ACTUARIAL RESEARCH CLEARING HOUSE
1998 VOL. 1

NEW SALARY FUNCTIONS FOR PENSION VALUATIONS

Jacques Carriere and Kevin Shand

ABSTRACT

This paper investigates salary functions as they are used in the valuation of pension plans. Pension actuaries may find many of the ideas in this article useful; moreover the paper may be interesting to researchers in actuarial science. The main conclusion of this paper is that salary functions, as derived from the parametric models, yield gains and losses that can be quite small and in some cases, less variable than non-parametric methods. This paper starts by defining the salary function as an accumulation function based on inflation and merit. Next, we investigate traditional estimation methods in the context of this definition. We then present a parametric age-based model for the salary function. Thereafter, we present a parametric service-based model and compare it to the age-based model. Finally, we apply real pension plan data to derive age and service based salary functions and through the use of two funding methods we investigate how these salary functions affect salary gains and losses.

1 Introduction

Consider the standard salary function, denoted as S_x where x is the age of an employee in a pension plan. Generally, S_x is a non-decreasing function in x that reflects increases in salary due to inflation and merit (seniority). In Actuarial Mathematics (3), we see that the usual purpose of the salary function is to estimate future salaries in pension valuations. For example, if AS_x is the actual salary of a person age x then the estimated future salary at age $y > x$ is

$$AS_x \times \frac{S_y}{S_x}. \tag{1.1}$$

Note that we only need the ratio S_y/S_x and so S_x can be arbitrarily rescaled. Before proceeding, we will define an age-based *valuation salary function* as an accumulation function with an inflation component and a merit component. It is well-known that accumulation functions can be expressed as $\exp\{\int_0^x \delta_z dz\}$ where δ_z is the force of accumulation. In our case we define $\delta_z = \xi + \psi_z$ as a force of salary accumulation where ξ is a constant inflation rate while ψ_z is the instantaneous rate of increase due to age. Therefore, we have

$$S_x = \exp\Big\{\xi \times x + \int_0^x \psi_z \, dz\Big\}, \tag{1.2}$$

and so

$$\frac{S_y}{S_x} = \exp\Big\{\xi(y-x) + \int_x^y \psi_z dz\Big\}. \tag{1.3}$$

As a result, the estimated future salary at age $y > x$ is equal to the actual salary at the age x times a function that accounts for inflation, denoted as $\exp\{\xi(y - x)\}$, and a function for merit, denoted as $\exp\{\int_x^y \psi_z dz\}$. In this article, we will assume that

$$\psi_z = \beta e^{-\lambda z},\ \beta > 0, \lambda > 0. \tag{1.4}$$

Other parametric formulas for the merit component can be used. The data that we used in this article is extremely variable; therefore no parametric formula will be much better than any other with respect to fit. The primary reason that we choose this function for ψ_z is that it is a decreasing function in z which results in merit increases that are smaller for older employees than younger employees. Note that ψ_z does not change with time, a hypothesis that we will test in the next section. Using equation (1.4) with equation (1.2) yields

$$S_x = \exp\left\{\xi \times x + \beta\lambda^{-1}(1 - e^{-\lambda x})\right\}. \tag{1.5}$$

In the second section of this paper, we look at two traditional estimation methods and introduce a statistical model for the observed data. The two traditional methods for estimating S_x are outlined in Marples (5); the first being the Current Average Method (CAM) and the second being the Increase Ratio Method (IRM). We will then show that where IRM yields an estimable regression function, CAM will not. In the third section, we introduce a parametric regression estimator of the age-based salary function and estimate it with actual pension plan data. We will also confirm that the merit increases are smaller for older employees than younger employees. In the fourth section, we will estimate a service-based salary function. And finally, in the last section, we will compare our parametric salary functions with a salary function estimated with the classical non-parametric IRM method and with one provided by the valuation actuaries who provided us with the pension plan data. Using two different pension valuation methods, we will show that our parametric functions can yield good predictions and result in small salary gains and losses relative to traditional estimation methods.

2 Traditional Estimation Methods

In this section, we compare the traditional CAM and IRM estimation techniques and conclude that IRM is a more complete method. In this discussion $AS_{[x]}$ is the actual salary of a person aged x at time $t = 0$ and $AS_{[x]+1}$ is the actual salary of that same person one year later at time $t = 1$. According to Schoenly (6) and Marples (5), the CAM estimate is derived by graduating the average salaries in quinquennial age groups or possibly single age groups if the data permits. In other words, S_x is estimated by a non-parametric regression function. For ease of discussion, let us suppose that we can group the observable data by single ages. Let N_x denote the number of people who are aged x at time $t = 0$. This will be the same number of people who are aged $x + 1$ at time $t = 1$ because we only have lives that had observable salaries during the whole study period. Let $AS^k_{[x]+t}$, $k = 1, 2, \ldots, N_x$ denote the actual salary of employee k, at age $x + t$. Also let s_k denote the amount of service that the employee has at $t = 0$ and let $AS_{[x]-s_k}$ denote the actual salary at hire. Then a realistic model of the progession of salaries is the relation

