**Replication file**

**“The Design of Teacher Assignment: Theory and Evidence”**

**by Combe, Tercieux and Terrier**

The empirical analysis of the paper is composed of 2 parts which have to be performed in the following order to replicate the analysis:

1. Estimate teachers’ preferences: this is done using files in the folder “Preference estimations” of the replication package.
2. Use the above estimations to perform counterfactual simulations. This is done using files in the folder “Counterfactual” of the replication package.

We describe the details of each part below. The analysis used Matlab (version R2020b) together with the toolboxes: parallel computing, statistics and optimization.

# TEACHER PREFERENCE ESTIMATION

This part presents the code we use to estimate teachers preferences and compute goodness of fit statistics:

* Tables 2, A1, A2, and S4 in the paper report teachers preference estimates.
* Table S1 presents fit measures for subgroups of teachers.
* Table S2 reports out-of-sample fit measures.

**MATLAB FILES**

The folder “Pref estimations” contains four matlab files:

|  |  |  |
| --- | --- | --- |
| 1 | Preference\_estimation.m | Runs the code to estimate teacher preferences |
| 2 | LL.m | Function that computes log-likelihood function |
| 3 | Fit\_measure\_TT.m | Function that computes goodness of fit statistics using estimated preferences under the truth-telling assumption |
| 4 | Fit\_measure\_ST.m | Function that computes goodness of fit statistics using estimated preferences under the stability assumption |

Note: We use three types of goodness of fit statistics in the paper:

1. The main measure is described in the paper and is reported at the bottom of Table 2 on teachers preference estimates. This statistic comes from the files Fit\_measure\_ST.m and Fit\_measure\_TT.m.
2. The second measure is identical to the main fit measure (mentioned in 1- above) but we only compute the statistic on the sample of teachers from Créteil and Versailles. This statistic also comes from the files Fit\_measure\_ST.m and Fit\_measure\_TT.m. It is reported in Table S1. “Fit measures for subgroups of teachers”
3. The third measure in an out-of-sample fit measure (reported in Table S2). This statistic compare the characteristics of the preferred vs. predicted feasible region of each teacher. This statistic is directly coded in the Preference\_estimation.m file under the section “Fit measures”

**INPUT FILES for *Preference\_estimation.m***.

We describe in this section, the input files that Preference\_estimation.m uses. The files that are available in the replication folder are indicated by an Asterix (\*). For the files that are not available in the replication folder, we explain at the end of this document how to apply to get access. In what follows, N refers to the number of teachers.

