README — Replication Package

Paper

The Effects of Immigration on Places and People – Identification and Interpretation. Journal of Labor Economics (accepted).

Version & Date

Version 1.0 — 2025-09-03

Authors

Christian Dustmann, Sebastian Otten, Uta Schönberg, and Jan Stuhler

Purpose of this Package

This README documents the contents and use of the replication materials accompanying the paper. It explains what is included, how the analysis pipeline is organized, what software environment is required, how to obtain the external/proprietary inputs, and how program files map to the paper's results.

Important Replicability Note (Restricted Data)

The empirical analysis relies on proprietary administrative microdata from the Institute for Employment Research (IAB) — a project-specific draw from the Integrated Employment Biographies (IEB, 2010). Due to German Social Code confidentiality, these data cannot be shared publicly and cannot be accessed by researchers who are not employed by the IAB. As a result, the results in the paper cannot be reproduced without access to the IEB project data. Where possible, we provide instructions for related FDZ/IAB datasets (e.g., SIAB) and for other sources used in ancillary analyses.

Summary of What This Package Contains

- Stata 14.2 program files, including a master script (master.do) that organizes the workflow.
- No microdata are included. Researchers must obtain the IEB project data (not publicly accessible), the SIAB dataset used in the correlation analysis, and the BIBB 1991/92 data.
- Two auxiliary data files are included and redistributable:
    1. bbr_2006.dta: district type classification (Bundesamt für Bauwesen und Raumordnung, BBR).
    2. wz73-wz28.dta: crosswalk mapping WZ73 3-digit industry groups into 28 sectors.
- A directory structure is created by master.do to store derived data, tables, and figures.

Software & Computing Environment

- Stata 14.2 (Windows, 64-bit). The analysis was developed and run on a 64-bit Windows server system.
- Randomness & bootstrap: We use wild-bootstrap standard errors with 500 replications. Runtime depends on hardware; bootstrap routines dominate overall runtime.
- User-written ado files: We do not provide a dependency list. If Stata flags a missing command, please install it when prompted.

How to Run

1. Adjust the working directory at the top of master.do: cd “mypath\sandbox/”
   Use a local path that contains the replication folder “sandbox” and its subfolder “do-files”.
2. Ensure that the external data (IEB project data, BIBB 1991/92, SIAB) are available in your environment following the access notes below.
3. Open Stata 14.2 (Windows 64-bit) and run: do master.do
4. The master script will create the folder structure, run data-management scripts, and execute the analysis scripts in the documented order.
5. Final assembly: Scripts write intermediate tables and figures to subfolders. Publication-ready tables and figures are assembled manually, as detailed in the result-mapping checklist below and in the notes at the top of each analysis do-file.

External data staging (required)

Place raw external data in the repository before running master.do:

- IEB project data (restricted): put provider files in data/orig
- SIAB 7519 v1 (restricted): put provider files in data/orig
- BIBB 1991/92 (ZA2565) (restricted): put provider files in data/orig
- Auxiliary data (included here): bbr_2006.dta and wz73-wz28.dta should sit in data/orig
- GIS inputs (not redistributed): if reconstructing, place GER_ao_gem_coord.dta, vg250_gem.dbf, vg250_gem.shp in data/orig

Notes:

- Scripts assume relative paths from the project root (set by the cd command at the top of master.do).
- Adjust provider filenames as used in the do-files here (do not rename scripts).
- No third-party microdata are redistributed in this package. Results cannot be reproduced without approved access to IEB, SIAB, and BIBB data.

