ACTUARIAL RESEARCH CLEARING HOUSE
1998 VOL. 1

Data Analysis for the Safest Annuity Rule Project: Implications for the Education of Actuarial Students

Matt Hassett, PhD, ASA
William Gundberg Jr., EA

Introduction. This paper is based on work the authors did for the Committee on Retirement Systems Research of the Society of Actuaries. Our task was to perform data analysis for that Committee in support of a study of the effects of the Safest Annuity Rule. The Committee's conclusions will be released in a separate report. This paper will describe their overall study only to put our own work in context. The real purpose of this paper is to discuss the data analysis skills and principles which were applied in this project, and to argue that these same skills and principles should be part of an actuarial student's education.

Background: the Safest Annuity Rule Project. In the early 1990s some major insurer insolvencies caused concern about the safety of annuities purchased by pension plans. In March of 1995, the Department of Labor issued Interpretive Bulletin 95-1, the Safest Annuity Rule (SAR) , as a response to this concern. (Although the bulletin was actually issued in March 1995, pension professionals were previously aware that a ruling would be issued.)

The Safest Annuity Rule stated that purchase of annuities to provide for pension liabilities is a fiduciary act. Thus *the safest available annuity provider is required.* The purchaser of an annuity must consider a number of factors, including the issuer's investment portfolio quality and diversification, size and administrative capability, capital and surplus levels, lines of business, liability exposure and contract guarantees.

At the time of issue of Interpretive Bulletin 95-1 the market for annuities had already been shrinking. A number of reasons were given for this. A decline in interest rates had raised the purchase cost of annuities. Tax code changes had restricted excess plan asset reversions. GATT rules had made lump sum distributions more attractive than annuity purchases. The purpose of the Safest Annuity Rule Project was to study whether or not that rule had also been a major influence on the narrowing of the annuity market.

A workgroup was formed to evaluate the impact of the Safest Annuity Rule on current practices of major players in the annuity market, with special emphasis on defined benefit plan terminations. The workgroup would also assess the relative impacts of the other factors mentioned in the preceding paragraph. The project had two phases:

1) <u>A study of termination data provided by the Pension Benefit Guarantee Corporation.</u> The termination data would include records from a time period before the Safest Annuity Rule and records from a later period when the SAR might have influenced the annuity market.

2) <u>Surveys of pension practitioners to determine trends and influences</u>.

The workgroup planned to issue a final report based on both phases of the project. Our task was to do the data analysis for phase 1.

The Phase 1 Data Analysis Project. The major goals of this project were:

1) To create usable databases from two original files containing information on PBGC standard terminations from a) calendar year 1990 (and possibly 1993) and b) the first half of calendar year 1995.

2) To clean bad data (as possible) from those two files.

3) To link to each record the actuarial firm involved in the termination.

4) To provide summaries for each file of total assets, liabilities and participant numbers and the frequency of involvement of each of the ten most active actuarial consulting firms.

We were provided with two large data files.

a) FITZ : A large (4906KB) RPT file containing reports on plans closed in 1990 with termination dates from 1986 to 1991. The file appeared to be a scan of a hard copy report file. Variables of importance for this project were:

EIN/PN (Plan ID number)
Date of Plan Termination
Enrolled Actuary Number
Assets
Liabilities
Number of Participants.

There was a field for actuary name, but this was blank in all but one of 8453 cases.

b) STDTERM: A spreadsheet file containing reports on plans with form 500 receipt dates in the first 6 months of 1995. Variables of importance for this project were:

EIN/PN (Plan ID number)
Date of Plan Termination
Date of receipt of form 500
Enrolled Actuary Number
Enrolled Actuary Name
Assets
Liabilities
Number of Participants

Phase 1 Analysis Problems and Personnel Requirements. The four major components of Phase 1 each presented special problems and demanded special skills.

