Replication Files for

"The Depth of Preferential Trade Agreements"

Pao-Li Chang, Wei Jin, and Kefang Yao

September 2025

Coding for this paper was done in MATLAB (R2021a) and Stata (MP 17), both run on the Windows 11 operating system. To replicate our results, please keep file folder structure unchanged and follow the instruction below step by step. If there is any mistake or typo, feel free to contact Kefang Yao (by email nbykf19@hotmail.com). Your feedback would be greatly appreciated by us.

Step 0

The replication starts with the Stata (.dta) data file "**main_raw_data.dta**" located in file folder "01_aggregate_raw_data", where we also store the PTA data as described in Section 2 in file folder "pta-bilateral-observations", the WDI unemployment rate data as described in Section B.3 in file folder "wdi_ue", the OECD-ICIO data "**OECD-ICIO.dta**" described in Online Appendix A.1.2, and the tariff data for years 1988-2015 "**WITS_tariff.dta**" described in Online Appendix A.1.1.

Due to the 2.5G limit per file in the online repository site, the zipped file ("**main_raw_data.zip**") is provided instead of the raw data file. Unzip this file to obtain the dataset "**main_raw_data.dta**" and store it in file folder "01_aggregate_raw_data". Delete the generated folders. "**main_raw_data.dta**" includes the bilateral trade flows described in Appendix A.1, the GDP, value-added share, and gross output data described in Appendix A.2, and the bilateral trade cost proxy variables described in Appendix A.4.

Step 1

Go to file folder "02_pseudoWorld_forRegression" to construct the pseudo world. First, run "**01_master_PseudoWorld.do**" to clean the data set and apply the first round of adjustment procedure for trade deficits and expenditures as described in paper "Appendix A.5 Pseudo World". Second, run "**02_master_ctylist.do**" to get a balanced panel of 147 economies for the period 1980-2015 and construct the pseudo world as described in paper "Appendix A.5 Pseudo World". Third, run "**03_master_pta_subcategory.do**" to construct the data set for PPML regressions "**ppml_trim.dta**". "**03_master_pta_subcategory.do**" merges "**pta_bilateral_updated.dta**" in the subfolder "pta-bilateral-observations" of file folder "01_aggregate_raw_data" and constructs the PTA indicators as described in Section 2.

Step 2

Go to file folder "03_pta_regression" to run PPML regressions of PTA effects. First, run "**Table2.do**". Table 2 in paper for PTA regression results can be found in "**Table2_ppml.csv**". Table 2 in paper for hypothesis tests can be found in "**Table2_lincom_result_p_x.csv**" and "**Table2_lincom_result_c_xnc.csv**". Second, run "**PPML_woRUS.do**" for the Soviet Union/Russia robustness checks. Online Appendix Table A.1 for PTA regression results can be found in "**ppml_woRUS.csv**". Online Appendix Table A.1 for hypothesis tests can be found in "**lincom_result_p_x _woRUS.csv**" and "**lincom_result_c_xnc _woRUS.csv**".

Third, go to file folder "03_pta_regression/heter", run "**PPML_HL.do**" to obtain PTA estimation results by development status of the exporter and the importer. Online Appendix Table A.5 result can be found in "**TableA5_online_appendix.csv**". Run "**PPML_goods_services.do**" to obtain PTA estimation results by merchandise/service trade. Online Appendix Table A.6 result can be found in "**TableA6_online_appendix.csv**".

Then, go to file folder "03_pta_regression/lag", run "**PPML_lag.do**" to estimate the phase-in effect of PTA. Online Appendix Table A.2 result can be found in "**TableA2_online_appendix.csv**".

Step 3

Go to file folder "03_pta_regression/wtPTA" for the weighted measures of PTA depth. First, run "**ppml_count_se.do**" to generate Table 5 and Table A.4 in paper. Table 5 for direct regression results can be found in "**wtPTA_ppml_count_w1_le2.csv**", and the mean effect of each PTA category can be found in "**wtPTA_mean_se.csv**". Part 1 result of Table A.4 can be found in "**wtPTA_summary_part1.csv**", Part 2 result of Table A.4 can be found in "**wtPTA_summary_part2.csv**", and Part 3 result of Table A.4 can be found in "**wtPTA_summary_part3.csv**".

Then, run "**ppml_count_se_goods_services.do**" for the effects on merchandise and service trade. Online Appendix Table A.7 for direct regression results on merchandise trade can be found in "**wtPTA_ppml_count_w1_le2_goods.csv**", and the mean effect of each PTA category can be found in "**wtPTA_mean_se_goods.csv**". Online Appendix Table A.8 for direct regression results on service trade can be found in "**wtPTA_ppml_count_w1_le2_service.csv**", and the mean effect of each PTA category can be found in "**wtPTA_mean_se_service.csv**".