Directory Structure (created by master.do)

sandbox/

├– do-files/

| ├– data_management/

| | ├– cr_consistenteduc.do

| | ├– cr_impute_genmeans.do

| | ├– cr_impute_tobits.do

| | ├– cr_impute_education.do

| | ├– cr_tasks.do

```
|   |   ├– cr_unempl_data.do
|   |   ├– cr_treat_control_sample.do
|   |   ├– cr_prepare_map.do
|   |   ├– cr_clean_data.do
|   |   ├– cr_region_data.do
|   |   ├– cr_data_siab_01.do
|   |   └– cr_data_siab_02.do
|   └– data_analysis/
|   |   ├– an_desc_stats_01.do
|   |   ├– an_reg_first_stage.do
|   |   ├– an_dec_employ_01.do
|   |   ├– an_dec_employ_02.do
|   |   ├– an_reg_occupg_task.do
|   |   ├– an_reg_employ_01.do
|   |   ├– an_dec_wage_01.do
|   |   ├– an_reg_wage_indv_01.do
|   |   ├– an_reg_wage_indv_02.do
|   |   ├– an_reg_wage_indv_03.do
|   |   ├– an_reg_wage_pseudo_panel..do
|       └– an_data_siab_01.do
├– data/
|   ├– orig/
|   ├– bootstrap/
|   |   ├– emp_app/1991_cz_shock/
|   |   ├– emp_decomp/1991_cz_shock/
```

| | ├– non_emp/

| | ├– pseudo_panel/

| | ├– occ_upg/1991_cz_shock/

| | ├– price_effect/1991_cz_shock/

| | └– wage_decomp/1991_cz_shock/

├– figures/

├– tables/

| ├– employ/

| ├– price_effect/

| └– wage/

Data Availability & Access

1. Integrated Employment Biographies (IEB, 2010)
   - Provider: Institute for Employment Research (IAB), Regensburger Str. 104, 90478 Nuremberg, Germany.
   - IEB (2010), project-specific draw: Due to SGB I §35 confidentiality, these data cannot be shared publicly and are not available to researchers outside IAB employment.
   - Replication impact: The main results of the paper cannot be reproduced without this project data.

2. SIAB "7519 v1" (used in Appendix correlation analysis)
   - Dataset: Sample of Integrated Labour Market Biographies (SIAB), factually anonymous, version 7519 v1 (DOI: 10.5164/IAB.SIAB-R7519.de.en.v1).
   - Access: Application via FDZ/IAB (see https://fdz.iab.de/en/pd_hd/factually-anonymous-version-of-the-sample-of-integrated-labour-market-biographies-siab-regionalfile-version-7519-v1/).

3. BIBB Acquisition and Application of Occupational Qualifications 1991/92 (ZA2565)
   - DOI: https://doi.org/10.4232/1.2565
   - Access: via GESIS (see DOI landing page for conditions and request workflow).

4. GIS inputs for border variables
   - Used: GER_ao_gem_coord.dta; vg250_gem.dbf; vg250_gem.shp
   - Researchers can reconstruct these inputs from public sources:

- BKG "Verwaltungsgebiete 1:250 000 (VG250)": https://gdz.bkg.bund.de/index.php/default/open-data/verwaltungsgebiete-1-250-000-stand-01-01-vg250-01-01.html
- Documentation: https://sg.geodatenzentrum.de/public/gdz/dokumentation/eng/vg250_01-01_eng.pdf
- Municipal coordinate files can be derived by computing centroids from VG250 Gemeinde polygons.

5. Auxiliary data included in this package
   - bbr_2006.dta: BBR district-type classification table (kreis-level). Source: Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR).
   - wz73-wz28.dta: mapping of WZ73 (3-digit) to 28 sectors (documentation notes embedded in the do-files).

Once access to the restricted data is approved and files are obtained, see "External data staging (required)" for where to place the data locally.

Program Flow & File Roles

1. Data management (inputs, sample construction):

- cr_consistenteduc.do — Consistent education variable.
- cr_impute_genmeans.do (+ cr_prepare_variables_genmeans_ado.do) — Controls for wage imputation.
- cr_impute_tobits.do (+ cr_impute_prepare_variables_ado.do; cr_generate-AMR50_ado.do) — Imputed wages.
- cr_impute_education.do — Imputed education variable.
- cr_tasks.do — Task variables at the occupation level.
- cr_unempl_data.do — Unemployment data.
- cr_treat_control_sample.do — Treatment and control sample (1980–2000).
- cr_prepare_map.do — Distance-to-border variable (IV construction).
- cr_clean_data.do (+ cr_real_wages_ado.do) — Clean, merge imputed education/wages and distance.
- cr_region_data.do (+ cr_dec_emp_groups_ado.do; cr_pseudo_panel_ado.do) — Final analysis data and groups.

