

Tutorial

Stata und Wharton Research Data Services (WRDS)

Inhaltsverzeichnis

1	Stata
2	Wharton Research Data Services (WRDS)

Inhaltsverzeichnis

1	Stata
2	Wharton Research Data Services (WRDS)

- 1 Was ist Stata?
- 2 Benutzeroberfläche
- 3 do-file
- 4 Daten einlesen und Datensätze mergen
- 5 Änderung der Variablentypisierung
- 6 Variablenberechnung
- 7 Bilden einer Stichprobe
- 8 Implementierung zeitverzögerter Effekte
- 9 Deskriptive Statistik
- 10 Regression
- 11 Wie behalte ich den Überblick?
- 12 Weiterführende Informationen

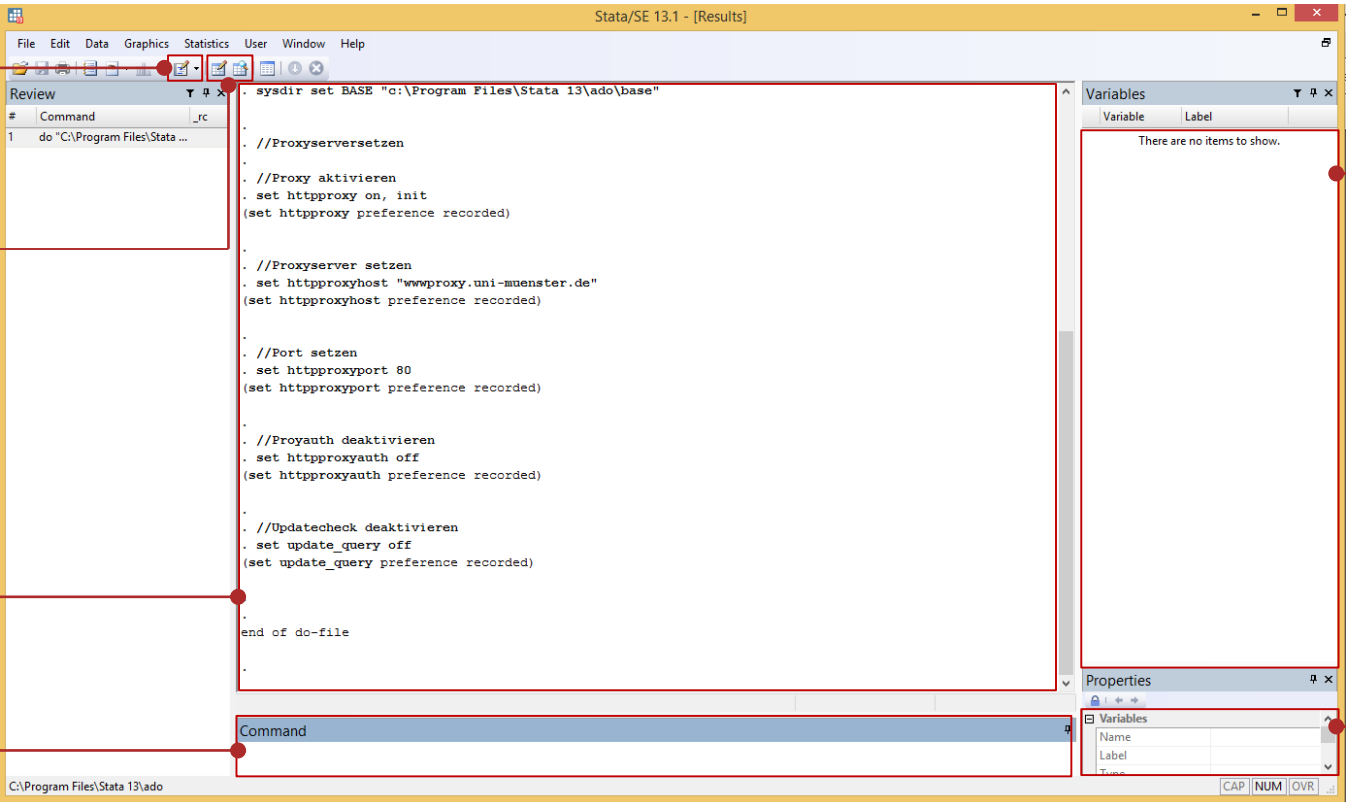
Stata

Was ist Stata?

- Stata ist eine kommerzielle Statistik-Software, die das Daten-Management insbesondere großer Datensätze vereinfacht
- Implementierung einer Vielzahl statistischer Modelle
- Implementierung einer Vielzahl grafischer Darstellungen
- Selbstverständnis von Stata als eigenständiges Betriebssystem mit einer recht simplen Programmiersprache
- Sehr hohe Transparenz durch die Offenlegung sämtlicher in Stata verwendeter Formeln und Funktionen

Stata

Benutzeroberfläche



The screenshot shows the Stata/SE 13.1 interface with the following components and annotations:

