

## ReadMe File for “The Efficiency of Slacking Off: Evidence from the Emergency Department,” *Econometrica*, 2018

This document describes files used in the empirical analysis of the paper. Files in this replication package are organized into the following folders:

- `Data` contains data files
- `Code` contains Stata (.do files) and MATLAB programs
- `ado` contains Stata .ado files
- `Estimates` contains Stata .ster files
- `Output` contains figure outputs created by above files, in .eps format

### Data

Data used in the paper are proprietary and protected by US Health Insurance Portability and Accountability Act (HIPAA) rules. They are proprietary in the sense that they contain personnel data from a large emergency department, including dates of hire which could uniquely identify personnel, and can be used to infer practice patterns and outcomes by physician, nurses, and residents. By agreement, the identity of the large emergency department is not to be disclosed. Data are also protected by HIPAA in two important ways. First, HIPAA prohibits the public disclosure of dates and times in patient care, which are central to the empirical analysis. Second, HIPAA prohibits disclosure of geographic and clinical identifiers that describe fewer than 20,000 possibly identifiable patients.

Interested readers should contact the author at [david.c.chan@stanford.edu](mailto:david.c.chan@stanford.edu) for further information on how to contact health care delivery systems for similar data and how the data should be structured. The following is a complete list of data files produced and used in the analysis:

- `add_u_lntdc_ptadj.dta`:\* Derived dataset produced by `make_table_a-5.1.do`, storing patient-adjusted length of stay utility flow for each visit and in each 5-minute time interval.
- `arrival_count_data.dta`:\* Derived dataset with shift identifiers, patient arrival times in 5-minute intervals, and counts of patient arrivals in each interval. This dataset is created by `prepare_arrival_count_data.do` and is used by `estimate_arrivals_count.do` to estimate the ordinal logit arrival function, with estimates stored as `ologit_*.ster`.
- `cf_sims_data.dta`: File of regression coefficients using simulated data of counterfactual assignment policies under structural model, produced by `prepare_cf_sims_data.do`. Later used in `make_figure_10-11.do`.
- `choice_hour_obs.dta`:\* Temporary dataset created by `schedule_in_intervals.ado` to transform physician schedules into intervals of time, with characteristics of shifts (e.g., overlap at beginning and at end of shift) and of these intervals relative to shift times. Dataset saves computational time if similar command using `schedule_in_intervals.ado` is invoked in a later program.
- `cost_data.dta`:\* Cost data for each ED visit, including both total costs and costs within spending categories.

- `cumhx_xwalk.dta`:\* Crosswalk of cumulative number of hours a given physician has worked with another physician up to a certain date. Created by `gen_cumhx_peer_mod.ado`.
- `dem_vit_data.dta`:\* Patient and demographic characteristics, including race, language, zip code of residence, and date of death (if applicable).
- `estimates_dp.dta`: Stata file saving estimates of dynamic programming model, estimated in MATLAB, stored and retrieved using the program `ext_estimates.ado`. The dynamic programming model can be estimated using code in `make_table_a-5.1.do`.
- `haselix_data.dta`:\* Elixhauser indices for each patient (and corresponding visit) with at least one positive index (out of 29 potential indices).
- `knots_record.dta`: File to save and retrieve knots for cubic splines, used by `make_spline.ado`. The knots are saved as notes and are viewable with the Stata command `notes`.
- `lcmat.txt`: Matrix for linear combinations to generate standard errors by delta method for dynamic programming parameter estimates, generated and used by `make_table_a-5.1.do`.
- `moments_dynamic.dta`: Moments of simulated data from dynamic programming model (produced and used in `make_figure_a-5.2.do`).
- `moments_static.dta`: Moments of simulated data from static model of patient discharges; moments of real discharge data also included (produced and used in `make_figure_a-5.2.do`).
- `noelix_data.dta`:\* List of patients (and corresponding visits) with no positive Elixhauser indices.
- `npid_fe_xwalk.dta`:\* Crosswalk of physician fixed effects in a regression of length of stay, controlling for patient characteristics, time dummies, and location dummies
- `npid_name_xwalk.dta`:\* Crosswalk of physician identifiers with their true names, in various forms in different data sources.
- `npid_xwalk_char.dta`:\* Crosswalk of physician identifiers with their characteristics, such as date of birth, gender, residency program, graduation year, and start year of work at the hospital.
- `nshiftid_sample_xwalk.dta`:\* Crosswalk of 10% random sample of shifts upon which to estimate and simulate dynamic programming structural model, given large size of data. The reduced sample still includes approximately 800,000 patient-interval observations.
- `pan_dta_xwalk.dta`:\* Crosswalk linking patient visit identifiers in the visit data (`sb`) with those in the physician order (`poeid`) and bed location (`visitid`) data. Used in `gen_working_data.ado` to create working dataset.
- `schedule_data.dta`:\* Dataset of physician schedules, including date of work, shift name, location of work, start of shift, and end of shift. Used by `schedule_in_intervals.ado`.
- `sim_int5_d_template.dta`:\* Template of 5-minute intervals corresponding to each physician shift, including characteristics of the shift and physician identities. Used by files that simulate patient arrivals and discharges.
- `stored_data.dta`:\* Main dataset of patient visits, including physician identifiers (`npid`), patient identifiers (`mrn`), visit identifiers (`sb`), resident or physician-assistant identifiers