$$AS^k_{[x]+t} = AS^k_{[x]-s_k} \times \exp\left\{\int_{-s_k}^{t} \xi_z \, dz + \int_{x-s_k}^{x+t} \psi_z \, dz\right\} \times \epsilon^k_t. \tag{2.1}$$

where the inflation factor is $\exp\{\int_{-s_k}^{t} \xi_z \, dz\}$, the merit factor is $\exp\{\int_{x-s_k}^{x+t} \psi_z \, dz\}$, and the error component is ϵ^k_t. Using equation (2.1) we find that

$$AS^k_{[x]+t+1} = AS^k_{[x]+t} \times \exp\left\{\int_{t}^{t+1} \xi_z \, dz + \int_{x+t}^{x+t+1} \psi_z \, dz\right\} \times \frac{\epsilon^k_{t+1}}{\epsilon^k_t}. \tag{2.2}$$

Using the *midpoint rule*, we can write $\int_t^{t+1} \xi_z dz \approx \xi_{t+1/2}$. This useful approximation will be used in the ensuing argument. Next, let

$$\overline{AS}_{[x]+t} = \frac{1}{N_x} \sum_{k=1}^{N_x} AS^k_{[x]+t}, \tag{2.3}$$

denote the average salary for the group aged $x+t$ at time t. By substituting (2.1) into (2.3) and employing the CAM technique, this would yield

$$\overline{AS}_{[x]+t} \approx \frac{1}{N_x} \sum_{k=1}^{N_x} \left[AS^k_{[x]-s_k} \times \exp\left\{\int_{-s_k}^{t} \xi_z \, dz + \int_{x-s_k}^{x+t} \psi_z \, dz\right\}\right], \tag{2.4}$$

which clearly does not allows us to extract a salary function like the one defined in equation (1.2). However, the IRM method does. To see why, first note that

$$\overline{AS}_{[x]+t+1} \approx \frac{1}{N_x} \sum_{k=1}^{N_x} \left[AS^k_{[x]+t} \times \exp\left\{ \int_t^{t+1} \xi_z \, dz + \int_{x+t}^{x+t+1} \psi_z \, dz \right\} \right] \\ \approx \overline{AS}_{[x]+t} \times \exp\left\{ \xi_{t+1/2} + \int_{x+t}^{x+t+1} \psi_z dz \right\}. \tag{2.5}$$

Using the IRM technique, as described by Schoenly (6), will give you a result similar to (1.2) because you must multiply the ratios

$$\overline{R}_{x+t} = \overline{AS}_{[x]+t+1} \Big/ \overline{AS}_{[x]+t}, \tag{2.6}$$

successively to generate the salary scale. As an intermediate step you could smooth these ratios before multiplying them successively. Therefore, the IRM method yields

$$\widehat{S}_x = \prod_{a=0}^{x-1} \overline{R}_a \approx \exp\left\{ \xi_{1/2} \, x + \int_0^x \psi_z \, dz \right\}, \tag{2.7}$$

which is exactly equal to (1.2), our definition of the salary function. In this last calculation, we assumed that the first term in the product was at age zero because it allows for an easy discussion. In reality, we only need the ratios $\widehat{S}_y / \widehat{S}_x$ and so the first term in the product definition of $\widehat{S}_x$ can start at any realistic age. As a result, the IRM method is preferable to CAM since it is a complete estimation method. Note that one weakness of the non-parametric estimator $\widehat{S}_x$ is that the inflation component is inherited from the data.

Next, we present an alternate nonparametric estimation method that is similar to the traditional IRM method. A parametric version of this method will be presented in the next section. Let $R_x^k = AS_{[x]+1}^k \Big/ AS_{[x]}^k$ for $k = 1, 2, \ldots, N_x$ denote the observed individual ratios. First, we calculate the average ratios

$$R_x^* = \frac{1}{N_x} \sum_{k=1}^{N_x} R_x^k. \tag{2.8}$$

Next, we find that

$$R_x^* \approx \exp\left\{ \xi_{1/2} + \int_x^{x+1} \psi_z \, dz \right\}. \tag{2.9}$$

Multiplying these ratios successively generates the estimate

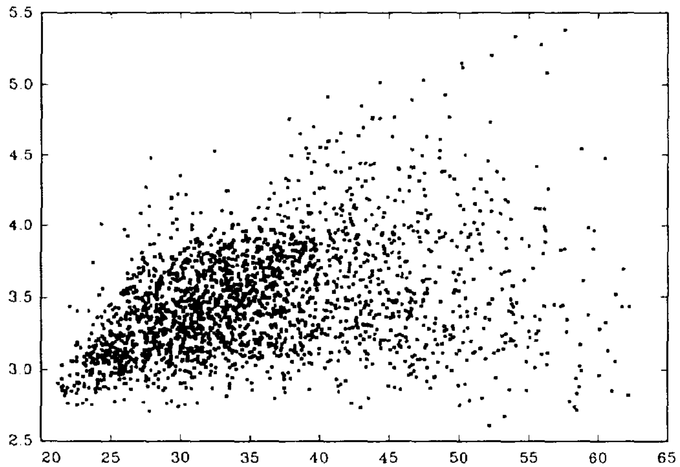
$$S_x^* = \prod_{a=0}^{x-1} R_a^* \approx \exp\left\{ \xi_{1/2}\, x + \int_0^x \psi_z \, dz \right\}, \tag{2.10}$$

which is exactly equal to (1.2), our definition of the salary function.

