EDUCATION COMMITTEE

OF THE

SOCIETY OF ACTUARIES

ADVANCED TOPICS IN GENERAL INSURANCE STUDY NOTE

CREDIBILITY WITH SHIFTING RISK PARAMETERS

Stuart Klugman, FSA, CERA, PhD

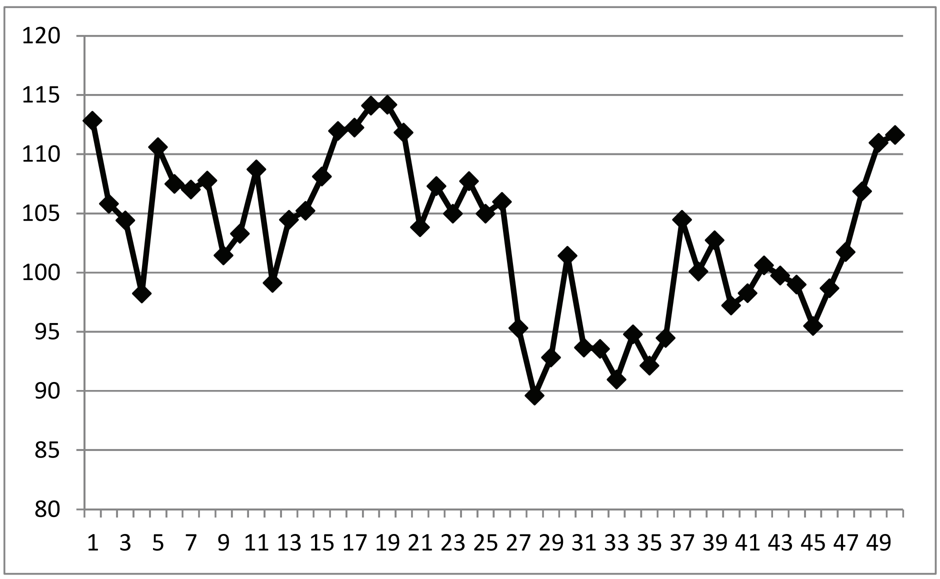
GIADV-101-14 **Printed in the U.S.A.**

Credibility with Shifting Risk Parameters

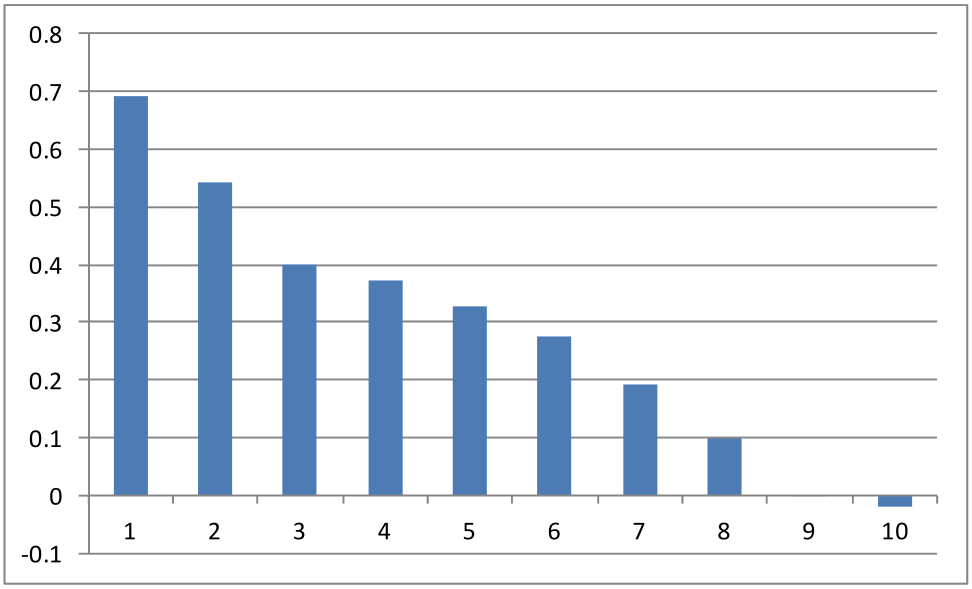
Stuart Klugman, FSA, CERA, PhD

Introduction

Consider the following problem. You have a sequence of observations taken over several time periods. For example, the observations might be observed pure premiums for a large employer's workers compensation coverage. Your goal is to forecast the next observation in the series. To simplify the problem, assume that all external influences have been removed (such as inflationary trend and business cycles). Further suppose the amount of data that produced each observation is relatively small. To fix the idea, consider the following time series of 50 points.



It appears there may be something going on here as the points are consistently above the average of 103.00 for the first half, then below, and then above again. Perhaps there is some correlation between adjacent points. The following chart shows the sample autocorrelations by lag.



This pattern implies that an autoregressive process of order one (AR(1)) may be appropriate (more on this later).

For a problem like this there are (at least) three approaches. The first is to assume independence and no other information. Then (assuming equal sample sizes) the best (as in minimum variance) unbiased linear estimator of the true mean is the sample mean, in this case 103.00.

If we believe that the observations are in fact dependent, and that an AR(1) process is appropriate, then the best prediction is a weighted average of the most recent observation and the sample mean, using the estimated lag one autocorrelation as the weight. This forecast, using the sample lag one autocorrelation as the estimate, is 0.69(111.64) + 0.31(103.00) = 108.96.

The time series approach is motivated by an assumption that observations are correlated (and typically that the correlation depends only on the time between the observations). Another way of looking at this is that the means are changing over time. As an example, if a factory has a culture of safety it may persist for a few years, but as there is turnover in management or workers, the culture may change. In general, if there is such correlation, it may make sense to give more weight to recent observations (as was the result with the AR(1) model).

Finally, because the sample size is small, a credibility approach may be appropriate. The credibility approach is motivated by two factors. First, with limited data, any forecast based only on that data may have considerable variability. Second, there may be multiple forecasts (e.g., for other insured groups) and credibility theory says that in aggregate, multiple forecasts are more accurate when the individual means are credibility-weighted with an overall mean. This leads to a reduction in weight on each observation.

