RESTORING TRUST TO AIR FORCE MATERIEL COMMAND'S PERSONNEL LOSS FORECAST LEGACY SOFTWARE[1,2]

Robert L. Sills, Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 937-255-4486, robert.sills.ctr@afit.edu
Alan W. Johnson[3], Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 937-255-3636, alan.johnson@afit.edu
Edward D. White, Department of Mathematics and Statistics, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, 937-255-3636, edward.white@afit.edu

ABSTRACT

The Command Human Resources Intelligence System is the Air Force Materiel Command (AFMC) Personnel Loss Forecasting solution. Accurate loss forecasts are critical to personnel recruitment, training and retention. Ordinary Least Squares (OLS) are classic methods to solve loss forecast problems. However, the OLS model at AFMC consistently underestimated staffing loss. The staff at AFMC was not confident in the software behind their OLS model - they did not trust that the software correctly implemented the OLS model, or that the OLS model itself was sufficiently responsive. In our study, we detected and corrected a variety of software errors and implemented a new moving average model that provides a more accurate forecast.

INTRODUCTION

The Command Human Resources Intelligence System (CHRIS) database collects and retains ten years of personnel actions in the 60,000-member workforce of Air Force Materiel Command (AFMC) to forecast future losses. An accurate forecast provides the most efficient use of resources for recruitment, retention, and training. As a shortage is foreseen in a specific skill set, new workers with that skill set can be recruited, or existing workers can be trained to fulfill the need. Also, retention policies can be developed that incentivize workers to stay, saving these recruitment and training costs. Alternatively, if reductions are needed, an accurate forecast can inform planners of the levels that may be reached through natural attrition, allowing efficient policies for early retirement incentives or reductions-in-force. A quality tool can, over the years, save the Air Force Materiel Command millions of dollars in misallocated personnel resources.
The CHRIS system applies a complex set of business rules to process data, such as calculating earliest retirement eligibility for a wide variety of employees. To gain confidence in the system it was essential to verify that these business rules were applied accurately. CHRIS used a fourth generation software system called Informatica which made a series of database operations on the data with complex Structured Query Language (SQL) calls, including Programmable Logic SQL (PL/SQL). It required knowledge of this system and these calls to verify how the business rules were applied by them. This verification was undertaken by a seasoned professional programmer

[1] This research is sponsored by the Air Force Logistics Transformation Office.

[2] The views expressed are those of the authors, and do not reflect the official policy or positions of the United States Air Force, Department of Defense, or the U.S. Government.

[3] Corresponding Author.

who had skills in SQL and PL/SQL, and who learned Informatica in the process. The poor CHRIS code documentation drove us to perform a line-by-line code analysis to understand the computational flow.

The legacy CHRIS system used an OLS model on a subset of the available data to find a line that best fit the past results, then forced all future results to conform to that line. This led to a great deal of smoothing and limited granularity, or fine detail, of the forecasts. There were questions about the assumptions implicit in the model, the limited data the model used, and whether an alternative model could provide better fidelity and accuracy. The programmer had no experience in modeling and so was joined first by an industrial engineer research assistant, then by a Ph.D. member in statistics with a specialization in data modeling. All of this was supervised by an industrial engineering Ph.D. faculty member. The project required approximately 2000 man-hours by research staff to interact with the sponsor, understand the requirements and examine the roughly 10,000 *Function Points* of software code. A function point is a measure of software complexity. It was more useful than lines of code estimation for complexity assessment for us, because we were dealing with a fourth generation modular software language.

VERIFICATION

During verification we found four software issues in the legacy model, involving *interpolation*, *retirement dates*, *earliest retirement dates*, and a *"magic" adjustment factor*. These issues in the production software system illustrate the potential errors in such systems.

Interpolation: We found a very complex series of undocumented database operations which, when re-written as equation 1, seemed to represent an interpolation using percentages and standard deviations. However it did not follow a known calculation of interpolation and was undocumented. To implement only the business rules given to us, we removed the equation 1 operations and replaced them with a much simpler method (equation 2) that uses a standard point interpolation, where the variable y is the unknown value, x is the known value, x_1 and y_1 are the coordinates below the x value, and x_2 and y_2 are the coordinates above the x value [1].

$$y = \frac{Lower\ Pct + \frac{Std\ Dev\ Pt2 - Lower\ Std\ Dev}{Mid\ Std\ Dev - Lower\ Std\ Dev}}{Mid\ Pct + \frac{Std\ Dev\ Pt1 - Mid\ Std\ Dev}{Upper\ Dev\ Pt2 - Mid\ Std\ Dev}} \tag{1}$$

$$y = y + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1} \tag{2}$$

Retirement Dates: We found the system did not follow business logic nor Air Force regulations in calculating retirement dates for Firefighters, Police and Air Traffic Controllers. For example, Air Force Regulations for firefighter s set mandatory retirement at age 55, qualified by 20 years of service [2]. The CHRIS system incorrectly determined that a firefighter had to mandatorily retire at age 62 with 5 years of service. The majority of retirement dates for these categories had similar errors, which led to incorrect calculations used in forecasting. We corrected this logic.

