Data documentation for

„Coaching, Councelling, Caseworking: Do They Help the Older Unemployed Out Of Benefit Receipt and Back Into the Labor Market?"

by Bernhard Boockmann and Tobias Brändle

Table of contents

I. Basic Data Preparation

0. Master and settings

	File name	Contents	Key variables generated	Data file start	Data file end
0.0	gen_p50p - 0.0 - settings	Central Do-File for Settings and Globals			
0.1	gen_p50p - 0.1 - master	Central Master Do-File			
0.2	gen_p50p - 0.2 - R - idgenerator	Creates uniform IDs and plays them to the records Some people have two dophash_id's (see delivery documentation "Doku_012209_IAW.pdf") and are duplicates except for the ID. These dophash_id's are specially transferred to iaw_id's, deleted and the duplicate rows of data removed. Associated dophash_id's can be identified by the variable n_paar. A list of duplicate dophash_id's can be found in "a022209_identical_dophash_iaw.dta".	iaw_id	a012209_`quelle'_iaw_level_1.dta	a012209_`quelle'_iaw_level_2.dta
0.3	gen_p50p - 0.3 - 1R01 - verbleibsnachweise - import	- Imports "Verbleibsnachweise" - Prepares data: missing obs are unified, string variables are converted, duplicates are removed - Spells are formed from key date values		a012395_verbleibe_ano_iaw_v2.dta	a012395_verbleibe_ano_iaw_v2_level_1.dta
	gen_p50p - 0.3 - R - `quelle' - import	- Rename variables to achieve consistency with previous do-files - Label variables - Unify missings obs - Transform string variables to numeric - Unless unemployed, which are interrupted only by illness, merge into one - Recode variables - Employment status are aggregated		a012209_`quelle'_iaw_level_1.dta	a012209_`quelle'_iaw_level_2.dta
	gen_p50p - 0.31 - 2R01 - TN - import	Imports participant data from .csv file into .dta file		Evaluierungliste13042011161850.csv	TN_level_1.dta
	gen_p50p - 0.32 - R - TN - id	As 0.2, but applied to participant data		TN_level_1.dta	TN_level_2.dta

1. Initial data preparation – by source

	File name *.do	Contents	Key variables generated	Data file start	Data file end
1.1	gen_p50p - 1.1 - R - `quelle' - hilfsvariablen	- Rename variables to achieve consistency with previous do-files - Label variables - Unify missing obs - Transform string variables to numeric - Unless unemployed, which are interrupted only by illness, merge into one - Employment status are aggregated - Form a current constant for each feature as of key date 1.1.2010 or latest value - LHG / XLHG are merged	*_`source': correct and final definition of a characteristic	a012209_`quelle'_iaw_level_2.dta	a012209_`quelle'_iaw_level_3.dta (Zwischenergebnis) a012209_`quelle'_iaw_level_4.dta
1.2	gen_p50p - 1.2 - R - ieb - hierarchisierung	- For the IEB dataset, priority rules are applied among the sources for each characteristic - Missings will automatically take the next most important source	*_ieb: correct and final definition of a characteristic	a012209_ieb_iaw_level_4.dta	a012209_ieb_iaw_level_5.dta
1.3	gen_p50p - 1.3 - R - `quelle' - bereinigung	- Keep only final constants - For each individual, the data is aggregated so that each characteristic per individual has only one value (which has been extracted in previous steps)	asu, mth, leh, lhg, ieb etc.: to later determine the origin of a person	a012209_`quelle'_iaw_level_4.dta a012209_ieb_iaw_level_5.dta	a012209_`quelle'_iaw_level_7.dta
1.4	gen_p50p - 1.4 - R - merge	- Aggregated data from the different sources are merged "; basic data set: LEH		a012209_`quelle'_iaw_level_7.dta	a012209_iaw_level_8.dta
1.5	gen_p50p - 1.5 - R - drop	- For variables that occur in multiple source records, a common variable is formed: if some sources indicate a missing obs, the variable is filled with information from other sources. This is especially often the case if the persons do not appear in all datasets. - For each variable composed in this way, a source variable is created which exactly states from which data record the information originates. Thus, one can only use information from a particular record.	*_source: returns the source record of the information	a012209_iaw_level_8.dta	a012209_iaw_level_9.dta
1.6	gen_p50p - 1.6 - R - label	- Both variables and their characteristics are labeled - Dummy variables are optionally recoded to 0 and 1		a012209_iaw_level_9.dta	a012209_iaw_final.dta