3 Parametric Estimation of the Age-Based Model

In this section, we present some parametric estimation methods and we apply them to real pension plan data from a medium-sized company that wishes to remain anonymous. We observed $N = 2231$ observations, which we denote as $[a_k, x_k, AS_{[x_k]}^k, AS_{[x_k]+1}^k, AS_{[x_k]+2}^k]$ for $k = 1, 2, \ldots, N$. As before, $AS_{[x_k]}^k$ is the actual salary of emplotee k who is aged x_k at time $t = 0$ and who was hired at age a_k at time $t = a_k - x_k$. $AS_{[x_k]+1}^k$ is the actual salary of that same person at time $t = 1$ while $AS^k_{[x_k]+2}$ is the employee's salary at time $t = 2$. The following discussion can be generalized to a situation where the salaries are observed over a longer period or to a situation where salaries might not be observed every year. As an example of typical salary data, Figure 1 shows a scatter plot of the log-salaries, $\log_e[AS^k_{[x_k]}]$ versus x_k from our $N = 2231$ observations at $t = 0$. The same pattern was found at $t = 1, 2$.

FIGURE 1
A Plot of the Log-Salaries $\log_e[AS^k_{[x_k]}]$ versus Attained Age x_k.



Using our statistical model from the last section, we find that

$$\frac{AS^k_{[x_k]+t+1}}{AS^k_{[x_k]+t}} = \exp\left\{\int_t^{t+1} \xi_z\,dz + \int_{x_k+t}^{x_k+t+1} \psi_z\,dz\right\} \times \frac{\epsilon^k_{t+1}}{\epsilon^k_t}, \quad t = 0, 1. \quad (3.1)$$

Next, we evaluate the integral $\int_{x_k+t}^{x_k+t+1} \psi_z\,dz$ when $\psi_z = \beta e^{-\lambda z}$. This yields

$$\int_{x_k+t}^{x_k+t+1} \psi_z\,dz = be^{-\lambda(x_k+t)},$$

where

$$b = \beta\lambda^{-1}(1 - e^{-\lambda}). \quad (3.2)$$

Next, we define

$$\begin{aligned} Y^k_t &= \log_e\left[AS^k_{[x_k]+t+1}/AS^k_{[x_k]+t}\right], \\ \varepsilon^k_t &= \log_e\left[\epsilon^k_{t+1}/\epsilon^k_t\right], \\ r_t &= \int_t^{t+1} \xi_z\,dz. \end{aligned} \quad (3.3)$$

Using these definitions, we find that (3.1) can be written as

$$Y^k_t = r_t + be^{-\lambda(x_k+t)} + \varepsilon^k_t, \quad (3.4)$$

for $t = 0, 1$ and $k = 1, 2, \ldots, N$. This last regression equation can be estimated many ways. Originally, we used a maximum likelihood approach assuming a normal distribution for ε^k_t. However, since the residual errors are not normal, we will instead use an *ordinary least squares* (OLS) approach to estimate the parameters. First, we will assume that the errors are uncorrelated. Even though this is not a realistic assumption, it does allow us to test for independence. More importantly, the parameter estimates of variances and correlations from this step will be used later in a *generalized least squares* (GLS) approach. To test the hypothesis that the merit component does not change with time we make b and λ functions of t. Let $\boldsymbol{\theta} = (r_0, r_1, b_0, b_1, \lambda_0, \lambda_1)'$ denote the parameters. Under OLS, we minimize

$$L(\boldsymbol{\theta}) = \sum_{t=0}^{1}\sum_{k=1}^{N}\left[Y_t^k - r_t - b_t e^{-\lambda_t(x_k+t)}\right]^2. \tag{3.5}$$

The OLS estimate of $\boldsymbol{\theta}$ is that value $\widehat{\boldsymbol{\theta}}$ such that $L(\widehat{\boldsymbol{\theta}}) \leq L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$. Using a numerical optimization program, we found that

$$\begin{aligned} \widehat{r}_0 &= .02518, \quad \widehat{r}_1 = .01499, \\ \widehat{b}_0 &= 1.6182, \quad \widehat{b}_1 = .24935, \\ \widehat{\lambda}_0 &= .15192, \quad \widehat{\lambda}_1 = .09364. \end{aligned} \tag{3.6}$$

