Code Documentation for: Wages and the Value of Nonemployment

Simon Jäger (MIT and NBER)
Benjamin Schoefer (UC Berkeley)
Samuel Young (MIT)
Josef Zweimüller (University of Zurich)

April 2020

Our study is mainly based on the Austrian Social Security Database (ASSD). Earnings are deflated using CPI data for Austria from FRED (`https://fred.stlouisfed.org/series/AUTCPIALLMINMEI`). Secondary analyses rely on Eurobarometer 66.1 (Sep-Oct 2006) (`http://dx.doi.org/10.4232/1.10980`), firm panel data from Bureau van Dijk from 2004 to 2016, and unemployment benefits data (AMS) from 1996 to 2000.

Software Requirements: Stata 16. There are a number of commands used throughout the code that must be installed prior to running it. We provide a comprehensive installation section in the "Installing Packages" section of `00a_SetPaths.do`. Figure 1 is run using MATLAB. Figures (main and appendix) are produced as they appear in the final draft. Tables may require additional manual formatting.

We ran our code on an 8-core Intel-based server with Windows Server 2012 R2 Standard.

General instructions to run the code

1. The main data we use in our analysis are confidential administrative ASSD data (Zweimüller et al. 2009), which we analyzed on a secure server based at UZH.[1] A subset of the data used in the present paper can be acquired from `https://arbeitsmarktdatenbank.at`. Please contact the authors for the replication of specific tables and figures in the paper.

2. Place all `.do`, `.ado`, and `utilities` files in appropriate folders. See `00a_SetPaths.do` description below.

3. Modify the home global in `00_MasterDoFile.do` to an appropriate folder.

4. `00a_SetPaths.do` sets the paths for:

   (a) User-provided data, such as the CPI, Eurobarometer data, Bureau van Dijk, and the meta-study data.

   (b) Temporary datasets, where we store intermediate datasets on which the final results are based.

   (c) Results data, where we store the output from the analysis files (which we describe below).

   (d) Figures, where we store our figures.

   (e) Tables, where we store our tables.

   (f) Do files, where we place all of the do-files for this paper.

   (g) Logs, where we store our log files.

   (h) Utilities, where do-files related to calculations spells and employment status in raw data are located.

[1] Zweimüller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philippe Wuellrich, Oliver Ruf, and Simon Buchi. 2009. "Austrian Social Security Database." Austrian Center for Labor Economics and the Analysis of the Welfare State Working Paper No. 0903.

(i) We set paths for additional `.ado` files, including scheme files and tax calculators.

5. `00a_SetGlobals.do` sets the globals used throughout the analysis. We provide a description for each of these globals in that do-file.

6. To replicate any part of the analysis, we suggest loading `00_MasterDoFile.do` and uncommenting the specific section of the analysis you wish to run. We have commented out all but the `00a_SetPaths.do` and `00a_SetGlobals.do` files for convenience.

Output conventions In `00_MasterDoFile.do`, figures and tables are run in the same order as they appear in the text and appendix. Figures are saved in `.pdf` format, while tables are saved in `.tex` format. One do-file is usually provided for each figure or table. Do-file names and for figures and tables match the names of the output files. Log files follow the same convention. File names for figures and tables begin with `FigXXX` or `TabXXX`, where `XXX` is the reference number in the main or Appendix text. An additional descriptive text which correlates with the titles used in the paper and appendix is usually appended to these file names.

In-text Calculations There are several computations mentioned in the text (yet not explicitly shown in Tables or Figures) that are carried out at different stages in our code. Two of these (rent-sharing in Austria and the validation exercixe excluding the 2001 reform) are carried out explicitly in the "In-text Calculations" section of `00_MasterDoFile.do`. We provide a list of in-text calculations and their relation to a specific `.do` file at the end of this document.

Code Structure

Data preprocessing Do-file `01a` processes files named `0010_panelBal_year‘y’_month‘m’.dta`. These files are generated by the following code:

- `0007_benr_ktkz.do` cleans establishments IDs.
- `0009_Beamte_SelEmpl.do` defines employment, unemployment, disability etc. status by spell.
- `0010_qualTOpanel.do` generates panel monthly files.
- `0010_balancePanel.do` balances the panel data above
- `0011_DefineRecall.do` generates recalls by svnr and benr.
- `0012_EmployerStaticVariables.do` generates employer characteristics

Building the data Do-files `01a` thru `01e` build monthly panels, spell variables, and labor market transitions. Lastly, they construct an annual balanced panel.