This note provides a brief review of time series and credibility approaches when treated separately as both are treated elsewhere in actuarial education. It then demonstrates a method of combining the two in a single analysis (though not all combinations will be covered). The notion of combining the two was explored in the paper by Mahler (1990). Similarities to his work will be noted as the model for combining the two is developed. The motivation for the method used for parameter estimation comes from two papers by Frees, Young, and Luo (1999 and 2001). In the remainder of this note they will simply be referred to as Mahler and Frees respectively.

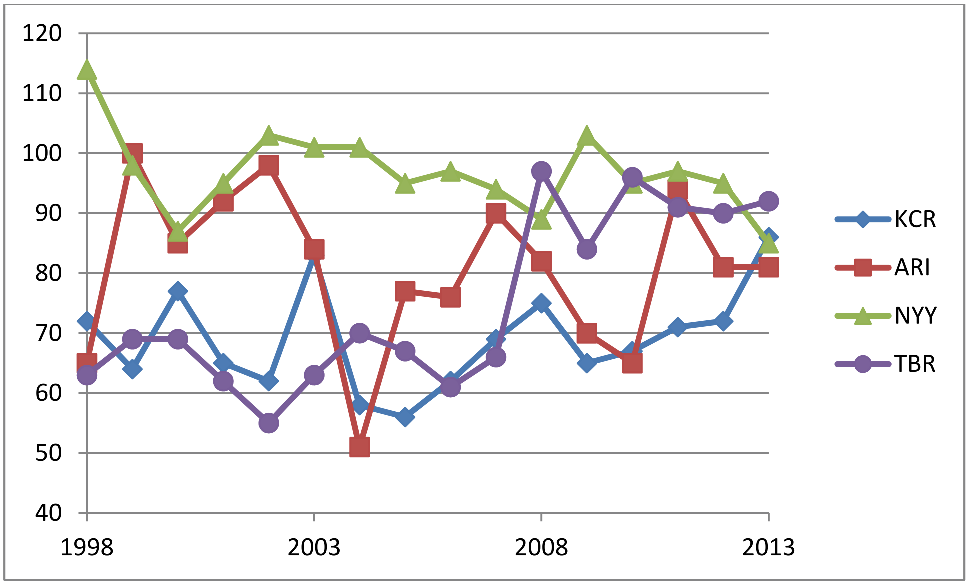
As noted in the simple example, it is important to understand what is not meant by "change over time." This does not refer to trend (there are other credibility methods for trend). It also does not refer to abrupt changes at specific time points (such as a change in marketing, underwriting, or regulation) nor to economic cycles. All of these should be accounted for with data adjustments before a time series or credibility approach is applied.

Mahler uses the losing percentage of Major League Baseball teams as an illustration. There is no overall trend as in total the collection of teams always loses half of its games. What is left is unaccounted for random fluctuations in the mean with some temporal connection. In baseball, many players return for the next season and thus there is likely a positive correlation between records of one year and the next. But over time there will be significant changes and later years are likely to be essentially independent of earlier years. A similar example is used here to illustrate the various approaches.

For the baseball example used here, to keep the calculations simple, the data are only from the sixteen years 1998-2013. The last expansion (to 30 teams) occurred in 1998 so the number of years of observation is the same for all teams (this is not a requirement in any of the formulas that will be presented). The observation is the number of wins in each season. While exposures are not exactly 162 for each team each year,[1] the differences are small enough that they can be ignored. In all the calculations that follow, exposures are taken as 1 rather than 162. This has no effect on any of the estimates other than the process variance estimate. The credibility formulas presented allow for exposures to differ for each observation.

As an illustration of the patterns, the following graph has the wins each year for the team with the fewest total wins (Kansas City Royals), the team closest to an average of 81 wins per year (Arizona Diamondbacks), the team with the most total wins (New York Yankees), and a team with an interesting pattern (Tampa Bay Rays).[2] It is interesting to note that the average sample lag one autocorrelation over the 30 teams is 0.26 (though the range is from –0.32 to 0.66) and the average sample autocorrelations for other lags are close to zero (though again there is considerable variation among teams). Autocorrelations have a high standard error (a rough approximation is the square root of the ratio of 1 and the sample size, in this case sqrt(1/16) = 0.25) and so this variation is not surprising.

[1] For example, in 2013 Tampa Bay and Texas had a playoff game for the final American League wild card spot. This 163rd game counted in the final standings. Or, if a rained out game has not been rescheduled by the end of the season and its outcome does not affect playoff eligibility, the game is not played.
[2] During this period there was a change in ownership and management that had a direct effect on the playing field (Keri, 2011). In this note's examples, forecasts will be given for each of these four teams.



Time Series Approach

This section is not an exhaustive discussion of time series modeling. Rather, three commonly used models will be discussed. Readers should consult one of the many texts on time series modeling to learn about the range of models and more formal methods of model selection and validation. In all cases in this note it is assumed that the observations have been transformed to become weakly stationary. That means the unconditional mean is the same at all times and that the covariance between two observations depends only on the time that separates them. One way to think about this is that based on first and second moments, if you are looking at a sequence of observations it is not possible to identify the time period from which they originated.

One model, the autoregressive model of order 1 has already been mentioned. The definition is that the observation at time *t* is related to the previous observation by

$$x_t = \rho x_{t-1} + (1-\rho)\mu + \varepsilon_t$$

where $-1 < \rho < 1$ and $\{\varepsilon_t\}$ is a sequence of independent random variables with a mean of zero and a standard deviation of σ_ε. Often, a normal distribution is assumed, but that is not necessary to ensure that this process is stationary. With these assumptions, $Var(x_t) = \sigma_x^2 = \sigma_\varepsilon^2 / (1-\rho^2)$ and $Corr(x_t, x_{t-s}) = \rho^s$. Parameter estimation can be done in an ad hoc manner (for example, using the sample mean for μ, the sample lag one correlation for ρ, and the sample standard deviation of the residuals, $\hat{\varepsilon}_t = x_t - \hat{\rho} x_{t-1} - (1-\hat{\rho})\hat{\mu}$, to estimate σ_ε). For the 50 observations presented in the introduction, those estimates are

$$\hat{\mu} = 103.00, \hat{\rho} = 0.6927, \text{and } \hat{\sigma}_\varepsilon = 4.559 .$$