These parameter estimates will be used later. For future reference, we note that $\widehat{\beta}_0 = \widehat{b}_0\,\widehat{\lambda}_0(1 - e^{-\widehat{\lambda}_0})^{-1} = 1.744$. Using these estimates, we calculated the residual errors, defined as

$$e_t^k = Y_t^k - \widehat{r}_t - \widehat{b}_t e^{-\widehat{\lambda}_t(x_k+t)}. \tag{3.7}$$

Using these residuals we constructed 95% confidence intervals for the following standard deviations and correlations:

$$\begin{aligned}\sigma_1 &= \sqrt{Var(\varepsilon_1^k)},\\ \sigma_2 &= \sqrt{Var(\varepsilon_2^k)},\\ \eta_1 &= \frac{Cov(\varepsilon_1^k, \varepsilon_1^l)}{\sigma_1^2}, k \neq l,\\ \eta_2 &= \frac{Cov(\varepsilon_2^k, \varepsilon_2^l)}{\sigma_2^2}, k \neq l,\\ \rho &= \frac{Cov(\varepsilon_1^k, \varepsilon_2^k)}{\sigma_1 \sigma_2}.\end{aligned} \tag{3.8}$$

Using standard estimators and confidence interval formulas as found in Manoukian (4), we investigated the null hypothesis that $\sigma_1 = \sigma_2$, $\eta_1 = 0$, $\eta_2 = 0$ and $\rho = 0$. The results are summarized in Table 1. We conclude that $\sigma_1 \neq \sigma_2$, $\eta_1 = 0$, $\eta_2 = 0$ and $\rho \neq 0$ since the confidence intervals for σ_1 and σ_2 do not overlap, the confidence intervals for η_1 and η_2 include zero while the confidence interval for ρ does not include zero.

Table 1
Confidence Intervals for Standard Deviations and Correlations

parameter	estimate	standard error	confidence interval
σ_1	.0568	.00085	(.0551, .0585)
σ_2	.0347	.00052	(.0337, .0357)
η_1	.0083	.0212	(−.0341, .0507)
η_2	.0126	.0212	(−.0298, .0550)
ρ	.1460	.0207	(.1046, .1874)

According to Seber and Wild (7), the OLS estimators are adequate for estimation purposes but the GLS estimates are better because it allows us to make inferences about the parameter models. Using our estimates for variances and correlations with the definition of the generalized sum of squares, as given in Seber and Wild (7), the loss function for GLS estimation is:

$$L(\boldsymbol{\theta}) = \frac{1}{2(1-\widehat{\rho}^2)} \left\{ \sum_{k=1}^{N} \left[\frac{\varepsilon_1^k}{\widehat{\sigma}_1}\right]^2 + \sum_{k=1}^{N} \left[\frac{\varepsilon_2^k}{\widehat{\sigma}_2}\right]^2 - 2\widehat{\rho} \sum_{k=1}^{N} \frac{\varepsilon_1^k}{\widehat{\sigma}_1} \frac{\varepsilon_2^k}{\widehat{\sigma}_2} \right\}, \qquad (3.9)$$

where

$$\varepsilon_t^k = Y_t^k - r_t - b_t e^{-\lambda_t (x_k + t)},$$

and $\widehat{\rho}$, $\widehat{\sigma}_1$, $\widehat{\sigma}_2$ are the estimates from the estimates based on the OLS residuals. The GLS estimate of $\boldsymbol{\theta}$ is that value $\widehat{\boldsymbol{\theta}}$ such that $L(\widehat{\boldsymbol{\theta}}) \leq L(\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$. Again, using a numerical optimization program, we found that

$$\widehat{r}_0 = .02515, \quad \widehat{r}_1 = .01537,$$
$$\widehat{b}_0 = 1.5968, \quad \widehat{b}_1 = .27704,$$
$$\widehat{\lambda}_0 = .15138, \quad \widehat{\lambda}_1 = .09811.$$

Note that the GLS estimates are very close to the OLS estimates, as expected. According to Seber and Wild (7),

$$Var(\widehat{\boldsymbol{\theta}}) \approx \left[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} L(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right]^{-1}.$$

Calculation of a numerical Hessian, and the corresponding variance-covariance matrix yields:

$$
\begin{aligned}
&Var(\widehat{r}_0) \approx 6.315 \times 10^{-6}, && Var(\widehat{r}_1) \approx 8.076 \times 10^{-6}, \\
&Var(\widehat{b}_0) \approx 1.3635, && Var(\widehat{b}_1) \approx .04746, \\
&Var(\widehat{\lambda}_0) \approx 8.751 \times 10^{-4}, && Var(\widehat{\lambda}_1) \approx 1.0448 \times 10^{-3}, \\
&Cov(\widehat{r}_0, \widehat{r}_1) \approx -7.732 \times 10^{-7}, && Cov(\widehat{b}_0, \widehat{b}_1) \approx -.02723, \\
&Cov(\widehat{\lambda}_0, \widehat{\lambda}_1) \approx -1.008 \times 10^{-4}.
\end{aligned}
$$

