Data and Replication Files for:

"Preference Discovery in University Admissions: The Case for Dynamic Multioffer Mechanisms"
The Journal of Political Economy

Julien Grenet, YingHua He and Dorothea Kübler

April 2022

Data Sources

The final datasets used in the paper are constructed from a dataset received from the Stiftung für Hochschulzulassung (SfH), the clearinghouse for university admissions in Germany. This dataset contains confidential information on students who applied to German universities for the winter term 2015/16.

We have signed an agreement with the Stiftung für Hochschulzulassung to use the dataset. This agreement prevents us from making the data available to third parties. The dataset containing the answers to the survey is subject to the same rules. Researchers interested in using the datasets should contact the Stiftung für Hochschulzulassung.

Stiftung für Hochschulzulassung (SfH)
Sonnenstraße 171
44137 Dortmund
GERMANY
Email: poststelle@hochschulstart.de

The final datasets used in the empirical analyses are described below.

Final Datasets

MAIN_DATA.DTA

Description: Main dataset used in the empirical analyses.
Unit of observation: Student (110,781) x program (379).
Number of observations: 318,368.

Variable	Type	Level	Description
stu_id	Integer	Student	Student identifier
stu_female	Boolean	Student	Whether student is a female
stu_age	Integer	Student	Age of student
stu_abitur_pctile	Numeric	Student	Student *Abitur* percentile (between 0 and 1, where a higher value indicates a better *Abitur* grade)
stu_length_init_rol	Integer	Student	Length of initial ROL (on July 15)
stu_reranked_before_phase1	Boolean	Student	Student reranked programs in Application Phase
stu_matched	Boolean	Student	Student was matched to a program
stu_active_acceptance	Boolean	Student	Student actively accepted an offer
stu_exit_phase1	Boolean	Student	Student left procedure in Phase 1
stu_particip_phase2	Boolean	Student	Student participated in Phase 2
stu_cancel_phase1	Boolean	Student	Student canceled applications in Phase 1
stu_date_exit	Date	Student	Date when the student left the procedure
univ_id	Integer	University	University identifier
pgm_id	Integer	Program	Program identifier
pgm_arch_design	Boolean	Program	Program field: architecture and design

MAIN_DATA.DTA (continued)

pgm_nat_sciences	Boolean	Program	Program field: natural sciences
pgm_law	Boolean	Program	Program field: law
pgm_teaching	Boolean	Program	Program field: teaching
pgm_engineer	Boolean	Program	Program field: engineering
pgm_math_cs	Boolean	Program	Program field: mathematics/computer science
pgm_medic_health	Boolean	Program	Program field: medicine and health
pgm_psycho	Boolean	Program	Program field: psychology
pgm_soc_work	Boolean	Program	Program field: social work
pgm_soc_sci	Boolean	Program	Program field: social sciences
pgm_lang_cult	Boolean	Program	Program field: language and culture
pgm_econ	Boolean	Program	Program field: business and economics
pgm_other	Boolean	Program	Program field: other
distance	Numeric	Student x university	Distance between student's municipality and university (km)
same_city	Boolean	Student x university	University is located in the student's city
same_land	Boolean	Student x university	University is located in the student's region (*Land*)
rk_init	Integer	Student x program	Student's initial ranking of program (on July 15)
rk_final	Integer	Student x program	Student's final ranking of program (on August 19)
rk_finalm	Integer	Student x program	Student's final ranking of program (on August 19) using only revealed preferences
rk_feas_init	Integer	Student x program	Student's initial ranking of feasible program (on July 15)
rk_feas_finalm	Integer	Student x program	Student's final ranking of feasible program (on August 19) using only revealed preferences
quota_pctile	Numeric	Student x program	Student's percentile rank among program applicants under the *Abitur* quota (between 0 and 1, with a higher value indicating a better rank)
ratio_cutoff1	Numeric	Student x program	Ratio of student's rank to Phase 1 admission cutoff
ratio_cutoff2	Numeric	Student x program	Ratio of student's rank to Phase 2 admission cutoff
date_feasible	Date	Student x program	Date when the program became feasible to the student
feasible	Boolean	Student x program	Program was ex post feasible to the student
early_feasible	Boolean	Student x program	Program became feasible to the student in Phase 1
first_early_feasible	Boolean	Student x program	Program was the first to become feasible to the student in Phase 1
feasible_added1	Boolean	Student x program	Extended feasibility: one feasible program is added for each student (used in simulations)
n_other_offers	Integer	Student x program	Number of other offers held by student when the program became feasible
date_offer	Date	Student x program	Date when program made offer to student
early_offer	Boolean	Student x program	Program made early offer to student in Phase 1
accepted_pgm	Boolean	Student x program	Student was matched with this program
early_offer_accepted_phase1	Boolean	Student x program	Early offer from this program was accepted by the student in Phase 1
lag_offer_acceptance	Integer	Student x program	Number of days between offer arrival and acceptance for this program
estimation_sample	Boolean	Student	Main estimation sample
extended_sample	Boolean	Student	Extended estimation sample (for robustness checks)
simulation_sample	Boolean	Student	Simulation sample