- `01a_Build_MonthlyPanel.do` creates worker-month panel from month files built out of the ASSD qual files. Creates variables for which all 12 months are needed but order is not important. Then saves a "spell data file" of all worker-months from 1972 to 2015 but fewer variables, which is used to construct variables in subsequent files. Code is run for each gender separately.
- `01c_Build_SpellVariables.do` creates variables requiring monthly observations across many years.
- `01d_Build_StayerMovers.do` generates labor market transitions. Classifications like stayers, movers, EE, EUE are defined here.
- `01e_Build_AnnualPanel.do` merges the data files above and constructs a balanced annual panel. Creates the underlying data file for `02a_Build_RegressionVariables.do` below.

Preparing the regression sample Do-files `02a` thru `02c` construct the regression variables, restrict the sample to only include certain years around each reform, and additionally build a pooled regression sample.

- `02a_Build_RegressionVariables.do` creates reform-specific worker-year panels. Creates all UIB and earnings variables needed for main analysis: generates earnings caps via the `earnings_min_austria.ado` and `earnings_max_austria.ado` files, generates earnings growth, generates earnings percentile bins, and generates UIB levels and replacement rates via `UIB_calc_austria.ado`. Creates the underlying data file for `02b_Build_ReformSamples.do` below.
- `02b_Build_ReformSamples.do` creates reform samples, i.e. isolates specific parts of the income distribution from the worker-year panels. Also creates actual db/w instruments at the percentile, firm, and industry level. Creates the underlying data files for `02c_Build_PooledSample.do` and `03a_Analysis_NonParametric.do` below.
- `02c_Build_PooledSample.do` pools reform samples together to create a pooled sample, renaming variables as needed for the main analysis.

Main analysis files Do-files `03a` thru `03d` run the main nonparametric, DiD, and IV analyses. These files generate the underlying data for multiple figures and tables throughout the text, which we specify below.

- `03a_Analysis_NonParametric.do` generates the non-parametric analysis of each reform. Creates the underlying data files for `Fig4_NonParametric.do` below.
- `03b_Analysis_DiD_Pooled.do` generates the difference-in-differences analysis for the pooled sample. **This is the main DiD analysis file, and, by altering its locals, generates the underlying data for Tables 3, 6, 7, A.2, A.3, and Appendix Figure A.12**.
- `03c_Analysis_IV_Pooled.do` creates IV estimates and directly creates the output for Table A.4. Note that the first stage here matches the validation exercise in Table A.3 (which is run in `03b_Analysis_DiD_Pooled.do` immediately above).
- `03d_Analysis_DiD_Pooled_NetTax.do` estimates difference-in-differences analysis for the pooled sample, with instrument in gross terms. Creates the underlying data files for `TabA7_DiD_Gross.do` below.

Diagnostics Do-files `04a` thru `04d` perform various diagnostics used in the main text and appendix.

- `04a_Diagnostic_BenefitChangeByPercentile.do` creates dataset of db/w by percentile for each reform. Creates the underlying data file for `Fig3_BenefitSchedules.do` below (Panel E).
- `04b_Diagnostic_Eurobarometer.do` creates dataset of predicted replacement rate (UIB/net income) from the 2006 Eurobarometer survey in Austria and the actual distribution from the AMS data. Creates the underlying data file for `FigA1_Eurobarometer.do` below.
- `04c_Diagnostic_Takeup.do` creates dataset of E-to-N spells to estimate take-up of UIB. Creates the underlying data file for `TabA1_Takeup.do` below.
- `04d_Diagnostic_WagePrediction.do` analyzes the predictive power of average wage growth to to predict individual wage growth. Calculates the total and by-percentile standard deviation of average wage growth. Creates the underlying data file for `FigA15_Quality_WagePrediction.do` below.