According to Seber and Wild (7), the estimators are approximately normally distributed with the above variances and covariances, even if the ε_t^k are not normally distributed. The following inferences are based on this approximation. First of all, note that we can reject the hypothesis of no inflation, $H_0 : r_0 = 0$ or $H_0 : r_1 = 0$, at a significance level of less than .0001. In testing the null hypothesis, $H_0 : b_0 = b_1$, we find that the hypothesis of a time-homogeneous parameter cannot be rejected because the $Z-$statistic is equal to

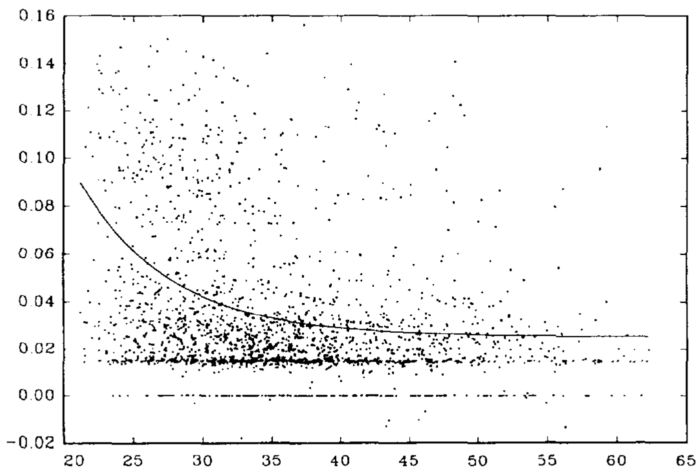
$$|Z| = \frac{\left|\widehat{b}_0 - \widehat{b}_1\right|}{\sqrt{Var(\widehat{b}_0) + Var(\widehat{b}_1) - 2\,Cov(\widehat{b}_0, \widehat{b}_1)}} = 1.09.$$

Note that in this case time-homogeneity in the merit function is restricted to the three year observation period. Any inferences on time-homogeneity over longer periods can only be done if longer data is available. However, the test formula for a longer term would be similar to the test statistic presented here. Next, we test the null hypothesis, $H_0 : \lambda_0 = \lambda_1$. Again, we find that we can not reject time-homogeneity in the parameter λ because $|Z| = 1.16$.

In conclusion, we find that the merit function, ψ_z, is homogeneous in time and so it does not vary from one period to the next.

This section concludes with a plot of the regression. Figure 2 shows a scatter plot of $\log_e[AS^k_{[x_k]+1}/AS^k_{[x_k]}]$ from the acquired pension plan data. Also shown is the regression function $\widehat{r} - \widehat{b}\,e^{-\widehat{\lambda}x}$. Note the extreme variability in this data.This means that no parametric formula will be much better than any other with respect to fit.

FIGURE 2
A Plot of the Log-Ratios $\log_e[AS^k_{[x]+1}/AS^k_{[x]}]$ and
the Regression Function $\widehat{r} - \widehat{b}\,e^{-\widehat{\lambda}x}$ versus Attained Age x.



4 A Service-Based Salary Function

In this section, we investigate an alternate approach to the traditional age-based salary function. Specifically, we let the salary function, denoted as S_s, be a function of attained service, denoted as $s \geq 0$. Again, S_s is a non-decreasing function in s that reflects increases in salary due to inflation and merit (seniority). If AS_s is the actual salary of a person with a service of s, then the predicted salary at $t > s$ is

$$AS_s \times \frac{S_t}{S_s}. \tag{4.1}$$

Before proceeding, we need to give a definition for a service-based *valuation salary function.* We write

$$S_s = \exp\left\{\xi \times s + \int_0^s \psi_z dz\right\}. \tag{4.2}$$

In this definition, ξ is a constant inflation rate while ψ_z is the instantaneous rate of increase due to service. As before, we assume that

$$\psi_z = \beta e^{-\lambda z}, \beta > 0, \lambda > 0. \tag{4.3}$$

Note that ψ_z does not change with time, a hypothesis that we will test. Moreover, ψ_z is a decreasing function in z and so relative merit increases are smaller for long-service employees than for short-service employees, another hypothesis that we will test. As before, we use the same $N = 2231$ observations, which we now denote as $[s_k, AS^k_{[s_k]}, AS^k_{[s_k]+1}, AS^k_{[s_k]+2}]$ for $k = 1, 2, \ldots, N$. In this case, $s_k = x_k - a_k$ denotes the service for employee k.