- do-file**: Points to the top toolbar icons.
- Data Editor**: Points to the Review window on the left side.
- Terminal Output**: Points to the main command window displaying the do-file content.
- Command Window**: Points to the Command input field at the bottom.
- Variablen Window**: Points to the Variables window on the right, which currently shows "There are no items to show."
- Eigenschaften Window**: Points to the Properties window at the bottom right, which shows the "Variables" section.

```

Review
# Command      _rc
1 do "C:\Program Files\Stata ...

.do file: C:\Program Files\Stata 13\ado\base.do
1 . sysdir set BASE "c:\Program Files\Stata 13\ado\base"
2 . //Proxyserver setzen
3 . //Proxy aktivieren
4 . set httpproxy on, init
5 (set httpproxy preference recorded)
6 .
7 . //Proxyserver setzen
8 . set httpproxyhost "wwwproxy.uni-muenster.de"
9 (set httpproxyhost preference recorded)
10 .
11 . //Port setzen
12 . set httpproxyport 80
13 (set httpproxyport preference recorded)
14 .
15 . //Proxyauth deaktivieren
16 . set httpproxyauth off
17 (set httpproxyauth preference recorded)
18 .
19 . //Updatecheck deaktivieren
20 . set update_query off
21 (set update_query preference recorded)
22 .
23 . end of do-file
24 .
25 .

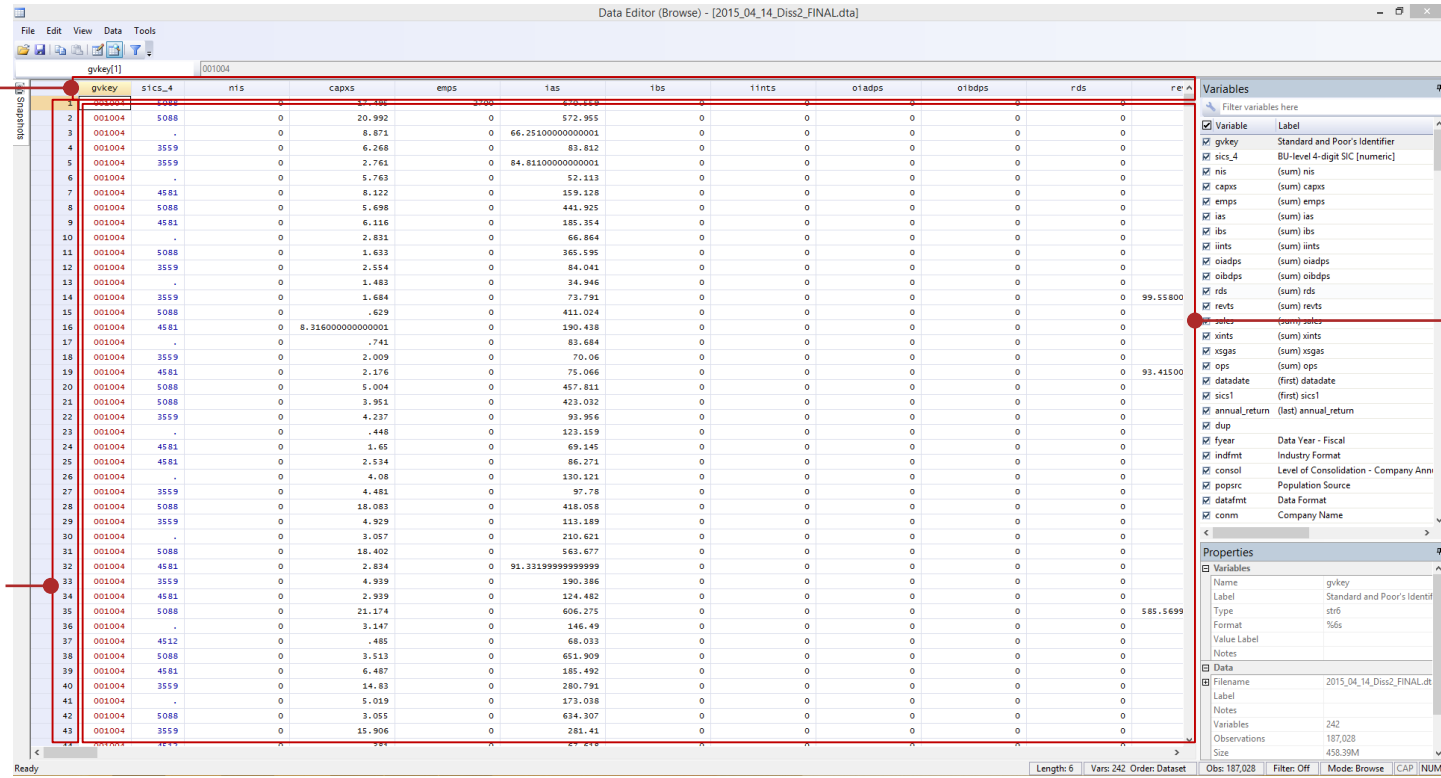
Command
C:\Program Files\Stata 13\ado
  
```

Stata

Benutzeroberfläche

Variablen-
bezeichnung

Observation



The screenshot shows the Stata Data Editor (Browse) window for the file [2015_04_14_Diss2_FINAL.dta]. The main window displays a grid of data with columns for variables: gvkey, s1cs_4, n1s, capxs, emp, l1s, l1bs, l1nts, o1adps, o1bdps, rds, and revts. The rows represent observations, with the first row highlighted in red. On the right side, there is a 'Variables' panel listing all variables in the dataset, including their labels and types. Below the variables panel is the 'Properties' panel, which shows details for the selected variable 'gvkey', such as its label, type, format, and value label. The status bar at the bottom indicates 'Ready', 'Length: 6', 'Vars: 242', 'Order: Dataset', 'Obs: 187,028', 'Filter: Off', 'Mode: Browse', 'CAP | NUM'.

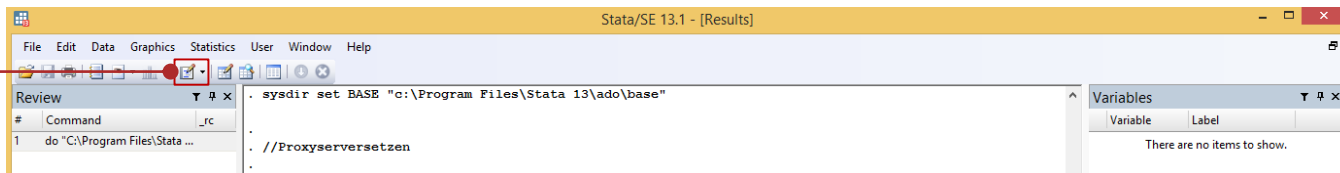
Daten

Stata

do-file

- Achtung: Direkte Bearbeitung kann zu unwiderruflichen Datenverlusten führen
- Daher: Nutzen eines do-files, der den verwendeten Roh-Datensatz immer wieder neu einliest und den bearbeiteten Datensatz in einer neuen Datei abspeichert