POTENTIAL_OFFERS.DTA

Description: Potential offers made by programs between July 16 and August 25, 2015.
Unit of observation: Date (39) x program (379).
Number of observations: 1,868.

Variable	Type	Level	Description
date	Date	Date	Date of potential offer
pgm_id	Integer	Program	Program identifier
pgm_selectivity	Numeric	Program	Program selectivity (*Abitur* percentile of last admitted student)
n_potential_offers	Integer	Program	Number of potential offers made by program that day

DOSV_ACTIVITIES.DTA

Description: Activities during the DoSV Procedure (Winter Term of 2015-16).
Unit of observation: Date (from April 15 to December 12, 2015).
Number of observations: 262.

Variable	Type	Level	Description
date	Date	Date	Date
weekday	Integer	Date	Day of the week (from 1: Monday to 7: Sunday)
registrations	Numeric	Date	Cumulative fraction of students who registered with the clearinghouse
finalized_rol	Numeric	Date	Cumulative fraction of students who finalized their rank-order list of programs
first_offers	Numeric	Date	Cumulative fraction of students who received at least one offer
exits_all	Numeric	Date	Cumulative fraction of students who exited the procedure
exits_accept	Numeric	Date	Cumulative fraction of students who actively accept an early offer received during Phase 1
exits_autom	Numeric	Date	Cumulative fraction of students on whose behalf the clearinghouse accepted their best offer during Phase 2
exits_cancel	Numeric	Date	Cumulative fraction of students who cancelled all applications
exits_no_offer	Numeric	Date	Cumulative fraction of students who participated in Phase 2 but received no offer
n_pgm_first_offers	Integer	Date	Number of programs sending their first batch of offers on that date
n_offers	Integer	Date	Number of offers made by programs on that date
n_acceptances	Integer	Date	Number of offers accepted by students on that date

RD_DATA.DTA

Description: Dataset used for regression discontinuity analysis (Appendix C).
Unit of observation: student (19,684) x program quota (208).
Number of observations: 40,216.

Variable	Type	Level	Description
stu_id	Integer	Student	Student identifier
pgm_id	Integer	Program	Program identifier
pgm_quota_id	Integer	Program quota	Program-quota identifier
distance_phase1_cutoff	Integer	Student x program quota	Distance between the student's rank under the program quota and the rank of the last student who received an offer in Phase 1 under that program quota
feasible	Boolean	Student x program	Program was feasible to the student
accepted_pgm	Boolean	Student x program	Student accepted offer from the program

STUDENT_SURVEY.DTA

Description: Student survey conducted online between July 7 and October 10, 2015.
Unit of observation: Survey respondent.
Number of observations: 8,995.

Variable	Type	Level	Description
respondent_id	Integer	Student	Survey respondent identifier
n_applications	Integer	Student	Number of programs to which the student applied
no_clear_ranking	Boolean	Student	Student did not have clear ranking of programs because needed to collect more information
wanted_postpone	Boolean	Student	Getting to a ranking was difficult and student wanted to postpone this decision
received_offer	Boolean	Student	Student received at least one offer
time_spent	Integer	Student	Time spent collecting information on universities that made an offer (1: more; 0: same; −1: less)
changed_ranking	Boolean	Student	Student changed ranking of programs since the start of the procedure
changed_perception	Boolean	Student	Some early offers changed the student's perception of the universities

Replication programs

The Stata .do files and Matlab scripts that are used to produce the results in the main text and the supplementary online Appendix are listed below. Each file generates the information for the table or figure referred to in the filename (e.g., figure_2.do generates Figure 2, xtable_B1.do generates Appendix Table B1, and so on). All files require proprietary data to be executed (see section "Data sources" below).