If we plot s_k versus $\log_e[AS^k_{[s_k]}]$, then we would find the same pattern as in Figure 1. Using the same statistical model as the age-based model, we get

$$Y_t^k = r_t + b\,e^{-\lambda(s_k+t)} + \varepsilon_t^k, \tag{4.4}$$

where

$$\begin{aligned} Y_t^k &= \log_e\left[AS^k_{[s_k]+t+1}/AS^k_{[s_k]+t}\right], \\ r_t &= \int_t^{t+1} \xi_z\,dz, \\ b &= \beta\lambda^{-1}(1-e^{-\lambda}). \end{aligned} \tag{4.5}$$

Using an ordinary sum of squares, we found that the OLS estimates were

$$\begin{aligned} \widehat{r}_0 &= .02418, \quad \widehat{r}_1 = .01608, \\ \widehat{b}_0 &= .08408, \quad \widehat{b}_1 = .04115, \\ \widehat{\lambda}_0 &= .22034, \quad \widehat{\lambda}_1 = .17408. \end{aligned} \tag{4.6}$$

These parameter estimates will be used later. For future reference, we note that $\widehat{\beta}_0 = \widehat{b}_0\,\widehat{\lambda}_0(1-e^{-\widehat{\lambda}_0})^{-1} = .09368$. Using the residual errors, we concluded that $\sigma_1 \neq \sigma_2$, $\eta_1 = 0$, $\eta_2 = 0$ and $\rho \neq 0$, which is the same result we got for the age-based model. The estimates and confidence intervals for these standard deviations and correlations were almost the same as those in the age-based model. Using our estimates for variances and correlations with the definition of the generalized sum of squares, the GLS method yielded the following estimates:

$$\widehat{r}_0 = .02134, \quad \widehat{r}_1 = .01612,$$
$$\widehat{b}_0 = .08356, \quad \widehat{b}_1 = .04134,$$
$$\widehat{\lambda}_0 = .21858, \quad \widehat{\lambda}_1 = .17524.$$

As expected, the GLS estimates are very close to the OLS estimates. Calculation of a numerical Hessian, and the corresponding variance-covariance matrix yields:

$$Var(\widehat{r}_0) \approx 8.608 \times 10^{-6}, \qquad Var(\widehat{r}_1) \approx 4.561 \times 10^{-6},$$
$$Var(\widehat{b}_0) \approx 2.911 \times 10^{-4}, \qquad Var(\widehat{b}_1) \approx 1.136 \times 10^{-4},$$
$$Var(\widehat{\lambda}_0) \approx 2.586 \times 10^{-3}, \qquad Var(\widehat{\lambda}_1) \approx 3.002 \times 10^{-3},$$
$$Cov(\widehat{r}_0, \widehat{r}_1) \approx -7.742 \times 10^{-7}, \quad Cov(\widehat{b}_0, \widehat{b}_1) \approx -2.155 \times 10^{-5},$$
$$Cov(\widehat{\lambda}_0, \widehat{\lambda}_1) \approx -3.061 \times 10^{-4}.$$

Note that we can reject the hypothesis of no inflation, $H_0 : r_0 = 0$, $H_0 : r_1 = 0$. Also, we accept the hypothesis $H_0 : \lambda_0 = \lambda_1$. Unlike the age-based model, we reject the hypothesis $H_0 : b_0 = b_1$ because $|Z| = 1.995$. A plot of the regression function versus the data would be almost identical to Figure 2. The fit for the service-based model was almost identical to the fit for the age-based model. It would be interesting to compare the fit of these models with other pension plan data.

5 The Effect on Salary Gains and Losses

In this section, we examine the age-based and service-based salary functions under the Projected Unit Credit (PUC) and Entry Age Normal - level %

(EAN) funding methods. Specifically, we will look at the variance of the individual salary gains and losses. We find that the variance from our aged-based parametric formula is smaller than the variance from the IRM method. Throughout this section, we use the 1983 GAM valuation mortality table, a valuation interest rate of 7%, and assume retirement at age 65. The only benefit in our pension plan example is a life annuity with a 5-year guarantee. Moreover, the benefit amount is equal to 2% of an employee's final salary times the number of years of credited service. We investigate four different salary functions. The first one will be parametric and age-based, while the second function will be parametric and service-based. The last two functions are tabulated age-based rates where one table was provided by from the valuation actuary of this plan and the other table was constructed by using the traditional IRM technique with quinquennial age groups.

Our parametric aged-based salary function is based on equation (1.5), where x is the age of the employee. For the first function, we let $\xi = .043$, $\beta = 1.744$ and $\lambda = .15192$. The parameter ξ is set equal to an inflation rate of 4.3% rather than the estimated inflation rate so that the resulting function is increasing at about the same rate as the salary function used by the valuation actuary. The parameter values β and λ are displayed in equation (3.6) and were found by using half the data. We did it this way because a salary function that is estimated from all the data would be unfairly competitive when we compare it with other salary functions. These parameters yield

$$S_x^{age} = \exp\left\{.043x \ + 11.480(1 - e^{-.15192x})\right\}. \qquad (5.1)$$

Our service-based salary function is also based on equation (1.5) except that x is replaced with s, the service of the employee. For the service-based function, we let $\xi = .043$, $\beta = .09368$ and $\lambda = .22034$. The parameter ξ is set equal to the inflation rate of 4.3% rather than the estimated inflation rate, which is consistent with the inflation rate in the age-based formula. The parameter values β and λ are displayed in equation (4.6) and were found again by using half the data. These parameters yield

