1 Folder1: prepare the data and define the parameter space to be evaluated

Step 1 Save locally the Matlab array `probs_allyears_data.mat`. Submit the shell script `cg_s.sh` to a computer cluster by using the command `qsub`. This file runs sequentially the Matlab scripts `main.m` and `f_s.m`. The Matlab script `f_s.m` is parallelised as an SGE array job. In this step, for each year $j = 1940, \ldots, 1967$, we construct the Matlab array `j.mat` which contains:

- `PYX_cond.mat`, reporting the conditional match probabilities for the men side previously stored in `cond_probsmenj_140.mat`;
- `PXY_cond.mat`, reporting the conditional match probabilities for the women side previously stored in `cond_probswomenj+1_140.mat`;
- `U_Gumbel.mat`, `V_Gumbel.mat`, and `Phi_Gumbel.mat`, reporting the estimates of U, V, and Φ, respectively, under the Logit assumption.

The Matlab script `f_s.m` calls the ancillary script `param_Gumbel_closedform.m` to compute the Logit estimates.

Step 2 Run the Matlab script `grid_construction.m` which defines the parameter space to be evaluated by the linear programs. As discussed in Appendix C of the paper, the grid is obtained by spanning hypercubes around `U_Gumbel.mat`, `V_Gumbel.mat`, and `Phi_Gumbel.mat`. In the same appendix, we also explain how we have tried several hypercubes sizes - by varying the parameter `dim` - to ensure that we do not leave unexplored some regions of the parameter space or neglect potential disconnections inside the identified sets. The parameter space is saved into the Matlab array `G_round.mat`.

2 Folder2: compute the estimates

The folder `Folder2` contains the replication codes for Specifications [A] (folder `SpecA`) and [B] (folder `SpecB`). MOSEK should be downloaded and installed before running the codes (`https://www.mosek.com/downloads/`).

2.1 SpecA

Step 1 Open the folder `M`. This folder contains four subfolders:

- `M_1`, which constructs the estimated identified set of $U_{1\cdot}$;
- `M_2`, which constructs the estimated identified set of $U_{2\cdot}$;

- `M_3`, which constructs the estimated identified set of $U_{3\cdot}$;
- `M_4`, which constructs the estimated identified set of $U_{4\cdot}$.

We now explain how to obtain such sets. Open the folder `M_1`. Save locally the Matlab arrays `j.mat` for $j = 1940, \ldots, 1967$ and `G_round.mat`. Submit the shell script `cg_s.sh` to a computer cluster by using the command `qsub`. This file runs sequentially the Matlab scripts `main.m` and `f_s.m`. The Matlab script `f_s.m` is parallelised as an SGE array job. For every year $j = 1940, \ldots, 1967$, each parallel worker *id* takes some rows of `G_round.mat` and saves the rows which belong to the estimated identified set of $U_{1\cdot}$ in the Matlab cell array `IdSetM.id{j}.mat`. The Matlab script `f_s.m` calls the ancillary scripts `changecoord_4.m`, `myind2ind.m`, `Oneside.m`, `pairIndices.m`, and `useful_anyparam.m`, which are used to run the linear programs. Repeat for `M_2`, `M_3`, and `M_4`, possibly in parallel.

Step 2 Open the folder `W`. This folder contains four subfolders:

- `W_1`, which constructs the estimated identified set of $V_{\cdot 1}$;
- `W_2`, which constructs the estimated identified set of $V_{\cdot 2}$;
- `W_3`, which constructs the estimated identified set of $V_{\cdot 3}$;
- `W_4`, which constructs the estimated identified set of $V_{\cdot 4}$.

We now explain how to obtain such sets. Open the folder `W_1`. Save locally the Matlab arrays `j.mat` for $j = 1940, \ldots, 1967$ and `G_round.mat` created in the folder `Folder1`. Submit the shell script `cg_s.sh` to a computer cluster by using the command `qsub`. This file runs sequentially the Matlab scripts `main.m` and `f_s.m`. The Matlab script `f_s.m` is parallelised as an SGE array job. For every year $j = 1940, \ldots, 1967$, each parallel worker *id* takes some rows of `G_round.mat` and saves the rows which belong to the estimated identified set of $V_{\cdot 1}$ in the Matlab cell array `IdSetW.id{j}.mat`. The Matlab script `f_s.m` calls the ancillary scripts `changecoord_4.m`, `myind2ind.m`, `Oneside.m`, `pairIndices.m`, and `useful_anyparam.m`, which are used to run the linear programs. Repeat for `W_2`, `W_3`, and `W_4`, possibly in parallel.