1. **\*DIS.xls**: [1x8] vector that contains the names of the subjects.
2. **Prio\_decomp\_critere.txt**: [Nx:] matrix that contains, for each teacher (1 row per teacher), the points he/she has for each priority criterion.
3. **\*REGIONS\_CHARACTERISTICS.txt:** [31x:] matrix that contains, for each region (1 row per region), the characteristics of the region, i.e., [the variables below are ordered in the same way as in our online data Appendix “S.2 Variable choice for demand estimation”]:
   * Student social background (Z\_Share\_disadvantaged)
   * Student academic level (Z\_Share\_Brevet)
   * The level of segregation in the region (Z\_Share\_ZEP)
   * Share of students in private schools (Z\_Share\_Private)
   * Teacher experience (Z\_Share\_Teacher\_NoExp)
   * Teacher age (Z\_Share\_Teacher\_YoungerThan30)
   * Whether the region is in the South of France (Z\_SudRegion)
   * Whether the region is urban or rural (Z\_Share\_urb)
   * Region size (Z\_Nb\_teachers)
   * Region unemployment rate (Z\_Share\_Unemp)
   * Cultural activities on offer in the region (Z\_theatres)
4. **TEACHERS\_CHARACTERISTICS.txt:** [Nx:] matrix that contains, for each teacher (1 row per teacher), the demographic characteristics of the teacher, i.e.,:
   * Experience (EXP)
   * Gender (FEMALE)
   * Marital status (MARIED)
   * Number of children (CDT\_ENFANT)
   * Teaching qualification (QUALIF\_AGREG)
   * Whether the current position is in a disadvantaged school (APV)
   * The region of birth (ACA\_NAIS)
5. **Pref\_profs\_inter\_ext.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), the rank order list of the teacher. The file contains the following variables:
   * Column 1: Region ranked first by the teacher
   * Column 2: Region ranked second by the teacher
   * Column 3: Region ranked third by the teacher
   * …
6. **Aca\_origine.txt:** [Nx1] matrix that contains, for each teacher (1 row per teacher), the code of the region of initial assignment.
7. **affectation\_finale\_DGRH.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the code of the region the teacher was assigned under the Ministry’s allocation.
8. **Prio\_aca\_inter\_TB.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), his/her priority in all the regions, i.e., the number of priority points of a teacher. The file contains the following variables:
   * Column 1: Teacher priority in region 1 (Paris)
   * Column 2: Teacher priority in region 2 (Aix-Marseille)
   * Column 3: Teacher priority in region 3 (Besancon)
   * …
9. **Prio\_aca\_origine.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the priority of the teacher in his/her region of initial assignment i.e., the number of priority points of a teacher.
10. **BAR\_CDT\_withTB.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the number of priority points in the initial region that are not teacher-specific, hence omitting the priority points that are teacher-region specific (such as the additional points a teacher has in a region if his/her partner is living in that region). A candidate’s priority incorporates the tie-break rule currently used by the French Ministry of Education: two teachers who have the same number of points are ranked based on their birth date (the older the teacher, the higher the priority).
11. **Dist\_ACA\_NAIS.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), the distance (in kilometers) between each region and the region of birth. The file contains the following variables:
    * Column 1: Distance between region of birth and region 1 (Paris)
    * Column 2: Distance between region of birth and region 2 (Aix-Marseille)
    * Column 3: Distance between region of birth and region 3 (Besancon)
    * …
12. **Dist\_ACAORI\_PSV.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), the distance (in kilometers) between each region and the region the teacher is initially assigned to (if any). The file contains the following variables:
    * Column 1: Distance between initial region and region 1 (Paris)
    * Column 2: Distance between initial region and region 2 (Aix-Marseille)
    * Column 3: Distance between initial region and region 3 (Besancon)
    * …
13. **Barres\_entree.txt**: [31x1] matrix that contains, for each region (1 line per region), the cutoff to enter the region.

**FROM OUTPUT FILES TO THE TABLES OF THE PAPER**

The matlab code Preference\_estimation.m exports Excel files called results\_ST\_dateResults.xls which contain the results for the preference estimations, i.e., for each variable (in row), the associated coefficient, standard error, and p-value. The file also exports two fit measures (the main fit measure for all teachers and for teachers from Créteil and Versailles), the number of observations, the subject ID (eg., “L0202” for French), and whether the estimation uses the stability or truth-telling assumption (“ST” vs. “TT”). The information reported in the Excel file is reported identically in tables 2, A1, A2, and S4 in the paper, except that, in the interest of space, these tables only report a subset of selected coefficients.

# COUNTERFACTUAL SIMULATIONS

Tables 3, 4, 5 and 6 report the results of the counterfactual simulations concerning respectively teacher welfare, fairness, regions’ welfare and administration’s objectives. This first set of simulations is performed on the set of tenured teachers (having an initial position).

Table 7 reports results of the counterfactual simulations performed over the full market of tenured and newly recruited teachers.

Figure A.6 reports, across 5000 draws of estimated preferences, the ratio of the number of teachers asking to enter over those requesting to leave a region with a log scale.

Additional simulation results are mentioned in the Conclusion (Section 7), footnote 85.

The simulations are run for each of the 9 subjects (8 subjects for Table 7) which are represented by their subject code. Results are saved in folders. For Tables 3, 4, 5 and 6, the folder “./Counterfactual results/” contains one folder for each subject + one folder for the aggregated results across subjects (“./Counterfactual results/Slices Aggreg Results/”). The structure is the same for the set of simulations of Table 7 in “./Counterfactual results newcomers/”.

The typical order in which to run the codes is:

1. Run the “run\_counter\_factualsXXX.m” script, outputs (matlab variables + excel of results) are saved in the folder of each subject in the folder “Counterfactual results” (or “Counterfactual results newcomers” depending on the simulation ran)
2. Run the “aggregate\_resultsXXX.m” script which aggregates the results across subjects and outputs the matlab variables and the excel of results in the folder “Slices Aggreg Results” of “Counterfactual results”

We describe below how to construct the tables from the excel outputs of each file.

