Research Protocol, Pension Investment Project

Final Version [April, 2013]

**Phase I — Download Data** (from <http://www.dol.gov/ebsa/foia/foia-5500.html>, the posted raw data, and <http://www.dol.gov/ebsa/publications/form5500dataresearch.html>, the Pension Research Files edited by EBSA)

General overview: Pension and welfare plans must file a Form 5500 annually. Various kinds of direct filing entities (DFEs) must also file this form. Specific types of DFEs include master trust investment accounts (MTIAs), common/collective trusts (CCTs), pooled separate accounts (PSAs) and 103-12 investment entities (103-12 IEs). The Form 5500 contains information about the characteristics of the filing entity. Schedule D, part 1 contains a list of DFEs in which a filing entity invests including the amount. Schedule H contains a breakdown of the filing entity’s assets by category. The raw data for all of these filings is available online in .csv files from 1999 to 2008 on the freedom of information pages of the EBSA website. Pension Plan Research Files are available with the information from all these schedules in an edited form with some helpful variables added, but only for pension plans, not welfare plans or DFEs.

**Step 1** - Download six (6) files for applicable year:

* + Form 5500 (filename: f\_5500\_*year*.csv)
  + Schedule D (filename: F\_SCH\_D\_*year*.csv)
  + Schedule D Part 1 (filename: F\_SCH\_D\_PART1\_*year*.csv)
  + Schedule D Part 2 (filename: F\_SCH\_D\_PART2\_*year*.csv)
  + Schedule H (filename: F\_SCH\_H\_*year*.csv)
  + Pension Research File
    1. Download the SAS file (filename: bulyr.sas7bdat)
    2. Use Stat Transfer to convert SAS file to a .csv file named form5500.year.clean.csv.

**Phase II — Organize Data**

General overview: The datasets are quite large and contain considerable information that is not needed, for example, variables we are not interested in and multiple filings. Furthermore, the existing raw datasets do not all contain clear designations of whether an entity is a plan or DFE (and which type of each) so dummy variables need to be created. These dummies then make it simple to pull out the relevant data. (Since the research files contain only plans, this only applies to the raw data files.)

**Step 2** — Create objects[[1]](#footnote-1) “dfe.form5500” and “welfare.form5500” using the raw data files

* Load the file “f\_5500\_year.csv” (rename as form5500) and strip out unnecessary columns (keep columns 1:11, 24, 35:37, 53:68, 81) [see 2008 Form 5500 File Layout, at <http://askebsa.dol.gov/FOIA%20Files/2008/F_5500_2008_layout.txt>]
* Create variables to reflect whether a filing entity is a plan or a DFE[[2]](#footnote-2)
  + “plan” = 1, if TYPE\_PLAN\_ENTITY\_IND is 1 or 2 or 3
    - [Source: Form 5500 Part I, item A, 2008 Form 5500 File Layout column 6.]
  + “DFE” = 1, if TYPE\_PLAN\_ENTITY\_IND is 4
    - [Source: same.]
* Create variables to reflect whether a plan is a pension plan and/or a welfare plan
  + “pension” = 1, if there is an entry in the variable PENSION\_BENEFIT\_PLAN\_IND
    - [Source: Form 5500 Part II, line 8a, 2008 Form 5500 File Layout column 63.]
  + “welfare” = 1, if there is an entry in the variable WELFARE\_BENEFIT\_PLAN\_IND
    - [Source: Form 5500 Part II, lone 8b, 2008 Form 5500 File Layout column 65.]
  + “penwel” = 1, if pension=1 and welfare=1[[3]](#footnote-3)
* Create variable named “dfe.type” to reflect what type of DFE (if any) a filing entity is[[4]](#footnote-4)
  + “dfe.type” = “mtia”, if TYPE\_DFE\_PLAN\_ENTITY is “M”
  + “dfe.type” = “cct”, if TYPE\_DFE\_PLAN\_ENTITY is “C”
  + “dfe.type” = “psa”, if TYPE\_DFE\_PLAN\_ENTITY is “P”
  + “dfe.type” = “103-12”, if TYPE\_DFE\_PLAN\_ENTITY is “E”
    - [Source (all four types): Form 5500 Part I, entry following item A(4), 2008 Form 5500 File Layout column 7.]
* Create a variable named “plan.maturity” that is equal to TOT\_ACTIVE\_PARTCP\_CNT divided by
  + - [Source: Form 5500 Part II, line 7a, 2008 Form 5500 File Layout column 54.]

