ReadMe for the Replication Package of "The Nature of Long-Term Unemployment: Predictability, Heterogeneity and Selection"

Andreas I. Mueller Johannes Spinnewijn

Journal of Political Economy, 2025

The data and code for this project are available on the *JPE*'s Dataverse Repository.

Contents

1 Data Availability and Provenance Statements **1**

2 Computational Requirements **2**

3 Instructions to Replicators **2**

4 List of Data Sets, Figures and Tables **3**
4.1 Raw Data Sets . . . 3
4.2 Figures . . . 3
4.3 Tables . . . 6

5 Description of Programs **7**
5.1 Data Management . . . 7
5.2 Output Generation . . . 11

1 Data Availability and Provenance Statements

Data Availability Statement. The majority of data for this project are confidential, but may be obtained from Statistics Sweden (SCB), the Swedish Unemployment Insurance Inspectorate (IAF) and the Swedish Public Employment Service (PES).

Researchers interested in access to the data can start the process by contacting SCB at mikrodata@scb.se. Access may be granted following a confidentiality assessment pursuant to the Public Access to Information and Secrecy Act. Note that affiliation with a Swedish research institution is required. It takes between 1 to 6 months from the time you are assigned a contact person to the delivery of microdata. Most orders cost between SEK 35,000 and SEK 65,000 plus VAT (approximately USD 3,100 to USD 5,800 as of January 2024). See https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/ for further details about the application to SCB. The data will be made available within MONA (Microdata Online Access), a secure remote-desktop environment provided by SCB to access confidential microdata.

Obtaining data from IAF and PES requires a separate application to each institution. Please contact statistik@iaf.se and statistik@arbetsformedlingen.se, respectively; for further information regarding IAF, see https://www.iaf.se/statistikdatabasen/a bestalla-individdata/. If the applications are successful, the data can be matched to the SCB registers and made available through MONA.

Data Citations:

Organization for Economic Co-operation and Development (OECD). 2024. "Infra-Annual Labor Statistics: Monthly Unemployment Rate Total: 15 Years or over for Sweden [LRHUTTTTSEA156S]", retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LRHUTTTTSEA156S (accessed 2024).

Statistics Sweden (SCB). 2024. "Consumer Price Index (CPI), Fixed Index numbers, total annual average, 1980=100. Year 1980 - 2024", https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__PR__PR0101__PR0101L/KPIFastAmed/ (accessed 2024).

Statistics Sweden (SCB). 2024. "Military Enlistment Tests (Monstringsdata/Riksarkivet) [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "Longitudinal Integrated Database for Health Insurance (LISA/Forvarvskalla), 1990-2017 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "Living Conditions Surveys (ULF/SILC), 2006-2019 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "UI Fund Membership Data (Akassa), 2005-2009 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "UI Fund Tax Data (AKAS), 2002-2006 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "Wage Structure Statistics (Lönerstrukturstatistiken), 1990-2017 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "Wealth Register (Förmögenhetsregistret), 1999-2007 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Statistics Sweden (SCB). 2024. "Swedish Industry Classification (SNI) Crosswalks", https://www.scb.se/dokumentation/klassifik och-standarder/standard-for-svensk-naringsgrensindelning-sni/ (accessed 2024).

Swedish Public Employment Service (PES). "Unemployment Register (HANDEL), 1990-2017 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

Swedish Unemployment Insurance Inspectorate (IAF). "Unemployment Funds Register (ASTAT), 1997-2014 [database]", Microdata Online Access At Statistics Sweden (MONA), (accessed 2024).

2 Computational Requirements

Software. STATA/MP 17.0 and R version 4.4.0.

System. The code was run on Batch Mode within MONA. Maximum memory usage in the server is 100GB, but most of the output-generating do files can be executed with under 16GB RAM. Users can expect the full replication to take around a week of computing time.

3 Instructions to Replicators

1. Upload the Replication Files to MONA.
2. Use Stata to open Programs/000_master.do.
3. Edit the macros with the absolute paths to the replication package folder (line 12) and the raw data (lines 17-34).
4. Run lines 1-95 in Programs/000_master.do.
5. Use R to open the files listed on lines 98-106 of Programs/000_master.do. Edit the absolute path to the replication package folder in the R scripts (usually between lines 30 and 32). Run the files in order.
6. Go back to Programs/000_master.do and run lines 108-125.
7. Use R to open the two files listed on lines 127 and 128 of Programs/000_master.do. Edit the absolute path to the replication package folder in the R scripts (lines 24 and 49, respectively). Run the files in any order.
8. Go back to Programs/000_master.do and run lines 130 to the end of the file.

4 List of Data Sets, Figures and Tables

4.1 Raw Data Sets

Table 1: Raw Data Sets

Data Set	Provided	Source	Used In
AF_Sokatper.dta, vy_Sok.dta, AF_Insper.dta, Insper.dta	No	Unemployment Register	001_1_CleaningPES.do
Lisa_1990.dta to Lisa_2017.dta	No	LISA	001_2_CleaningLISA.do
Inflation.dta	Yes	CPI	001_2_CleaningLISA.do
WealthR_1999.dta to WealthR_2007.dta	No	Wealth Register	001_4_CleaningWealth.do
centralgovt_wages1990.dta to centralgovt_wages2017.dta	No	Wage Structure Statistics	001_5_CleaningWages.do
localgovt_wages1990.dta to localgovt_wages2017.dta	No	Wage Structure Statistics	001_5_CleaningWages.do
private_bluecollar_wages1990.dta to private_bluecollar_wages2017.dta	No	Wage Structure Statistics	001_5_CleaningWages.do
private_whitecollar_wages1990.dta to private_whitecollar_wages2017.dta	No	Wage Structure Statistics	001_5_CleaningWages.do
regionalgovt_wages1990.dta to regionalgovt_wages2017.dta	No	Wage Structure Statistics	001_5_CleaningWages.do
IAF_Utbetalning_ny.dta	No	Unemployment Funds Register	001_7_CleaningUnemplBenefits.do
AKAS_2002.dta to AKAS_2006.dta	No	UI Fund Tax Data	001_8_CleaningUnemplInsurance.do
UI_fund.dta	No	UI Fund Membership Data	001_8_CleaningUnemplInsurance.do
Enlistment_old.dta, Enlistment_new.dta	No	Enlistment Tests	001_10_CleaningIQ.do
silc_2006.dta to silc_2019.dta	No	ULF/SILC	001_11_CleaningSILC.do
SNI_2002_92.csv, SNI_2002_2007.csv	Yes	SNI Crosswalks	006_CleaningIndustryCodes.do
Unemployment_Official.dta	Yes	OECD Infra-Annual Labor Statistics	101_Unemployment_rate.do