A second ad hoc approach is to note that the AR(1) model is a regression formula. The least squares estimate is

$$x_t = 0.71746x_{t-1} + 29.030$$

with 47 degrees of freedom from which a mean of 102.75 can be inferred. The standard error is 4.555. A third approach is to make a distributional assumption and use maximum likelihood. Running those same 50 observations through the fpp package in R (which defaults to the normal distribution and MLE) produces

$$\hat{\mu} = 103.94, \ \hat{\rho} = 0.7381, \text{ and } \hat{\sigma}_\varepsilon = 4.510 .$$

The forecast for future observations is a simple matter of applying the AR(1) model formula with the error term for future values set equal to zero (its expected value). The forecast error is also easy to obtain (noting that this version reflects only process variability and not estimation error). Using the R estimates,

$$\begin{aligned}
&\hat{x}_{51} = \hat{\rho}x_{50} + (1-\hat{\rho})\hat{\mu} = 0.7381(111.64) + 0.2619(103.94) = 109.62 \\
&Var(x_{51} \mid x_{50}) = Var[\rho x_{50} + (1-\rho)\mu + \varepsilon_{51} \mid x_{50})] = \sigma_\varepsilon^2 \\
&V\hat{a}r(x_{51} \mid x_{50}) = \hat{\sigma}_\varepsilon^2 = 20.34.
\end{aligned}$$

The formulas are conditional upon the most recent observation because it is known and not random.

The baseball data is more difficult to analyze in that 16 observations are too few to obtain reliable estimates of the autocorrelations. For this analysis we will use the average sample lag one autocorrelation presented earlier and begin with the AR(1) model. Using the average sample lag one autocorrelation of 0.26155,[3] the results for the four teams (where the forecast is 0.26155 times the 2013 wins plus 0.73845 times the average wins) are:

Team	2013 wins	Average wins	Forecast
Kansas City	86	69.00	73.45
Arizona	81	80.69	80.77
Tampa Bay	92	74.69	79.22
New York Yankees	85	96.81	93.72

When only the lag one autocorrelation is essentially nonzero an MA(1) model may be appropriate. This model is

$$x_t = \mu + \varepsilon_t - \theta\varepsilon_{t-1}, -1 < \theta < 1.$$

It then follows that

$$\begin{aligned}
&Var(x_t) = Var(\varepsilon_t - \theta\varepsilon_{t-1}) = Var(\varepsilon_t) + Var(-\theta\varepsilon_{t-1}) = (1+\theta^2)\sigma_\varepsilon^2 \\
&Cov(x_t, x_{t-1}) = Cov(\varepsilon_t - \theta\varepsilon_{t-1}, \varepsilon_{t-1} - \theta\varepsilon_{t-2}) = Cov(-\theta\varepsilon_{t-1}, \varepsilon_{t-1}) = -\theta\sigma_\varepsilon^2 \\
&Corr(x_t, x_{t-1}) = \frac{-\theta}{1+\theta^2}.
\end{aligned}$$

[3] An alternative would be to use each team's own sample lag one autocorrelation for its forecast. The approach taken here is more in line with the models presented later in this note.

With the sample lag one autocorrelation of 0.26155, it can be inferred that $\theta = -0.28241$. You may recall that any MA model can be rewritten as an AR model, but with an infinite number of terms. In this case, it is

$$x_t = \mu - \theta(x_{t-1} - \mu) - \theta^2(x_{t-2} - \mu) - \cdots + \varepsilon_t .$$

This places geometrically decreasing weights on the deviations from the mean. Because the parameter estimate is negative, the signs of the weights alternate. When applied to the four teams, the results are:

Team	2013 wins	Average wins	Forecast
Kansas City	86	69.00	73.61
Arizona	81	80.69	81.13
Tampa Bay	92	74.69	78.59
New York Yankees	85	96.81	93.65

These forecasts are not much different. With this model the average wins receives a weight of 0.7798, slightly more than with the AR(1) model. The most recent season also gets more weight, at 0.2824. The main difference is that the second most recent season receives a weight of –0.0798.

Instead of the actual observations having an MA(1) model, let the differences of the observations have this model (in time series terminology, this is an ARIMA(0,1,1) model) with no constant term. Having no constant makes sense as otherwise a team would have an ever increasing (or decreasing) expected number of wins. The model is

$$y_t = x_t - x_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}, -1 < \theta < 1.$$

For our baseball data, the average sample lag one correlation of the differences is –0.29451 which can be solved for $\hat{\theta} = 0.32577$.

This model can also be written as a pure AR model:

$$x_t = (1-\theta)[x_{t-1} + \theta x_{t-2} + \theta^2 x_{t-3} + \cdots] + \varepsilon_t.$$

This model is called *simple exponential smoothing* and is one of the models examined by Mahler (though he approached it as a reasonable way to impose decreasing emphasis on older observations rather than from an ARIMA modeling perspective). In the case of baseball forecasting, the weights are 0.6742, 0.2196, 0.0716,…, 0.00000003. Because the infinite series ends at 16 terms, the weights do not sum to exactly 1 (though in this case they come extremely close). In practice, the weights should be rescaled.

Consider predicting the 2014 win total for the Tampa Bay Rays. Their wins for the 16 years were 63, 69, 69, 62, 55, 63, 70, 67, 61, 66, 97, 84, 96, 91, 90, and 92. Applying the weights leads to a forecast of 91.50. The other teams have forecasted values of KCR – 81.20, ARI – 81.48, and NYY – 88.45.

Note that in all ARIMA models, forecasts more than one period ahead can be made by repeated applications of the formula with future error terms set equal to zero.

Incorporating Credibility

We first turn to credibility methods that do not take into account correlation over time. Only greatest accuracy credibility will be discussed. Being model-based, this method can be extended to the time-varying situation in a formal manner. In this section we review the Bühlmann-Straub empirical Bayes approach and apply it to the baseball data. To keep this section simple, only the case of equal exposures will be covered. The generalization to arbitrary exposures will be provided in the next section.