TOT\_ACT\_RTD\_SEP\_BENEF\_CNT.

* + - [Source: Form 5500 Part II, line 7f, 2008 Form 5500 File Layout column 59.]
* Create a variable to reflect whether a return is amended or not
  + “amended.return” = 1, if there is a “2” in the variable

TYPE\_PLAN \_FILING \_IND

* + - [Source: Form 5500 Part I, item B2, 2008 Form 5500 File Layout column 8.]
* Create a variable “EINPN” which is a concatenation of the filing entity’s EIN and plan number (SPONS\_DFE\_EIN and SPONS\_DFE\_PN).
  + - [Source: Form 5500 Part II, lines 2b and 1b (respectively), 2008 Form 5500 File Layout columns 2 & 3.]
* Create an object named “dfe.form5500” with only DFE filings.
* Create a variable “last.filing” which equals 1 for the filing of each EINPN which has the largest FILING\_ID (and is presumably filed last) and zero otherwise.[[5]](#footnote-5)
* Drop multiple filings by the same DFE from dfe.form5500 (keep only those rows for which last.filing=1.
* Create an object named “welfare.form5500” with only welfare plans (that are not also pension plans).[[6]](#footnote-6)

**Step 3** — Create object “pension.form5500” using the research files [For the source of the variables referred to in Step 3, and in the edited Pension Research File generally, consult Actuarial Research Corporation, User Guide to 2008 Form 5500 Private Pension Plan Research File, Contract DOLJ089327412 (Dec. 2010), pp. 23-33 (codebook and list of plan characteristics codes), at <http://www.dol.gov/ebsa/pdf/2008-5500-researchfileuserguide.pdf>.]

* + Load the research file named “form5500.year.clean.csv”
  + Strip out unnecessary columns.
  + Extract only rows where the BEST\_FOR\_PLAN variable equals 1.
  + Extract only large pension plans (where the LARGE variable equals 1).
  + Drop rows with no entry in the FILING\_ID variable. (Since this is how plans are identified, the absence of a filing id makes it impossible to use these data.)[[7]](#footnote-7)
  + Create a variable “add.filing” to designate when a filing is an additional filing of an entity (as indicated by the addition of a letter to the end of the OPR\_PN).[[8]](#footnote-8) Presumably a filing which is both the best for a plan and an additional filing (i.e., not the first filing by the plan) is an amended return.
  + Modify the existing variable “EINPN” by reformatting the OPR\_EIN and OPR\_PN variables and concatenating them. (This is necessary because the EINPN variable in the dataset originally has formatting problems that result in the numbers being incorrect.) Note: the letters appended to the OPR\_PN for additional filings are not included for purposes of creating the EINPN.
  + Create a variable “last.filing” which equals 1 for the filing of each EINPN which has the largest FILING\_ID (and is presumably filed last) and zero otherwise.[[9]](#footnote-9)
  + Drop multiple filings by the same pension plan from pension.form5500 (keep only those rows for which last.filing=1.[[10]](#footnote-10)
  + Create a variable named “plan.maturity” that is equal to TOT\_ACTIVE\_PARTCP\_CNT divided by TOT\_ACT\_RTD\_SEP\_BENEF\_CNT.

**Step 4** — Create object “D” using the raw data files

* Strip out unnecessary columns (keep columns 1:3) from schedule D file.
  + [Source: Form 5500 Schedule D, items D and B, 2008 Schedule D File Layout columns 1-3, at <http://askebsa.dol.gov/FOIA%20Files/2008/F_SCH_D_2008_layout.txt>.]