do-file



Stata

Daten einlesen und Datensätze mergen

Löscht den internen Speicher

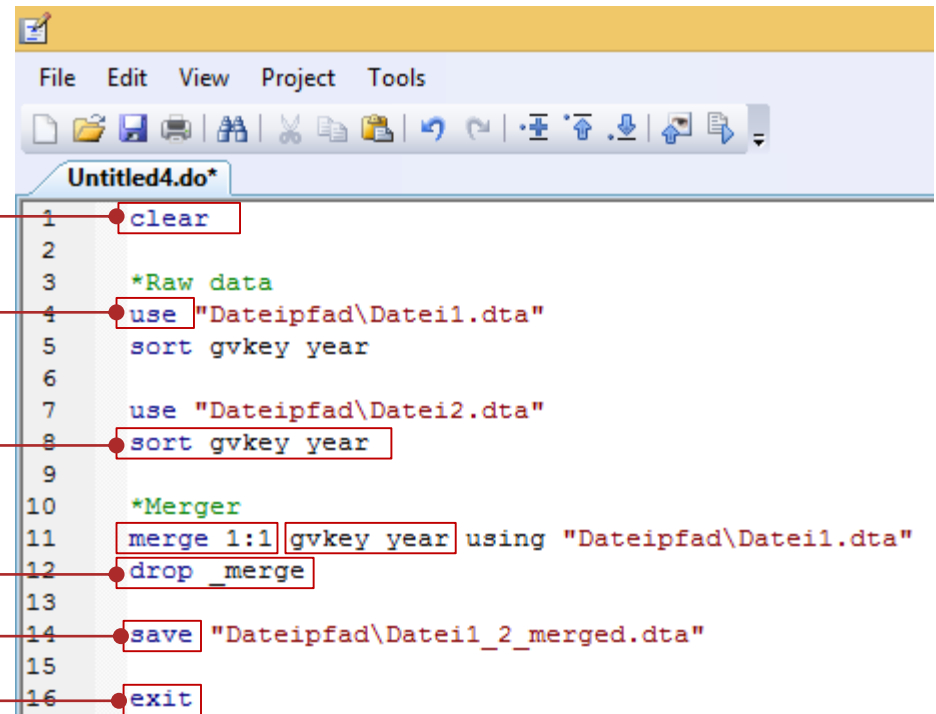
Daten einlesen

Die zu mergenden Datensätze müssen vorab
entsprechend ihrer **EINDEUTIGEN** und **EINMALIGEN**
Variablenkombination sortiert werden

Löscht die nicht benötigte Variable `“_merge“`

Speichern des neuen Datensatzes

Lässt Stata wissen, dass der do-file beendet ist



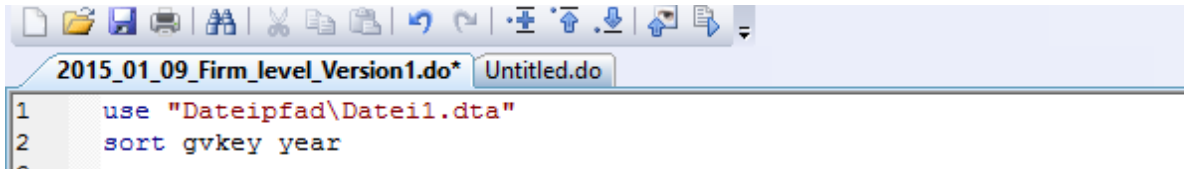
```
1 clear
2
3 *Raw data
4 use "Dateipfad\Datei1.dta"
5 sort gvkey year
6
7 use "Dateipfad\Datei2.dta"
8 sort gvkey year
9
10 *Merger
11 merge 1:1 gvkey year using "Dateipfad\Datei1.dta"
12 drop _merge
13
14 save "Dateipfad\Datei1_2_merged.dta"
15
16 exit
```

Weitere merger: one-to-many (1:m); many-to-one (m:1); many-to-many (m:m)

Stata

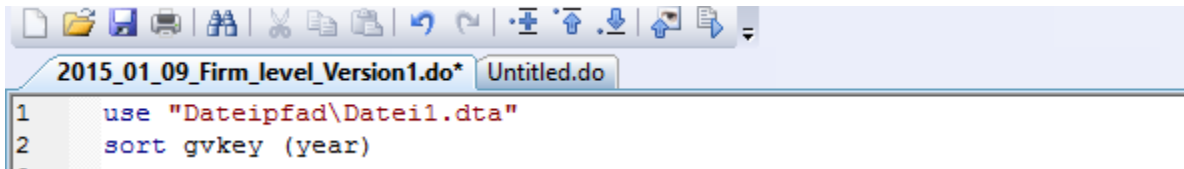
Daten einlesen und Datensätze mergen

- Sortieren der Datensätze



```
2015_01_09_Firm_level_Version1.do* Untitled.do
1 use "Dateipfad\Datei1.dta"
2 sort gvkey year
```