Earliest Retirement Dates: We found inaccuracy in determining eligibility for retirement. The system compared three different types of retirement: Regular, Mandatory, and Minimum Retirement Age (MRA). It sought the earliest date that a person was eligible for one of these retirements. The system did not properly compare them. It used this logic:

```
IIF (YRS_REG >YRS_MAND
OR  YRS_REG > YRS_MRA,
        YRS_REG,
```

Instead, it should have used this logic:

```
IIF (YRS_REG >= YRS_MAND
AND  YRS_REG >= YRS_MRA,
        YRS_REG,
```

This first logic could erroneously choose "years regular" as the earliest date in cases where it is actually later than one of the other retirement dates. We corrected this and several similar errors.

"Magic" Adjustment Factor: We found an undocumented operation which involved a multiplicative factor. After the software had executed to obtain a probability of loss, this result was then multiplied by a value in the database, currently set to 1.0 to return a final loss probability. This operation seemed to support only one possible purpose: to arbitrarily adjust the model results to a desired value. Anecdotally, our sponsor reported one incident where this adjustment may have occurred. Needless to say, we removed this operation.

MODEL

One OLS model limitation is that when smoothing the data to conform to a line, fidelity is lost for sub-populations that might have different behaviors. For example, in resignations, the OLS method sampled data from people in their first ten years of service, and ignored data from anyone with longer service. In reality, the data revealed that many people resign with more than ten years of service. Also, in the way data was sampled, half of the data used populations with five years of service or less, which in-turn rendered data sets with excessive sparsity. Finally, the model seemed to contain a bias to always underestimate loss.

We chose to instead implement a model that makes use of more data, can retain and analyze the behavior of sub-populations without smoothing, and does not exhibit a systemic bias. We chose a moving average approach [3]. We captured all the data from a preceding X number of years, with groups based on gender, employment type, and years of service (for resignations) or years of eligibility (for retirements). The groupings allowed us to capture fine sub-populations that might exhibit very different behavior, for example resignations rates at three years of service versus fifteen years, and capture that exact behavior.

After calculating and comparing Mean Average Percent Error (MAPE) [4] for different moving averages, we arrived at a three-year moving average as being the most accurate and responsive to changes in behavior. What we do is look at the past three years of actual behavior for a specific group, do an average of the loss of population in that time, and then project that the next year's

loss will be that average. Consider an actual example for two groups in the 2010 data, Professional Females who resigned with three years and fifteen years of service. Table 1 depicts the data for these groups over three years:

Table 1: Professional females, three years of service.

Year	Number of Personnel	Number that Resigned
2008	103	6
2009	141	10
2010	125	10

We average the resignation rate for this group to obtain a resignation probability of .07046. Table 2 depicts the data for fifteen years of service:

Table 2: Professional females, fifteen years of service.

Year	Number of Personnel	Number that Resigned
2008	52	1
2009	53	1
2010	38	2

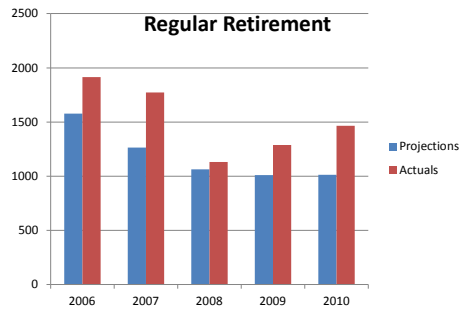
For this group the resignation rate is much lower, .027777, just over one-third the earlier rate.

Our new model applies the appropriate rate to a person who belongs in each of these groups, and projects that her resignation loss probability will be exactly that of her peer group. It is not influenced by behavior of other groups, as happens when fitting to a line. If for some reason there is a spike in resignations from people with many years of service, this model will capture that spike and include it in projections, where the old model would totally miss it. Interestingly, there was such a spike for thirteen years of service, with a resignation rate of .071856, the second-highest rate in all the years of service for that employment type-gender combination. That does not fit an assumption that resignations decline with increased years of service.