**Step 5** — Create object “D1” using the raw data files

* Strip out unnecessary columns (keep columns 1, 7:11) from schedule D1 file.
  + [Source: Form 5500 Schedule D, Part I, items (a)-(e), 2008 Schedule D Part 1 File Layout, at <http://askebsa.dol.gov/FOIA%20Files/2008/F_SCH_D_PART1_2008_layout.txt>.]
* Merge with “D” to add identifying information (EIN and PN)

**Step 6** — Create object “D2” using the raw data files

* Strip out unnecessary columns (keep columns 1, 7:10) from schedule D2 file.
  + [Source: Form 5500 Schedule D, Part II, items (a)-(c), 2008 Schedule D Part 2 File Layout, at <http://askebsa.dol.gov/FOIA%20Files/2008/F_SCH_D_PART2_2008_layout.txt>.]
* Merge with “D” to add identifying information

**Step 7** — Create object “H” using the raw data files

* Strip out unnecessary columns (keep columns 1:5, 22, 30, 37:67) from schedule H file.
  + [Source: Form 5500 Schedule H, Part I, 2008 Schedule H File Layout, at <http://askebsa.dol.gov/FOIA%20Files/2008/F_SCH_H_2008_layout.txt>.]

**Step 8** — Create object “pension.form5500l” by adding DFE link info from Schedule D to pension.form5500.

* Merge D1 with pension.form5500 to add the variables DFE\_P1\_PLAN\_01\_EIN\_PN and DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT.[[11]](#footnote-11)
  + [Source: Form 5500 Schedule D, Part I, items (c) & (e), 2008 Schedule D Part 1 File Layout columns 9 &11.]

**Step 8a** – Create object “dfe.form5500l” by adding info from Schedule D2 to dfe.form5500 in order to create a variable for how many plans each DFE reports as investors.

* Merge D2 with dfe.form5500 to create “dfe.form5500l.D2”
  + [Source: Form 5500 Schedule D, Part 2]
* Drop all non-unique combination of the DFE’s EINPN, the plan’s EINPN, and the plan’s plan number. What remains is one row for each plan that invests in each DFE.
* Using “form5500”, create a list of all unique EINPN’s of pension plans which filed a Form5500.
* Using “pension5500”, create a list of all unique EINPN’s of large pension plans which have a filing in the cleaned database of Form5500 data.
* In the dataset “dfe.form5500l.D2” create a variable named “pension.plan” which equals one if the plan investing in the DFE is a pension plan (and zero otherwise).
* Create a similar variable named “lpension.plan” which equals one if the plan investing in the DFE is a large pension plan (and zero otherwise).
* For every row where “lpension.plan” equals one, make sure that “pension.plan” equals one as well.
* Create a variable named “num.D2.plans” which sums up the number of rows for each unique DFE. In other words, this is the total number of plans each DFE identifies in their Schedule D2 as investing in them.
* Create a variable named “num.D2.pen.plans” which sums up the number of rows for each unique DFE where “pension.plan” equals one. In other words, this is the total number of pension plans each DFE identifies in their Schedule D2 as investing in them.
* Create a variable named “num.D2.lpen.plans” which sums up the number of rows for each unique DFE where “lpension.plan” equals one. In other words, this is the total number of large pension plans each DFE identifies in their Schedule D2 as investing in them.
* Merge these three new variables into the “dfe.form5500” dataset and rename it “dfe.form5500l”.

Note: At the end of phase II the data for each year which is needed for the project is located in the following set of objects:

* dfe.form5500l
* welfare.form5500
* D1
* D2
* H
* pension.form5500l

**Phase III — Reconfigure Data**

General overview: Now that the datasets contain only the pertinent information, the data they contain must be manipulated to permit the calculation of the breakdown of each pension plan’s assets held indirectly in a direct filing entity (“DFE”). The key to this process is linking data between plans and DFEs. Plans and DFEs are uniquely identified by a 12-digit combination of their EIN and plan number (PN). (The IRS itself uses this 12-digit combination to uniquely identify filing entities.) The pension plan is ultimately the unit of analysis with relevant information from the DFE incorporated to permit the calculation of a plan’s second level assets. Since a plan can invest in multiple DFEs, the list of unique plan-DFE pairings can contain the same plan multiple times. The data must be collapsed so that each plan appears only once with the sum of its second level holding in each asset and liability category. Once these second level holdings are calculated it is, of course, a simple matter to calculate the total holding in each asset and liability category.

