Read me file for *Where do Students Go When For-Profit Colleges Lose Federal Aid?*  
Stephanie R. Cellini, Rajeev Darolia, and Lesley J. Turner

Summary

Included are (1) programs used to create the analysis data sets (in “create\_analysis\_datasets” folder), (2) programs used to create the tables and figures for the paper and appendices (in “create\_tables\_figures” folder), and (3) publicly available data and intermediate data sets used in the creation of the analysis data sets, excluding Pell Grant recipient data. Data on Pell Grant recipients and disbursements by institution for academic years 2000 – present is available at <https://www2.ed.gov/finaid/prof/resources/data/pell-institution.html>. Data for earlier years can be requested from the Department of Education using the contact information listed on the above website.

Description of programs and data

For a detailed description of the data sets and construction of the analysis data set, please see the included Online Appendix A.

Programs:

1. Data creation programs:
   * create\_analysis\_data.do: calls several programs that clean and format raw data on Pell Grant recipient enrollment, cohort default rates, sanctions, and school-level information to create an analysis data set with one observation per postsecondary institution-academic year.
   * stack\_data\_rev.do: creates main analysis data set with one observation per institution-academic year-sanction year.
   * stack\_data\_radius15.do and stack\_data\_radius30.do: create supplemental analysis data sets using alternate definitions of local markets and competitors (schools within a 15-mile radius and schools within a 30-mile radius, respectively). Distances between institutions was calculated using ArcGIS.
2. Figure and table creation programs: program names correspond to tables/figures for which output is created.

Data:

1. Original data:
   * Cohort default rate (CDR) data: for cohorts entering repayment between 1992 and 2009. Default rates are calculated for the cohort that has been in repayment for two fiscal years (e.g., the default rate reported in 1992 applies to the cohort of borrowers that entered repayment in FY 1990).There is one original file for every year, which contains information on the institution (name, address, type) and the default rates for the three preceding years. We use the most recent information on default rates. These data are available in the “2-Year Press Packages” links at <https://ifap.ed.gov/DefaultManagement/press/>. Also included in this folder is data on sanctions due to CDR violations. for the academic years 1991 through 2010. This data includes the sanction trigger and penalty by institution and cohort year.
   * Postsecondary Education Participants System data: Department of Education administrative data. Includes information about an institution’s location, sector, participation in Title IV programs, closure date (if participating in Title IV programs at the time of closure) and provides a crosswalk between earlier institution identifiers (“Pell IDs”) and modern institution identifiers (“OPEIDs”). These data cover all institutions that ever participated in Title IV. We use the June 28, 2013 PEPS extract. The most recent PEPS data is available at: <https://www2.ed.gov/offices/OSFAP/PEPS/dataextracts.html>.
   * FOIA request data: this folder contains the results from two Freedom of Information Act requests regarding information on sanction appeals (supplied) and cohort default rates for earlier cohorts (not supplied). Appeals data in .csv files are also included in the corresponding folder.
   * Zip code data: contains several crosswalks between zip codes (which were generally available in the institution-level data we received) and county FIPS codes (which were used as a proxy for local higher education markets in our main analyses).
   * Postsecondary Career School Survey (PCSS) data: obtained from ICPSR (<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2382>). Used to supplement the main data sets to obtain information on location and sector when otherwise missing.
2. Intermediate data sets:
   * Data on for-profit chains (chain\_ids.dta): Unfortunately, there is not a reliable indicator for whether an institution belongs to a chain during the period we examine. We assume that two or more institutions belong to a chain if they have the same name, allowing for variation in the spelling of common words (e.g., Sch versus School, Tech versus Technology) and taking into account the fact that school names that include many very common words (e.g., American Beauty College) will not represent two locations of the same chain unless other criteria are met (e.g., all instances are within the same state or group of adjoined states).
   * Imputed Pell Grant recipients and student loan borrowers for systems of institutions that switched between reporting these at the institution-level and at the system-level over the years that our analyses focus on. This primarily occurs in the public sector (e.g., the Indiana Ivy Tech community college system, which includes 14 separate Title IV institutions, reports all recipients under one OPEID before 2006), but there are two large for-profit chains that suffer from this issue (ITT Tech and DeVry). To deal with the fact that this variation in reporting leads to large fluctuations in enrollment that are unrelated to actual enrollment changes, we impute recipients for locations in years in which Title IV enrollment is not separately reported. We use the most recent year in which recipients are reported for specific institutions and calculate the share of all recipients contained within each institution relative to the total number of recipients summed over the set of schools that are reported under a single OPEID in other years. Then, in years in which recipients are not reported for separate institutions, we allocate the total number of recipients reported under the single OPEID based on these baseline shares. At most, less than 3 percent of schools have imputed recipients in a given year.
   * Information on an institution’s local competitors within 15- and 30-mile radii.