- Sortiert nach Variable 1 und Variable 2 gleichzeitig



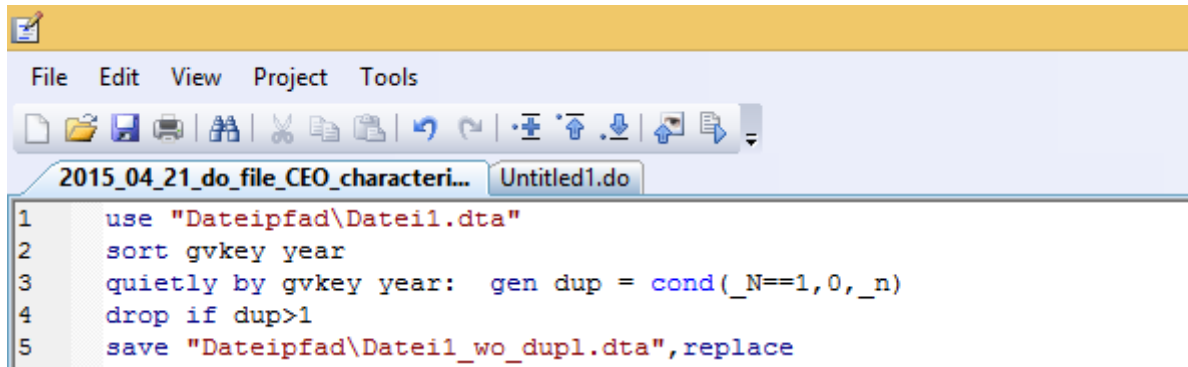
```
2015_01_09_Firm_level_Version1.do* Untitled.do
1 use "Dateipfad\Datei1.dta"
2 sort gvkey (year)
```

- Sortiert erst nach Variable 1 und dann nach Variable 2
- Sinnvoll bei Datensatz mit einer verschachtelten Struktur (Bsp.: Datensatz mit sowohl Daten auf Segment- als auch Unternehmensebene)

Stata

Daten einlesen und Datensätze mergen

- Duplikate auffinden und aus dem Datensatz löschen



```
1 use "Dateipfad\Datei1.dta"  
2 sort gvkey year  
3 quietly by gvkey year: gen dup = cond(_N==1,0,_n)  
4 drop if dup>1  
5 save "Dateipfad\Datei1_wo_dupl.dta",replace
```

- Variable “*dup*” hilft falsche/fehlerhafte Dateneingaben ausfindig zu machen
- Lediglich die erste Observation eines Duplikats wird zur weiteren Berechnung herangezogen
- Zum mergen von zwei Datensätze obligatorisch

Stata

Daten einlesen und Datensätze mergen

- Beim mergen von Datensätzen wird die Variable “*_merge*” von Stata automatisch generiert
 - “*_merge==1*”: Daten nur in Masterdatensatz vorhanden
 - “*_merge==2*”: Daten nur in zu mergenden Datensatz vorhanden
 - “*_merge==3*”: Daten sind im kombinierten Datensatz vorhanden

Result	# of obs.	
not matched	139,720	
from master	131,089	(<i>_merge==1</i>)
from using	8,631	(<i>_merge==2</i>)
matched	55,074	(<i>_merge==3</i>)

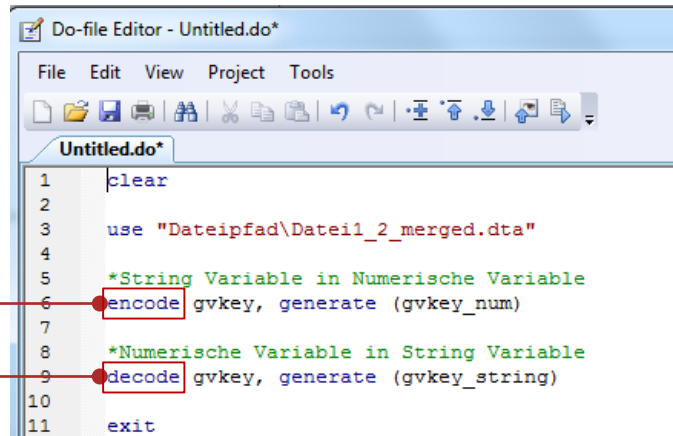
- Lediglich der kombinierte Datensatz wird zur weiteren Berechnung herangezogen

Stata

Änderung der Variablentypisierung

Wandelt eine String
Variable in eine
numerische Variable um

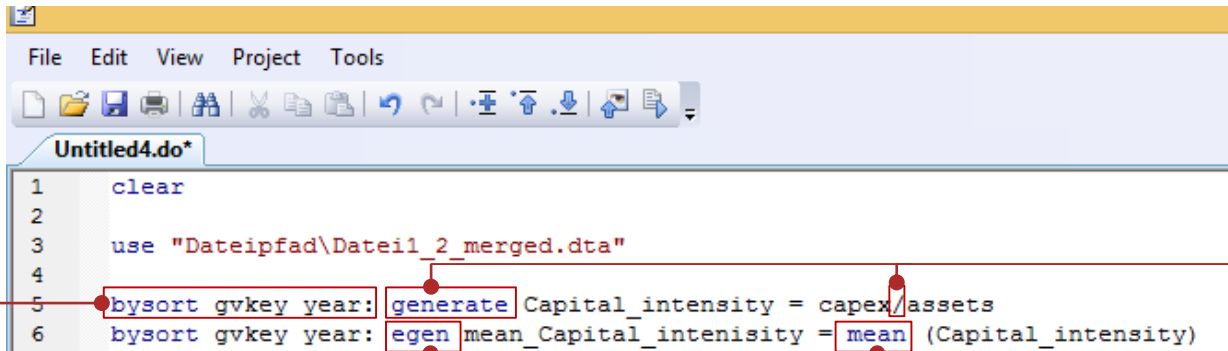
Wandelt eine
numerische Variable in
eine String Variable um



```
Do-file Editor - Untitled.do*
File Edit View Project Tools
Untitled.do*
1 clear
2
3 use "Dateipfad\Datei1_2_merged.dta"
4
5 *String Variable in Numerische Variable
6 encode gvkey, generate (gvkey_num)
7
8 *Numerische Variable in String Variable
9 decode gvkey, generate (gvkey_string)
10
11 exit
```

Stata

Variablenberechnung



```
1 clear
2
3 use "Dateipfad\Datei1_2_merged.dta"
4
5 bysort gvkey year: generate Capital_intensity = capex/assets
6 bysort gvkey year: egen mean_Capital_intenensity = mean (Capital_intensity)
```

Sortieren des
Datensatzes

Simple
Operatoren (+, -, * etc.)