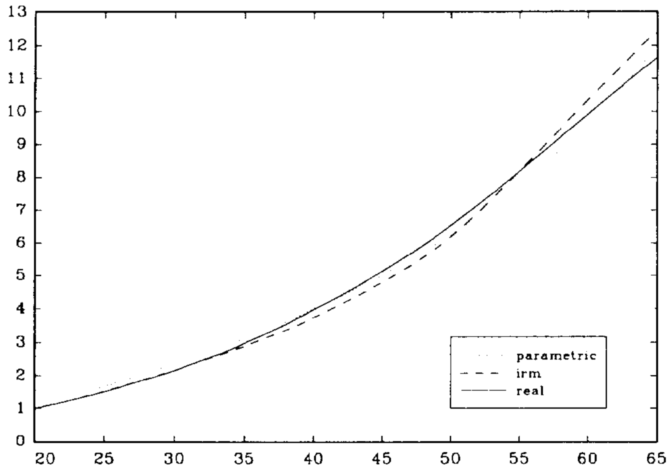
$$S_s^{srv} = \exp\left\{.043s + .61664(1 - e^{-.22034s})\right\}. \tag{5.2}$$

Next, the salary function supplied by the valuation actuary is an age-based function. It is denoted as S_x^{real} and the function values are given in Table 2. Finally, we used the traditional IRM technique to generate the last age-based salary function. The estimate was constructed by first grouping the data in quinquennial age groups, taking the weighted average by salary in each group, and performing a linear interpolation to get the single age values. The estimate is denoted as S_x^{irm} and the function values are in Table 2. However, estimating S_x in this way yields an estimate that inherits its inflation rate from the data. Since our objective is to give a fair comparison of this estimate with S_x^{age} and with S_x^{real}, we adjusted the IRM estimate so that the resulting function is increasing at about the same rate as the salary function used by the valuation actuary, S_x^{real}. Figure 3 shows a plot of our three age-based salary functions, S_x^{age}, S_x^{real} and S_x^{irm}. Note that all three functions are increasing at about the same rate.

TABLE 2
A Real Salary Function vs The Traditional IRM Salary Function

x	S_x^{real}	S_x^{irm}	x	S_x^{real}	S_x^{irm}	x	S_x^{real}	S_x^{irm}
20	1.000	1.000	35	2.974	2.890	50	6.539	6.190
21	1.089	1.089	36	3.159	3.046	51	6.859	6.555
22	1.184	1.187	37	3.350	3.206	52	7.186	6.956
23	1.284	1.290	38	3.548	3.744	53	7.518	7.366
24	1.391	1.399	39	3.753	3.551	54	7.855	7.785
25	1.504	1.515	40	3.965	3.736	55	8.196	8.210
26	1.622	1.637	41	4.183	3.931	56	8.540	8.641
27	1.747	1.764	42	4.409	4.135	57	8.887	9.075
28	1.878	1.896	43	4.643	4.347	58	9.235	9.510
29	2.016	2.031	44	4.884	4.568	59	9.584	9.942
30	2.159	2.167	45	5.133	4.798	60	9.933	10.37
31	2.308	2.306	46	5.390	5.037	61	10.28	10.79
32	2.464	2.445	47	5.658	5.286	62	10.63	11.21
33	2.627	2.589	48	5.939	5.559	63	10.97	11.64
34	2.797	2.737	49	6.232	5.860	64	11.31	12.08

FIGURE 3
A Plot of the Age-Based Functions: S_x^{age}, S_x^{real} and S_x^{irm} versus x.



Using the same notation as before, we describe the salary gain/loss under both the PUC and EAN funding methods. Remember that we had salary data at time $t = 0, 1, 2$ and that the age x_k of employee k was at the time $t = 0$. The salary gain/loss analysis is done at $t = 2$. First, we define the accrued liability at $t = 2$, denoted as $AL_{[x_k]+2} > 0$, for an employee aged $x_k + 2$ who was hired at age a_k. Under the PUC method, the accrued liability is equal to

$$AL^{PUC}_{[x_k]+2} = \frac{x_k + 2 - a_k}{65 - a_k} PVFB_{[x_k]+2}, \tag{5.3}$$

where $x_k + 2 - a_k$ is the past credited service and $65 - a_k$ is the total of past and future credited service. The quantity $PVFB_{[x_k]+2}$ represents the present value of the future benefits for an employee aged $[x_k]+2$ with a salary of $AS^k_{[x_k]+2}$. See Anderson (1) or Berin (2) for details on the calculation of $PVFB$. Next, the accrued liability under the EAN method is

$$AL^{EAN}_{[x_k]+2} = PVFB_{[x_k]+2} - PVFB_{a_k}\frac{PVFSAL_{[x_k]+2}}{PVFSAL_{a_k}}, \tag{5.4}$$

The quantity $PVFSAL_{[x_k]+2}$ represents the present value of the future salaries for an employee aged $[x_k] + 2$ with a current salary of $AS^k_{[x_k]+2}$. See Anderson (1) or Berin (2) for details on the calculation of $PVFSAL$.

