Uwe Blien, Joachim Möller, Phan thi Hong Van, Stephan Brunow:

Long-lasting Labour Market Consequences of German Unification

Paper prepared for the Jahrbücher of Nationalökonomie und Statistik

Stephan.Brunow@iab.de, Van.Phan-thi-Hong@iab.de

Appendix: Data issues

Most of the data is taken from IAB data sets and from data provided by the Federal Employment Agency:

- The SIAB data set of 1993-2010 provided by the Research Data Center (FDZ) of the Federal Employment Agency in the IAB. This is a 2-% sample of the Employment Statistics
- the IAB Establishment Panel
- a special draw of the IAB Employment Statistics
- the Federal Employment Agency Data Warehouse (DWH).

This data are so-called "Sozialdaten" (social data) which are protected against free access by German law. However, access to this data is possible via different ways for non-members of the IAB. Access to this data is possible when a guest researcher contract is signed. For the SIAB the use is possible by approaching the FDZ, because it is a scientific use file (SUF). Data from the Establishment Panel can be used by visits at the FDZ. Please contact the Research Data Department of the IAB, the FDZ by email iab.fdz@iab.de.

Other data used in this study is taken from the Federal Statistical Office (destatis) and the "Patentatlas Deutschland" (DPMA). This data can be freely obtained from webpages. In the excel file (public_data.xls) this material is presented.

The next sections provide a detailed description of specific IAB data used in the analysis and reports contact persons for the confidential data.

The SIAB data set

The SIAB data set is provided by the Research Data Department of the IAB. A general description of this data set at individual level can be found here:
http://fdz.iab.de/en/FDZ_Individual_Data/integrated_labour_market_biographies.aspx.

For support regarding this article and the SIAB data please contact Van Phan thi Hong.

Individual inflation-adjusted wages (variable tentgelt) from the SIAB are used for all wage analysis. To avoid double counting of people, only those employees of the SIAB are considered that were employed at the 30th of June each year. Unemployed that are registered at a job-creation scheme at the 30th of June are excluded from the sample (i.e. we only consider variable erwstat=101 or 102).

Some wage information does not seem to be reliable; we exclude those cases where wages of less than 6 Euro per day and more than 200 Euro per day are reported (tentgelt>5 & tentgelt<201). We make use of the inflation rate provided by the Destatis to deflate wages, using the year 2005 as basic year. We control for:

- gender (variable frau), individual age (it is computed as survey year-year of birth (gebjahr) and include only persons of age 15 to 65)
- occupational status and working hours (stib):
    - we include dummy indicators for stib=1 (unskilled workers), 2 (skilled workers), 3 (Master craftsman and foreman), 4 (employees)
    - we exclude stib=0 (vocational training), 7 (home workers), 8 and 9 (parttime workers), missing values (.n (which relate to unemployed person), .z (no valid entry))
- education (variable bild): We aggregate the values bild=1 and 3, and bild=2 and 4 to two distinct groups; no further groups were aggregated. Missing values are treaded as separate group. All education information is controlled for by means of dummy variables.
- occupation (variable beruf): We consider 14 aggregated occupational groups based on the occupational classification 1988/92 by means of dummy variables.
- industry (variable w93_5, w03_5, w08_5): We manually aggregate 5 digit industries to 31 distinct industries that take the changes in the industry classification into account. Finally, dummy indicators for each industry are introduced in the regression model. (the industry classification is included in the file WZ_BHP_7510_w73_w93_v1.dta)
- establishment size measured as the number of employees in 9 distinct categories (0-4, 5-9, 10-49, 50-99, 100-199, 200-499, 500-999, 1000-4999, 5000 and more). This information is not available in the original SIAB file and has to be imported from another establishment specific SIAB file (siabs_btr_1975_2010.dta).
- regional heterogeneity is taken into account by nine different regional groups ("Siedlungsstrukturelle Raumtypen") that are taken from a classification provided by the Federal Institute of Research on Building, Urban Affairs and Spatial Development, BBSR. Additionally, we introduce a dummy for East Germany. Regions relate to NUTS3 regions, the German "Kreise" and "Kreisfreie Städte" and for individuals it relates to their respective workplace.

Data from the Data Warehouse of the German Federal Employment Agency

The unemployment rate is taken from the Data Warehouse of the Federal Employment Agency. It relates to the unemployment rate based on all employees and unemployed that are registered in June each year. We further make use of data from the Data Warehouse on the regional stock of employees and unemployed. Self-employed people or civil servants are not considered.

This data is not freely accessible but can be provided by the Statistical office of the Federal Employment Agency upon request (Statistik-Datenzentrum@arbeitsagentur.de). For support related to this study, please contact Van Phan thi Hong.

IAB Establishment Panel

The IAB Establishment Panel is an annual survey. First, the data of all waves is put together to achieve a panel data set. For this task, variable names are harmonized among waves. The original variable names consist of a letter, which define the survey year followed by a number that represents the question and, if appropriate, another letter representing the sub-question. In the following table, information is provided, which variables were used and how they are renamed to a unique panel variable.