Komplexere Operatoren
(median, sum, total etc.)

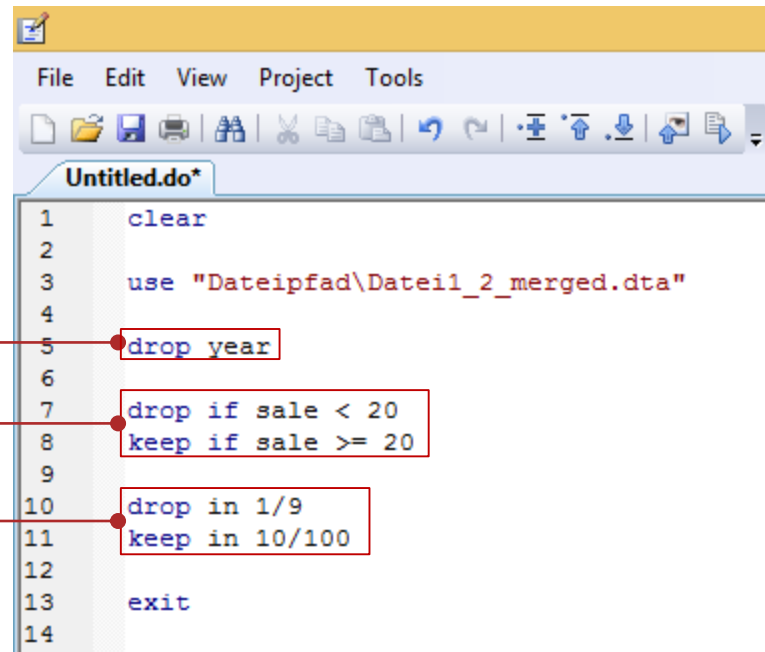
Stata

Bilden einer Stichprobe

Alle Observations der Variable „*year*“ fallen lassen

Alle Observations, bei der die Variable „*sale*“ kleiner als \$20 Mio., fallen lassen

Die ersten neun Observations über alle Variablen fallen lassen



```
1  clear
2
3  use "Dateipfad\Datei1_2_merged.dta"
4
5  drop year
6
7  drop if sale < 20
8  keep if sale >= 20
9
10 drop in 1/9
11 keep in 10/100
12
13 exit
14
```

Stata

Implementierung zeitverzögerter Effekte (time-lag/lead)

```
Do-file Editor - Untitled.do*
File Edit View Project Tools
Untitled.do*
1 clear
2
3 use "Dateipfad\Datei1_2_merged.dta"
4
5 *Lag
6 bysort gvkey year: generate capx_lag= (capx[_n-1]) if year==year[_n-1]+1
7
8 *Lead
9 bysort gvkey year: generate capx_lead= (capx[_n+1]) if year==year[_n+1]-1
10
11 exit
```

Bezeichnung der
(neu generierten)
Variable

Gibt den Wert des
Folgejahrs wieder

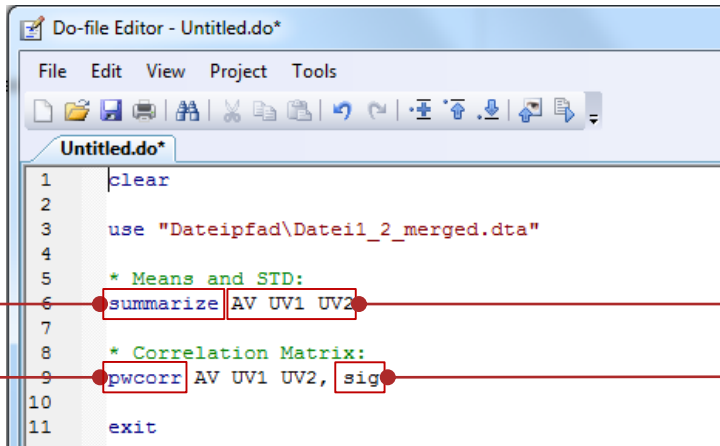
Gibt den Vorjahreswert wieder

Überprüft, ob ein Vorjahreswert
existiert

Überprüft, ob ein Wert im
Folgejahr existiert

Stata

Deskriptive Statistik



```
1 clear
2
3 use "Dateipfad\Datei1_2_merged.dta"
4
5 * Means and STD:
6 summarize AV UV1 UV2
7
8 * Correlation Matrix:
9 pwcorr AV UV1 UV2, sig
10
11 exit
```

Mittelwerte und
Standardabweichung

Paarweise Korrelation
der Koeffizienten

Variablenliste

Signifikanzlevel der einzelnen Korrelationen

Stata

Regression

Paneldaten-
Deklaration:
Panelvariable,
Zeitvariable

Internes Speichern
der Koeffizienten

Hausman-Test

Regression von
Paneldaten

```

1  clear
2
3  use "Dateipfad\Datei1_2_merged.dta"
4
5  *Paneldaten
6  xtset gvkey year
7
8  *Hausman-Test
9  quietly xtreg AV UV1 UV2 UV3, fe
10 estimates store fixed_effects
11 quietly xtreg AV UV1 UV2 UV3, re
12 estimates store random_effects
13 hausman fixed_effects random_effects
14
15 *Finales Regressionsmodell
16 xtreg AV UV1 UV2 UV3, fe
17
18 exit
19

```

Unterdrücken des
Terminal-Outputs

fixed-effects

random-effects

Stata

Wie behalte ich den Überblick?

