1. Introduction

- This data description provides the details about the data used in the paper "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences" by Annette Bergemann, Bernd Fitzenberger, Stefan Speckesser which is forthcoming in the Journal of Applied Econometrics.

2. Data used for the analysis:

- The data set used for our analysis is confidential, i.e. we are legally not allowed to provide here the data we use for our analysis. However, the data can be ordered by any academic researcher from the address below. This document allows a researcher to follow our data compilation steps once the researcher has obtained legal access to the data set.

- Our analysis uses the Labor Market Monitor Sachsen-Anhalt (Arbeitsmarktmonitor Sachsen-Anhalt, LMM-SA) for the years 1997, 1998, and 1999. The LMM-SA is a panel survey of the working-age population of the state (Bundesland) of Sachsen-Anhalt with 7,100 participants in 1997, 5,800 in 1998, and 4,760 in 1999. 1999 is the last year in which the survey was conducted.

- Data were supplied as a merged file by the Center for Social Research Halle (zsh). Key contact at the data supplier is Thomas Ketzmerick, Zentrum für Sozialforschung Halle an der Martin-Luther-Universität Halle-Wittenberg, Neuwerk 11, 06108 Halle/Saale (Germany), Tel: *49 345 / 388 08 70, Fax: *49 345 / 388 08 72, Email ketzmerick@zsh.uni-halle.de. The data use agreement was originally signed on May 3rd, 1999 and amended July 18th, 2000.

- In the three survey years used for the analysis (1997-1999), retrospective questionnaires on the monthly employment status between 1990 and the interview date were included, covering employment, unemployment, or participation in a program of ALMP, as well as periods in the education system, inactivity, or in the military. Individuals who did not participate in the 1998 survey are recorded until at least September 1997, those who dropped out in 1999 at least until October 1998

- In the three survey years used the categories of the labor market status information differ. For compatibility, the data set also includes a combined monthly calendar for the three survey years. This calendar distinguishes the following categories: Education, full-time employed, part-time employed, unemployed, job creation scheme, training, retirement, pregnancy/maternity leave, not in active workforce. Additional information on the individuals that goes beyond the monthly labor market status since 1990 can be retrieved from the cross-sectional dimension of the survey for the years 1997 to 1999. We use static individual characteristics, such as education, area of residence at the time of the interview, and year of birth.

3. Case selection

- We only consider individuals with uninterrupted information on their labor market history between January 1990 and at least September 1997 (i.e. individuals who completed the retrospective question in 1997). The individuals are between 25 and 50 years old in January 1990 and employed before the start of the Economic and Social Union" in June 1990.

- Only individuals are included who belonged to the active labor force of the GDR, who therefore are fully hit by the transformation shock, and who are not too close to retirement. Individuals who are later on in education, on maternity leave or retired are excluded completely from the analysis. The goal is to construct a consistent data base excluding

individuals who have left the labor market completely. In addition, we exclude a small number of individuals without valid information on those individual characteristics, on which we build our matching estimator.

- The following table shows the changes in the sample size according to the selection rules applied.

Table 1: Sample selection rules and number of observations

Selection Criteria	Resulting Number of Observations
Fully observed labor market history and year of birth	10,715
Aged between 25 and 50 years in January 1990	6,088
Employed in June 1990	5,529
Not in Education after June 1990	5,480
Not in Maternity Leave after June 1990	5,334
Not retired after June 1990	5,224
Final sample: with valid information on relevant covariates	5,165

4. Programme participation:

- Table 2 summarizes participation in ALMP based on our data. The two most important programs, Training (TR) and Job Creation Schemes (JC), were implemented on a large scale.

- In total, 27% of our sample participated at least once in one of the two programs. 13% (689 cases) participated at least once in JC, however, TR was by far the most important program with a rate of 20% (1,021 cases).

- Multiple participation is common in East Germany. After a first TR, a second treatment in TR or JC occurred in 326 cases, i.e. in more than 36 % of the 889 cases in a first treatment in TR.

- Here, we focus on TR as the first treatment in ALMP. We observe 9.8% (495 cases) of our sample to participate in a first TR during the first period from 1990 to 1993 and 7.6% (394 cases) to participate during the second period from 1994 to 1999.

Table 2: Program Participation in the LMM-SA during 1990 and 1999

One Program	Job Creation Scheme		Training	
	%	N	%	N
At least once	13.3	689	19.8	1,021
As first program	9.4	484	17.2	889
	Training in 1990-1993		Training in 1994-1999	
	%	N	%	%
As first program	9.8	495	7.6	394
%: percentage of the population of interest				

5. Descriptive statistics of the samples used in the analysis

- The following tables show fundamental descriptive statistics for the samples of participants and non-participants of the periods 1990-93 and 1994-99.

Table 3: Descriptive statistics for participants and non-participants 1990-93

		Participants 1990-1993		Non-participants 1990-1993	
Variable name	Variable label	Mean	Std Dev	Mean	Std Dev
Personal characteristics					
AGE90	Age 1990	36.00	6.95	37.77	7.34
SEX	Female	0.66	0.48	0.46	0.50
County					
DES	Dessau	0.16	0.37	0.11	0.32
HBS	Halberstadt	0.08	0.28	0.09	0.29
HAL	Halle	0.14	0.35	0.19	0.39
MD	Magdeburg	0.23	0.42	0.24	0.43
MER	Merseburg	0.14	0.34	0.13	0.34
SGH	Sangerhausen	0.11	0.31	0.10	0.30
SDL	Stendal	0.08	0.27	0.08	0.27
WB	Wittenberge	0.06	0.23	0.05	0.22
Level of education					
BATFARB	Un- and Semi-skilled worker	0.01	0.12	0.03	0.16
BAFARB	Skilled worker	0.46	0.50	0.43	0.49
BAMEIS	Master craftsmen	0.05	0.22	0.08	0.27
BAFACHS	Technical college	0.20	0.40	0.19	0.39
BAHOCH	University	0.27	0.44	0.27	0.45
Number of persons			495		4670

Table 4: Descriptive statistics for participants and non-participants 1994-99

		Participants 1994-1999		Non-participants 1994-1999	
Variable name	Variable label	Mean	Std Dev	Mean	Std Dev
Personal characteristics					
AGE90	Age 1990	36.94	6.98	37.66	7.35
SEX	Female	0.52	0.50	0.47	0.50
County					
DES	Dessau	0.11	0.31	0.12	0.32
HBS	Halberstadt	0.10	0.30	0.09	0.29
HAL	Halle	0.21	0.41	0.19	0.39
MD	Magdeburg	0.24	0.43	0.24	0.43
MER	Merseburg	0.13	0.34	0.13	0.34
SGH	Sangerhausen	0.12	0.32	0.10	0.29
SDL	Stendal	0.05	0.22	0.08	0.27
WB	Wittenberge	0.04	0.19	0.05	0.22
Level of education					
BATFARB	Un- and semi-skilled worker	0.02	0.13	0.03	0.16
BAFARB	Skilled worker	0.50	0.50	0.43	0.49
BAMEIS	Master craftsmen	0.08	0.27	0.08	0.27
BAFACHS	Technical college	0.17	0.38	0.19	0.40
BAHOCH	University	0.23	0.42	0.28	0.45
Number of persons			394		4771