PBD Extension Do-files `05a` thru `05c` analyze the data for Appendix Section F.

- `05a_PBDExtension_Build.do` constructs worker-year panel of workers based on ages and years for the PBD extension reform in 1989. Creates the underlying data file for `05b_PBDExtension_Analysis_NonParametric.do` and `05c_PBDExtension_Analysis_DiD.do` below.
- `05b_PBDExtension_Analysis_NonParametric.do` generates non-parametric analysis of PBD extension reform by age. Creates the underlying data file for `FigA17_PBD_NonParametric.do` below.
- `05c_PBDExtension_Analysis_DiD.do` performs difference-in-differences analysis of PBD extension reform. Creates the underlying data file for `FigA18_PBD_DiD.do` below.

Heterogeneity Analysis Do-files `06a` thru `07b` prepare the underlying data for Tables 4 and 5.

- `06a_Heterogeneity_Build_LocalUnempRate.do` calculates local unemployment rates.
- `06b_Heterogeneity_Build_AKMFirmEffects.do` calculates AKM firm effects.
- `06c_Heterogeneity_Build_FirmCharacteristics.do` calculates a variety of firm characteristics: firm size, share on UI in last two years, within-firm dispersion in earnings and earnings growth.
- `06d_Heterogeneity_Build_IndOccUnempRate.do` calculates industry-occupation separation measures.
- `06e_Heterogeneity_Build_IndustryGrowthRate.do` calculates LOM industry growth measures. (LOM = Leave Out My).
- `07a_Heterogeneity_Build_Panel.do` merges the output data files from `06a` thru `06e` and constructs worker-year panel for heterogeneity analysis using the regular regression sample. Creates groups for the "heterogeneity variables" and for the labor market transitions. Creates the underlying data file for `07b_Heterogeneity_Analysis.do` and `Tab4_Transition.do` below.
- `07b_Heterogeneity_Analysis.do` performs heterogeneity analysis using regression sample merged with heterogeneity groups. Creates the underlying data file for `Tab4_Transition.do` and `Tab5_Heterogeneity.do` below.

Robustness Checks Do-files `08a` thru `08d` generate the underlying data for Appendix Figures A.8 thru A.11.

- `08a_Robustness_Cluster.do` analyzes the sensitivity of the main DiD results to changes in the level of clustering. Creates the underlying data file for `FigA8_Cluster.do` below.
- `08b_Robustness_Winsor.do` analyzes the sensitivity of the main DiD results to changes in the level of winsorization (no winsorization, 1 %, and 5%). Creates the underlying data file for `FigA9_Winsor.do` below.
- `08c_Robustness_Controls.do` analyzes the sensitivity of the main DiD results to the choice of earnings control (log earnings, linear earnings, and earnings percentile). Creates the underlying data file for `FigA10_Control.do` below.
- `08d_Robustness_DonutHole.do` analyzes the sensitivity of the main results to the choice of donut hole size for both fixed nominal and percentile ranges. Creates the underlying data file for `FigA11_DonutHole.do` below.

τ Analysis Do-files `09a` thru `09d` prepare the underlying data and estimate the results for Table 1, Table A.5, and Figure 7.

- `09_TauMaster.do` is the master do-file for the τ analysis. It sets globals and calls the do-files below.
- `09a_Build_SpellData.do` builds the spell data used in the τ analysis.
- `09b_Build_TauEstimates_FullSample.do` builds τ dataset for all individuals (not just separators). Calculates counting ending date for a given age (set in the τ master file).
- `09b_Build_TauEstimates_Separators.do` builds τ dataset for separators. Calculates counting ending date for a given age (set in the τ master file).
- `09c_Calculate_Mean_Tau_FullSample.do` estimates τ for all individuals (not just separators).
- `09c_Calculate_Mean_Tau_Separators.do` estimates τ for separators.
- `09d_Heterogeneity_Tau.do` estimates difference-in-differences analysis of pooled sample interacting with τ heterogeneity. Creates the underlying data file for `Fig7_TauHeterogeneity.do` below.