2. Analysis (results referenced to the paper):

- an_desc_stats_01.do — Descriptive statistics → Table A.3
- an_reg_first_stage.do — First-stage regressions → Table A.2; and Figure A.3
- an_dec_employ_01.do — Employment decomposition → Figure 1 & Table 1; task- and age-group results → Table 3 & 4; Figure 2
- an_dec_employ_02.do — Non-employed workers (employment regressions) → Table 3
- an_reg_occupg_task.do — Share of task-group employment → Figure 2
- an_reg_employ_01.do — Apprenticeships employment → Figure 3

- an_dec_wage_01.do — Wage decomposition → Figure 1; Table 2; Table A.4
- an_reg_wage_indv_01.do — Pure wage effect → Figure 1; Table 2; Table A.4; by task & age → Table 3
- an_reg_wage_indv_02.do — Age-selection coefficient → Table 2
- an_reg_wage_indv_03.do — Pure wage effect for non-employed → Table 3
- an_reg_wage_pseudo_panel.do — Wage growth (pseudo-panel) → Table A.4

3. Correlation analysis based on SIAB data:

- cr_data_siab_01.do — Prepare SIAB data
- cr_data_siab_02.do — Create sample for correlation analysis
- an_data_siab_01.do — SIAB correlation analysis → Table A.1

Outputs

- Figures are written to: /figures
- Tables are written to: /tables and subfolders (/tables/employ, /tables/wage, /tables/price_effect)

Overview Result Mapping

Paper element	Do-file(s)	Output path
Figure 1	an_dec_employ_01.do; an_dec_wage_01.do; an_reg_wage_indv_01.do	tables/employ tables/wage tables/price_effect
Figure 2	an_dec_employ_01.do; an_reg_occupg_task.do	tables/employ tables/employ
Figure 3	an_reg_employ_01.do	tables/employ
Table 1	an_dec_employ_01.do	tables/employ
Table 2	an_dec_wage_01.do; an_reg_wage_indv_01.do; an_reg_wage_indv_02.do	tables/wage tables/price_effect log-file
Table 3	an_dec_employ_01.do; an_dec_employ_02.do; an_reg_wage_indv_01.do; an_reg_wage_indv_03.do	tables/employ tables/employ tables/price_effect tables/price_effect
Table 4	an_dec_employ_01.do	tables/employ
Table A.1	an_data_siab_01.do	log-file
Table A.2	an_reg_first_stage.do	tables/
Table A.3	an_desc_stats_01.do	tables/
Table A.4	an_dec_wage_01.do; an_reg_wage_indv_01.do; an_reg_wage_pseudo_panel.do	tables/wage tables/price_effect tables/wage
Figure A.3	an_reg_first_stage.do	figures/

Reproducibility

- Numerical tolerance: small differences can arise across Stata versions/OS; these should be negligible.

Licensing

- Code: Provided for academic replication and review; other uses require permission.
- Data: All third-party data are not redistributed.

How to Cite

Please cite the article and this replication package as follows:

(1) Article: Dustmann, Christian, Sebastian Otten, Uta Schönberg, and Jan Stuhler (2025). "The Effects of Immigration on Places and People – Identification and Interpretation." Journal of Labor Economics, forthcoming.

(2) Replication package (Dataverse): Dustmann, Christian, Sebastian Otten, Uta Schönberg, and Jan Stuhler (2025). "Replication data for: The Effects of Immigration on Places and People – Identification and Interpretation." Harvard Dataverse.

Contact

For data-related questions, please contact: Sebastian Otten (sebastian.otten@uni-due.de)