To develop the model, let X_{it} be the random observation from group i at time t. For this simple setting we will assume k groups and n time periods for each group. Here and throughout, the term group will be used to describe the entity for which we want to make a forecast. In the baseball example, a group is a team, k = 30, and n = 16.

The model does not specify a distribution, only the first two moments. The model has two levels. The top level is a model for an unobservable random quantity, the mean for group i. Let ξ_i be that mean and let $E(\xi_i)=\mu$ and $Var(\xi_i)=\tau^2$. It is further assumed that these k random variables are independent. This process assigns the means to each group. It appears that for baseball teams this process has assigned a high mean to the New York Yankees and a low mean to the Kansas City Royals. The second level provides the moments of the actual observations. Given the (unobservable) mean, the moments are $E(X_{it}\,|\,\xi_i)=\xi_i$ and $Var(X_{it}\,|\,\xi_i)=\sigma^2$. The objective is to use the available data to estimate the k unknown means.

The empirical Bayes approach determines the estimates in two steps. The first is to assume that the top level mean and the two variances are known and that the estimator must be a linear combination of the observations from the group. The estimate is then the value that minimizes mean squared error. The formula will not be derived here. It is produced in all credibility textbooks and is a special case of the formula derived in the next section. The solution is

$$\hat{\xi}_i = Z\bar{X}_i + (1-Z)\mu, \quad \bar{X}_i = \frac{1}{n}\sum_{t=1}^{n} X_{it}, \quad Z = \frac{n}{n+\dfrac{\sigma^2}{\tau^2}}.$$

The second step is to use the data to estimate the three parameters. The usual estimates are derived by positing reasonable estimators and then making adjustments to make them unbiased. The formulas are

$$\hat{\mu} = \bar{X} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{t=1}^{n} X_{it}$$

$$\hat{\sigma}^2 = \frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{t=1}^{n}(X_{it}-\bar{X}_i)^2, \quad \hat{\tau}^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\bar{X}_i-\bar{X})^2 - \frac{1}{n}\hat{\sigma}^2.$$

The formulas are easily modified when there are differing numbers of time periods per group or differing exposures per observation.

When applied to the baseball data, the results are

$$\hat{\mu} = 80.9646$$
$$\hat{\sigma}^2 = 104.513$$
$$\hat{\tau}^2 = 35.3285.$$

With this model, and data, the same credibility factor is used for all 30 teams. For Tampa Bay, the calculation is

$$Z = \frac{16}{16 + \frac{104.513}{35.3285}} = 0.84396$$
$$\hat{\xi}_{27} = 0.84396(74.6875) + 0.15604(80.9646) = 75.67.$$

Applying the same weights to the other teams yields KCR – 70.87, ARI – 80.73, and NYY – 94.34.

There is an alternative way to get to the same result. The advantage is that unlike the empirical Bayes approach, which relies on manipulations of arbitrarily selected sums of squares, this approach leads directly to the solution. As such, it is easier to generalize, as will be done in the next section.

The alternative is to set up the problem as a random effects linear model. More details about random effects models and actuarial applications can be found in Frees and also in Klinker (2011). The model is

$$X_{it} = \mu + \alpha_i + \varepsilon_{it}, \quad \alpha_i \sim N(0, \tau^2), \quad \varepsilon_{it} \sim N(0, \sigma^2).$$

What distinguishes this model from the more traditional fixed effects model is that the middle term is a random variable. First note that if we did use a fixed effects regression model, the estimator for each group would be the sample mean and hence not apply credibility. As Klinker notes, the motivation that led to the random effects model is not directly applicable to insurance settings. However, as we shall see, it does exactly what we want, shrinking the estimates toward the overall mean. Another key point is that unlike the Bühlmann-Straub approach, a distributional assumption has been incorporated. While we may not believe that the normal distribution is the correct model, we do note that there is a correspondence between the normal distribution and least squares estimation. So it will not be surprising if this alternative approach produces similar results.

We will now derive the estimates for this model. Because the distributions are specified, maximum likelihood estimation could be used. However, variance estimates tend to be biased (recall that for a sample from the normal distribution, the MLE of the variance has n in the denominator rather than the n – 1 that produces an unbiased estimator). A technique known as restricted maximum likelihood (REML) has been developed to address this problem. Rather than derive the REML loglikelihood function (which is done in the 1999 Frees paper), we present the regular loglikelihood function and then point out the difference.

Writing the regular loglikelihood function is a challenge because the observations within a group are dependent (though observations from different groups are independent). For observations in group i, we have

$$E(X_{it}) = E(\mu + \alpha_i + \varepsilon_{it}) = \mu$$
$$Var(X_{it}) = Var(\mu + \alpha_i + \varepsilon_{it}) = \tau^2 + \sigma^2$$
$$Cov(X_{it}, X_{is}) = Cov(\mu + \alpha_i + \varepsilon_{it}, \mu + \alpha_i + \varepsilon_{is}) = Var(\alpha_i) = \tau^2, \quad t \neq s.$$

Then the vector of observations from group *i* has the multivariate normal density function

$$f(x_{i1}, \ldots, x_{in} \mid \mu, \sigma^2, \tau^2) = \frac{1}{(2\pi)^{n/2}[\det(\mathbf{V})]^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \right]$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{in})'$ is a vector of the *n* observations from group *i*, $\boldsymbol{\mu} = (\mu, \ldots, \mu)'$ is a vector containing the mean, and **V** is a matrix with $\sigma^2 + \tau^2$ on the diagonal and τ^2 everywhere else. The loglikelihood for the entire set of observations is then

$$\ln L = -\frac{nk}{2}\ln(2\pi) - \frac{k}{2}\ln[\det(\mathbf{V})] - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{x}_i - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}).$$

Taking the derivative with respect to the mean is relatively simple. The solution is

$$\hat{\mu} = \frac{\sum_{i=1}^{k} sum(\mathbf{V}^{-1}\mathbf{x}_i)}{k[sum(\mathbf{V}^{-1})]} = \frac{\sum_{i=1}^{k}\sum_{t=1}^{n} x_{it}}{kn}$$

due to the fact that the inverse of **V** has the same structure as **V**, with a constant value on the diagonal and a different, but constant, value off the diagonal. The *sum* function adds the elements of the matrix or vector to which it is applied. It can be shown that maximizing the likelihood with respect to the two variance terms results in a formula similar to the empirical Bayes formula, but the estimator of τ^2 has a different denominator. The REML estimator changes the loglikelihood function to

$$\ln L = -\frac{nk}{2}\ln(2\pi) - \frac{k}{2}\ln[\det(\mathbf{V})] - \frac{1}{2}\ln[k \times sum(\mathbf{V}^{-1})] - \frac{1}{2}\sum_{i=1}^{k}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})'\mathbf{V}^{-1}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}).$$

The basis of the estimator is to obtain a likelihood function that does not depend on the mean (note that the mean is replaced by its estimate here). The likelihood is essentially based on the residuals after subtraction of the estimated mean. It turns out that the values that maximize this function exactly match the empirical Bayes estimates.