**Step 9** — Create a summary of DFE assets named “dfe.assets.year.csv”

* Merge H with dfe.form5500l to add schedule H information and create an object named dfe.assets.[[12]](#footnote-12)
* Rename “FILING\_ID” to “FILING\_ID.dfe”, and rename “EINPN”, to “EINPN.dfe”.
* Drop variables which are not needed (plan, DFE, pension, welfare, penwel, and plan.maturity).
* Create a variable “sum.assets” which is the total of all the reported assets in each category and the analogous variable for “sum.liab”.
* Create a variable for the difference between the reported total assets and liabilities and the calculated total assets and liabilities.
  + slop.assets = TOT\_ASSETS\_EOY\_AMT – sum.assets[[13]](#footnote-13)
    - [Source: Form 5500 Schedule H, Part I, line 1f col. (b), 2008 Schedule H File Layout column 61.]
  + slop.liab = TOT\_LIABILITIES\_EOY\_AMT – sum.liab[[14]](#footnote-14)
    - [Source: Form 5500 Schedule H, Part I, line 1k col. (b), 2008 Schedule H File Layout column 66.]
* Drop DFEs which have zero total assets (according to ether the reported or calculated total).[[15]](#footnote-15)
* Drop DFEs when the total assets minus total liabilities does not match the reported NET\_ASSETS\_EOY\_AMT and is negative.[[16]](#footnote-16)
  + [Source: Form 5500 Schedule H, Part I, line 1l col. (b), 2008 Schedule H File Layout column 67.]
* Write dfe.assets to a .csv file.

**Step 10** — Use D1 to link DFEs to pension plans, create object D1.link.a

* Extract pension plan filings from pension.form5500l which report a non-zero interest in a DFE.[[17]](#footnote-17)
* Create an object “pension.dfe.links” which contains only five variables from pension.form5500l which pertain to DFE links: FILING\_ID, EINPN, DFE\_P1\_ENTITY\_01\_NAME, DFE\_P1\_PLAN\_01\_EIN\_PN, and DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT.
* Rename “FILING\_ID” to “FILING\_ID.pension”, and rename “EINPN”, to “EINPN.pension”.
* Merge dfe.assets with pension.dfe.links using EINPN.dfe for the former and DFE\_P1\_PLAN\_ 01\_EIN\_PN for the latter. Keep all pension-DFE links reported by pension plans even if a link cannot be made to the filing of the relevant DFE.
* Create a variable named “dfe.type” which is based on the plan’s Schedule D: “DFE\_P1\_ENTITY\_01\_CODE”. Using the data from the plan’s filing enables categorizing the DFE even when a link is not successful or not possible.
* Create a variable called “trip.zero” which equals 1 if the DFE’s plan number is “000” and zero otherwise. The triple zero plan number indicates that a CCT or PSA reported on Schedule D, Part I did not file a Form 5500 for the year and therefore is not a DFE, so there will not be a plan-DFE link.[[18]](#footnote-18) [See 2008 Instructions to Form 5500, pp. 12, 25, at <http://www.dol.gov/ebsa/pdf/2008-5500inst.pdf>.]
* All triple zero DFEs should be CCTs or PSAs. Recode the small handful that are mistakenly reported as other types of DFEs as “unknown.”
* Create a variable called “trip.zero.cct” which equals 1 if the DFE a plan reports investing in is a triple zero CCT and zero otherwise. Also create “trip.zero.psa”, and “trip.zero.unknown.”
* Create a variable called “bad.link” which equals 1 if the pension plan’s reported investment in a DFE should be able to be linked, but cannot.[[19]](#footnote-19)

