke (2022)

Nachtrag zu Konvergenzkonzepten

- **Die mehrmals genutzte Konsistenz ist die schwache Konsistenz**, d.h. Schätzer ß konvergiert in WS gg b, **formal** lim $_{n\to inf}$ P(|b-ß$_n$|>d) = 0 für d > 0 (beliebig klein!)
- I.a.W. Die WS ß nahe b, also in einer noch so kleinen d Umgebung um ß herum, zu finden, geht gegen 1, formal Komplementärereignis: lim $_{n\to inf}$ P(|b-ß$_n$|<d) = 1

- **Strenge Konsistenz** ist stärker, nämlich, dass mit WS 1 ß auf b zuläuft. D.h. das Ereignis, dass ß neben b landet, hat WS von Null.
- Formal: lim $_{n\to inf}$ P(ß$_n$->b) = 1

- **Deterministische Konvergenz** ist irrelevant, d.h. wir können nicht sicher sein, dass ß$_n$->b mit n gegen unendlich, weil wir ja nicht wissen, was für eine Stipo wir haben

- „Consistency and unbiasedness are simply different concepts“ (Davidson & MacKinnon, 2009, 96).
- „Estimators may be biased but consistent“ (Davidson & MacKinnon, 2009, 96).
- Example (Davidson & MacKinnon, 2009, 97): Estimate the true population mean µ by $\frac{\sum_{t=1}^{n} x_t}{n+1}$. We see that it comes close to the unbiased estimator „sample mean“ in the limit for large n, because then n is very similar to n+1. Hence, it is consistent but misses the true µ always slightly.