(`nrid`), nurse identifiers (`nurse`), location identifiers, times of the visit (triage, ED bed, discharge order, discharge, and possible admission to inpatient), and emergency severity index (ESI) score. Produced from raw data by `prepare_stored_data.do`, and used by `gen_working_data.ado`.

- `stored_dp_data.dta`:\* Dataset derived `gen_working_data.ado`, produced by `prepare_stored_dp_data.do`. Saved as a dataset in preparation for any analyses related to dynamic programming estimation and simulation, including estimation of arrival (ordinal logit) and static (conditional logit) discharge models.
- `stored_dp_data_supp.dta`:\* Supplementary dataset used in dynamic programming estimation and simulation, also produced by `prepare_stored_dp_data.do`.
- `sum_poe_data.dta`:\* Dataset of physician orders, identified by `poeid` and used by `gen_working_data.ado`.
- `temp_data_dp.dta`:\* Derived dataset of approximately 800,000 patient-interval observations, with utility and value function splines necessary for dynamic programming estimation. Produced in `make_figure_9.do` using `prep_dp_data.ado`.
- `temp_data_fig_a-1.2.dta`:\* Derived dataset produced by `make_figure_1.2.do`.
- `temp_data_table_a-1.2.dta`:\* Derived dataset produced by `make_table_1.2.do`.

Files with an asterisk are not included in the replication package because they violate confidentiality of proprietary data or are protected under HIPAA rules, as described above.

### Programs in code folder

The replication package includes programs to generate all tables and figures from prepared datasets. All tables and figures can be generated by a master program `master_tables_figures.do`, which calls other programs. Prior to running any program, the reader should inspect and modify the file `setdir.do` to declare correct folders names. If the folder organization is kept as it is in the replication package, then only the path of parent folder in which the replication package resides need be specified.

As in the paper, any of mention of “Alpha” or “Bravo” in the programs does not necessarily imply the real name of locations in the original data.

In order to understand the programs, the following are important identifiers in the prepared datasets:

- Visit identifier: `sb`
- Patient identifier: `mrn`
- Visit date: `visit_td`
- Physician identifier: `npid`
- Resident or physician-assistant identifier: `nrid`
- Nurse identifier: `nurse`
- Physician-resident-nurse trio identifier: `team`

The following are important times:

- Arrival at triage: `trriage_tc`
- Arrival at pod: `ed_tc`
- Discharge order: `dcord_tc`

In addition to `master_tables_figures.do`, the reader may easily find and inspect programs to generate an individual table or figure in the paper or appendix. The names are self-explanatory, and the programs are written so that they can also be run on their own, assuming all the data listed above exist. If not running `master_tables_figures.do`, the reader must first run programs with the `prepare_` prefix, as described below, to generate some intermediate datasets. In many of the table-specific files, relevant results are copied and pasted as comments in the `.do` files. The following is a list of table- or figure-specific Stata programs:

- `make_table_1.do`
- `make_table_2.do`
- `make_table_3.do`
- `make_table_4.do`
- `make_table_a-1.1.do`
- `make_table_a-1.2.do`
- `make_table_a-1.3.do`
- `make_table_a-5.1.do`
- `make_table_a-7.1.do`
- `make_table_a-7.3.do`
- `make_figure_2.do`
- `make_figure_3.do`
- `make_figure_4.do`
- `make_figure_5.do`
- `make_figure_6.do`
- `make_figure_7.do`
- `make_figure_8.do`
- `make_figure_9.do`
- `make_figure_10-11.do`
- `make_figure_a-1.1.do`
- `make_figure_a-1.2.do`
- `make_figure_a-2.1.do`
- `make_figure_a-3.1.do`
- `make_figure_a-5.1.do`
- `make_figure_a-5.2.do`
- `make_figure_a-6.do`
- `make_figure_a-7.1-7.2.do`
- `make_figure_a-7.3.do`
- `make_figure_a-7.4.do`

The following are programs to generate intermediate datasets. Descriptions of the `.dta` files are given above in the `Data` section. Datasets referred to in the code as located in a `{Raw}` folder represent raw datasets that are not included in this package due to confidentiality reasons.