We do this calculation for every employment type-gender-(years of service/years of eligibility) combination found in AFMC, and project future behavior based on actual past behavior for that group. It has more data behind it, and a lower error in the projections tested. We believe it to be a superior model to OLS. It is also open to further refinement to make it more accurate, and perhaps account for more factors, such as geographic location, early retirement incentives or unemployment rate changes.

Bias: Figure 1 depicts projection differences in the models:

OLS



Moving Average

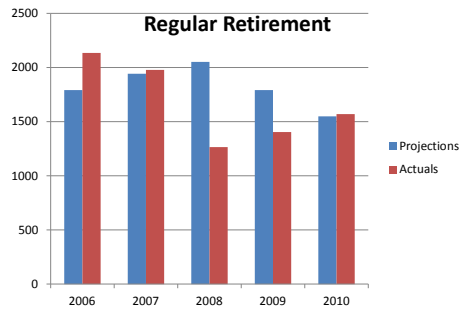


Figure 1: Regular Retirement Totals

The OLS model shows a bias by consistently under-estimating retirement losses. The moving average forecasts did not show this bias, as it both under- and over-estimates losses, based on changes in recent behavior. Sudden spikes or drops in behavior could cause error in the moving average forecast, but it would respond to this past behavior and gradually revert to more accurate estimates. OLS showed little response to actual past behavior.

Benchmarks: Our investigation provided the first benchmarks to measure the accuracy of the CHRIS projections. Equation 3 is the MAPE formula that we applied, where A_t is an actual result, F_t is a forecast result, and M is the measurement of error:

$$M = \sum_{t=0}^{n} \left| \frac{A_t - F_t}{A_t} \right| \tag{3}$$

Table 3 shows the forecasts for regular retirement losses in the year 2011 and associated MAPE scores, based on the same 2010 data:

Table 3. Model accuracy compared.

	Total Loss	MAPE Score
Actual	1450	
OLS Method	1003.2	30.6
Moving Average Method	1508.1	4.0

Sample Sizes: Our moving average approach used more data points than the OLS model, and roughly equal samples for each sub-group examined. These sizes are compared in Table 4.

Table 4. Sample sizes compared

Type	Current Model	New Model
Resignations	35,092	172, 235
Regular Retirement	30,565	78,889

CONCLUSIONS & LESSONS

One important success factor was a dedicated development environment. A place where we could modify the code to try out new approaches without jeopardizing AFMC's work in progress was essential. We safely made many changes before we arrived at the final system. There were several less tangible factors that also contributed to this study's success. Firstly, people with the right skills were essential to make sense of the code and to find buried errors, as well as to understand the model's math. Second, collaboration with the AFMC Subject Matter Experts (SMEs) was critical. We depended on their knowledge and cooperation to help us understand and improve the system. At first some SMEs were very willing to help but others less-so. As we developed relationships and provided some preliminary results, the teamwork improved.

We worked hard to keep our project sponsor updated on our methods and progress so that they could make timely adjustments to our work when needed. The result is that they remained up-to-speed and comfortable with our work and were willing to implement our code recommendations at project end. Before this work began, our sponsor said they did not have confidence in the CHRIS software and could not defend it when questioned. They now believe that CHRIS a much better tool and are marketing it throughout their organization's personnel offices. They further said they would use our work to help them respond to a Government Accountability Office query on Air Force depot staffing. Large organizations often have custom legacy software systems whose value erodes with time. Careful periodic investments in operation and verification must be done for the software to retain user trust.

REFERENCES

[1] Beyer, W.H. (Editor), CRC Standard Mathematical Tables and Formulae, 29th Edition. Boca Raton: CRC Press, 1991.

[2] *Civil Service Retirement System (CSRS) and Federal Employees Retirement System (FERS) Handbook for Personnel and Payroll Offices, Special Retirement Provisions for Law Enforcement Officers, Firefighters, Air Traffic Controllers, and Military Reserve Technicians, Section 46A3.3-2 Mandatory Separation (series: 0081, 0083),* United States Office of Personnel Management,1998.

[3] Kachigan, S.K., Statistical Analysis, An Interdisciplinary Introduction to Univariate and Multivariate Methods. New York: Radius Press, 1986.

[4] Ragsdale, Cliff T., Spreadsheet Modeling and Decision Analysis, Revised 5th Edition. Mason: Thomson/South-Western, 2008.