Kommentare

```

380 *Industry Growth capability
381 bysort sics_4 year: egen salesgrowth_industry_4 = median (salesgrowth_single) if count_seg4 > 4
382 label var salesgrowth_industry_4 "Industry adjusted salesgrowth on 4-digit SIC"
383 bysort sics_3 year: egen salesgrowth_industry_3 = median (salesgrowth_single) if count_seg4 < 5
384 label var salesgrowth_industry_3 "Industry adjusted salesgrowth on 3-digit SIC"
385 bysort sics_2 year: egen salesgrowth_industry_2 = median (salesgrowth_single) if count_seg4 < 5 & count_seg3 < 5
386 label var salesgrowth_industry_2 "Industry adjusted salesgrowth on 2-digit SIC"
387
388 generate ia_salesgrowth = salesgrowth_industry_4
389 replace ia_salesgrowth = salesgrowth_industry_3 if count_seg4 < 5
390 replace ia_salesgrowth = salesgrowth_industry_2 if count_seg4 < 5 & count_seg3 < 5
391 replace ia_salesgrowth = . if diversification < 1
392 label var ia_salesgrowth "Industry adjusted salesgrowth [at least 5 business units within the same industry of focal firms]"
393 *winsorize variable at the 1st and 99th percentiles
394 winsor ia_salesgrowth, generate (ia_salesgrowth_w) p(0.01)
395
396 *****
397 /*Financial constraints?*/
398 generate dividends = (dvc_w+dvp_w)/at
399 *winsorize variable at the 1st and 99th percentiles
400 winsor dividends, generate (dividends_w) p(0.01)
401 generate debt = dltt_w+dlc_w
402 *winsorize variable at the 1st and 99th percentiles
403 winsor debt, generate (debt_w) p(0.01)
404

```

Eindeutige Variablenbezeichnungen

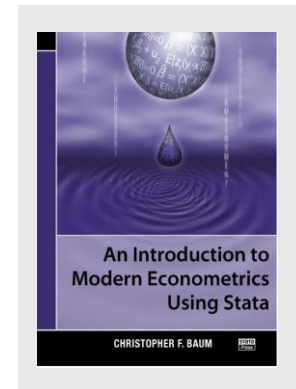
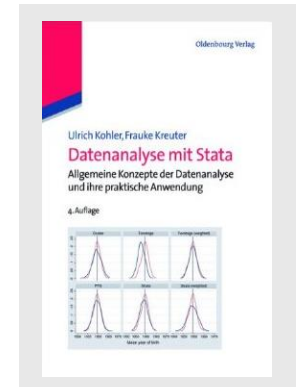
Grafische Elemente

Variablenlabel

Stata

Weiterführende Informationen

- help-Funktion über command-Fenster (Bsp.: „*help pwcrr*“)
- Internet-Foren
- Bedienungsanleitung
- Stata-Lernbücher:
 - Christopher F. Baum: An Introduction to Modern Econometrics Using Stata
 - Ulrich Kohler & Frauke Kreuter: Datenanalyse mit Stata



Inhaltsverzeichnis

1	Stata
2	Wharton Research Data Services (WRDS)

- 1 Was ist WRDS?
- 2 Benutzeroberfläche
- 3 Datenzugriff
- 4 Beispieldatensatz
- 5 Zugang
- 6 Weiterführende Informationen

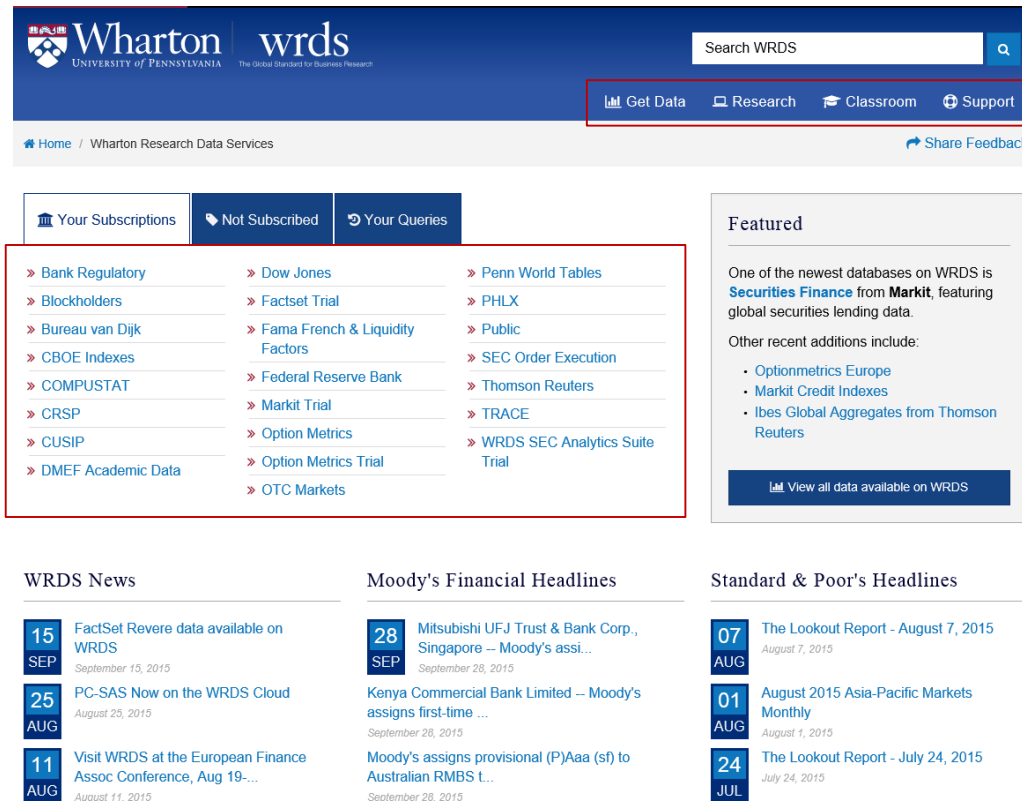
WRDS

Was ist WRDS?