2. Defining sub-periods for employment histories – by source

	File name *.do	Contents	Key variables generated	Data file start	Data file end
2.1	gen_p50p - 2.1 - R - ieb - periodisierung	- Separated for each individual according to employment state: o Overlapping spells are adjusted so that there is no overlap o The employment biography is divided into 327 half-month intervals o For each interval it is indicated how long the respective state was attained o Record is transformed from "wide" to "long" - For each employment state, a separate record is saved - The data record already used for the cleansing (original data + subsequent deliveries) is used as the initial data record	ccfep_`n': Number of days with employment state `n` in the half-month interval	a012209_ieb_iaw_level_4.dta	p50p_ieb_zustand`z'.dta
2.2	gen_p50p - 2.2 - R - ieb - merge	Data records belonging to the individual employment states are merged via the variables iaw_id fep		p50p_ieb_zustand`z'.dta	p50p_period_2.dta
2.3	gen_p50p - 2.3 - R - ieb - hierarchisierung	- The longest 4 spells per person and interval are extracted lengthwise - Spells of equal length are selected randomly	zustand`x': x. longest employment spell z`x'days: cumulative spell length of the x-th longest state	p50p_period_2.dta	p50p_period_3.dta
2.4	gen_p50p - 2.4 - R - ieb - inkonsistenzregeln	- Inconsistencies are removed - If this creates gaps, information is filled from shorter states		p50p_period_3.dta	p50p_period_4.dta
2.5	gen_p50p - 2.5 - R - ieb - priorisierung	- Prioritization rules are applied. For this purpose, first state 4 is compared with state 3, then the prioritized state with state 2, etc. - Aggregated as well as disaggregated final states are created - Solving problems with spell transitions	zustand`x'_orig: zustand`x' from file „p50p_period_4.dta" z`x'days_orig: z`x'days from file „p50p_period_4.dta" endzust_dis: disaggregate final state endzust_agg: aggregate final state endzust_days: duration of final state	p50p_period_4.dta	p50p_period_5.dta
2.6	gen_p50p - 2.6 - R - ieb - label	- Create labels for states		p50p_period_5.dta	p50p_period_6.dta

#	File name *.do	Contents	Key variables generated	Data file start	Data file end
2.7	gen_p50p - 2.7 - R - ieb - reshape	- Reshape data, so that each person has only 1 data line; every period gets its own variables	endzust_dis`fep': disaggregate final state in period `fep' endzust_agg`fep': aggregate final state in period `fep' endzust_days`fep': duration of final state in period `fep'	p50p_period_6.dta	p50p_period_7.dta
2.8	gen_p50p - 2.8 - R - ieb - indikatoren	- Calculated a multitude of variables from which the final indicators are calculated - All variables except for iaw_id and the final indicators are deleted		p50p_period_7.dta	ieb_indikatoren.dta

3. Employment histories – by source

#	File name *.do	Contents	Key variables generated	Data file start	Data file end
3	gen_p50p - 3 - R - `quelle' - `statistik' history	- If necessary, "Verbleibsnachweise" and XSozial data are added - Only spells that are valid from the censoring date will be kept - Parallel spells are deleted - For partially overlapping spells, the end date of the previous spell is set to one day before the start date of the subsequent spell - Gaps between two spells are closed if gap <= 3 days (BeH / ASU: 7 days) - Creates a spell history by making the record wide format - Each person is now described by single a data line - The individual spells are each represented by specific variables		a012209_`quelle'_iaw_level_3.dta ggf. a012395_verbleibe_iaw_level_2.dta TN_level_3.dta	`quelle'_history.dta