**Step 11** — Use Schedule D2 data to fix bad links where possible. (There is still a separate row for each link.)

* Calculate the total number of bad links for each unique EINPN.pension.
* Merge this total into the D1.link.a data and add the variable named “num.bad.link”.
* Extract only the pension plans which have exactly one bad link.[[20]](#footnote-20) (If there is more than one bad link it is not possible to tell which schedule D2 entries are linked to which schedule D1 entries.)
* Create a variable named “D1.good.link” which contains the concatenation of the EINPN of the pension plan and DFE for all initially successful links.
* Load the D2 data and create a variable named “D2.links” which contains the concatenation of the EINPN of the DFE and all the pension plans which it reports invest in them (but with the pension plan’s EINPN listed first).
* If a D2.link matches an entry from the list of D1.good.links, drop it from the D2 data. (Since there is a good link from the D1 data already, that link has been accounted for, so it is dropped which will prevent it from erroneously being used to fix a bad link from the D1 data.)
* Using the remaining D2 data, extract only the schedule D2 filings from DFEs which report that a pension plan with one, and only one, bad link invests in them.
* Calculate the total number of schedule D2 filings for each unique EINPN.pension.
* Extract only the schedule D2 links for which exactly one DFE reports being invested in by a pension plan and that pension plan has exactly one bad link.[[21]](#footnote-21)
* Replace the EINPN.dfe from the bad link with the EINPN.dfe from relevant schedule D2 link.
* Rerun Step 10 with slight modifications as follows:
  + Create an object “pension.dfe.links2” which contains only five variables: FILING\_ID.pension, EINPN.pension, DFE\_P1\_ENTITY\_01\_NAME, EINPN.dfe, and DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT.
  + Merge dfe.assets with pension.dfe.links2 using EINPN.dfe.
  + Create a variable named “dfe.type” which is based on the plan’s Schedule D: “DFE\_P1\_ENTITY\_01\_CODE”.
  + Create a variable called “trip.zero” which equals 1 if the DFE’s plan number is “000” and zero otherwise.[[22]](#footnote-22)
  + All triple zero DFEs should be CCTs or PSAs. Recode the small handful that are mistakenly reported as other types of DFEs as “unknown.”
  + Create a variable called “trip.zero.cct” which equals 1 if the DFE a plan reports investing in is a triple zero CCT and zero otherwise. Also create “trip.zero.psa”, and “trip.zero.unknown.”
  + Create a variable called “bad.link” which equals 1 if the pension plan’s reported investment in a DFE should be able to be linked, but cannot.[[23]](#footnote-23)
* Save as D1.link.a2.

**Step 12** — Calculate proportion of DFE holdings contributed by each plan investing in a DFE, create object D1.link.b

* Create a variable named “net.assets” which is the calculated net assets of the DFE and is calculated using the calculated “sum.assets” and “sum.liab” variables created in Step 9. “net.assets” = “sum.assets” – “sum.liab”.
* For each pension-DFE link create the variable “pension.portion” which is the amount a pension plan reports investing in a DFE (DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT) divided by the net assets of the linked DFE (NET\_ASSETS\_EOY\_AMT).[[24]](#footnote-24)
* Create a variable named trip.zero.cct.amt which is the dollar amount invested in a CCT which does not file a form5500. Do the same for PSAs and unknown DFEs.
* For the pension-DFE links where the amount reported by the plan is greater than the net assets of the DFE, create a variable “unmatched.overflow” with the difference between the two.
* For the pension-DFE links where the amount reported by the plan is greater than the net assets of the DFE, make the pension.portion variable equal 1.
* For pension-DFE links that are bad links, put the dollar amount invested in a DFE in the variable “unmatched.overflow”.

**Step 13** — Create variables with second-level assets and second-level liabilities for each pension plan for each pension-DFE link. (There is still a separate row for each link whether it is a triple zero DFE or a bad link or a good link.)

* Create a variable for each category called SECLEV\_category which is the product of the portion of the DFE contributed by the pension plan and the amount the DFE reports in the category (i.e., pension.portion \* category).
* Delete all the category variables since they are no longer necessary. (This will prevent later confusing them with pension plan first level holdings in each category.) Also drop categories with the form5500 characteristics of the DFE (except dfe.type) so they are not confused with pension plan characteristics.
* If a pension-DFE link is successful, but the DFE type cannot be identified, set the variable dfe.type to “unknown”.[[25]](#footnote-25)
* Save as D1.link.

**Step 14** — Create an object named “D1.link.agg2” which contains one row for each unique EINPN.pension[[26]](#footnote-26) with a sum of each second level asset and liability category as well as the following variables:

* “mtia.total.amt” — the total dollar amount a pension plan said on the Schedule D (DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT) they invested in a MTIA (good, bad, and triple zero (for CCT and PSA) links). Also create analogous categories for CCT, PSA, 103-12 IE, and unknown DFE type investments.
* “linked.mtia.amt” — the total dollar amount a pension plan said on the Schedule D (DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT) they invested in an MTIA which has been successfully linked. Also create analogous categories for CCT, PSA, 103-12 IE, and unknown DFE type investments.
* “D1.amt.agg” — The sum of all amounts a pension plan reports in schedule D1 filings (whether successfully linked to a DFE or not).
* “total.trip.zero.cct.amt” — The sum of the “trip.zero.cct.amt” variable (from D1.link) for all of a plan’s investments in CCTs. Do the same for PSAs and unknown DFEs.
* “subtotal.unmatched.amt” — The sum of the “unmatched.overflow” variable (from D1.link) for all the reported DFE investments which could not be matched to a DFE at all. In other words, the total dollar amount of unsuccessful links.
* “total.unmatched.amt” — The sum of the “unmatched.overflow” variable whether the plan-DFE link was successful or not.
* “SECLEV\_category.agg” — The sum of a second level asset or liability category for of a plan’s (successful) links.
* “all.mtia.links” — The number of links between a particular pension plan and an MTIA (good, bad, and-if applicable-triple zero). Also create analogous variables for the number of links to a CCT, PSA, 103-12 IE, or DFE of an unknown type.
* “good.mtia.links” — The number of successful links between a particular pension plan and MTIAs. Also create analogous variables for the number of links to CCTs, PSAs, 103-12 IEs, and DFEs of unknown type. These variables exclude both the number of bad links and the number of triple zero links (for CCTs and PSAs).
* “num.bad.link” — The number of DFEs a plan reported investing in which could not be successfully linked, but should have been able to.[[27]](#footnote-27)
* “num.trip.zero” — The number of triple zero DFEs a plan reported investing in.

**Step 15** — Compile a summary of plan assets with direct holdings, indirect holdings, total holdings, and plan characteristics.

* Create an object named “plan.assets” by merging pension.form5500 with D1.link.agg2 to compile direct holdings, plan characteristics, and indirect holdings.
* Create a variable called “imputed.assets” for the common situation where a plan[[28]](#footnote-28) only reports a non-zero holding in the PARTCP\_LOANS\_EOY\_AMT category (and not any others) and the total reported assets is larger. This imputed category accounts for the improperly unreported assets.
  + imputed.assets = TOT\_ASSETS\_EOY\_AMT – PARTCP\_LOANS\_EOY\_AMT (if participant loans is the only assets category with a reported holding).[[29]](#footnote-29)
* Create a variable for the difference between the reported total assets and liabilities and the calculated total assets and liabilities.
  + slop.assets = TOT\_ASSETS\_EOY\_AMT – sum.assets[[30]](#footnote-30)
  + slop.liab = TOT\_LIABILITIES\_EOY\_AMT – sum.liab[[31]](#footnote-31)
* Drop all pension plans that report a negative number for TOT\_LIABILITIES\_EOY\_AMT or OTHER\_LIAB\_EOY\_AMT, unless the absolute value of the reported negative number is less than one percent of TOT\_ASSETS\_EOY\_AMT.[[32]](#footnote-32)
* Create totals for each asset and liability category called TOT\_category, by adding the first level and second level holdings in that category. For the categories for MTIA, PSA, CCT, and 103-12 IEs, the dollar amount which was successfully linked to a DFE (and thus attributed at the second level) must be subtracted from the final totals. For example:

TOT\_INT\_MASTER\_TR\_EOY\_AMT = INT\_MASTER\_TR\_EOY\_AMT + SECLEV\_INT\_MASTER\_TR\_EOY\_AMT- mtia.total.amt

* Create variables to reflect whether a plan is single employer, multiple employer, or multiemployer.
  + “single.employer” = 1, if TYPE\_PLAN\_ENTITY\_IND is “2”[[33]](#footnote-33)
  + “multiple.employer” = 1, if TYPE\_PLAN\_ENTITY\_IND is “3”[[34]](#footnote-34)
  + “multiemployer” = 1, if TYPE\_PLAN\_ENTITY\_IND is “1”[[35]](#footnote-35)
* Create a variable named “cct.tot.less.000” which = “cct.total.amt” – “total.trip.zero.cct.amt. Create a similar variable for PSAs.
* Create two variables to designate the asset decile for defined benefit and defined contribution plans separately. For each type of plan, the relevant asset decile variable will equal zero if the plan reports a total of zero assets and otherwise takes on a number 1 through 10 indicated which asset decile the plan is in.[[36]](#footnote-36)
* Create a variable named “mega” which equals 1 if a plan’s assets are equal to or greater than 2.5 billion dollars and 0 otherwise.[[37]](#footnote-37)

**Step 16** — Classify plans into three groups based on whether they invest in any non-triple-zero DFEs and, if so, whether they file consistent totals on their Schedules H and D.