- WRDS: Wharton Research Data Services
- Führende Data Research-Plattform mit über 200 Terabyte an Daten
- Datenbanken decken die verschiedensten Bereiche ab (Bsp.: Accounting, Banking, Finance und Marketing)
- WRDS bündelt die Datenbanken in einem standardisierten Format
- Goldstandard in den Wirtschaftswissenschaften

WRDS

Benutzeroberfläche



The screenshot shows the Wharton WRDS website interface. At the top, there is a navigation bar with the Wharton University of Pennsylvania logo and the WRDS tagline "The Global Standard for Business Research". A search bar is located on the right side of the navigation bar. Below the navigation bar, there are several tabs: "Your Subscriptions", "Not Subscribed", and "Your Queries". The "Your Subscriptions" tab is active, displaying a grid of subscription options such as "Bank Regulatory", "Blockholders", "Bureau van Dijk", "CBOE Indexes", "COMPUSTAT", "CRSP", "CUSIP", "DMEF Academic Data", "Dow Jones", "Factset Trial", "Fama French & Liquidity Factors", "Federal Reserve Bank", "Markit Trial", "Option Metrics", "Option Metrics Trial", "OTC Markets", "Penn World Tables", "PHLX", "Public", "SEC Order Execution", "Thomson Reuters", "TRACE", and "WRDS SEC Analytics Suite Trial". A "Featured" section on the right highlights the "Securities Finance" database from Markit. Below the featured section, there are three columns of news headlines: "WRDS News", "Moody's Financial Headlines", and "Standard & Poor's Headlines".

Wharton **wrds**
UNIVERSITY OF PENNSYLVANIA The Global Standard for Business Research

Search WRDS

Get Data Research Classroom Support

Home / Wharton Research Data Services Share Feedback

Your Subscriptions Not Subscribed Your Queries

- » Bank Regulatory
- » Blockholders
- » Bureau van Dijk
- » CBOE Indexes
- » COMPUSTAT
- » CRSP
- » CUSIP
- » DMEF Academic Data
- » Dow Jones
- » Factset Trial
- » Fama French & Liquidity Factors
- » Federal Reserve Bank
- » Markit Trial
- » Option Metrics
- » Option Metrics Trial
- » OTC Markets
- » Penn World Tables
- » PHLX
- » Public
- » SEC Order Execution
- » Thomson Reuters
- » TRACE
- » WRDS SEC Analytics Suite Trial

Featured

One of the newest databases on WRDS is **Securities Finance** from **Markit**, featuring global securities lending data.

Other recent additions include:

- Optionmetrics Europe
- Markit Credit Indexes
- Ibes Global Aggregates from Thomson Reuters

View all data available on WRDS

WRDS News

- 15 SEP** FactSet Revere data available on WRDS
September 15, 2015
- 25 AUG** PC-SAS Now on the WRDS Cloud
August 25, 2015
- 11 AUG** Visit WRDS at the European Finance Assoc Conference, Aug 19-...
August 11, 2015

Moody's Financial Headlines

- 28 SEP** Mitsubishi UFJ Trust & Bank Corp., Singapore -- Moody's assi...
September 28, 2015
- Kenya Commercial Bank Limited -- Moody's assigns first-time ...
September 28, 2015
- Moody's assigns provisional (P)Aaa (sf) to Australian RMBS t...
September 28, 2015

Standard & Poor's Headlines

- 07 AUG** The Lookout Report - August 7, 2015
August 7, 2015
- 01 AUG** August 2015 Asia-Pacific Markets Monthly
August 1, 2015
- 24 JUL** The Lookout Report - July 24, 2015
July 24, 2015

WRDS

Benutzeroberfläche



The screenshot displays the Wharton WRDS website interface. At the top, there is a blue header with the Wharton University of Pennsylvania logo and the WRDS logo. A search bar is located in the top right corner. Below the header, there is a navigation menu with links for 'Get Data', 'Research', 'Classroom', and 'Support'. A breadcrumb trail shows the current path: 'Home / Get Data / Compustat / North America - Annual Updates / Fundamentals Annual'. A 'Share Feedback' link is also present.

The main content area is divided into two columns. The left column contains a list of navigation links under the heading 'Compustat' and 'North America - Annual Updates'. The right column contains the main content area, which is currently displaying the 'Fundamentals Annual' dataset page. This page includes a title 'Compustat Annual Updates - Fundamentals Annual', a description, and a search form.

Navigation Links (Left Column):

- Fundamentals Annual
- Fundamentals Quarterly
- Index Constituents
- Index Fundamentals
- Index Prices
- Industry Specific Annual
- Industry Specific Quarterly
- Pension Annual
- Pension Quarterly
- Ratings
- Security Monthly
- Segments (Non-Historical)
- Supplemental Short Interest File

Main Content Area (Right Column):

Compustat Annual Updates - Fundamentals Annual

For more about this dataset, see the [Variable Descriptions](#), [Dataset List](#), [Manuals and Overviews](#) or [FAQs](#).

Click here to [preview this dataset](#) (Beta)

Step 1: What date range do you want to use?

Date Variable:

I would like data from start date: to end date: (yyyy-mm).

Step 2: How would you like to search this dataset?