It should be noted that this exact match occurs only when the total exposures are identical across groups. When total exposures differ among groups, the REML estimates will be biased. What we lose in bias we gain by having an approach that is easy to generalize. That is the subject of the next section.

Before moving on, there is one more element to address and that is how the unknown means are estimated when using a random effects model. It turns out this is done using the same credibility formula as used earlier and so no new formulas are needed.

A General Model

A model that allows for both time dependence and credibility will now be developed. As in the previous section, it will be done two different ways. One is a traditional credibility approach. This will provide the formula for the credibility factors. The second is a random effects model approach. This will provide the method for estimating the parameters.

To begin the credibility discussion, the group subscript will be dropped. The other groups are only needed for parameter estimation. The following discussion mirrors that presented in Klugman, et al. (2012).

Let $X_1, X_2, \ldots, X_n$ be the random variables that represent observations at times 1 through n. The goal is to use these observations to estimate a future random observation, X_{n+d}. Only linear functions will be used. Thus

$$\hat{X}_{n+d} = Z_0 + Z_1 X_1 + \cdots + Z_n X_n .$$

The values that minimize the expected squared error are the solutions to the equations

$$E[X_{n+d}] = Z_0 + \sum_{t=1}^{n} Z_t E(X_t)$$

$$Cov(X_t, X_{n+d}) = \sum_{s=1}^{n} Z_s Cov(X_t, X_s), t = 1, \ldots, n.$$

The solution requires a model with sufficient specification to enable calculation of the required expected values and covariances. One such model will now be specified. In particular, it incorporates autocorrelations. It is a three-level model. The top level states that there is a mean for the group that is stable over time. Let ξ denote that mean. That does not mean each year has this mean but that without additional information (such as from nearby years) this mean is the best representation of a given year's expected value. That mean is drawn from a distribution with mean μ and variance τ^2. This is a standard feature of credibility models. It allows each group to have its own mean.

The second level reflects the dependency structure of the means over time. Let the means for the observations at times 1 through n be $(\theta_1, \ldots, \theta_n)$. These values are drawn from a multivariate distribution where each value has a mean of ξ. To allow for dependence, let $Cov(\theta_t, \theta_s) = \delta_{ts}$. For now, no restrictions will be placed on these values (other than that the covariance matrix must be positive definite). The idea is that the mean at time t is random, but related to the mean at other times. So if one period has a large mean, it may persist for several future periods.

The third level is the observation itself. The value X_t is drawn from a distribution with mean θ_t and variance σ^2 / w_t where w_t is a known weight. This part is similar to the Bühlmann-Straub model.

The next step is to determine the elements of the equations for the credibility weights. In these formulas, t and s can be any of the values $1, 2, \ldots, n, n + d$.

$$
\begin{aligned}
&E[X_t] = E[E(X_t \mid \theta_t)] = E[\theta_t] = E[E(\theta_t \mid \xi)] = E[\xi] = \mu \\
&Var(X_t) = E[Var(X_t \mid \theta_t)] + Var[E(X_t \mid \theta_t)] = E[\sigma^2 / w_t] + Var(\theta_t) \\
&\qquad = \sigma^2 / w_t + E[Var(\theta_t \mid \xi)] + Var[E(\theta_t \mid \xi)] = \sigma^2 / w_t + E(\delta_{tt}) + Var(\xi) \\
&\qquad = \sigma^2 / w_t + \delta_{tt} + \tau^2 \\
&Cov(X_t, X_s) = E[X_t X_s] - \mu^2 = E[E(X_t X_s \mid \theta_t, \theta_s)] - \mu^2 = E[\theta_t \theta_s] - \mu^2 \\
&\qquad = E[E(\theta_t \theta_s \mid \xi)] - \mu^2 = E[\delta_{ts} + \xi^2] - \mu^2 = \delta_{ts} + \tau^2 + \mu^2 - \mu^2 = \delta_{ts} + \tau^2, t \neq s.
\end{aligned}
$$

The equations are then

$$
\begin{aligned}
&\mu = Z_0 + \mu \sum_{t=1}^{n} Z_t \\
&\delta_{t,n+d} + \tau^2 = Z_t \sigma^2 / w_t + \sum_{s=1}^{n} Z_s (\delta_{ts} + \tau^2), t = 1, \ldots, n.
\end{aligned}
$$

Once the second set of equations is solved, the first equation is easily solved as $Z_0 = \left(1 - \sum_{t=1}^{n} Z_t\right)\mu$ which applies the complement of credibility to the overall mean, if known, or its estimate.