**MATLAB FILES**

The folder “Counterfactual” contains the following Matlab files:

|  |  |  |
| --- | --- | --- |
| 1.a | run\_counter\_factuals.m | Runs the simulations for the market composed of tenured teachers and exports slices of results and excel results for each subject |
| 1.b | aggregate\_results.m | Aggregates the simulation results of code 1.a across subjects and outputs the main excel file of results (Tables 3, 4, 5 and 6) |
| 2.a | run\_counter\_factuals\_add\_better.m | Runs the simulations mentioned in Footnote 85 of Section 7 where we bootstrap teachers and add them into the market |
| 2.b | aggregate\_results\_add\_better.m | Aggregates the simulation results of code 2.a across subjects and outputs the excel file of results |
| 3.a | run\_counter\_factuals\_delete\_worse.m | Runs the simulations mentioned in Footnote 85 of Section 7 where we iteratively delete teachers being worse-off and run the mechanisms again. |
| 3.b | aggregate\_results\_delete\_worse.m | Aggregates the simulation results of code 3.a across subjects and outputs the excel file of results |
| 4.a | run\_counter\_factuals\_newcomers.m | Runs the simulations for the market composed of both tenured and newly recruited teachers and exports slices of results and excel results for each subject |
| 4.b | aggregate\_results\_newcomers.m | Aggregates the simulation results of code 4.a across subjects and outputs the main excel file of results (Table 7) |
| 5 | test\_attractiveness.m | Runs the excel table used to plot Figure A.6 |
| 6 | nb\_new\_deleted.xlsx | Table which reports, for each field, the number of newcomers deleted from the original dataset during the cleaning. This allows us to delete accordingly the number of empty seats. |

The folder “FUNCTIONS” contains the following Matlab files which are called by the main scripts above:

|  |  |  |
| --- | --- | --- |
| 1.a | be.m | A Block Exchange mechanism with a random selection of cycles |
| 1.b | be\_step.m | A step of the BE mechanism called by the function 1.a above |
| 2 | DAs.m | The DA\* mechanism |
| 3 | iter\_pref2.m | Runs a given batch of simulations for the script run\_counter\_factuals.m |
| 4 | iter\_pref2\_add\_better.m | Runs a given batch of simulations for the script run\_counter\_factuals\_add\_better.m |
| 5 | iter\_pref2\_delete\_worse.m | Runs a given batch of simulations for the script run\_counter\_factuals\_delete\_worse.m |
| 6 | iter\_pref2\_newcomers.m | Runs a given batch of simulations for the script run\_counter\_factuals\_newcomers.m |
| 7.a | one\_sbe.m | A One Sided Block Exchange mechanism with a random selection of cycles |
| 7.b | one\_sbe\_step.m | A step of the 1SBE mechanism called by the function 7.a above |
| 8 | Opti\_teach.m | The IR mechanism maximizing mobility which results are used in Fact 0 (mentioned in footnote 56) |
| 9.a | tobe.m | The Teacher Optimal Block Exchange mechanism |
| 9.b | tobe\_step.m | A step of the Teacher Optimal Block Exchange mechanism called by the function 9.a above |
| 10.a | tobeThenDAs\_newcomers.m | The version of the TOBE mechanism when newly recruited teachers are present |
| 10.b | tobe\_newcomers.m | The mechanism used in the main round of the function 10.a above (the 2nd round uses the function DAs.m above in 2) |
| 10.c | tobe\_step\_newcomers.m | A step of the function used above in 10.b |

Each function in the above table contains a commented headline in the file describing its inputs required and their matlab format. Since these functions’ inputs are automatically built by the main script calling all the functions, we only describe below the data inputs used by the main scripts.

As mentioned, the main scripts are divided into 2 scripts: one running the simulations for each subject and one aggregating the results across subjects.

**INPUT FILES** for ***run\_counter\_factuals.m, run\_counter\_factuals\_add\_better.m, run\_counter\_factuals\_delete\_worse.m, run\_counter\_factuals\_newcomers.m and and test\_attractiveness.m***

We describe in this section, the input files of the “run\_counter\_factualXXX.m” types of codes which have the same structure and almost the same inputs (each script as a commented headline to describe the important steps to run it). The files that are available in the replication folder are indicated by an Asterix (\*). For the files that are not available in the replication folder, we explain at the end of this document how to apply to get access. In what follows, N refers to the number of teachers.