Step 3 Go back to the folder `SpecA`. Run the Matlab script `compose_results_HPC.m` which composes all the files saved in Steps 1 and 2 into the Matlab arrays `IdSetMj_final_x.mat` and `IdSetWj_final_y.mat` for $j = 1940, \ldots, 1967$, $x = 1, \ldots, 4$, and $y = 1, \ldots, 4$. For instance, `IdSetM1941_final_2.mat` is a matrix with 5 columns representing the estimated identified set of $U_{2\cdot}$ in 1941. `IdSetW1967_final_4.mat` is a matrix with 5 columns representing the estimated identified set of $V_{\cdot 4}$ in 1967.

Step 4 Open the folder `output`. Save locally the following Matlab arrays:

- `j.mat` for $j = 1940, \ldots, 1967$;
- `IdSetMj_final_x.mat` and `IdSetWj_final_y.mat` for $j = 1940, \ldots, 1967$, $x = 1, \ldots, 4$, and $y = 1, \ldots, 4$;
- `probs_allyears_data.mat`;

Run the Matlab script `main_core.m` which computes the estimated identified sets of $D(\Phi)$ and stores it in the Matlab array `D_R1.mat`. This array contains the matrices `D_21.mat`, `D_31.mat`, `D_32.mat`, `D_41.mat`, `D_42.mat`, and `D_43.mat`. For instance, `D_21.mat` is a 28×2 matrix, whose j-th row reports the estimated identified set of $D_{22,11}(\Phi)$ in year j, for $j = 1940, \ldots, 1967$.

Run the Matlab script `main_core_MW_contributions.m` which computes the estimated identified sets of the men and women's contribution to $D(\Phi)$ and stores them in the Matlab arrays `D_R1_M.mat` and `D_R1_W.mat`, respectively. `D_R1_M.mat` contains the matrices `D_21_M.mat`, `D_31_M.mat`, `D_32_M.mat`, `D_41_M.mat`, `D_42_M.mat`, and `D_43_M.mat`. For instance, `D_21_M.mat` is a 28×2 matrix, whose j-th row reports the estimated identified set of $U_{22} + U_{11} - U_{21} - U_{12}$ in year j, for $j = 1940, \ldots, 1967$. Similarly, `D_R1_W.mat` contains the matrices `D_21_W.mat`, `D_31_W.mat`, `D_32_W.mat`, `D_41_W.mat`, `D_42_W.mat`, and `D_43_W.mat`. For instance, `D_32_W.mat` is a 28×2 matrix, whose j-th row reports the estimated identified set of $V_{33} + V_{22} - V_{32} - V_{23}$ in year j, for $j = 1940, \ldots, 1967$.

Run the Matlab script `main_marital.m` which computes the estimated identified sets of $C(U)$ and $C(V)$ and stores them in the Matlab arrays `Cm_R1.mat` and `Cw_R1.mat`, respectively. `Cm_R1.mat` contains the matrices `C_21_m.mat`, `C_32_m.mat` `C_43_m.mat`. For instance, `C_21_m.mat` is a 28×2 matrix, whose j-th row reports the estimated identified set of $C_{21}(U)$ in year j, for $j = 1940, \ldots, 1967$. Similarly, `Cw_R1.mat` contains the matrices `C_21_w.mat`, `C_32_w.mat` `C_43_w.mat`. For instance, `C_43_w.mat` is a 28×2 matrix, whose j-th row reports the estimated identified set of $C_{43}(V)$ in year j, for $j = 1940, \ldots, 1967$.