The second set of equations is linear and so has the matrix solution

$$
\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix} = \begin{bmatrix} \sigma^2 / w_1 + \delta_{11} + \tau^2 & \delta_{12} + \tau^2 & \cdots & \delta_{1n} + \tau^2 \\ \delta_{21} + \tau^2 & \sigma^2 / w_2 + \delta_{22} + \tau^2 & \cdots & \delta_{2n} + \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} + \tau^2 & \delta_{n2} + \tau^2 & \cdots & \sigma^2 / w_n + \delta_{nn} + \tau^2 \end{bmatrix}^{-1} \begin{bmatrix} \delta_{1,n+d} + \tau^2 \\ \delta_{2,n+d} + \tau^2 \\ \vdots \\ \delta_{n,n+d} + \tau^2 \end{bmatrix}.
$$

The matrix in the middle can be divided into three components. Let **D** be the first component, a diagonal matrix with $\sigma^2 / w_1, \ldots, \sigma^2 / w_n$ on the diagonal. The second component is $\mathbf{\Sigma}$, a matrix with the δ_{ts} values. The final matrix is **T**, in which every element is τ^2. Then, with **S** denoting the right-hand vector,

$$
\mathbf{Z} = (\mathbf{D} + \mathbf{\Sigma} + \mathbf{T})^{-1}\mathbf{S} = \mathbf{V}^{-1}\mathbf{S}.
$$

It will be useful to have the mean squared error both as a function of the general value of **Z** and for the value that minimizes it. The development is (substituting $Z_0 = \left(1 - \sum_{t=1}^{n} Z_t\right)\mu$):

$$E\left[\left(X_{n+d}-Z_0-\sum_{t=1}^{n}Z_tX_t\right)^2\right]=E\left[\left(X_{n+d}-\mu+\sum_{t=1}^{n}Z_t\mu-\sum_{t=1}^{n}Z_tX_t\right)^2\right]=E\left[\left(X_{n+d}-\mu-\sum_{t=1}^{n}Z_t(X_t-\mu)\right)^2\right]$$

$$=Var(X_{n+d})+\sum_{t=1}^{n}\sum_{s=1}^{n}Z_tZ_sCov(X_t,X_s)-2\sum_{t=1}^{n}Z_tCov(X_{n+d},X_t)$$

$$=\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2+\sum_{t=1}^{n}\sum_{s=1}^{n}Z_tZ_s(\delta_{ts}+\tau^2)+\sum_{t=1}^{n}Z_t^2\sigma^2/w_t-2\sum_{t=1}^{n}Z_t(\delta_{n+d,t}+\tau^2)$$

$$=\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2+\mathbf{Z}'(\mathbf{D}+\mathbf{\Sigma}+\mathbf{T})\mathbf{Z}-2\mathbf{Z}'\mathbf{S}$$

$$=\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2+\mathbf{Z}'\mathbf{V}\mathbf{Z}-2\mathbf{Z}'\mathbf{S}.$$

At the optimal value of **Z** the value is

$$\begin{aligned}\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2+\mathbf{Z}'\mathbf{V}\mathbf{Z}-2\mathbf{Z}'\mathbf{S}&=\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2+\mathbf{S}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{S}-2\mathbf{S}'\mathbf{V}^{-1}\mathbf{S}\\&=\sigma^2/w_{n+d}+\delta_{n+d,n+d}+\tau^2-\mathbf{S}'\mathbf{V}^{-1}\mathbf{S}.\end{aligned}$$

The Bühlmann-Straub model is the same but with $\mathbf{\Sigma}=\mathbf{0}$ and all elements of $\mathbf{S}$ equal to τ^2. It should also be noted that the current model is overspecified (adding a constant to τ^2 and subtracting the same amount from each δ produces the same solution).

This model has far too many parameters to be useful. In fact, there would never be enough data points to estimate all of them. For the remainder of this note, the following simplifications will be imposed:

- All groups have the same parameters.
- The covariance terms depend only on the difference between the subscripts. This is consistent with the stationarity requirement when performing time series analysis.

As a result, when the model is extended to multiple groups, there will be common values for

$$\sigma^2,\tau^2,\mathbf{\Sigma}=\begin{bmatrix}\delta_0&\delta_1&\cdots&\delta_{n-1}\\\delta_1&\delta_0&\cdots&\delta_{n-2}\\\vdots&\vdots&\ddots&\vdots\\\delta_{n-1}&\delta_{n-2}&\cdots&\delta_0\end{bmatrix},\text{ and }\mathbf{S}=\begin{bmatrix}\delta_{n+d-1}+\tau^2\\\delta_{n+d-2}+\tau^2\\\vdots\\\delta_d+\tau^2\end{bmatrix}.$$

The total number of parameters is 1 + 1 + n + d = n + d + 2. This model is still overspecified. There are at most n + d + 1 free parameters.

Consider the baseball example with 30 groups and 16 years of observation. We want to forecast the value 1 year ahead. There are 480 data points and only 18 parameters to estimate. But there is a problem. The first term of **S** requires δ_{16}, but there is no data available to estimate the covariance of observations 16 years apart. There are two possible solutions:

1. Base the credibility estimate on only the most recent 15 years, using the first year's observations only to estimate the needed covariance.

2. Propose a further simplification similar to those seen in the time series models. Then extrapolation can be used.

The first solution may be appropriate if there are enough years of data so that the correlations between distant observations are very low. Dropping the first year may make little difference. The second solution enforces a pattern on the covariances that can be extended.

Now that we can obtain the forecasts given the parameter values, it is time to turn to parameter estimates. For that we use the random effects model formulation. For group *i*:

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{in} \end{bmatrix} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \text{ with all vectors independent of each other}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \boldsymbol{\alpha}_i = \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} \sim N(\mathbf{0}, \mathbf{T}), \quad \boldsymbol{\gamma}_i = \begin{bmatrix} \gamma_{i1} \\ \vdots \\ \gamma_{in} \end{bmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{D}_i)$$

The two middle terms can be combined:

$$\mathbf{X}_i = \begin{bmatrix} X_{i1} \\ \vdots \\ X_{in} \end{bmatrix} = \boldsymbol{\mu} + \boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \boldsymbol{\theta}_i = \begin{bmatrix} \theta_{i1} \\ \vdots \\ \theta_{in} \end{bmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma} + \mathbf{T}), \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{D}_i)$$

In Frees the general model with random effects is presented as

$$\mathbf{X}_i = \mathbf{U}_i \boldsymbol{\theta}_i + \mathbf{W}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\theta}_i \sim N(\mathbf{0}, \mathbf{R}), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{D}_i).$$