4.2 Figures

Table 2: Main Text Figures

Figure	Program	Output File
Figure 1, Panel A	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins.pdf

Figure 1, Panel B	112_1_DistributionsGraphs_byYear_Full_vs_Basic.do	112_Predicted6MJFR_Distribution_0M_Full_2006_histogram.pdf
Figure 2, Panel A	112_1_DistributionsGraphs_byYear_Full_vs_Basic.do	112_Predicted6MJFR_Distribution_0Mvs6M_Full_2006_kdensity.pdf
Figure 2, Panel B	112_2_DistributionsGraphs_byYear_2006_vs_2009.do	112_Predicted6MJFR_Distribution_0M_Full_2006vs2009_kdensity.pdf
Figure 2, Panel C	112_3_DistributionsGraphs_byYear_DifferentJobFindingHorizons.do	112_PredictedJFR_Distribution_DifferentJobFindingHorizons_Full_2006_kdensity.pdf
Figure 2, Panel D	112_1_DistributionsGraphs_byYear_Full_vs_Basic.do	112_Predicted6MJFR_Distribution_0M_FullvsBasic_2006_kdensity.pdf
Figure 3	109_DynamicSelection.do	109_SelectionInJobFinding_Full.pdf
Figure 4	119_5_DurationDependence_Heterogeneity_Visualization.do	119_5_Visualizing_Heterogeneity_DurationDependence_ShrunkInd.pdf
Figure 5	128_Timelines_Compositional_Cyclicality.do	128_LTU_Risk_byYear_emplAft6M_0M_In.pdf
Figure 6	117_Cyclicality_R_squared.do	117_6MJFR_relative_R_squared_AllIndiv_2006Model_Full_from1995_3models.pdf
Figure 7	114_5_Cyclicality_Heterogeneity_Visualization.do	114_5_Visualizing_Heterogeneity_Cyclicality_ShrunkInd.pdf

Table 3: Appendix Figures

Figure	Program	Output File
Figure A1, Panel A	112_4_DistributionsGraphs_ALMPs.do	112_Predicted6MJFR_Distribution_0M_ALMPS_Full_ALMPs_Narrow_First6M_2006_kdensity.pdf
Figure A1, Panel B	112_4_DistributionsGraphs_ALMPs.do	112_Predicted6MJFR_Distribution_0M_ALMPS_Full_NoTraining_ALMPs_Narrow_First6M_2006_kdensity.pdf
Figure A2	134_2_TwoSpell_Analysis.do	134_TwoSpells_histogram_timeDiff_years.pdf
Figure A3, Panel A	128_Timelines_Compositional_Cyclicality.do	128_LTU_Risk_byYear_emplAft6M_6M_In.pdf
Figure A3, Panel B	128_Timelines_Compositional_Cyclicality.do	128_LTU_Risk_byYear_emplAft6M_12M_In.pdf
Figure A4	101_Unemployment_rate.do	101_Unemp_rate_SSvsOECD.pdf
Figure A5, Panel A	119_4_DurationDependence_Distributions.do	119_4_DurationDependence_Full_2006_Distribution_ExpBeta0Shrunken.pdf
Figure A5, Panel B	119_4_DurationDependence_Distributions.do	119_4_DurationDependence_Full_2006_Distribution_BetaDShrunken.pdf
Figure A6, Panel A	114_4_Cyclicality_Distributions.do	114_4_Cyclicality_Distribution_Beta0Shrunken.pdf
Figure A6, Panel B	114_4_Cyclicality_Distributions.do	114_4_Cyclicality_Distribution_BetaUShrunken.pdf
Figure A7	114_5_Cyclicality_Heterogeneity_Visualization.do	114_5_Visualizing_Heterogeneity_Cyclicality_ShrunkInd_Rel.pdf

Figure A8	133_Heterogeneity_Regressions.do	133_Visualizing_Heterogeneity_Correlates_Reg_-Coefplot.pdf
Figure A9	133_Heterogeneity_Regressions.do	133_Visualizing_Heterogeneity_Correlates_-PartR2_Heatplot.pdf
Figure A10	133_Heterogeneity_Regressions.do	133_Visualizing_Heterogeneity_Correlates_-PartR2_Heatplot.pdf
Figure B1, Panel A	136_Prop_Hazard_Test.do	136_PropHazTest_Histogram_6Mvs0M.pdf
Figure B1, Panel B	136_Prop_Hazard_Test.do	136_PropHazTest_Histogram_2009vs2006.pdf
Figure B2	136_Prop_Hazard_Test.do	136_PropHazTest_Regression.pdf
Figure C1	142_ML_Robustness_Tuning.do	142_ML_Models_Tuning_CV_results_Full_-2006.pdf
Figure C2	142_ML_Robustness_Tuning.do	142_ML_Models_Tuning_Time_Series.pdf
Figure C3	140_ML_Robustness_Weights.do	140_Weights_Scatterplot.pdf
Figure C4, Panel A	107_ROC_Curves_2006.do	107_ROCPlot_emplAft6M_0M_In_2006_Full_ensemble.pdf
Figure C4, Panel B	107_ROC_Curves_2006.do	107_ROCPlot_emplAft6M_6M_In_2006_Full_ensemble.pdf
Figure C4, Panel C	107_ROC_Curves_2006.do	107_ROCPlot_emplAft6M_0M_In_2006_Full_ensemble_vs_other.pdf
Figure C4, Panel D	107_ROC_Curves_2006.do	107_ROCPlot_emplAft6M_6M_In_2006_Full_ensemble_vs_other.pdf
Figure C5	141_ML_Robustness_Correlation.do	141_ML_Models_Correlation_Heatmap_2006_-emplAft6M_0M_In.pdf
Figure C6, Panel A	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins.pdf
Figure C6, Panel B	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_6M_Full_-EqualSizedBins.pdf
Figure C6, Panel C	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_12M_Full_-EqualSizedBins.pdf
Figure C7	113_3_Scatterplot_empiricalJFR_on_predicted-JFR_ML_Robustness.do	113_3_Predicted_emplAft6M_0M_In_onEmpirical_ML_Robustness_2006.pdf
Figure C8	113_2_Scatterplot_empiricalJFR_on_predicted-JFR_LinearModel.do	113_Predicted6MJFR_onEmpirical_0M_Linear_EqualSizedBins.pdf
Figure C9	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins_Subsample_median_L_TotInc_-adj.pdf
Figure C10	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins_Subsample_Gender.pdf
Figure C11	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins_Subsample_foreign.pdf
Figure C12	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-EqualSizedBins_Subsample_median_DaysUnemp_5Years.pdf