Run the Matlab script `main_Table1.m` which computes:

- the Logit estimates of Φ averaged across cohorts 1940, 1941, and 1942 ("early cohorts");
- the Logit estimates of Φ averaged across cohorts 1965, 1966, and 1967 ("late cohorts");
- the estimated identified set of Φ averaged across early cohorts;
- the estimated identified set of Φ averaged across late cohorts.

The results are stored in the Matlab array `Table1_R1.mat`. This array contains:

- `Phi_Gumbel_early.mat`, which is a 4×4 Matlab array cell whose (xy)-th element reports the Logit estimate of Φ_{xy} averaged across early cohorts, for $x = 1, \ldots, 4$ and $y = 1, \ldots, 4$;
- `Phi_Gumbel_late.mat`, which is a 4×4 Matlab array cell whose (xy)-th element reports the Logit estimate of Φ_{xy} averaged across late cohorts, for $x = 1, \ldots, 4$ and $y = 1, \ldots, 4$;

- `Delta_Gumbel_late.mat`, which is a 4×4 Matlab array cell whose (xy)-th element reports the difference between the Logit estimates of Φ_{xy} for late and early cohorts, for $x = 1, \ldots, 4$ and $y = 1, \ldots, 4$;

- `Phi_early_R1.mat` `Phi_late_R1.mat`, and `Delta_late_R1.mat`, which are 4×4 Matlab array cells reporting the estimated identified sets of the quantities above.

The Matlab script `main_Table1.m` calls the ancillary script `param_Gumbel_closedform_nosingles.m`, which is used to compute the Logit estimates. To compute `Phi_Gumbel_early.mat` and `Phi_Gumbel_late.m` the estimates are weighted by the same weights used by Chiappori, Salanié, and Weiss (2017) to account for different variances across cohorts. Such weights are taken from Chiappori, Salanié, and Weiss (2017) and saved in the Matlab arrays `weightj_4g.mat` for $j = 1, \ldots, 28$.

2.2 SpecB

Repeat the steps outlined in Section 2.1 for the folder `SpecB`. The output files that will be used in the next steps are the Matlab arrays `D_R2.mat`, `D_R2_M.mat`, `D_R2_W.mat`, `Cm_R2.mat`, `Cw_R2.mat`, `Table1_R2.mat`.

3 Folder3: create Figures 3, 4, 5, 6, D.1, D.2, Table 1

Step 1 Save locally the following Matlab arrays:

- `D_R1.mat`, `D_R1_M.mat`, `D_R1_W.mat`, `Cm_R1.mat`, `Cw_R1.mat`, `Table1_R1.mat`;

- `D_R2.mat`, `D_R2_M.mat`, `D_R2_W.mat`, `Cm_R2.mat`, `Cw_R2.mat`, `Table1_R2.mat`

- `probs_allyears_data.mat`.

Step 2 Run the Matlab script `main_core.m` which computes the Logit estimates of $D(\Phi)$ and plots them together with the identified set in Figures 3, 4. The Matlab script `main_core.m` calls the ancillary script `param_Gumbel_closedform.m` to compute the Logit estimates.

Step 3 Run the Matlab scripts `main_marital_m.m` and `main_marital_w.m` which compute the Logit estimates of $C(U)$ and $C(V)$, respectively, and plot them together with the identified sets in Figures 5, 6. The Matlab scripts `main_marital_m.m` and `main_marital_w.m` call the ancillary script `param_Gumbel_closedform.m` to compute the Logit estimates.

Step 4 Run the Matlab scripts `main_core_m.m` and `main_core_w.m` which compute the Logit estimates of the men and women's contribution to $D(\Phi)$, respectively, and plot them together with the identified sets in Figures D.1, D.2. The Matlab scripts `main_core_m.m` and `main_core_w.m` call the ancillary script `param_Gumbel_closedform_MW.m` to compute the Logit estimates.

Step 5 Run the Matlab script `gen_table` which reproduces Table 1. The script calls the ancillary scripts `fill_function.m` and `fill_gunction_Gumbel.m` which help fill the table, and `matrix2latex.m` which creates a LaTex file and can be downloaded from `https://uk.mathworks.com/matlabcentral/fileexchange/4894-matrix2latex?s_tid=mwa_osa_a`. Table 1 in the paper has been slightly reformatted to fit the paper format.