Our model fits this specification with

$$\mathbf{U}_i = \mathbf{I}, \text{ the identity matrix}$$

$$\mathbf{W}_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \text{ a column vector of 1s}$$

$$\boldsymbol{\beta} = \mu, \text{ a scalar}$$

$$\mathbf{R} = \boldsymbol{\Sigma} + \mathbf{T}$$

$$\mathbf{D}_i = \mathbf{D}_i.$$

The REML estimates are found by maximizing the following function, first stated with the general random effects model, where $\mathbf{V}_i = \mathbf{U}_i \mathbf{R} \mathbf{U}_i' + \mathbf{D}_i$:

$$\ln L = -0.5\left[\sum_{i=1}^{k}\ln\det(\mathbf{V}_i) + \ln\sum_{i=1}^{k}\det(\mathbf{W}_i'\mathbf{V}_i^{-1}\mathbf{W}_i) + \sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{W}_i\hat{\boldsymbol{\beta}})'\mathbf{V}_i^{-1}(\mathbf{X}_i - \mathbf{W}_i\hat{\boldsymbol{\beta}})\right]$$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{k}\mathbf{W}_i'\mathbf{V}_i^{-1}\mathbf{W}_i\right)^{-1}\sum_{i=1}^{k}\mathbf{W}_i'\mathbf{V}_i^{-1}\mathbf{X}_i.$$

For the particular model in this note, where $\mathbf{V}_i = \boldsymbol{\Sigma} + \mathbf{T} + \mathbf{D}_i$ and *sum* means to add all the elements of the given matrix or vector:

$$\ln L = -0.5\left[\sum_{i=1}^{k}\ln\det(\mathbf{V}_i) + \ln\sum_{i=1}^{k} sum(\mathbf{V}_i^{-1}) + \sum_{i=1}^{k}(\mathbf{X}_i - \mathbf{1}\hat{\mu})'\mathbf{V}_i^{-1}(\mathbf{X}_i - \mathbf{1}\hat{\mu})\right]$$

$$\hat{\mu} = \frac{\sum_{i=1}^{k} sum(\mathbf{V}_i^{-1}\mathbf{X}_i)}{\sum_{i=1}^{k} sum(\mathbf{V}_i^{-1})}.$$

All the formulas are in place to make forecasts that combine time series and credibility.

Forecasting with both Time Series and Credibility

The ultimate goal is to make a forecast that is a weighted average of observations from the group plus weight on an external value. In an ideal situation, the weights are determined using these two formulas:

$$\mathbf{Z} = (\mathbf{D} + \boldsymbol{\Sigma} + \mathbf{T})^{-1}\mathbf{S} = \mathbf{V}^{-1}\mathbf{S} \qquad Z_0 = \left(1 - \sum_{t=1}^{n} Z_t\right)\mu.$$

The formulas depend on knowing the correlation structure and having estimates of the mean and the parameters that define the correlation structure.

The estimates can be obtained via REML using the formulas in the previous section.

The challenge when using time series models is that with limited data (and often in actuarial analyses there are few years of data) many ARIMA models will fit well. Hence there may be a variety of correlation structures that make sense. Judgment may be required in selecting a model that leads to a set of weights that makes sense.

Mahler offers a solution that separates the two steps. He separately poses a correlation structure (using one that is as general as possible while still being amenable to estimation) and a pattern of weights. The weights are then selected to minimize the least squares criterion

$$\sigma^2 / w_{n+d} + \delta_{n+d,n+d} + \tau^2 + \mathbf{Z}'\mathbf{V}\mathbf{Z} - 2\mathbf{Z}'\mathbf{S}$$

where the weights are constrained to follow the set pattern. This will lead to a suboptimal result with regard to the least squares criterion, but may come close to being optimal. It has the advantage of producing weights that are easier to explain.

This approach will be illustrated with the baseball example.

Baseball Example

We begin with an analog of the AR(1) model. We know the pattern of correlations for that model. Thus the additional specification is $\delta_g = \delta\rho^g$ where g is the number of periods separating the two observations. This adds two parameters to the Bühlmann-Straub model.

The REML maximization results are[4]

$$\hat{\sigma}^2 = 30.49, \quad \hat{\tau}^2 = 14.77, \quad \hat{\delta} = 95.80, \quad \hat{\rho} = 0.6672, \quad \hat{\mu} = 80.97\,.$$

Because the exposures are the same for all observations, the credibility weights are the same for each team. The vector of weights is:

Year	1998	1999	2000	2001	2002	2003	2004	2005
Weight	0.0185	0.0102	0.0084	0.0080	0.0079	0.0079	0.0079	0.0079
Year	2006	2007	2008	2009	2010	2011	2012	2013
Weight	0.0079	0.0079	0.0081	0.0090	0.0127	0.0300	0.1085	0.4664

The remaining credibility is 0.2728 to be applied to the overall mean.

When applied to the four teams, the 2014 forecasts are: KCR – 80.42, ARI – 81.03, TBR – 86.21, and NYY – 87.07.

Note that the credibility weights do not exactly reflect an AR(1) model for each team with the resulting estimate credibility-weighted with an overall mean of 81. Doing so would have constant weights for the first 15 years and a higher weight for 2013. Other than the additional weight on 2012, the results are pretty close.

This brings us to the approach taken by Mahler. Having posited a correlation structure and estimated the required variances and correlations, any credibility structure can be imposed. This can be done by returning to the mean squared error formula. We know it is minimized at the values in the previous table. However, suppose we accept a suboptimal result that has the nicer AR(1) pattern. That is, minimize the error but with the constraints

$$Z_1 = Z_2 = \cdots = Z_{15} = a, \quad Z_{16} = b, \quad a, b \geq 0, \quad 15a + b \leq 1\,.$$

Before doing that, note that the mean squared error when optimized is 94.47. The optimal value with the constraints occurs at a = 0.01380 and b = 0.5174. This was obtained by setting up the vector of Z values as a function of a and b. Then a minimization program can find the (a, b) pair that achieves the minimum mean squared error (all other values in the formula were previously estimated). The mean squared error increases to 95.53. This appears to be a good tradeoff for a cleaner result. The forecasts are now KCR – 80.86, ARI – 80.92, TBR – 85.14, and NYY – 86.50. The main difference is that the first version places additional weight on the 2012 value. We could try optimizing with unique values for the last two weights rather than the last one. Of course, this must produce a smaller mean squared error.