Figure C13	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_-Full_EqualSizedBins_Subsample_median_-DaysOnDI_5Years.pdf
Figure C14, Panel A	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-36Groups.pdf
Figure C14, Panel B	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_6M_Full_-36Groups.pdf
Figure C14, Panel C	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_12M_Full_-36Groups.pdf
Figure C15, Panel A	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_0M_Full_-144Groups.pdf
Figure C15, Panel B	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_6M_Full_-144Groups.pdf
Figure C15, Panel C	113_1_Scatterplot_empiricalJFR_on_predicted-JFR.do	113_Predicted6MJFR_onEmpirical_12M_Full_-144Groups.pdf

4.3 Tables

Table 4: Main Text Tables

Table	Program	Output File
Table 1	-	-
Table 2	106_BasicStatistics.do	106_BasicStatistics_table.tex
Table 3	111_1_Explanatory_power_of_variables.do, 111_-3_Explanatory_power_of_variables_Expanded.do	111_1_Explanatory_power_of_variables_Sub-models_table.tex
Table 4	134_3_TwoSpell_Table_Main.do	134_TwoSpell_Table_Main.pdf
Table 5	108_1_MainStatistics_xMonthPred_yMonth-Model.do	108_1_MainStatistics_Table_Full_2006_xMonth-Pred_yMonthModel.tex

Table 5: Appendix Tables

Table	Program	Output File
Table A1	108_2_MainStatistics_2006_DifferentJobFind-ingHorizons.do	108_2_MainStatistics_Table_Full_2006_Differen-tJobFindingHorizons.tex
Table A2	108_3_MainStatistics_2006_Robustness.do	108_3_MainStatistics_Table_Full_2006_Robust-ness.tex
Table A3	111_1_Explanatory_power_of_variables.do	111_1_Explanatory_power_of_variables_MLvs-Linear_table.tex
Table A4	111_2_Explanatory_power_of_variables_Employ-mentHistory.do	111_2_Explanatory_power_of_variables_Employ-mentHistory_Combined_table.tex
Table A5	120_Regression_SILC_variables.do	120_Regression_JFR_SILC_Appendix_table.tex
Table A6	111_3_Explanatory_power_of_variables_Ex-panded.do	111_3_Explanatory_power_of_variables_Basic-Expanded_table.tex

Table A7	134_2_TwoSpell_Analysis.do	134_TwoSpells_Table_Appendix.pdf
Table A8	134_4_TwoSpell_Table_LTU.do	134_TwoSpell_Table_LTU.pdf
Table A9	134_5_TwoSpell_Table_SplitByBetaD.do	134_TwoSpells_Table_SplitByBetaD.tex
Table A10	-	-
Table A11	108_1_MainStatistics_xMonthPred_yMonthModel.do	108_1_MainStatistics_Table_Full_SeqDrop_incIndiv_2006_xMonthPred_yMonthModel.tex
Table A12	108_1_MainStatistics_xMonthPred_yMonthModel.do	108_1_MainStatistics_Table_Full_Pooled_2006_2007_xMonthPred_yMonthModel.tex
Table A13	108_1_MainStatistics_xMonthPred_yMonthModel.do	108_1_MainStatistics_Table_Full_Pooled_2009_2010_xMonthPred_yMonthModel.tex
Table A14	128_Timelines_Compositional_Cyclicality.do	128_Reg_LTU_risk_on_unemp_rate.tex

5 Description of Programs

Note that the programs are listed in alphabetical order, which does not necessarily coincide with the order of execution. For the latter, see Programs/000_master.do.

5.1 Data Management

- **001_1_CleaningPES.do**. This do file creates the dataset of the unemployment spells.
  - Inputs:
    - AF_Sokatper.dta
    - vy_Sok.dta
    - AF_Insper.dta
    - Insper.dta
  - Outputs:
    - 001_1_UnemploymentSubspells_Initial.dta
    - 001_1_UnemploymentSubspells_Final.dta
    - 001_1_UnemploymentSpells_Save.dta
    - 001_1_UnemploymentSpells_temp.dta
    - 001_1_UnemploymentSubspells.dta
    - 001_1_UnemploymentSubspells_2.dta
    - 001_1_UnemploymentSpells.dta
- **001_2_CleaningLISA.do**. This do file cleans the data from LISA.
  - Inputs:
    - Lisa_1990.dta to Lisa_2017.dta
    - 001_1_UnemploymentSpells.dta
    - Inflation.dta
    - 006_SNI92_new.dta
    - 006_SNI2002_new.dta
    - 006_SNI2007_new.dta
  - Outputs:
    - 001_2_Lisa_1990_tidy.dta to 001_2_Lisa_2017_tidy.dta
    - 001_2_Lisa_allYears.dta
    - 001_2_LISA_clean.dta
    - 001_2_LISA_clean2.dta
- **001_4_CleaningWealth.do**. This do file creates the dataset with wealth variables.