4. Preparation of program-specific data

#	File name *.do	Contents	Key variables generated	Data file start	Data file end
4.1	gen_p50p - 4.1 - R - TN - hilfsvariablen	- Rename variables to achieve consistency with previous do-files - Label variables - Unify Missings - Transform string variables to numeric		TN_level_2.dta	TN_level_3.dta
4.2	gen_p50p - 4.2 - TN - label	- Variables and values are labeled - Dummy variables are optionally recoded to 0 and 1		TN_level_3.dta	TN_final.dta

5. Merging the data

	File name *.do	Contents		Data file start	Data file end
5	gen_p50p - 5 - R - merge	Records are merged		a012209_iaw_final.dta	p50plus.dta
				TN_history.dta	
				`quelle'_history.dta	
				ieb_indikatoren.dta	

II. Specific Data Preparation

0. Master and settings

gen_p50p - 6.0 - settings analyse.do

This do-file contains the global settings for the do-files of the analysis. These settings supplement those from gen_p50p - 0.0 - settings.do. The global "days" contains the daily intervals for which the regressions of the diff-in-diff analysis are performed. The global "siglev" contains the significance level for the confidence intervals of the diff-in-diff analysis graphs.

1. Definition of outcome variables, cleaning and recoding

gen_p50p - 6.1 - 2R10 - ergebnisvariablen.do

This do-file is used to calculate the different outcome variables for the analysis. First of all, the persons according to the file a012209_verw_iaw_id.dta are assigned to the different groups and persons with conflicting group affiliations are deleted (these are persons with double dophash_id.).

Subsequently, the LHG spell to be considered is determined first, which must meet the requirements that it

1. does not end before 1.1.2007,

2. does not start after 30/9/2007 and

3. coincides with an ASU spell.

The earliest date which meets all conditions is stored in the variable datelastlhg07. Similarly, this variable is determined for 2010 and then stored in the variable datelastlhg depending on the group affiliation of the person.

Subsequently, the outcome variables are calculated. In this case, it is checked whether or not there was a BeH spell in the daily intervals stored in the global days after the date stored in datelastlhg. The result can be found as a dummy in the variable svpb (employment subject to social security contributions). If the date is after 30.09.2011, the variable is set as missing. The variable nsvpb (sustainable employment subject to social security contributions) is also supplemented by the condition that the BeH-Spell must have a duration of at least 180 days. The variables iea (integration in the primary labor market) and niea (sustainable integration into the primary labor market) are calculated in the same way, but with the additional requirement that there is no MTH spell for the requested date. If there is an MTH-Spell the respective variable is 0.

gen_p50p - 6.1 - 1R05 - aufbereitung.do

This do-file runs within the diff-in-diff or matching formats. Its purpose is to ensure that the variables listed here are calculated identically when processing the Diff-in-Diff and Matching records. First, different variables are generated from the GSS number, then age variables are formed, then certain variables are coarsened and finally formed for certain variable missing categories.

III. Estimation

1. DiD estimation

gen_p50p - 6.2 - 2R05 - diff-in-diff aufbereitung.do

In this do-file the diff-in-diff-specific preparation is carried out. To do this, first all observations are removed from the record that do not belong to one of the diff-in-diff samples. Subsequently, the correct values of the covariate are determined for each observation. This is done with information from the LHG or the ASU on the basis of the spells used in 6.1, for information from other sources, the corresponding value is selected for variables that were calculated according to the cut-off date method on the basis of group affiliation. The periodization variables are also created by group membership. All variables created are labeled. Then a number of variables are coarsened and the do-file gen_p50p-6.1-R-reparation.do is called. Then, a dummy variable for P50plus participation is created and all persons over 53 are removed from the P50 sample. Finally, the variables that are no longer needed are removed from the data set and the data from the GSS / Pakt survey as well as the regional data and the results of the typing are added.