1) <u>To create usable databases from two original files</u> The STDTERM file for 1995 was a useable spreadsheet file which presented no special problems. However, the FITZ file of termination data from the pre-SAR period was not a useable database. It was a text file which appeared to be a scan of hardcopy output. Each page had a page number, a title and an underline for the title consisting of asterisks printed on the line below the title. Each record had a line of field titles with data elements printed on successive lines below the field titles. A record consisted of a sequence of alternating title lines and data lines. *Our team needed a database specialist who was experienced in dealing with such messy files and could rapidly translate the text file into a useable database file with no extraneous information and a standardized field structure.*

2) <u>To clean bad data (as possible) from those two files.</u> Once a useable database file was obtained, records containing clearly invalid information had to be purged or corrected. *Identification of records that did not make sense required our team to have an experienced pension actuary who easily could identify bad data such as invalid EA numbers.*

3) <u>To link to each record the actuarial firm involved in the termination</u>. This task involved multiple file linking. The FITZ file identified the actuaries who terminated plans by their EA numbers only. The project team was also provided with two more files –the first gave a name for each EA number, and the second gave a company name for each actuary name. Linking files was messy, since actuary names were often given in slightly different forms in the last two files. For example, a first name might be Elizabeth in one file and Betty in the next. *The computer database specialist was needed to link files efficiently.*

4) <u>To provide summary reports</u>. Once databases were useable, cleaned and linked, statistical analysis and reports were required. *The team needed a statistical analyst who could also serve as team leader by focussing all prior work on the final goal of providing the required reports.*

The Project Team. The project team had three members.

1) *Team leader and statistician.* Matt Hassett, the first author of this paper had recently completed a large analysis project involving 100,000 records on paramedic activity for the Emergency Medical Services company of Phoenix. He had previously led small consulting teams which cleaned and analyzed the files of failed thrifts.
2) *Pension specialist.* William Gundberg is an enrolled actuary with twelve years of pension experience. Mr. Gundberg also has a Master of Arts in Statistics.
3) *Computer specialist.* John Hassett is the lead database programmer for Emergency Medical Services and has extensive experience in the types of file conversion needed for this project.

An important principle in forming such a team is to use experienced specialists who can perform tasks rapidly and efficiently and avoid the temptation to have work done by someone who is learning on the job. For example, the team leader or the pension specialist could have performed the data conversion tasks –given a few weeks of learning time. The database specialist converted the FITZ file in an afternoon on a laptop in Newark Airport. (He is a pilot for Continental Airlines.)

The Interactive Team Effort. The team was designed to work well together. (It must be clear to the reader that the team consisted of an actuarial professor, a former student who was a pension actuary and the professor's son who was a computer specialist.) Such database projects require constant interaction between the computer specialist who is creating data structures, the industry expert who knows what the data should look like and the analyst who will compile the final reports using the database.

The project required the following steps:

1) All three team members browsed the original files, using EXCEL for the STDTERM spreadsheet file and Wordpad for the original FITZ text file. This was done to anticipate problems which might arise in the file conversion and data cleaning processes.
2) The computer specialist made a first pass conversion of the FITZ text file to an ACCESS database file.
3) The team leader took a random sample of records and compared the database file to the original text for those records. A trailing zero truncation problem was discovered and fixed.
4) The file was converted again and re-checked by random sampling. This time no problems were found.
5) The resulting working files from steps 1-4 above were then browsed for obvious errors and subjected to systematic checks. The browsing revealed the presence of a few hundred duplicate records which were removed from the database. The systematic checks dealt with searches for such events as blank fields or wildly inconsistent asset and liability figures. The records uncovered by systematic checks were analyzed and either corrected or purged. The team leader discussed corrections not only with team members, but also with the client representative, Tom Edwalds, research actuary at the Society of Actuaries office.
6) The computer specialist linked the cleaned termination data files with the actuary name and company name files. The pension expert assisted in resolving name ambiguities
7) The corrected versions of the working files were analyzed, and preliminary reports were generated for the client.
8) Analysis was modified based on client requests.
9) Final reports were completed.

Copies of two pages from the final report are included as an Appendix to indicate the type of summary data tables that were produced. However, the content of the final report is not the subject of this paper. The real point is this project illustrates a wide range of skills and principles which would be useful for actuarial students.

What should we teach our students? To better prepare our students for their careers, we need to teach them a number of things which appeared in this project but do not appear on any examination syllabus.