* Create a variable names “D1.non.000.amt” which equals “D1.amt.agg” minus the dollar amount invested in triple zero CCTs or PSAs.
* Create a variable named “group”.
* If “D1.non.000.amt” equals zero, then set the value of the “group” variable as “ONE”.
* For each DFE category, create a variable which is the difference between the dollar amount reported invested in that category on Schedule H and on Schedule D as follows:
  + mtia.diff = INT\_MASTER\_TR\_EOY\_AMT – mtia.total.amt
  + cct.diff = INT\_COMMON\_TR\_EOY\_AMT – cct.total.amt – total.trip.zero.cct.amt
  + ie.diff = INT\_103\_12\_INVST\_EOY\_AMT – ie.total.amt
  + psa.diff = INT\_POOL\_SEP\_ACCT\_EOY\_AMT – psa.total.amt – total.trip.zero.psa.amt
* If all four difference categories are between -10 and 10, then set the value of the “group” variable as “TWO”. If one or more have a discrepancy greater than $10, set the value of the “group” variable as “THREE”.
* Save “plan.assets2” as a .csv file.

1. For purposes of this project an “object” can be considered a spreadsheet, i.e., a rectangular set of data with a variable in each column and an observation in each row. [↑](#footnote-ref-1)
2. For 2008, 2,201 observations (out of 795,022) could not be classified as either a plan or a DFE. Of the filing entities that were classified, 7,352 were DFEs, and the remainder were plans. [↑](#footnote-ref-2)
3. In the form5500 file for 2008, there are 6,398 filing entities that indicate they are both a pension plan and a welfare plan. Only 63 of these are multiemployer plans. (Only 55 are multiple employer plans.) [↑](#footnote-ref-3)
4. For 2008, only 7,247 (of 7,352) DFEs were classified in one of the four categories. The breakdown by category is as follows: 1,652 MTIAs; 3,115 CCTs; 2,048 PSAs; and 432 103-12 IEs. [↑](#footnote-ref-4)
5. For 2008, there were a total of 7,174 unique EINPN’s associated with DFEs. In other words, 73 duplicate filings were dropped. [↑](#footnote-ref-5)
6. There are 78,023 welfare plan filings in 2008. This includes some multiple filings because this is from the raw data, but there are still a total of 76,734 unique EINPN’s. [↑](#footnote-ref-6)
7. For 2008, there were 378 plans which did not have a filing id (and 80,670 which did). [↑](#footnote-ref-7)
8. The 2008 Pension Research File User Guide defines this variable (at p. 11) s follows: “OPR\_PN adds a fourth character which distinguishes among filings with the same EIN/PN. The “best” filing will have the traditional three-character PN, while additional filings have a character (A, B, C, …) appended to their plan number.” For 2008, there were 243 filings that were not the first for a large plan, yet were still designated as the best for the plan. (Under these circumstances there was typically also an original filing designated as best for the plan). [↑](#footnote-ref-8)
9. The majority of the time last filing and best filing correspond. (They are the same for 80,494 filings). There are 250 which are best, but not last. There are 315 which are last, but not best. [↑](#footnote-ref-9)
10. For 2008, there were a total of 80,541 unique EINPN’s. In other words, 129 duplicate filings were dropped. [↑](#footnote-ref-10)
11. For 2008, pension.form5500l has 501,858 rows. Each row is a pension plan — DFE link. If a plan has not reported any investments in a DFE, it appears in one row only. There are 35,843 pension plans without any DFE links at all. The remaining 44,698 have at least one link, but 1,420 only have link(s) with a zero dollar amount. (This is why the number in note 26 is 43,278.) [↑](#footnote-ref-11)
12. For 2008, there were 71 DFEs which did not have a Schedule H filed under the same FILING\_ID. So there are a total of 7,103 DFEs in dfe.assets. [↑](#footnote-ref-12)
13. For 2008, there were 38 DFEs with a non-zero amount in slop.assets. [↑](#footnote-ref-13)
14. For 2008, there were 12 DFEs with a non-zero amount in slop.liab. [↑](#footnote-ref-14)