DiD estimation files

The operation of all diff-in-diff estimation do-files is identical: Before each calculation, the record is saved, then a regression is performed for each day interval and results of interest are stored in separate variables. These are in particular the level of significance (on _p_ ending), parameters themselves (on _c_ ending) as well as the upper and lower end of the confidence interval (on _u_ or _l_ ending), also the number of observations (n_), the general integration probability (intprob_) and the proportion of P50plus subscribers (p50plus_). Then the data file is restructured so that it only contains these results in long format with one line per day interval. This data set is saved and used for the creation of graphs. Finally, the original record is restored (using restore).

The do-files "gen_p50p-6.2-R-diff-in-diff `result_variable'.do" calculate an unconditional DiD model for each outcome variable, the single sink estimates (ie models of all interesting variables that are conditioned on just one variable) and the conditioned DiD model as well as the same for the DiDiD models.

The do-file "gen_p50p-6.2-R-diff-in-diff sample split.do" calculates the conditional DiD and DiDiD models determined in gen_p50p-6.3-R-diff-in-diff iea.do once only for women and once only for men. The do-file is not automatically adjusted if the model was changed in the previous do-file.

gen_p50p - 7 - 1R03 – publikationsergebnisse

gen_p50p - 7 - 1R01 - neue publikationsergebnisse

These files creates the graphs in Figure 7 of the paper.

2. Matching and DiD Matching

gen_p50p - 6.3 - 1R01 - matching_master.do

This dofile is used to call all do-files in the matching routine. It indicates the order and a short description of the dofiles.

gen_p50p - 6.3 - 3R25 - DiD matching aufbereitung.do

In this do-file the matching-specific data preparation is carried out. For this purpose, first all observations are removed from the record that do not belong to one of the matching samples (NB: Some observations are both in the DiD and in the matching record.). Then the correct values of the covariate are determined for each observation (see DiD processing, with the difference that there is only one point in time). Then certain variables are coarsened and the do-file gen_p50p-6.1-R-reparation.do is called. Dummy variables are then created for the various samples and all persons below 48 or over 65 are removed from the sample (NB: this is different from the DiD sample). Then the variables that are no longer needed are removed from the data set and the data from the GSS / Pakt survey as well as the regional data and the results of the typing are added.

Thereafter, the assignment of the entry date in P50P and various measures from the participant data (previously only coaching, activation and type of assignment). Finally, a correction of the treatment group to persons who are not yet in P50plus or in an active labor market measure.

gen_p50p - 6.5 - 2R17 - DiD matching_iea_balancing_nur_iea_Angrist_Imbens clean.do

This do-file contains the specifications for the results shown in the paper.

First, descriptive statistics and the unconditional integration probability are calculated.

Similar to the DiD do-files, an OLS and a matching regression is performed for each daily interval and results of interest are stored in separate variables. Then the data file is reshaped so that it only contains these results in long format with one line per day interval. This data set is saved and used for the creation of graphics.

Next, the propensity scores are estimated and the balancing tests are performed. The various control variables are selected. Pscores are estimated for all three models as well as for the different treatments. Histograms are followed by Leuven-Sianesi and Smith-Todd Balancing Tests.

At the end, the implementation of the matching takes place.

This file generates the upper graphs in Figures 4 and 5.

gen_p50p - 6.5 - 2R22 - DiD matching_iea_balancing_nur_iea clean.do

This file is derived from DiD matching_iea_balancing_nur_iea_Angrist_Imbens.do.

In addition, it performs matching over time and among the non-participating job centres, so that there are three matching steps. The DiD step is based on the matched samples and is produced by a regression.

This file generates the lower graphs in Figures 4 and 5.

gen_p50p - 6.5 - 2R18 - DiD matching_iea_balancing_nur_iea_MALE

gen_p50p - 6.5 - 2R18 - DiD matching_iea_balancing_nur_iea_FEMALE

This file generates the graphs in Figure 6.