  - Inputs:
    * WealthR_1999.dta to WealthR_2007.dta
    * 001_1_UnemploymentSpells.dta
  - Outputs:
    * 001_4_Wealth_clean.dta

- **001_5_CleaningWages.do**. This do file creates a dataset with wages.
  - Inputs:
    * centralgovt_wages1990.dta to centralgovt_wages2017.dta
    * localgovt_wages1990.dta to localgovt_wages2017.dta
    * private_bluecollar_wages1990.dta to private_bluecollar_wages2017.dta
    * private_whitecollar_wages1990.dta to private_whitecollar_wages2017.dta
    * regionalgovt_wages1990.dta to regionalgovt_wages2017.dta
  - Outputs:
    * 001_5_Wages_AllYears.dta
    * 001_5_Wages_clean.dta

- **001_7_CleaningUnemplBenefits.do**. This do file creates a dataset with the level of unemployment benefits.
  - Inputs:
    * IAF_Utbetalning_ny.dta
  - Outputs:
    * 001_7_UnempBenefits_clean.dta
    * 001_7_UnempBenefits_clean_alt.dta

- **001_8_CleaningUnemplInsurance.do**. This do file creates the dataset with variables indicating that an individual bought additional unemployment insurance and was a member of a union.
  - Inputs:
    * AKAS_2002.dta to AKAS_2006.dta
    * 001_1_UnemploymentSpells.dta
    * UI_fund.dta
  - Outputs:
    * 001_8_UI_clean.dta
    * 001_8_unionMember.dta
    * 001_8_UI_clean2005-2009.dta
    * 001_8_UI_final.dta

- **001_9_MergingData.do**. This do file creates the main dataset used for the analysis of long term unemployment.
  - Inputs:
    * 001_1_UnemploymentSpells.dta
    * 001_2_LISA_clean.dta
    * 001_5_Wages_clean.dta
    * 001_4_Wealth_clean.dta
    * 001_8_UI_final.dta
    * 001_7_UnempBenefits_clean_alt.dta
    * 001_10_IQData_clean.dta
    * 001_8_unionMember.dta
  - Outputs:
    * 001_9_LISA_PES_cleaned.dta
    * 001_9_Share_non_missing_allYears.dta
    * 001_9_FinalMainDataset.dta

    - 001_9_FinalMainDataset_expanded.dta

- **001_10_CleaningIQ.do**. This do file creates the dataset with IQ variables.
  - Inputs:
    - Enlistment_old.dta
    - Enlistment_new.dta
  - Outputs:
    - 001_10_IQData_old.dta
    - 001_10_IQData_new.dta
    - 001_10_IQData_clean.dta

- **001_11_CleaningSILC.do**. This do file creates the dataset with SILC variables.
  - Inputs:
    - silc_2006.dta to silc_2019.dta
    - 001_1_UnemploymentSpells.dta
  - Outputs:
    - 001_11_SILCData.dta

- **002_1_DataForR_byYear.do**. This do file creates the datasets that are to be fed into R to create the ensemble predictions, separately for every year.
  - Inputs:
    - 001_9_FinalMainDataset.dta
  - Outputs:
    - 002_DataForR_Full_1992.dta to 002_DataForR_Full_2016.dta
    - 002_DataForR_Full_SeqDrop_*.dta
    - 002_DataForR_Full_DropPastSpells_*.dta
    - 002_DataForR_Full_Marg_*.dta

- **002_2_DataForR_byYear_Expanded.do**. This do file creates the datasets that are to be fed into R to create the ensemble predictions for the expanded models, separately for every year.
  - Inputs:
    - 001_9_FinalMainDataset_expanded.dta
    - 002_DataForR_Full_2006.dta
  - Outputs:
    - 002_DataForR_EX_FullSample_*.dta
    - 002_DataForR_EX_Ba_FullSample_*.dta

- **002_3_DataForR_byYear_EmploymentHistory.do**. This do file creates the datasets that are to be fed into R to create the ensemble predictions for the models with the expanded employment history variables.
  - Inputs:
    - 001_9_FinalMainDataset.dta
  - Outputs:
    - 002_DataForR_BasicWithEmplHist_2006.dta
    - 002_DataForR_BasicWithEmplHist_*.dta
    - 002_DataForR_BasicWithEmplHist_IndivVars_Seq_*.dta
    - 002_DataForR_BasicWithEmplHist_IndivVars_Marg_*.dta

- **002_4_DataForR_Pooled.do**. This do file creates the datasets that are to be fed into R to create the ensemble predictions, but pooling several years of data.
  - Inputs:
    - 001_9_FinalMainDataset.dta

  - Outputs:
    * 002_DataForR_Full_Pooled_*.dta

- **002_5_DataForR_NoTraining.do**. This do file creates the datasets that are to be fed into R to create the ensemble predictions for 2006 excluding spells that included training or recalls.
  - Inputs:
    * 002_DataForR_Full_2006.dta
    * 005_UnemploymentCategoryStats.dta
    * 001_1_UnemploymentSpells_Save.dta
  - Outputs:
    * 002_DataForR_Full_NoTraining_2006.dta
    * 002_DataForR_Full_NoRecalls_2006.dta

- **003_0_CleaningPredictions_program.do**. This do file creates the programs that clean the predictions coming from R, obtain the weights for the algorithms, and produce the final Ensemble predictions.
  - Inputs:
    * None.
  - Outputs:
    * None.

- **003_1_CleaningPredictions_byYear.do**. This do file cleans the predictions coming out of R, obtains the ensemble model and saves the final dataset. This do file does it separately for each year.
  - Inputs:
    * 003_0_CleaningPredictions_program.do
    * 103_predictionsR_*.csv
    * 002_DataForR_*.dta
    * 001_9_FinalMainDataset.dta
  - Outputs:
    * 003_PredEnsemble_*.dta
    * 003_MainWithEnsemblePred_*.dta
    * 003_Weights_*.dta
    * 003_Means_*.dta

- **003_2_CleaningPredictions_byYear_xMonthPred_yMonthModel.do**. This do file cleans the predictions coming out of R, obtains the ensemble model and saves the final dataset, like 003_1, but for the predictions for a given unemployment duration generated using a model trained at a different duration.
  - Inputs:
    * 003_0_CleaningPredictions_program.do
    * 103_predictionsR_*.csv
    * 002_DataForR_*.dta
    * 001_9_FinalMainDataset.dta
  - Outputs:
    * 003_PredEnsemble_*.dta
    * 003_MainWithEnsemblePred_*.dta
    * 003_Weights_*.dta
    * 003_Means_*.dta

- **003_3_CleaningPredictions_YearIndividuals_TrainedOnOtherYears.do**. This do file cleans the predictions coming out of R, obtains the ensemble model and saves the final dataset, but for the predictions generated using models trained in different years.
  - Inputs:

    * 003_0_CleaningPredictions_program.do
    * 103_predictionsR_*.csv
    * 002_DataForR_*.dta
    * 001_9_FinalMainDataset.dta
  - Outputs:
    * 003_PredEnsemble_*.dta
    * 003_MainWithEnsemblePred_*.dta
    * 003_Weights_*.dta
    * 003_Means_*.dta