Step 4

Go to file folder "04_pta_DID_matching/DID-PTA" for Difference-in-differences matching estimations of PTA effects. First, run "**01_master_BB_data.do**" to approximate the multilateral terms by a first-order Taylor-series expansion. Second, run "**02_ForMatlab_real.do**" to get "**primal_MR.csv**". Third, convert "**primal_MR.csv**" to "**primal_MR.mat**" using MATLAB (On the Home tab, in the Variable section, select Import Data, rename the variable "year1" to "year"). The "**primal_MR.mat**" will be used to run matching.

Step 5

Go to file folder "04_pta_DID_matching/DID-PTA". First, run "**main_test_pta.m**" to conduct the DID matching and permutation tests as described in Appendix B.3. Second, run "**main_test_pta_collect.m**" to collect results. Third, run "**03_report_matchingl.do**" to report matching estimates of PTA. Table 4 results in paper can be found in **"04_pta_DID_matching/DID-PTA/a1b5/A_latex2_0.dta", "04_pta_DID_matching/DID-PTA/a5b5/A_latex2_0.dta"**, and "**04_pta_DID_matching/DID-PTA/a10b5/A_latex2_0.dta**".

Then, run "**04_CSDID.do**" to conduct pre-trend test for the matched sample as described in Online Appendix A.2.3. Online Appendix Figure A.2 summarizes the pdf files **"04_pta_DID_matching/DID-PTA/a1b5/csplot_pre_pta.pdf", "04_pta_DID_matching/DID-PTA/a5b5/csplot_pre_pta.pdf",** and **"04_pta_DID_matching/DID-PTA/a10b5/csplot_pre_pta.pdf".**

Run "**SMD_pta.m**" to conduct balance test for the matched sample as described in Online Appendix A.2.3. Online Appendix Figure A.1 summarizes the jpg files **"04_pta_DID_matching/DID-PTA/a1b5/SMD-pta-histogram_a1b5.jpg", "04_pta_DID_matching/DID-PTA/a5b5/SMD-pta-histogram_a5b5.jpg",** and **"04_pta_DID_matching/DID-PTA/a10b5/SMD-pta-histogram_a10b5.jpg"**. Online Appendix Table A.3 can be found in **"04_pta_DID_matching/DID-PTA/a1b5/ SMD_Balance_Table_pta.csv", "04_pta_DID_matching/DID-PTA/a5b5/ SMD_Balance_Table_pta.csv",** and **"04_pta_DID_matching/DID-PTA/a10b5/ SMD_Balance_Table_pta.csv"**.

Step 6

Go to file folder "04_pta_DID_matching/DID-PTA_tariff" for Difference-in-differences matching estimations with controlling for tariffs. First, run "**01_master_BB_data_tariff.do**" to approximate the multilateral terms by a first-order Taylor-series expansion and merge with the tariff data set "**WITS_tariff.dta**" from file folder "01_aggregate_raw_data". Second, run "**02_ForMatlab_real_tariff.do**" to get "**primal_MR_tariff.csv**". Third, convert "**primal_MR_tariff.csv**" to "**primal_MR_tariff.mat**" using MATLAB (On the Home tab, in the Variable section, select Import Data, rename the variable "year1" to "year"). The "**primal_MR_tariff.mat**" will be used to run matching.

Step 7

Go to file folder "04_pta_DID_matching/DID-PTA_tariff". First, run "**main_test_pta.m**" to conduct the DID matching and permutation tests. Second, run "**main_test_pta_collect.m**" to collect results. Third, run "**03_report_matchingl.do**" to report matching estimates of PTA. Table 4 results (control for tariffs) in paper can be found in **"04_pta_DID_matching/DID-PTA_tariff /a1b5/A_latex2_0.dta", "04_pta_DID_matching/DID-PTA_tariff /a5b5/A_latex2_0.dta",** and **"04_pta_DID_matching/DID-PTA_tariff /a10b5/A_latex2_0.dta"**.