The local path to the folder containing the replication files is specified in the scripts current_directory.do and current_directory.m.

Figures and Tables (in folder /figures_tables)

Main text

Do file	Input data	Description
figure_2.do	DoSV_activites.dta	Activities during the DoSV Procedure (Winter Term of 2015-16)
figure_3.do	DoSV_activites.dta	Reasons for Exiting the DoSV Procedure (Winter Term of 2015-16)
figure_4.do	main_data.dta potential_offers.dta	Offer Arrival, Program Selectivity, and Program Desirability
figure_6.do	simulation_results.xlsx (generated by script simulations.m)	Comparing the DA, M-DA, and BM-DA Mechanisms: Simulation Results
table_1.do	main_data.dta	Summary Statistics of DoSV Application Data for 2015-16 (Winter Term)
table_2.do	main_data.dta potential_offers.dta	Offer Arrival, Program Selectivity, and Program Desirability: Regression Analyses
table_3.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Conditional Logit
table_4.do	main_data.dta	Initial vs. Final Ranking of Feasible Programs: Rank-Ordered Logit
table_5.do	main_data.dta	Evidence from a Survey on Students in the DoSV for 2015-16

Online Appendix

Do file	Input data	Description
xfigure_B1.do	DoSV_activites.dta	First Batch of Offers Sent Out by Programs
xfigure_B2.do	main_data.dta	Accepted Offer: Cumulative Distribution of Number of Days Elapsed between Offer and Acceptance
xfigure_B3.do	DoSV_activites.dta	Distribution of Early Offers and Acceptances across the Days of the Week
xtable_B1.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Heterogeneity Analysis
xtable_B2.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Robustness to Contracting Students' Feasible Sets
xtable_B3.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Robustness to Expanding Students' Feasible Sets
xtable_B4.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: By Week in which Program Became Feasible
xtable_B5.do	main_data.dta	Acceptance among Feasible Programs and Final ROLs: Controlling for How Students Initially Rank Programs

xtable_B6.do	main_data.dta	Initial vs. Final Ranking of Feasible Programs: Students who Submitted an Initial ROL that they Reranked in the Application Phase
xtable_B7.do	main_data.dta	How Long do Students Wait before Accepting an Offer?
xtable_B8.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Using Flexible Controls for a Program's Ranking of the Student
xtable_B9.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Students who only Applied to Programs Located in their Municipality of Residence
xtable_B10.do	main_data.dta	Early Offer and Acceptance among Feasible Programs: Students who Did not Accept an Early Offer until at least Halfway Through Phase 1
xfigure_C1.do	rd_data.dta	Phase 1 Cutoff: Density Test
xfigure_C2.do	rd_data.dta	Probability of Receiving and Accepting an Early Offer at a Program's Phase 1 Cutoff
xtable_C1.do	rd_data.dta	Impact of (Potential) Early Offers on Acceptance Probability: RD Estimates

Macros (in subfolder /figures_tables)

File	Description
macros.do	Macro functions

Required Stata packages: estout, lpdensity, rddensity, rdrobust, unique.

Simulations (in folder /simulations)

Prepare data for simulations (in subfolder /simulations/do)

File	Input data	Description
data_simulation.do	main_data.dta	Creates the dataset simulation_data.csv that is used to perform the Monte Carlo simulations (Section IV.D)

Main program (in subfolder /simulations/m)

File	Input data	Description
numerical_example.m		Costly Preference Discovery and Early-offer Effect: A Numerical Example (Section IV.C). Results are stored in the /output folder (example.m, table_6.csv, figure_5a.eps, and figure_5b.eps)
simulations.m	simulation_data.csv	Simulation-based comparison of mechanisms (Section IV.D). Results are stored in /output/simulation_results.xlsx.

Matlab Functions (in subfolder /simulations/m/func)

File	Description
f_cond_eu.m	Function for calculating the expected utility conditional on the realization of the first learned X (called by numerical_example.m)
DA_func.m	Implements Gale-Shapley Program-proposing Deferred Acceptance algorithm
feasible_func.m	Determines each student's set of ex-post feasible programs
rank_by_group_func.m	Computes rank order by group
rank_feasible_func.m	Computes relative rank order of program among the student's feasible set
ParforProgressbar.m	To monitor progress of parallel for-loops
progressbar.m	Called by ParforProgressbar.m function
set_path.m	Set paths to the different subfolders