- **004_Combining_weights.do**. This do file combines the weights of different models and employment durations into a single summary dataset.
  - Inputs:
    * 003_Weights_Full_emplAft6M_0M_In_1992.dta to 003_Weights_Full_emplAft6M_12M_In_2016.dta
  - Outputs:
    * 004_Weights_combined.dta

- **005_UnemploymentCategoryStats.do**. This do file analyses the different categories of unemployment subspells (e.g., training, subsidized work, etc.).
  - Inputs:
    * 001_1_UnemploymentSubspells.dta
  - Outputs:
    * 005_UnemploymentCategoryStats.dta

- **006_CleaningIndustryCodes.do**. This do file cleans the crosswalks for industry codes.
  - Inputs:
    * SNI_2002_92.csv
    * SNI_2002_2007.csv
  - Outputs:
    * 006_SNI92_new.dta
    * 006_SNI2002_new.dta
    * 006_SNI2007_new.dta

5.2 Output Generation

- **101_Unemployment_rate.do**. This do file calculates and plots unemployment rate, separation rate and JFR.
  - Inputs:
    * 001_2_Lisa_allYears.dta
    * 001_1_UnemploymentSpells.dta
    * Unemployment_OECD.dta
  - Outputs:
    * 101_EmploymentData.dta
    * 101_Employed.dta
    * 101_Unemployed_allyears.dta
    * 101_Unemployment_N.dta
    * 101_Employed_Unemployed.dta
    * 101_Unemp_rate_SSvsOECD.pdf

- **102_0_caret_parameter_tuning_function.R**. This code defines the function that performs parameter tuning for our three machine learning algorithms.
  - Inputs:

    - None
  - Outputs:
    - None

- **103_0_caret_predictions_function.R**. This code defines the functions that perform training and prediction for our three machine learning algorithms.
  - Inputs:
    - None
  - Outputs:
    - None

- **104_1_caret_execution_byYear.R**. This code trains the baseline model for all years in the sample period.
  - Inputs:
    - 002_DataForR_Full_*.dta
    - 102_0_caret_parameter_tuning_Function.R
    - 103_0_caret_predictions_Function.R
  - Outputs:
    - 102_rfgrid_*.csv
    - 102_boostgrid_*.csv
    - 102_lassogrid_*.csv
    - 102_rfgrid_search_*.csv
    - 102_boostgrid_search_*.csv
    - 102_lassogrid_search_*.csv
    - 103_Models_*.rda
    - 103_predictionsR_*.csv

- **104_2_caret_execution_byYear_xMonthPred_yMonthModel.R**. This code creates predictions using models trained on other unemployment durations.
  - Inputs:
    - 002_DataForR_Full_*.dta
    - 103_Models_*.rda
    - 103_0_caret_predictions_Function.R
  - Outputs:
    - 103_predictionsR_*.csv

- **104_3_caret_execution_byYear_YearIndividuals_TrainedOnOtherYears.R**. This code takes models trained on years X and makes predictions for individuals from year Y.
  - Inputs:
    - 002_DataForR_Full_*.dta
    - 103_Models_*.rda
    - 103_0_caret_predictions_Function.R
  - Outputs:
    - 103_predictionsR_*.csv

- **104_4_caret_execution_Pooled.R**. This code trains the models using pooled data from several years.
  - Inputs:
    - 002_DataForR_Full_Pooled_*.dta
    - 102_0_caret_parameter_tuning_Function.R
    - 103_0_caret_predictions_Function.R
  - Outputs:

    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv
    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv

- **104_5_caret_execution_2006_OtherModels.R**. This code trains the extended and restricted models.
  - Inputs:
    * 002_DataForR_Full_*.dta
    * 102_0_caret_parameter_tuning_Function.R
    * 103_0_caret_predictions_Function.R
  - Outputs:
    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv
    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv
- **104_6_caret_execution_EmploymentHistory.R**. This code trains the submodels with more employment history variables.
  - Inputs:
    * 002_DataForR_BasicWithEmplHist_*.dta
    * 102_0_caret_parameter_tuning_Function.R
    * 103_0_caret_predictions_Function.R
  - Outputs:
    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv
    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv
- **104_7_caret_execution_NoTraining.R**. This code trains the model that excludes individuals who entered into training programs.
  - Inputs:
    * 002_DataForR_Full_NoTraining_2006.dta
    * 102_0_caret_parameter_tuning_function.R
    * 103_0_caret_predictions_function.R
  - Outputs:
    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv

    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv

- **104_8_caret_execution_ML_Robustness.R**. This code trains the baseline 2006 model with alternative hyperparameter tuning settings.
  - Inputs:
    * 002_DataForR_Full_*.dta
    * 102_0_caret_parameter_tuning_Function.R
    * 103_0_caret_predictions_Function.R
  - Outputs:
    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv
    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv
- **104_9_caret_execution_NoRecalls.R**. This code trains the model that excludes recalled employees.
  - Inputs:
    * 002_DataForR_Full_NoRecalls_2006.dta
    * 102_0_caret_parameter_tuning_function.R
    * 103_0_caret_predictions_function.R
  - Outputs:
    * 102_rfgrid_*.csv
    * 102_boostgrid_*.csv
    * 102_lassogrid_*.csv
    * 102_rfgrid_search_*.csv
    * 102_boostgrid_search_*.csv
    * 102_lassogrid_search_*.csv
    * 103_Models_*.rda
    * 103_predictionsR_*.csv
- **106_BasicStatistics.do**. This do file calculates basic descriptive statistics about the baseline sample.
  - Inputs:
    * 001_9_LISA_PES_cleaned.dta
    * 005_UnemploymentCategoryStats.dta
    * 001_2_Lisa_allYears.dta
    * Inflation.dta
  - Outputs:
    * 106_Basic_Stats_SampleAnalysis.dta
    * 106_Basic_Stats_SampleAll.dta
    * 106_Basic_Stats_SampleAnalysis_and_SampleAll.dta
    * 106_BasicStatistics_table.tex
- **107_ROC_Curves_2006.do**. This do file creates ROC curves for particular models.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta

    - 116_Linear_Predictions_Full_2006.dta
  - Outputs:
    - 107_ROCPlot_*.pdf

- **108_0_MainStatistics_Program.do**. This do file creates functions that help automate the process of calculating the main statistics of interest for the predictions.
  - Inputs:
    - None.
  - Outputs:
    - None.