Miscellaneous in-text calculations

- `10a_Rent_Sharing_Austria_BvD.do` see Table 1
- `10b_ValidationExcluding2001.do` see Table 1

Tables

- `Tab1_TauPrediction.do`
- `Tab2_SumStats.do`
- `Tab3_6_7_DiD.do`
- `Tab4_Transition.do`
- `Tab5_Heterogeneity.do`

Appendix Tables

- `TabA1_Takeup.do`
- `TabA2_DiD_Takeup.do`
- `TabA3_Validation.do`
- `TabA5_Tau_Actual.do`
- `TabA6_PredWageChanges.do`
- `TabA7_DiD_Gross.do`

Figures

- `Fig1a_graph_PHI.m`
- `Fig1b_graph_TAU.m`
- `Fig2_MetaStudy.do`
- `Fig3_BenefitSchedules.do`
- `Fig4_NonParametric.do`
- `Fig5_Scatter.do`
- `Fig6-A6_BinScatter.do`
- `Fig7_TauHeterogeneity.do`

Appendix Figures

- `FigA1_Eurobarometer.do`
- `FigA2-5_Additional_NonParametric.do`
- `FigA7_Distribution_IndOccInstrument.do`
- `FigA8_Cluster.do`
- `FigA9_Winsor.do`
- `FigA10_Control.do`

- FigA11_DonutHole.do
- FigA12_AlternateOutcomes.do
- FigA13_ImpliedSigma.do
- FigA14_UI_rule_validation.do
- FigA15_Quality_WagePrediction.do
- FigA16_PBD_Reform.do
- FigA17_PBD_NonParametric.do
- FigA18_PBD_DiD.do
- FigA19_NonParametric_Gross.do
- FigB1-4_RR_Schedule_Flipbook.do

.ado and scheme files ado files and scheme.

- ado/earnings_max_austria.ado used in 02a for earnings cap
- ado/earnings_min_austria.ado
- ado/gross_to_net.ado used in 04b
- ado/rr_rate_austria.ado used in Fig3 and FigB1-B4
- ado/run_specification_hetcuts.ado used in 07b, 09d
- ado/run_specification_nettax.ado used in 03d
- ado/run_specification.ado used in 03b, 08a-08d, 10b
- ado/UIB_calc_austria.ado used in 02a
- ado/scheme-wagebarg.scheme

Utilities files in the utilities folder.

- utillities/connectspells.do
- utillities/definestatus.do
- utillities/dropintspells.do
- utillities/onestatusaday.do

Table 1: In-text Calculations

Number	Source / log file
The R^2 of the model is 9%, and the unexplained variation captures a combination of unobservables, model misspecification, and likely also ex-post stochastic realized spell durations unrelated to the quality of the model.	`09d_Heterogeneity_Tau.do`
Among these studies, we focus on worker-level specifications to net out composition effects. We calculate an average of 0.099, hence setting $\phi = 0.1$.	`Fig2_MetaStudy.do`, line 40
We use firm panel data from Bureau van Dijk from 2004 to 2016 and regress wages per employee on value-added per employee, controlling for firm and industry-by-year effects, estimating a level-on-level coefficient of 0.046 (SE 0.009).	`10a_Rent_Sharing_Austria_BvD.do`, line 100
Excluding the 2001 reform from this validation exercise (because the reform occurred at a time when benefits were determined based on lagged years' wages) yields quantitatively very similar results with a 0.755 (SE 0.013) coefficient at the one-year horizon and of 0.481 (SE 0.028) at the two-year horizon.	`10b_ValidationExcluding2001.do`
The average worker's rate is 64.03% (SE 0.72) in the survey, close to the 65.29% among actual recipients.	Figure; `FigA1_Eurobarometer.do`, line 94
Appendix Figure A.12 note: The base rates for the outcome variables averaged across all the pre-reform years are: *Movers: 0.086*, *Recall: 0.040*, *ENE: 0.049*, *EUE: 0.029*, *Mth NE: 0.044*, *Mth UI: 0.017*, and *Mth Sick: 0.006*. (Also added base rates for the 2-year horizon)	`FigA12_AlternateOutcomes.do`, line 82
Appendix Section E.1.3 Percentile Ranges of Reform Samples	`02b_Build_ReformSamples.do`