15. For 2008, there were 414 DFEs which reported zero total assets. [↑](#footnote-ref-15)
16. There are only two of these for 2008. [↑](#footnote-ref-16)
17. For 2008, there are 411,971 non-zero pension-DFE links. The 35,843 rows with pension plans that had no reported DFE investments were dropped. Also, 54,044 pension-DFE links were dropped because the reported dollar amount was zero These 54,044 links are presumably reported on Schedule D because the plan had an interest in the MTIA, CCT, PSA, or 103-12 IE at some point in the year, but it was zero at the end of the year. However, there is no beginning of year number reported that corresponds to “DFE\_P1\_PLAN\_INT\_EOY\_01\_AMT”, so it’s impossible to say for sure. (See page 25 of the 2008 Form 5500 instructions). [↑](#footnote-ref-17)
18. For 2008, there are 184,089 reported links that have a DFE with a triple zero plan number. [↑](#footnote-ref-18)
19. For 2008, there were 42,493 bad pension-DFE links. [↑](#footnote-ref-19)
20. For 2008, 6,190 plans have exactly one bad link. (5,150 plans have more than one bad link.) [↑](#footnote-ref-20)
21. For 2008, 1,544 bad links meet these criteria (and thus, can potentially be fixed using schedule D2). There are 3,565 bad D1 links where the pension plan is not listed by any DFE in the D2 filings. There are 1,081 bad D1 links where the pension plan is listed by more than one DFE in the D2 filings. [↑](#footnote-ref-21)
22. For 2008, there are 184,089 reported links that have a DFE with a triple zero plan number. [↑](#footnote-ref-22)
23. For 2008, there are now 41,173 bad pension-DFE links. Note: Even a plan-DFE link which was corrected previously in this step using Schedule D2 data may still not be successful (i.e., linked to a DFE’s schedule H filing) at this stage. There are 224 of these for 2008. Most of them (209) are associated with a DFE that was dropped from the dfe.assets dataset because of reporting zero assets (see Step 9). [↑](#footnote-ref-23)
24. For 2008, the amount reported as invested by the pension plan was greater than the net assets of the DFE for 677 pension-DFE links. For 419 of these, the DFE reported zero net assets. [↑](#footnote-ref-24)
25. For 2008, there are 51 links that fit this description. [↑](#footnote-ref-25)
26. For 2008, there were 43,278 pension plans. [This means one row for each pension plan that reports an investment in one or more DFEs of any kind. That’s how we went from 80,541 plans (in note 10) to only 43,278 here. That is, about 37,000 large pension plans had no money in DFEs in 2008.] [↑](#footnote-ref-26)
27. For 2008, of the 43,278 pension plans that report DFE investments, 33,133 have no bad plan-DFE links, 4,995 have only one bad link, and 5,150 have more than one bad link. [↑](#footnote-ref-27)
28. I double-checked that this problem does not occur at all in DFE filings, so a similar step is not necessary for DFEs. As a result, there is no SECLEV\_imputed.assets variable. [↑](#footnote-ref-28)
29. For 2008, there are 6,731 plans with non-zero imputed assets. (Note: Most of them are participant directed plans.) [↑](#footnote-ref-29)
30. For 2008, there are 5,011 plans which have a non-zero dollar amount of slop.assets. [↑](#footnote-ref-30)
31. For 2008, there are only 82 pension plans which have a non-zero dollar amount of slop.liab. [↑](#footnote-ref-31)
32. For 2008, 32 plans were dropped according to this criterion. [↑](#footnote-ref-32)
33. For 2008, there are 75,755 single employer plans. [↑](#footnote-ref-33)
34. For 2008, there are 2,222 multiple employer plans. [↑](#footnote-ref-34)
35. For 2008, there are 2,526 multiemployer plans. [↑](#footnote-ref-35)
36. For 2008, there are 1,047 defined benefit plans in each non-zero decile category (and 39 with zero assets). There are 6,960 defined contribution pension plans in each non-zero decile category (and 405 with zero assets). [↑](#footnote-ref-36)
37. For 2008, there are 232 mega plans. [↑](#footnote-ref-37)