- **108_1_MainStatistics_xMonthPred_yMonthModel.do**. This do file produces statistics based on the ensemble predictions at different unemployment durations (e.g., Table 5).
  - Inputs:
    - 108_0_MainStatistics_Program.do
    - 003_MainWithEnsemblePred_*.dta
    - 116_Linear_Predictions_Full_2006_xMonthPred_yMonthModel.dta
    - 119_3_DurationDependenceBetas_Full_2006.dta
    - 101_Employed_Unemployed.dta
  - Outputs:
    - 108_1_MainStatistics_Table_*.tex
    - 108_1_MainStatistics_Table_*.csv
    - 108_1_MainStatistics_TimeSeries_Selection.pdf
    - 108_1_MainStatistics_TimeSeries_Selection_Relative.pdf

- **108_2_MainStatistics_2006_DifferentJobFindingHorizons.do**. This do file produces statistics based on the ensemble predictions generated for different job finding horizons in the year 2006.
  - Inputs:
    - 108_0_MainStatistics_Program.do
    - 003_MainWithEnsemblePred_Full_2006.dta
  - Outputs:
    - 108_2_MainStatistics_Table_Full_2006_DifferentJobFindingHorizons.tex

- **108_3_MainStatistics_2006_Robustness.do**. This do file produces Table A2.
  - Inputs:
    - 108_0_MainStatistics_Program.do
    - 003_MainWithEnsemblePred_Full_2006.dta
    - 003_MainWithEnsemblePred_Full_NoTraining_2006.dta
    - 003_MainWithEnsemblePred_Full_NoRecalls_2006.dta
    - 116_Linear_Predictions_Full_2006.dta
    - 001_1_UnemploymentSpells_Save.dta
    - 003_MainWithEnsemblePred_Full_ML_Robustness_1020Split_2006.dta
    - 003_MainWithEnsemblePred_Full_ML_Robustness_2020Split_2006.dta
    - 003_MainWithEnsemblePred_Full_ML_Robustness_1040Split_2006.dta
    - 003_MainWithEnsemblePred_Full_2006_withSpline.dta
    - 003_MainWithEnsemblePred_Full_2006_PositiveWeights.dta
  - Outputs:
    - 108_3_MainStatistics_Table_Full_2006_Robustness.tex

- **109_DynamicSelection.do**. This do file produces figure 3.

  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
  - Outputs:
    * 109_SelectionInJobFinding_Full.dta
    * 109_SelectionInJobFinding_Full.pdf

- **111_1_Explanatory_power_of_variables.do**. This do file generates panels A and B of Table 3 and Appendix Table A3 in full.
  - Inputs:
    * 003_MainWithEnsemblePred_*.dta
    * 116_Linear_Full_SeqDrop_*.dta
  - Outputs:
    * 111_1_Explanatory_power_of_variables_Submodels_table.tex
    * 111_1_Explanatory_power_of_variables_MLvsLinear_table.tex

- **111_2_Explanatory_power_of_variables_EmploymentHistory.do**. This do file generates appendix Table A4.
  - Inputs:
    * 003_MainWithEnsemblePred_*.dta
  - Outputs:
    * 111_2_Explanatory_power_of_variables_EmploymentHistory_Combined_table.tex

- **111_3_Explanatory_power_of_variables_Expanded.do**. This do file generates panel C of Table 3 and Appendix Table A6.
  - Inputs:
    * 003_MainWithEnsemblePred_*.dta
  - Outputs:
    * 111_1_Explanatory_power_of_variables_Submodels_table.tex
    * 111_3_Explanatory_power_of_variables_BasicExpanded_table.tex

- **112_1_DistributionsGraphs_byYear_Full_vs_Basic.do**. This do file generates distribution graphs comparing Full vs Basic models by year.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
  - Outputs:
    * 112_Predicted6MJFR_Distribution_0M_FullvsBasic_2006_kdensity.pdf
    * 112_Predicted6MJFR_Distribution_0M_Full_2006_histogram.pdf
    * 112_Predicted6MJFR_Distribution_0Mvs6M_Full_2006_kdensity.pdf

- **112_2_DistributionsGraphs_byYear_2006_vs_2009.do**. This do file generates distribution graphs comparing the years 2006 vs 2009.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
    * 003_MainWithEnsemblePred_Full_2009.dta
  - Outputs:
    * 112_Predicted6MJFR_Distribution_0M_Full_2006vs2009_kdensity.pdf

- **112_3_DistributionsGraphs_byYear_DifferentJobFindingHorizons.do**. This do file produces panel C of Figure 2.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta

  - Outputs:
    * 112_PredictedJFR_Distribution_DifferentJobFindingHorizons_Full_2006_kdensity.pdf

- **112_4_DistributionsGraphs_ALMPs.do**. This do file produces Figure A1.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
    * 005_UnemploymentCategoryStats.dta
    * 003_MainWithEnsemblePred_Full_NoTraining_2006.dta
  - Outputs:
    * 112_Predicted6MJFR_Distribution_0M_ALMPS_Full_NoTraining_ALMPs_Narrow_First6M_2006_kdensity.pdf
    * 112_Predicted6MJFR_Distribution_0M_ALMPS_Full_NoTraining_ALMPs_Broad_First6M_2006_kdensity.pdf

- **113_1_Scatterplot_empiricalJFR_on_predictedJFR.do**. This do file produces binned scatterplots of empirical JFR on predicted JFR for the baseline model.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
    * 001_9_FinalMainDataset.dta
  - Outputs:
    * 113_Predicted6MJFR_onEmpirical_0M_Full_EqualSizedBins.pdf
    * 113_Predicted6MJFR_onEmpirical_6M_Full_EqualSizedBins.pdf
    * 113_Predicted6MJFR_onEmpirical_12M_Full_EqualSizedBins.pdf

- **113_2_Scatterplot_empiricalJFR_on_predictedJFR_LinearModel.do**. This do file produces binned scatterplots of empirical JFR on predicted JFR for the linear model.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
    * 116_Linear_Predictions_Full_2006.dta
    * 001_9_FinalMainDataset.dta
  - Outputs:
    * 113_Predicted6MJFR_onEmpirical_0M_Linear_EqualSizedBins.pdf
    * 113_Predicted6MJFR_onEmpirical_6M_Linear_EqualSizedBins.pdf
    * 113_Predicted6MJFR_onEmpirical_12M_Linear_EqualSizedBins.pdf