- `prepare_arrival_count_data.do` produces `arrival_count_data.dta`, which is later used in `estimate_arrivals_count.do` and `make_figure_a-5.2.do`.
- `prepare_cf_sims_data.do` produces `cf_sims_data.dta`, which is later used in `make_figure_10-11.do`.
- `prepare_stored_data.do` produces `stored_data.dta`.
- `prepare_stored_dp_data.do` produces `stored_dp_data.dta` and `stored_dp_data_supp.dta`.

Some programs require Stata estimate (`.ster`) files. These are provided in the `Estimates` folder. However, the reader would be able to replicate these files using the following `.do` files:

- `estimate_arrivals_count.do` estimates ordinal logit models of patient arrival counts. These estimates are specific to a “shift type,” which is the triple of shift length (in hours), prior overlap (in hours), and post overlap (in hours). The estimates are stored as `ologit_len`shiftlen'`nshifttype'.ster`, where ``shiftlen'` is a shift length in hours, and ``nshifttype'` is the shift-type identifier.
- `estimate_dc_clogit.do` estimates static multinomial logit models of patient discharges, for each shift length. The estimates are stored as `clogit_len`shiftlen'.ster`.
- `estimate_dp.do` estimates the dynamic programming model, as a shell that calls the MATLAB program `dp_main_fun.m`.
- `estimate_cf_valfn.do` estimates the counterfactual value functions, holding fixed utility parameters previously estimated in `estimate_dp.do`, under counterfactual patient assignment policies. This program is a shell that calls the MATLAB program `dp_main_fun.m`.
- `estimate_xbd_lntdc.do` estimates a linear model of expected length of stay for patient arrivals, for each shift length. The estimates are stored as `xbd_lntdc_len`shiftlen'.ster`.

Finally, maximum likelihood estimation of the dynamic programming model occurs in MATLAB. The main MATLAB program, `dp_main_fun.m`, is called from Stata in `make_table_a-5.1.do`.

- `dp_clogit.m` is called by `dp_main_fun.m` to calculate the conditional (multinomial) logit log likelihood and gradient.
- `dp_confn.m` is called by `dp_main_fun.m` to calculate constraints and constraint gradients.
- `dp_main_fun.m` is the main function to perform maximum likelihood estimation, described in Appendix Section A-5.1. The algorithm follows Barwick and Pathak (2015).
- `dp_nlls.m` is called if the purpose is to estimate value function parameters, holding utility parameters fixed, under counterfactual assignment policies. This is done by non-linear least squares, as in Step 3 in the algorithm in Appendix Section A-5.1.2 and also described in Appendix Section A-5.2.

## Programs in `ado` folder

Stata `.do` programs often draw on `.ado` functions. They are listed below, along with their purposes:

- `binned_scatter.ado` creates binned scatterplots.
- `cat_shift_types.ado*` categorizes shifts into types, based on the location name, time of the day, day of the week, and date.
- `discretize_dta.ado` discretizes data to estimate structural models and simulate counterfactuals in discrete time intervals.
- `ext_estimates.ado` provides a suite of functions to save and retrieve parameter vectors from the structural model as a Stata dataset (`estimates_dp.dta`).
- `gen_census.ado` generates the patient census (i.e., the number of patients being served) for each physician or location.
- `gen_census_dense.ado` is an alternative program to generate the patient census but is tailored for shorter time-intervals, which can cause `gen_census.ado` to be inefficient.
- `gen_cumhx_peer_mod.ado` calculates the cumulative number of hours that two physicians have worked together in the same location in the past.
- `gen_revertimeext.ado` generates variables that describe visit-specific times (e.g., a patient's arrival time) relative to shift-specific times (e.g., the end of shift time).
- `gen_var.ado` is a core program to generate or import variables of interest in the analytical dataset. Variables include demographics, vital statistics, costs, outcomes, Elixhauser indices, and ED location dummies.
- `gen_working_data.ado` is a core program to generate the analytical dataset for almost all analyses.
- `make_logit_dta.ado` generates data necessary for estimation of discrete choice discharge model. This program transforms the data in memory from visit-level format to visit-time-interval format, creates an outside option observation of "no discharge" for each time interval, and includes other variables relevant to the discharge model.
- `make_spline.ado` creates cubic splines using Stata's `mk spline` command, while adding the functionality of saving the knots or retrieving the knots (if the command has been run previously for a given variable and model). The knots are stored as Stata notes in `knots_record.dta`.
- `merge_np_id_char.ado` merges in physician characteristics, stored in `np_id_xwalk_char.dta`.
- `name_pods.ado*` merges in location identifiers, based on the name of a physician's shift and the date of a visit.
- `prep_dp_data.ado` prepares the data in working memory, in visit-time-interval format, for dynamic programming estimation, specifically adding utility-relevant characteristics and expectations of the future state in characteristic-space. This program can be used on simulated data as well as on actual data. The default is to use actual data, which is based on `stored_dp_data.dta` and `nshiftid_sample_xwalk.dta`.
- `prep_est_clogit.ado` prepares the data for multinomial logit estimation of discrete choice discharge model. It relies on `make_logit_dta.ado` and `make_spline.ado`.
- `proc_dp_varlist.ado` processes data, with specified variables (and interactions of them) related to utility flow or expected future-state characteristic, to have standard variable names in a dataset that can then be exported to MATLAB for estimation of

dynamic programming model. Relevant Stata and MATLAB variable names are saved as macros, along with a macro for the name of the model specification. When estimates are returned and processed with `ext_estimates.ado`, this program is also called to record the Stata variables (or their interactions) that correspond to parameter estimates.

- `run_sim.ado` is a main program that runs simulations. Patient discharges are simulated depending on static models or dynamic models of discharges. In its most basic functionality, it simulates patient discharges over a specified time range relative to end of shift in each shift. It has additional capability to simulate arrivals.
- `schedule_in_intervals.ado` loads scheduling data and transforms the data to a time-interval format.
- `set_global_knots.ado` is a trivial program to record and use knots to make splines, where the knots are passed to memory as global macros.
- `sim_arrivals_multi.ado` is called by `run_sim.ado` to simulate arrivals.
- `sim_xbd_lntdc_multi.ado` is called by `run_sim.ado` to simulate the expected length of stay of (simulated) arrivals.
- `sum_sim_data.ado` summarizes simulated (and potentially real) data as statistics according to patient arrivals, length of stay, patient-adjusted length of stay, and census, at each time interval relative to end of shift.

Files with asterisks are not included in the replication package because they contain information that identifies the institution or individual physicians within the institution.

## Estimates

Stata estimates (`.ster` files) are stored in the `Estimates` folder. Four types of estimates are included:

- `clogit_len`shiftlen'.ster` are estimates for conditional (multinomial) logit models of the probability of patient discharge for a given visit and time interval. These models are estimated separately for observations belonging to shifts of different lengths, denoted by ``shiftlen'`. The estimates are produced in `estimate_dc_clogit.do`.
- `ologit_len`shiftlen'`shifttype'.ster` are estimates for ordinal logit models of the count of patient arrivals on a given shift and time interval. These models are estimated separately for observations belonging to shifts of different lengths (``shiftlen'`) and types (``shifttype'`). The estimates are produced in `estimate_arrivals_count.do`.
- `xbd_lntdc_len`shiftlen'.ster` are estimates for linear models of patient length of stay depending on the time of day, day of week, month-year, physician identity, and location in the ED. These models are estimated separately for observations belonging to shifts of different lengths (``shiftlen'`). The estimates are produced in `estimate_xbd_lntdc.do`.

## Output

The following list of figure `.eps` files are contained in this folder:

- Figure\_2.eps
- Figure\_3.eps
- Figure\_4.eps
- Figure\_5.eps
- Figure\_6.eps
- Figure\_7.eps
- Figure\_8.eps
- Figure\_9.eps
- Figure\_10.eps
- Figure\_11.eps
- Figure\_A-1.1.eps
- Figure\_A-1.2.eps
- Figure\_A-2.1.eps
- Figure\_A-3.1.eps
- Figure\_A-5.1.eps
- Figure\_A-5.2.eps
- Figure\_A-6.1.eps
- Figure\_A-6.2.eps
- Figure\_A-7.1.eps
- Figure\_A-7.2.eps
- Figure\_A-7.3.eps
- Figure\_A-7.4.eps

Each file is generated by its respective .do file in the Code folder.