Then, run "**04_CSDID_tariff.do**" to conduct pre-trend test for the matched sample. Online Appendix Figure A.2 (control for tariff) summarizes the pdf files

"**04_pta_DID_matching/DID-PTA_tariff /a1b5/csplot_pre_pta _tariff.pdf", "04_pta_DID_matching/DID-PTA_tariff /a5b5/csplot_pre_pta _tariff.pdf",** and **"04_pta_DID_matching/DID-PTA_tariff /a10b5/csplot_pre_pta _tariff.pdf".**

Run "**SMD_pta_tariff.m**" to conduct balance test for the matched sample. Online Appendix Table A.4 can be found in "**04_pta_DID_matching/DID-PTA_tariff /a1b5/SMD_Balance_Table_pta_tariff.csv", "04_pta_DID_matching/DID-PTA_tariff /a5b5/SMD_Balance_Table_pta_tariff.csv",** and **"04_pta_DID_matching/DID-PTA_tariff /a10b5/SMD_Balance_Table_pta_tariff.csv".** Run **"SMD_pta_figure_tariff.m".** Online Appendix Figure A.1 (control for tariff) summarizes the jpg files **"04_pta_DID_matching/DID-PTA_tariff /a1b5/SMD-pta-histogram_a1b5_tariff.jpg", "04_pta_DID_matching/DID-PTA_tariff /a5b5/ SMD-pta-histogram_a5b5_tariff.jpg",** and **"04_pta_DID_matching/DID-PTA_tariff /a10b5/ SMD-pta-histogram_a10b5_tariff.jpg"**.

Step 8

Go to file folder "05_iteration_bootstrap" for the iterative estimation procedure. Further go to file folder "bs_baseline". First, run "**bs_baseline_preMATLAB.do**" to prepare for the datasets used in the iterative estimation procedure. Second, run "**02_bs_baseline.m**" to conduct the iteration procedure as described in Appendix B.1. Third, run "**bs_baseline_postMATLAB.do**" to collect the GATT/WTO membership effect estimates. Table 3 in paper for the point estimates can be found in "**Table3_baseline_report.csv**".

Then, for the bootstrap procedure as described in Appendix B.2, further go to file folder "bs_multiplier". First, run "**01_bs_ multiplier_preMATLAB.do**" to prepare for the datasets used in the 1000 bootstraps. Second, run "**01_bs_ multiplier.m**" to conduct the multiplier bootstrap as described in Appendix B.2. Third, run "**03_bs_baseline_postMATLAB.do**" to generate "**bs_coefficients_report.csv**". It includes the standard error and the 90% confidence interval of the GATT/WTO membership effect estimates in Table 3.

Step 9

Go to file folder "06_counterfactuals", where we conduct the main counterfactual analysis using the PTA and WTO estimates from Table 2 and Table 3. Further go to file folder "06_ counterfactuals /CA". Run "**CA_run.do**" and "**CA_run_deepwto.do**". We will get a list of csv files with names starting with "A". The csv files are inputs for the counterfactual analysis. Copy these csv files from file folder "06_ counterfactuals /CA" to file folders "06_counterfactuals/CA/Melitz" and "06_counterfactuals/CA/BKL".

Step 10

Go to file folder "06_CA_analysis /CA/Melitz" for counterfactual analysis under Melitz model. Run "**CA_main_test_pta_Melitz.m**" to conduct counterfactual analysis and then run "**CA_main_test_pta_Melitz_post.m**" to collect results. Copy the generated txt files with names starting with "CA" from file folder "06_counterfactuals/CA/Melitz" to file folder "06_counterfactuals/CA_ublb".

Next, run "**CA_collect_Melitz.do**" to plot the results in figures. Figure 2 summarizes the pdf files **"frequency_xW1_1980_5_ca0v1_2_M_fixed"** and **"frequency_xW1_2015_5_ca0v1_2_M_fixed".** Figure 3 summarizes the pdf files **"frequency_xW1_1980_5_ca0v21_2_M_fixed"** and **"frequency_xW1_2015_5_ca0v21_2_M_fixed".** Figure 4 summarizes the pdf files **"frequency_xW1_1980_5_ca0v31_2_M_fixed"** and **"frequency_xW1_2015_5_ca0v31_2_M_fixed".** Online Appendix Figure A.3, Online Appendix Figure A.4, and Online Appendix Figure A.5 summarizes the complete set of pdf files for the period of 1980-2015 (in every five years). Figure 5 summarizes the pdf files **"scatter_sr_xW1_1980_5_ca0v1_2_M_fixed"** and **"scatter_sr_xW1_2015_5_ca0v1_2_M_fixed".** Figure 6 summarizes the pdf files **"scatter_sr_xW1_1980_5_ca0v21_2_M_fixed"** and **"scatter_sr_xW1_2015_5_ca0v21_2_M_fixed".** Figure 7 summarizes the pdf files **"scatter_sr_xW1_1980_5_ca0v31_2_M_fixed"** and **"scatter_sr_xW1_2015_5_ca0v31_2_M_fixed".** Online Appendix Figure A.6, Online Appendix Figure A.7, and Online Appendix Figure A.8 summarizes the complete set of pdf files for the period of 1980-2015 (in every five years).