- **113_3_Scatterplot_empiricalJFR_on_predictedJFR_ML_Robustness.do**. This do file produces binned scatterplots of empirical JFR on predicted JFR for the underlying ML models.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
  - Outputs:
    * 113_Predicted6MJFR_onEmpirical_0M_Full_ML_Robustness_EqualSizedBins.pdf

- **114_1_Cyclicality_DataPreparation_beforeR.do**. This do file prepares the data for the individual regressions of predictions on the unemployment rate to investigate the cyclicality of JFR.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
    * 101_Employed_Unemployed.dta
  - Outputs:
    * 114_RPanelRegression_Full_2006_relative_trend.dta

- **114_2_Cyclicality_IndividualRegression.R**. This R script runs individual bivariate regressions and extracts the coefficients.
  - Inputs:

    - 114_RPanelRegression_Full_2006_relative_trend.dta
  - Outputs:
    - 114_RPanelRegression_Full_2006_relative_trend_rel.csv

- **114_3_Cyclicality_DataPreparation_afterR.do**. This do file processes the results from the cyclicality regressions and saves them into Stata format.
  - Inputs:
    - 114_RPanelRegression_Full_2006_relative_trend_rel.csv
    - 114_RPanelRegression_Full_2006_relative_trend.dta
  - Outputs:
    - 114_3_CyclicalityBetas_Full_2006.dta

- **114_4_Cyclicality_Distributions.do**. This do file plots the distribution of the estimated parameters in the cyclicality regressions.
  - Inputs:
    - 114_3_CyclicalityBetas_Full_2006.dta
  - Outputs:
    - 114_4_Cyclicality_Distribution_Beta0Shrunken.pdf
    - 114_4_Cyclicality_Distribution_BetaUShrunken.pdf
    - 114_4_Cyclicality_Distribution_BetaTShrunken.pdf
    - 114_4_Cyclicality_Distribution_TvsU_heatplot.pdf

- **114_5_Cyclicality_Heterogeneity_Visualization.do**. This do file produces graphs that depict the heterogeneity by cyclicality in our sample.
  - Inputs:
    - 114_3_CyclicalityBetas_Full_2006.dta
  - Outputs:
    - 114_5_Visualizing_Heterogeneity_Cyclicality_ShrunkPop.pdf
    - 114_5_Visualizing_Heterogeneity_Cyclicality_ShrunkInd.pdf
    - 114_5_Visualizing_Heterogeneity_Cyclicality.pdf
    - 114_5_Visualizing_Heterogeneity_Cyclicality_ShrunkInd_Rel.pdf

- **116_1_Linear_reference_model.do**. This file trains the linear model.
  - Inputs:
    - 002_DataForR_Full_*.dta
    - 003_MainWithEnsemblePred_*.dta
  - Outputs:
    - 116_Linear_*.dta
    - 116_Linear_Predictions_*.dta

- **116_2_Linear_expanded_models.do**. This file trains the expanded linear models.
  - Inputs:
    - 002_DataForR_EX_FullSample_*.dta
    - 003_MainWithEnsemblePred_*.dta
  - Outputs:
    - 116_Linear_*.dta
    - 116_Linear_Predictions_*.dta

- **117_Cyclicality_R_squared.do**. This do file investigates how the predictive power changes when we use a model from year Y to make predictions for individuals from year X.

- Inputs:
  - 108_0_MainStatistics_Program.do
  - 003_MainWithEnsemblePred_*.dta
- Outputs:
  - 117_6MJFR_R_squared_AllIndiv_2006Model_Full_from1995.pdf
  - 117_6MJFR_relative_R_squared_AllIndiv_2006Model_Full_from1995_3models.pdf

- **119_1_DurationDependence_DataPreparation_beforeR.do**. This do file prepares the predictions for the duration dependence analysis, which is performed in R.
  - Inputs:
    - 003_MainWithEnsemblePred_*.dta
  - Outputs:
    - 119_1_DataForR_DurationDependenceBetas_Full_*.dta
    - 119_1_DataForR_DurationDependenceBetas_Full_*.dta
    - 119_1_DataForR_DurationDependenceBetas_Full_*.dta

- **119_2_DurationDependence_IndividualRegressions.R**. This R script runs individual bivariate regressions and extracts the coefficients.
  - Inputs:
    - 119_1_DataForR_DurationDependenceBetas_Full_*.dta
  - Outputs:
    - 119_2_DurationDependenceBetas_Full_*.csv

- **119_3_DurationDependence_DataPreparation_afterR.do**. This file processes the results from the duration dependence regressions and saves them in Stata format.
  - Inputs:
    - 119_2_DurationDependenceBetas_Full_*.csv
    - 119_1_DataForR_DurationDependenceBetas_Full_*.dta
  - Outputs:
    - 119_3_DurationDependenceBetas_Full_*.dta

- **119_4_DurationDependence_Distributions.do**. This do file produces the histograms of the duration dependence estimated parameters.
  - Inputs:
    - 119_3_DurationDependenceBetas_Full_*.dta
  - Outputs:
    - 119_4_DurationDependence_*.pdf

- **119_5_DurationDependence_Heterogeneity_Visualization.do**. This do file produces graphs that depict the heterogeneity by duration dependence in our sample.
  - Inputs:
    - 119_3_DurationDependenceBetas_Full_*.dta
  - Outputs:
    - 119_5_Visualizing_Heterogeneity_DurationDependence_ShrunkPop.pdf
    - 119_5_Visualizing_Heterogeneity_DurationDependence_ShrunkInd.pdf
    - 119_5_Visualizing_Heterogeneity_DurationDependence.pdf

- **120_Regression_SILC_variables.do**. This do file investigates the predictive power of SILC variables.
  - Inputs:
    - 001_11_SILCData.dta

    * 003_MainWithEnsemblePred_Full_*.dta
  - Outputs:
    * 120_Regression_JFR_SILC_Appendix_table.tex

- **126_Bootstrapping.do**. This do file does bootstrapping on the hold-out sample to check the variability of R-squared.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_2006.dta
  - Outputs:
    * 126_Bootstrapping_Full_2006.dta