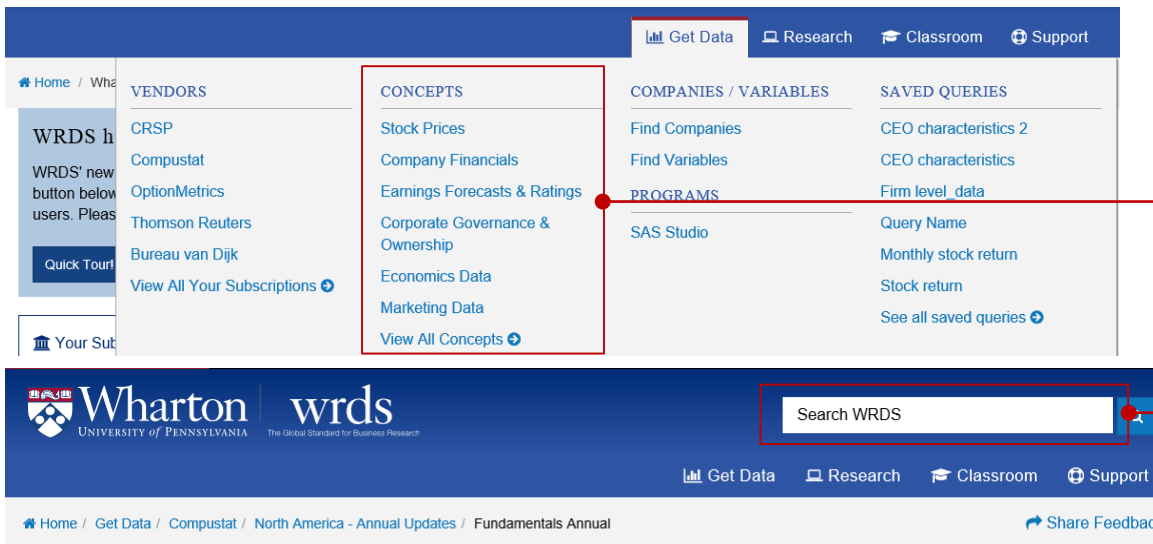
What format are your company codes?

- TIC
- GVKEY
- CUSIP
- SIC
- NAICS
- CIK

Manually enter company codes

WRDS

Datenzugriff



The screenshot shows the WRDS website navigation menu. The 'Get Data' tab is highlighted in red. Below it, the 'CONCEPTS' menu is also highlighted in red, listing various data categories like Stock Prices, Company Financials, and Earnings Forecasts & Ratings. The 'COMPANIES / VARIABLES' menu is also visible, listing options like Find Companies and Find Variables. The 'SAVED QUERIES' menu is also visible, listing options like CEO characteristics 2 and Firm level_data.

Home-Button: Get Data

Search WRDS

Compustat

Compustat Annual Updates - Fundamentals Annual

North America - Annual Updates

Fundamentals Annual
Fundamentals Quarterly
Index Constituents
Index Fundamentals

For more about this dataset, see the [Variable Descriptions](#), [Dataset List](#), [Manuals and Overviews](#) or [FAQs](#).

Click here to [preview this dataset!](#) (Beta)

Variablenbeschreibung des
ausgewählten Datensatzes



wrds business unit ROA



Internetrecherche

WRDS

Beispieldatensatz

- WRDS-Webinar (ZIV-Zugangsdaten benötigt)



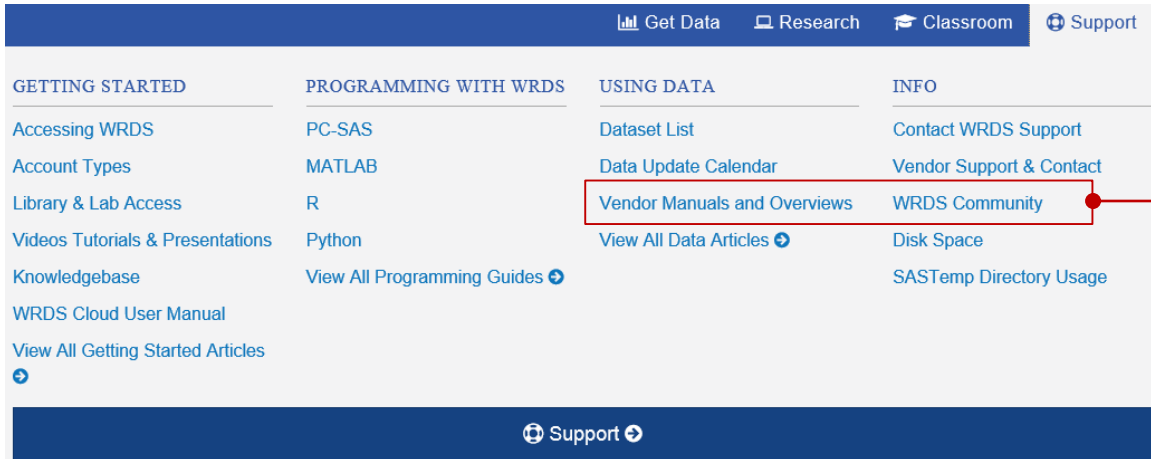
WRDS

Zugang

- Ansprechpartner: Betriebliche Datenverarbeitung
- Voraussetzungen:
 - Der Antragssteller muss über eine gültige ZIV-Kennung verfügen
 - Regelungen für den Zugriff auf das Datenbankangebot
 - Erklärung zur Freischaltung
 - Antrag auf Freischaltung → Betreuer muss eingetragen werden

WRDS

Weiterführende Informationen



The screenshot shows the WRDS website navigation menu. The menu is organized into four columns: GETTING STARTED, PROGRAMMING WITH WRDS, USING DATA, and INFO. The 'Vendor Manuals and Overviews' link in the USING DATA column is highlighted with a red box. A red arrow points from this link to the text 'Manuals and Overview; WRDS-Forum' on the right side of the slide.

GETTING STARTED	PROGRAMMING WITH WRDS	USING DATA	INFO
Accessing WRDS	PC-SAS	Dataset List	Contact WRDS Support
Account Types	MATLAB	Data Update Calendar	Vendor Support & Contact
Library & Lab Access	R	Vendor Manuals and Overviews	WRDS Community
Videos Tutorials & Presentations	Python	View All Data Articles	Disk Space
Knowledgebase	View All Programming Guides		SASTemp Directory Usage
WRDS Cloud User Manual			
View All Getting Started Articles			

Manuals and Overview; WRDS-Forum

- Weitere Foren (Google-Suche)
- Journal-Paper