Harmonized variable name	Original variable names from the IAB Establishment Panel for single waves.	Short description	Variables in econometric model
ownership	d77 ez4 f76 g85 h80 i81 j82 k84 l91 m91 n92 o92 p94 q91 r87 s80b t79b u86b	Contains information of the ownership structure	d_foreign_owned (Foreign ownership) d_own_west_in_east (West owner of establ. in East)
frac	hr1993q hr1994q hr1995q hr1996q hr1997q hr1998q hr1999q hr2000q hr2001q hr2002q hr2003q hr2004q hr2005q hr2006q hr2007q hr2008q hr2009q hr2010q	Survey weight to achieve representativeness	frac
bula	bula1993 bula1994 bula1995 bula1996 bula1997 bula1998 bula1999 bula2000 bula2001 bula2002 bula2003 bula2004 bula2005 bula2006 bula2007 bula2008	Bundesland, from which also a dummy variable is constructed, which indicates Eastern Germany.	d_ost (Dummy for Eastern Germany)

	bula2009 bula2010		
innovpro	o27 p27 q25 r28 s23 t27 u24	Process innovation	y_innovpd4
innovpd1	a18 bz10 f23 i14 l20 o24 p24 q22 r25 s20 t24 u21	Product innovation: improvement	y_innovpd1
innovpd3	a20 bz12 f25 i16 l22 o26 p26 q24 r27 s22 t26 u23	Product innovation: introduction	y_innovpd3
Innovpd2	a19 bz11 f24 i15 l21 o25 p25 q23 r26 s21 t25 u22	Innovation: adoption of technologies/ products	y_innovpd2
tech	a10 bz13 c24 d25 e30 f22 g26 h09 i30 j19 k22 - m26 n20 o23 p23 q21 r24	Information of the state of the art of machinery and equipment	d_tech1-d_tech4 (Dummy for the state-of-the-art)
legal_form	a74 bz05 cz05 d79 e70 f69 g78 h78 i80 j81 k83 l90 m89 n91 o87 p89 q86 r80	Legal form of the establishment	d_einzel (Sole trader), d_persges (Private partner-ship company)
inputs	a08 b15 c15 d17 e17 f14 g14 h17 i12 j11 k14 l14 m13 n10 o12 p13 q10 r13	Intermediate inputs in production as a proportion on revenues	s_vorleistung
Y_lag	a06b b13 c13 d12 e12 f11 g11 h12 i07 j06 k09 l09 m08 n08 o09 p10 q07 r10	Revenues/ Turnover in previous year	revenues and lab_prod= revenues/ptvza
Y_type	a06a b12 c12 d11 e11 f10 g10 h11 i06 j05 k08 l08 m07 n07 o08 p09 q06 r09	Type of income in previous year (e.g. turnover)	
org	a73 bz04 cz04 d78 e69 f79 g77 h77 i79 j80 k82 l89 m88 n86 o90 p91 q88 r83	Organizational type (e.g. single-site-plant, headquarters, branch office)	d_zentral (headquarters) d_niederl (branch office)

From the variables mentioned above the set of (control) variables is constructed. Also, some of the variables were time-lagged to achieve a consistent data set (Y_lag, Y_type, inputs, innovpd1-innovpd3, innovpro). This is necessary as the question is, for example, "Did you perform innovation in the previous year? yes-no". Missing values for time constant variables are imputed (org, legal_form).

Access to this data is possible for guest researchers by using the Research Data Department, FDZ of the IAB. For detailed questions on the project on Germany's reunification, please contact Stephan Brunow.

The special draw of the IAB Employment Statistics

This data source is a special draw from all German employees' subject social security contributions. At establishment level it contains information on the employment levels of

occupations in fulltime, large/ small parttime work. Among others, these are the number of employment days in the respective category and the wage sum. Using the occupational classification scheme of Brunow and Blien (2015), total work volume (ptvza) and a human capital measure (sH) can be constructed by computing the fulltime equivalent workers in the respective skill groups. From annual total employment, also firm-size (i.e. employment size) groups can be derived. Fulltime equivalents (ptvza) are computed by (Person days full time) + 0.6*(Person days large part time) + 0.3 (Person days small part time). Using the wage sum of all employees and employment levels, the average wage rate of the establishment can be computed, avlohn = (wage sum/ptzva). Additionally, in combination with the IAB Establishment Panel, a productivity measure is computed (turnover per employee in fulltime equivalents).

The IAB Employment Statistics covers information on the industry level, from which we used the NACE Rev.1.1 (German classification scheme: WZ 2003) at the 2-digit level. Industry fixed effects are taken into account by means of dummy variables (dind2x, where x represents the industry number). Additionally, there is information on the date of the first appearance of an establishment identifier. This information is used to generate dummy indicators for the age of the establishment (young_est: Establishment age of 5-14 years, old_est: Establishment age of 15+ years, Establishments of age <5 years serve as reference category).

Because this data is highly confidential, no access via the IAB Research Data Department (FDZ) is possible. Access to that data is only possible by means of a Guest Research Visit. Please contact Stephan Brunow.

Patentatlas Deutschland (DPMA)

Data from the Patentatlas can be freely obtained upon request, presse@dpma.de. More information on that specific data is available on the DPMA webpage, http://www.dpma.de/ This data is available at the regional level.