How to work in teams. Employers who come to Arizona State University tell us that the perceived ability to work in teams is an important factor in the hiring decision. It is not adequate to try to prepare students for this by simply giving a project and requiring that three people do the project together. In many cases the strongest student simply takes over and does the entire project. In an industrial setting management would soon learn that the two non-participants are not needed.
The students need to learn to divide tasks in pieces and then to integrate the results at the end. They need to learn to coordinate and communicate. One possible solution is to require the students working on a project to submit a work plan for the group members and keep time sheets recording what was done by each member. (The students enjoy this more when we tell them that they are really preparing a consulting bill.)

How to work in integrated computer environments. All of our graduates are required to use spreadsheets in their first jobs. An increasing number of them are now also doing jobs which start with a messy data files which must be converted and cleaned before the spreadsheet analysis begins. In our project database set-up work was done in ACCESS, analysis in EXCEL and SPSS 7.0 for Windows and report writing in Word (with liberal pasting of tables from EXCEL and SPSS.)

How to clean data files The first step is to recognize that data files can be imperfect –unlike the data sets in textbooks. The next step is to learn how to do error checking and reasonableness tests.

One possible response to the observations above is to recommend that the schools stick to actuarial mathematics and let the students learn about teamwork, computer skills and data quality on the job. Unfortunately this strategy can lead to actuarial students with bad habits –one employer has noted that he has workers who resist starting out with data cleaning.

Our own strategy at Arizona State University will be to create a data file which was obtained by scanning a textbook file into text and then inserting mistakes which must be cleaned. Our class project will require conversion, cleaning, analysis and a clearly defined work plan with final time sheets for individuals.

1995 STDTERM File					
Plan Size		**Assets**	**Liabilities**	**A/L Ratio**	**Participants**
Large	Number of Cases	474	474	471	474
	Sum	1,303,188,843	1,125,117,892		73,839
	Mean	2,749,344	2,373,666	1.088	156
	Minimum	30,000	0	0.950	0
	Maximum	126,000,000	67,000,000	5.100	8,336
Small	Number of Cases	1301	1301	1290	1301
	Sum	499,798,488	479,268,668		21,575
	Mean	384,165	368,385	1.077	17
	Minimum	0	0	0.730	0
	Maximum	997,825	997,825	6.010	98
Total	Number of Cases	1775	1775	1761	1775
	Sum	1,802,987,331	1,604,386,560		95,414
	Mean	1,015,768	903,880	1.079	54
	Minimum	0	0	0.730	0
	Maximum	126,000,000	67,000,000	6.010	8,336
Notes	a) 17 cases had Assets = Liabilities =0, and were not included above b) 1 case had blank asset and liabilities, and was not included. That case had one participant.				

1990 FITZ File					
Plan Size		**Assets**	**Liabilities**	**A/L Ratio**	**Participants**
Large	Number of Cases	1675	1675	1675	1677
	Sum	5,779,896,502	4,576,994,977		333,855
	Mean	3,450,684	2,732,534	1.389	199
	Minimum	10,903	7,700	1.000	0
	Maximum	1,055,000,000	622,000,000	260.740	6,291
Small	Number of Cases	6751	6751	6732	6751
	Sum	2,221,158,478	2,004,631,003		111,682
	Mean	329,012	296,938	1.431	17
	Minimum	0	0	0.750	0
	Maximum	999,232	999,232	1374.170	99
Total	Number of Cases	8426	8426	8407	8428
	Sum	8,001,054,980	6,581,625,980		445,537
	Mean	949,567	781,109	1.422	53
	Minimum	0	0	0.750	0
	Maximum	1,055,000,000	622,000,000	1374.170	6,291
Notes	a) 25 cases had Assets = Liabilities =0, and were not included above b) 2 cases had missing values in all above fields and were not included. c) The mean Asset/Liability ratios were excessively influenced by outliers. When all cases with A/L > 10 were excluded, the mean ratios were: *Large Plans* *1.2259* *Small Plans* *1.1923* *All Plans* *1.199*				

Most Active Firms: All Plans

1990 FITZ file

Rank	1989 Company	Number of Cases
1	**A**	218
2-3 tie	**B**,C	119
4	D	115
5	**E**	113
6	**F**	110
7-8 tie	G,H	81
9	I	78
10	J	75

1995 STDTERM file

Rank	1994 Company	Number of Cases
1	**B**	48
2	**E**	45
3	**A**	43
4	K	33
5	**F**	27
6	L	25
7	M	19
8-9 tie	N,O	16
10-14 tie	P,Q,R,S,T	15