Then, run "**CA_collect_Melitz_table.do**" to collect the results in tables. Table 7 results are stored in "**xW1_5_2 _mean_M_d1.dta", "xW1_5_2 _mean_M_d2.dta",** and **"xW1_5_2 _mean_M_d3.dta**". Table 8 results are stored in **"xW1_5_2 _mean_M_d2_svd.dta"** and **"xW1_5_2 _mean_M_d3_svd.dta".** Table 6 result is stored in **"Region_GLZ_xW1_5_2_M_All.dta**". Table A.5 results in paper are stored in **"xW1_5_2_mean_M_d1_negflag.dta", "xW1_5_2_mean_M_d2_negflag.dta", "xW1_5_2_mean_M_d3_negflag.dta", "xW1_5_2_mean_M_d2_negflag.dta", "xW1_5_2_mean_M_d2_svd_negflag.dta",** and **"xW1_5_2_mean_M_d2_negflag.dta", "xW1_5_2_mean_M_d3_svd_negflag.dta".**

Last, run "**CA_main_test_pta_Melitz_deepWTO_full.m**" to compare the welfare effects of deepWTO versus shallowWTO, and then run "**CA_collect_Melitz_deepWTO.do**" to collect the results. Table 9 results are stored in "**xW1_5_2 _mean_M_deepwto_d2.dta"** and **"xW1_5_2 _mean_M_deepwto_d3.dta**". Copy the generated mat files with names starting with "Solution" from file folder "06_counterfactuals/CA/Melitz" to file folders "06_counterfactuals/CA_ublb".

Step 11

Go to file folder "06_CA_analysis /CA/BKL" for counterfactual analysis under BKL model. Run "**CA_main_test_pta_BKL_full.m**" and then run "**CA_collect_BKL.do**" to collect results. Table 10 results are stored in **"xW1_1950_sigma_diffrati_mean_BM.csv", "xW1_2015_sigma_diffrati_mean_BM.csv", "xN_1950_sigma_diffrati_mean_BM.csv",** and **"xN_2015_sigma_diffrati_mean_BM.csv**" (under Melitz Model). Online Appendix Table A.13 and Online Appendix Table A.14 results also can be found in the above four csv files for the comparison between Melitz and BKL models.

Step 12

Go to file folder "06_CA_ analysis /CA_ublb" for the counterfactual analysis based on the bootstrap GATT/WTO membership effect estimates. Run "**CA_run_ublb.do**". We will get a list of csv files with names starting with "A". Run "**CA_main_test_pta_ub_Melitz.m**" and then run "**CA_main_test_pta_ub_Melitz_post.m**" for the counterfactual analysis based on the 95th-percentile bootstrap GATT/WTO membership effect estimates. Run "**CA_main_test_pta_lb_Melitz.m**" and then run "**CA_main_test_pta_lb_Melitz_post.m**" for the counterfactual analysis based on the 5th-percentile bootstrap GATT/WTO membership effect estimates). Then run "**CA_collect_ublb_Melitz.do**" to collect the results. Online Appendix Table A.9 results are stored in "**xW1_5_2 _mean_lb_M_d1.dta", "xW1_5_2 _mean_lb _M_d2.dta",** and **"xW1_5_2 _mean_lb _M_d3.dta**". Online Appendix Table A.10 results are stored in "**xW1_5_2 _mean_lb _M_d2_svd.dta"** and **"xW1_5_2 _mean_lb _M_d3_svd.dta**". Online Appendix Table A.11 results are stored in "**xW1_5_2 _mean_ub_M_d1.dta", "xW1_5_2 _mean_ub _M_d2.dta",** and **"xW1_5_2 _mean_ub _M_d3.dta".** Online Appendix Table A.12 results are stored in "**xW1_5_2 _mean_ub _M_d2_svd.dta"** and **"xW1_5_2 _mean_ub _M_d3_svd.dta"**.

Step 13

Go to file folder "07_data_summary", where we provide summary statistics tables. Run "**01_agreement_list.do**". Table A.3 result in paper is stored in "**TableA3_agreements_list.dta**". Column (c) of Table A.2 in paper can be found in "**TableA2_c.dta**". Run "02_Statistics_pseudoWorld.do". Table A.1 in paper is stored in "**TableA1_appendix.dta**". Table A.2 results in paper can be found in "**TableA2a_appendix.dta"** and **"TableA2b_appendix.dta**" (except Column c).