In Table 3, we present a summary of our valuation results at time $t = 2$. In this table, you will find the *actual total accrued liabilities* for the EAN and PUC valuation methods. These are calculated as follows:

$$\sum_{k=1}^{N} AL^{EAN}_{[x_k]+2} \quad \text{and} \quad \sum_{k=1}^{N} AL^{PUC}_{[x_k]+2}. \tag{5.5}$$

Also shown are the *expected total accrued liabilities* for time $t = 2$ which were calculated with the data at time $t = 1$. Next is the *total salary gain* and *total mortality loss* which reconcile the difference between the actual and expected liabilities. See Anderson (1) or Berin (2) for details on these calculations. The individual salary gain is calculated as

$$SALGAIN_{[x_k]+2} = AL_{[x_k]+2}\left\{\frac{S_{[x_k]+2}}{S_{[x_k]+1}} \times \frac{AS^k_{[x_k]+1}}{AS^k_{[x_k]+2}} - 1\right\}. \tag{5.6}$$

This last formula is the same for either the EAN and PUC methods. If you inspect Table 3, you will find that the total accrued liabilities for the age-based function is slightly smaller than the service-based function.

TABLE 3
A Comparison of Total Accrued Liabilities, (expressed in units of 1,000)

	S_x^{age} EAN	S_s^{srv} EAN	S_x^{age} PUC	S_s^{srv} PUC
estimated liability	143,131	141,092	118,558	118,113
actual liability	139,539	137,358	115,599	114,987
salary gain	3,861	4,001	3,193	3,361
mortality loss	269	267	234	235

Next, note that if we define

$$Q_{[x_k]+2} = \frac{S_{[x_k]+2}}{S_{[x_k]+1}} \times \frac{AS^k_{[x_k]+1}}{AS^k_{[x_k]+2}} = 1, \qquad (5.7)$$

then $SALGAIN_{[x_k]+2} = 0$ and we have perfect predictions. To measure the closeness of Q_x to 1, we will calculate the mean-squared-error (MSE), variance (Var), and absolute bias ($Bias$) of Q_x for the four salary functions under the two funding methods. We found that changing the inflation affected the bias directly but had little effect on the variance. Therefore, the criteria for comparing the salary functions will be the variance. Using the weights

$$w_{[x_k]+2} = \frac{AL_{[x_k]+2}}{\sum_{k=1}^{N} AL_{[x_k]+2}},$$

we define

$$\begin{aligned} m_1 &= \sum_{k=1}^{N} w_{[x_k]+2} \times Q_{[x_k]+2}, \\ m_2 &= \sum_{k=1}^{N} w_{[x_k]+2} \times [Q_{[x_k]+2}]^2, \\ Var &= m_2 - [m_1]^2, \\ Bias &= |m_1 - 1|, \\ MSE &= Var + [Bias]^2. \end{aligned}$$

Using these measures we find that $MSE = 0$ if and only if $SALGAIN_{[x_k]+2} = 0$ for all $k = 1, \ldots, N$. This is true because

$$MSE = \sum_{k=1}^{N} w_{[x_k]+2} \frac{[SALGAIN_{[x_k]+2}]^2}{[AL_{[x_k]+2}]^2}.$$

Moreover, $MSE = 0$ if and only if $Var = 0$ and $Bias = 0$. Table 4 shows MSE, Var, and $[Bias]^2$ for the four different salary functions under EAN and PUC. We found that

$$Var[S_x^{age}] < Var[S_x^{irm}] < Var[S_x^{real}],$$

concluding that the age-based parametric model did very well. The service-based model did not do a very good job in this case. Maybe, the service-based model could do better than the age-based model in other cases.

TABLE 4
A Comparison of MSE, Var, and $[Bias]^2$ (expressed in units of 100)

	EAN			PUC		
Function	MSE	Var	$[Bias]^2$	MSE	Var	$[Bias]^2$
S_s^{srv}	.2350	.1502	.0848	.2383	.1528	.0855
S_x^{age}	.2233	.1467	.0766	.2227	.1464	.0763
S_x^{irm}	.2523	.1472	.1051	.2525	.1473	.1052
S_x^{real}	.2578	.1509	.1069	.2561	.1503	.1058

References

1. Anderson, A.W. *Pension Mathematics for Actuaries*, 2nd Edition. Winsted, Connecticut: ACTEX Publications (1992).
2. Berin, B.N. *The Fundamentals of Pension Mathematics*. Schaumburg, Illinois: Society of Actuaries (1989).
3. Bowers N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. *Actuarial Mathematics*. Schaumburg, Illinois: Society of Actuaries (1989).
4. Manoukian, E.B. *Modern Concepts and Theorems of Mathematical Statistics*. New York, NY: Springer-Verlag (1985).
5. Marples, W.F. "Salary Scales," *TSA* XIV, Part I (1962): 1-50.
6. Schoenly, S.G. *Pension Topics-Course P-460C Study Note*. Schaumburg, Illinois: Society of Actuaries (1991).
7. Seber, G.A.F. and Wild, C.J. *Nonlinear Regression*. New York, NY: John Wiley (1989).