1. **\*DIS.xls**: [1x9] vector that contains the administrative codes of the 9 subjects used in the analysis.

1. \***nb\_new\_deleted.xlsx**: [9x1] vector containing the number of newly tenured teachers deleted for each subject during the pre-processing of the raw data.
2. **\*estim\_LL\_ST\_FE\_SUBJECT\_DATE.mat :** mat file containing all the matlab variables which were obtained as outputs of the preference estimation codes (see the folder “Pref estimations” in the replication package). In particular, the code uses the parameters estimated (LL\_TT\_est). There is one mat file per SUBJECT and the name of this file has a DATE (in the simulations of the paper it is “06-Jan-2020”). The subject’s administrative codes are fixed in a loop using the xls file DIS.xls described above
3. **Prio\_decomp\_critere.txt**: [Nx:] matrix that contains, for each teacher (1 row per teacher), the points he/she has for each priority criterion.
4. **Pref\_profs\_inter\_ext.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), the rank order list of the teacher. The file contains the following variables:
   * Column 1: Region ranked first by the teacher
   * Column 2: Region ranked second by the teacher
   * Column 3: Region ranked third by the teacher
   * …
5. **Aca\_origine.txt:** [Nx1] matrix that contains, for each teacher (1 row per teacher), the code of the region of initial assignment.
6. **affectation\_finale\_DGRH.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the code of the region the teacher was assigned under the Ministry’s allocation.
7. **Prio\_aca\_inter\_TB.txt**: [Nx31] matrix that contains, for each teacher (1 row per teacher), his/her priority points in all the regions The file contains the following variables:
   * Column 1: Teacher priority in region 1 (Paris)
   * Column 2: Teacher priority in region 2 (Aix-Marseille)
   * Column 3: Teacher priority in region 3 (Besancon)
   * …
8. **Prio\_aca\_origine.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the priority points of the teacher in his/her region of initial assignment.
9. **BAR\_CDT\_withTB.txt**: [Nx1] matrix that contains, for each teacher (1 row per teacher), the number of priority points that are not teacher-specific, hence omitting the priority points that are teacher-region specific (such as the additional points a teacher has in a region if his/her partner is living in that region).

**INPUT FILES for *aggregate\_results.m, aggregate\_results\_add\_better.m, aggregate\_results\_delete\_worse.m, aggregate\_results\_newcomers.m***

These scripts use the matlab outputs generated by the previous scripts “run\_counterfactualsXXX.m” described above. In addition, they also import the same data inputs as those described above.

They output an excel file with all the results inside.

**FROM OUTPUT FILES TO THE TABLES OF THE PAPER**

Once the scripts run\_counterfactualsXXX.m and the aggregate\_resultsXXX.m are run. The final output will be an excel file saved in the folder “Slices Aggreg Results” (the original output file of the simulations of the paper is in this folder of the replication package).

We list below, for each script of simulations, the output excel file and where the tables of the paper are:

* aggregate\_results.m => AggregResultsCF\_06-Jan-2020ST\_FE\_.xlsx in /Counterfactual results/Slices Aggreg Results
  + Fact 0 point i) in Section 5.4.1 is found in the sheet “DAMoveIR”
  + Fact 0 point ii) in Section 5.4.2 is found in the column “Opti\_teach”, line 52 of sheet “CumRk”
  + Fact 1 Section 5.4.2 about two-sided maximality of DA\* can be found in sheet “DAs2or1sMax”, The mobility of DA\* and TOBE(DA\*) can be found in sheet “CumRk”
  + Information on the % of cycles of TOBE(DA\*) blocked (last paragraph of Section 5.4.2) can be found in the sheet blockedCy
  + Fact 2 Section 4.2.3 can be found in sheets “testSDRk” and “testsSDrk2” for the stochastic dominance and “CumRk” for the number of teachers moving (line 52) and the number obtaining their first choice (line 2)
  + The numbers in footnote 63 on the teachers worse-off or better-off between DA\* and TO-BE can be found in sheet “worse-better-stats”
  + Fact 3 Section 4.2.3 can be found in sheets “testSDBP” and “testSDBP2” for the stochastic dominance and “CumBP” for the number of teachers not part of a blocking pairs (line 2)
  + Fact 4 Section 4.2.3 can be found in “testSDregions” and testSDregions2” for the stochastic dominance and “welf\_regions” for the number of regions with more than 10% of their seats having a teacher with a net lower priority (line 6)
  + Table 3 on Teacher Welfare is found in sheet “CumRk” for the corresponding algorithms (columns B, C, E, J and K). The table of the excel goes up to rank 25, the one in the paper pools all the values above 7 together (which sum up to the total number of teachers)
  + Table 4 on Fairness is found in sheet “CumBP” for the corresponding algorithms (columns B, C, E, J and K). As before, the ranks above 7 are pooled together in the table of the paper.
  + Table 5 on Regions’ Welfare is found in sheet “welf\_regions”
  + Table 6 on Fullfillment of the administration’s objectives can be found in sheet “admin\_obj”
  + The numbers appearing in paragraph “Teacher participation decisions” of the conclusion (Section 7) about the number of worse-off teachers under TO-BE not moving from their initial position can be found in sheet “worse-better” by comparing line 2 and 6 for column C
* aggregate\_results\_newcomers => AggregResultsCF\_06-Jan-2021ST\_FE\_.xlsx in /Counterfactual results newcomers/Slices Aggreg Results
  + Table 7 is reconstructed using sheet “CumRk” (columns C, D and E) for the mobility of tenured teachers and in sheet “move\_ten” for the movement in the disadvantaged regions (which are regions 24 and 25) and non-disadvantaged ones.
  + The fraction of inexp. Teachers in disadvantaged regions can be found in sheet “admin\_obj”
* aggregate\_results\_add\_better.m => AggregResultsAddBetterCF\_06-Jan-2021ST\_FE\_.xlsx in /Counterfactual results/Slices Aggreg Results
  + These results are mentioned in footnote 86 of the conclusion, we do not report any specific number in the paper. The statement “the gap in terms of teachers’ mobility (…) widens” can be seen in sheet “CumRk”
* aggregate\_results\_delete\_worse.m => AggregResultsDeleteWorseCF\_06-Jan-2021ST\_FE\_.xlsx in /Counterfactual results/Slices Aggreg Results
  + The numbers appearing in paragraph “Teacher participation decisions” of the conclusion (Section 7) about the mobility under or iterative process can be found line 52 of sheet “CumRk”

# DATA SOURCES AND REFERENCES

The datasets used presented above was constructed from several administrative datasets owned by the DEPP (Direction de l’Evaluation de la Prospective et de la Performance) of the French Ministry of Education (Ministère de l’Education Nationale, de l’Enseignement Supérieur, et de la Recherche, MENESR-DEPP).

**The references of the two datasets we use are:**

1. **Teacher-level data on teachers reported preferences**:

Direction de l’Evaluation de la Prospective et de la Performance (DEPP), France, “ Bases sur la mouvement national inter-académique des enseignants, 2013”. Note: As we write this document, there is unfortunately no publicly available documentation for this dataset, and no public link to access it.

1. **Region-level data on regions characteristics**:

Direction de l’Evaluation de la Prospective et de la Performance (DEPP), France, “Géographie de l'Ecole, 2014”, <https://www.data.gouv.fr/fr/datasets/geographie-de-lecole/>

# DATA AVAILABILITY STATEMENT

The data that contains information on teachers transfers is restricted-use microdata that is only available by signing a data access agreement with the Direction de l’Evaluation de la Prospective et de la Performance (DEPP), France. We signed an data access agreement which prevents us from making the data available to third parties.

Researchers interested in using these datasets should contact directly the statistical offices of the MENESR-DEPP:

Ministère de l’Education Nationale, de l’Enseignement Supérieur, et de la Recherche Direction de l’Evaluation de la Prospective et de la Performance (MENESR-DEPP)

61-65, rue Dutot

75231 Paris Cedex 05

FRANCE

[depp\_recherche@education.gouv.fr](mailto:depp_recherche@education.gouv.fr)