- **128_DataPreparation_Compositional_Cyclicality.do**. This do file analyses the ensemble predictions generated using a model from a fixed year for individuals from all years.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_*.dta
  - Outputs:
    * 128_Statistics_Full_1992_2016.dta
    * 128_Statistics_Full_1992_2016Individuals_TrainedOn_Full2006.dta

- **128_Timelines_Compositional_Cyclicality.do**. This do file plots the LTU risk over time both using predictions obtained year by year and predictions from a fixed year to investigate the compositional aspect of cyclicality of LTU.
  - Inputs:
    * 128_Statistics_Full_1992_2016.dta
    * 128_Statistics_Full_1992_2016Individuals_TrainedOn_Full2006.dta
    * 101_Employed_Unemployed.dta
  - Outputs:
    * 128_LTU_Risk_byYear_emplAft6M_0M_In_data.csv
    * 128_LTU_Risk_byYear_emplAft6M_0M_In.pdf
    * 128_LTU_Risk_byYear_emplAft6M_6M_In.pdf
    * 128_LTU_Risk_byYear_emplAft6M_12M_In.pdf
    * 128_Reg_LTU_risk_on_unemp_rate.tex

- **133_Heterogeneity_Regressions.do**. This do file runs regressions of LTU risk, duration dependence and cyclicality on observables in the baseline 2006 sample.
  - Inputs:
    * 119_3_DurationDependenceBetas_Full_2006.dta
    * 114_3_CyclicalityBetas_Full_2006.dta
    * 002_DataForR_Full_2006.dta
  - Outputs:
    * 133_Visualizing_Heterogeneity_Correlates_Corr_Heatplot.pdf
    * 133_Visualizing_Heterogeneity_Correlates_Reg_Coefplot.pdf
    * 133_Visualizing_Heterogeneity_Correlates_PartR2_Heatplot.pdf

- **134_1_TwoSpell_DataPreparation.do**. This do file creates the multiple spells sample.
  - Inputs:
    * 003_MainWithEnsemblePred_Full_*.dta
  - Outputs:
    * 134_MultipleSpellSample.dta
    * 134_TwoSpellSample.dta

- **134_2_TwoSpell_Analysis.do**. This do file produces statistics about individuals with at least two spells in our sample.
  - Inputs:
    - 001_9_FinalMainDataset.dta
  - Outputs:
    - 134_MultipleSpells_SummaryStats.csv
    - 134_ReferenceSample_VCov.csv
    - 134_MultipleSpellsSample_VCov.csv
    - 134_MultipleSpellsxLTUSample_VCov.csv
    - 134_MultipleSpells_SummaryStats_SplitByBetaD.csv
    - 134_betaD_kdensity.pdf
    - 134_TwoSpell_Analysis.ster
    - 134_TwoSpell_Analysis_LTU.ster
    - 134_TwoSpell_Analysis_SplitByBetaD.ster
- **134_3_TwoSpell_Table_Main.do**. This do file generates the main table for the two-spell analysis.
  - Inputs:
    - 134_TwoSpell_Analysis.ster
  - Outputs:
    - 134_TwoSpell_Table_Main.pdf
    - 134_TwoSpell_Table_Main.csv
    - 134_TwoSpells_Table_Appendix.tex
- **134_4_TwoSpell_Table_LTU.do**. This do file generates the LTU table for the two-spell analysis.
  - Inputs:
    - 134_TwoSpell_Analysis_LTU.ster
  - Outputs:
    - 134_TwoSpell_Table_LTU.pdf
    - 134_TwoSpell_Table_LTU.csv
- **134_5_TwoSpell_Table_SplitByBetaD.do**. This do file produces Appendix Table A9.
  - Inputs:
    - 134_TwoSpell_Analysis_SplitByBetaD.ster
  - Outputs:
    - 134_TwoSpells_Table_SplitByBetaD.tex
    - 134_TwoSpells_Table_SplitByBetaD_NonMissing.csv
    - 134_TwoSpells_Table_SplitByBetaD_BelowMedian.csv
    - 134_TwoSpells_Table_SplitByBetaD_AboveMedian.csv
- **136_Prop_Hazard_Test.do**. This do file carries out the proportional hazards tests in Appendix B.3.
  - Inputs:
    - 003_MainWithEnsemblePred_Full_2006.dta
    - 003_MainWithEnsemblePred_Full_2006_xMonthPred_yMonthModel.dta
    - 003_MainWithEnsemblePred_Full_2006Individuals_TrainedOn2009modelIndividuals_Full.dta
    - 002_DataForR_Full_2006.dta
  - Outputs:
    - 136_PropHazTest_Histogram_6Mvs0M.pdf
    - 136_PropHazTest_Histogram_2009vs2006.pdf
    - 136_PropHazTest_Regression.pdf

- **140_ML_Robustness_Weights.do**. This do file generates descriptive statistics and scatterplots about individual predictor weights in the ensemble.
  - Inputs:
    - 004_Weights_combined.dta
  - Outputs:
    - 140_Weights_Scatterplot.pdf
- **141_ML_Robustness_Correlation.do**. This do file creates a correlation matrix of the different prediction algorithms at different unemployment durations.
  - Inputs:
    - 003_MainWithEnsemblePred_Full_*.dta
    - 116_Linear_Predictions_Full_2006.dta
  - Outputs:
    - 141_ML_Models_Correlation.dta
    - 141_ML_Models_Correlation_Scatterplot.pdf
    - 141_ML_Models_Covariance_Scatterplot.pdf
    - 141_ML_Models_Correlation_Heatmap_2006_emplAft6M_0M_In.pdf
- **142_ML_Robustness_Tuning.do**. This do file creates diagnostics plots for the tuning of the ML models.
  - Inputs:
    - 102_rfgrid_Full_emplAft6M_0M_In_*.csv
    - 102_boostgrid_Full_emplAft6M_0M_In_*.csv
    - 102_lassogrid_Full_emplAft6M_0M_In_*.csv
    - 102_rfgrid_search_Full_emplAft6M_0M_In_*.csv
    - 102_boostgrid_search_Full_emplAft6M_0M_In_*.csv
    - 102_lassogrid_search_Full_emplAft6M_0M_In_*.csv
  - Outputs:
    - 142_ML_Models_Tuning_Time_Series.pdf
    - 142_ML_Models_Tuning_CV_results_Full_2006.pdf
    - 142_ML_Models_Tuning_2006.tex