[4] No optimization method can be assured to provide perfect results. The calculations in this note were done in both Excel and R. At times there were differences in the fourth or fifth significant digit.

Recall that for the baseball data only the average sample lag one autocorrelation coefficient is not close to zero. This implies an MA(1) model may be appropriate. The correlation structure is that δ_0 and δ_1 have unique values and all other covariances are zero. Because the sample sizes are equal, in this formulation σ^2 and δ_0 always appear together as $\sigma^2+\delta_0$. The REML estimates are

$$\widehat{\sigma^2+\delta_0}=104.25, \quad \hat{\tau}^2=31.55, \quad \hat{\delta}_1=31.42, \quad \hat{\mu}=80.97.$$

The resulting weights are:

Year	1998	1999	2000	2001	2002	2003	2004	2005
Weight	0.0459	0.0305	0.0357	0.0339	0.0345	0.0343	0.0344	0.0344
Year	2006	2007	2008	2009	2010	2011	2012	2013
Weight	0.0342	0.0348	0.0329	0.0387	0.0213	0.0733	–0.0819	0.3811

The remaining credibility is 0.1820, to be applied to the overall mean.

When applied to the four teams, the 2014 forecasts are: KCR – 76.87, ARI – 81.24, TBR – 80.30, and NYY – 90.29.

The ARIMA(0,1,1) model used earlier has geometrically decreasing weights (which may be more satisfactory) but does not extend to a credibility context. After taking differences, all 30 teams now have an expected value of zero. There is no way to model the variation in the hypothetical means when they are known to be the same.

However, the Mahler approach suggests a way to get the result we are looking for (geometrically decreasing weights with weight also being placed on the overall mean). The constraints can be changed to become

$$Z_{16}=a, \quad Z_t=ab^{16-t}, t=1,\ldots,15, \quad a>0, 0\le b\le 1, a\frac{1-b^{16}}{1-b}\le 1.$$

Because this pattern does not relate to a usable ARIMA model, consider an arbitrary correlation structure (as Mahler did). Rather than force a pattern on the correlations, they will only be restricted to be the same for a given lag. Two previously mentioned issues need to be addressed. The first is that no estimate of the correlation for observations 16 years apart can be obtained. If it turns out that the estimated correlations at lags 13, 14, and 15 are very small, the one at lag 16 can be taken as zero. The second issue is that in this general model there is a lack of specificity. Recall that the key matrices are

$$\mathbf{V}=\begin{bmatrix} \sigma^2+\delta_0+\tau^2 & \delta_1+\tau^2 & \cdots & \delta_{15}+\tau^2 \\ \delta_1+\tau^2 & \sigma^2+\delta_0+\tau^2 & \cdots & \delta_{14}+\tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{15}+\tau^2 & \delta_{14}+\tau^2 & \cdots & \sigma^2+\delta_0+\tau^2 \end{bmatrix}, \text{ and } \mathbf{S}=\begin{bmatrix} \delta_{16}+\tau^2 \\ \delta_{15}+\tau^2 \\ \vdots \\ \delta_1+\tau^2 \end{bmatrix}.$$

Note that there are 16 unique entries in **V** but 18 parameters. Let the *ij*th entry in the matrix be $\gamma_{|i-j|}$. It is not possible to separately estimate the three contributors to the matrix. The REML estimates are generally decreasing. The last value is negative, which indicates that setting $\gamma_{16} = \delta_{16} + \tau^2$ in the first element of **S** to zero may not be appropriate. Instead, it is set equal to the estimate of γ_{15}. The credibility weights are a mixture of positive and negative values, which may not be reasonable. The mean squared error (which is minimized, given this model) is 83.22. The forecasts are KCR – 84.50, ARI – 79.09, TBR – 92.14, and NYY – 83.01.

Enforcing geometrically decreasing credibility weights increases the mean squared error from 83.22 to 95.39. Because this is based on a model with a large number of estimated parameters, this increase may not be meaningful. The revised *Z* values are (only values that are at least 0.0001 are presented):

Year	2006	2007	2008	2009	2010	2011	2012	2013
Weight	0.0001	0.0003	0.0010	0.0033	0.0113	0.0388	0.1334	0.4583

The overall mean receives a weight of 0.3534.

The forecasts here are KCR – 81.47, ARI – 81.28, TBR – 87.81, and NYY – 85.56.

The approach just taken is just one way to exploit the fact that the variance parameters can be estimated separately from the determination of the credibility weights. Thus, any correlation structure could have been proposed and then geometrically decreasing credibility weights derived from that structure. Similarly, for any correlation structure, any pattern of credibility weights can be imposed.

Final Comments

A variety of approaches was employed to analyze a single problem. For a particular problem it is important to understand which (or both) of the two factors needs attention (the time varying mean and the credibility adjustment). Baseball wins are a case where both are important. In 1983, Bill James postulated two forces that affect teams from year to year. He named one the “Plexiglass Principle,” noting “If a team improves in one season, it will likely decline in the next.” The magnitude of this effect can be calibrated with a time varying mean. The other he named the “Whirlpool Principle,” noting “All teams are drawn forcefully toward the center.” This effect can be calibrated with a credibility adjustment.[5] As with any model building exercise, results should combine expert knowledge of the phenomenon and statistical analysis.

Suppose you had the task of forecasting Tampa Bay’s win total for 2014. The various estimates produced in this note are presented along with commentary.

[5] James (1983), p. 220. The clear acrylic substance is spelled with both one and two “s”s. James likely chose to name his principle Plexiglas(s) because teams that improved see a clear pathway to further success, but more likely than not bounce backward the next year.

Method	Forecast	Comments
AR(1) using own data	79.22	Too much weight on the average over past 16 years
MA(1) using own data	78.59	Similar to AR(1) plus negative weight on 2012
ARIMA(0,1,1) using own data	91.50	Perhaps too much emphasis on recent results
Bühlmann-Straub	75.67	Weights all prior years equally
AR(1) with random effects	86.21	Weight is shifted to overall mean of 81 rather than first 15 years
AR(1) with Zs forced to be constant for years 1-15	85.14	Similar to previous
MA(1) with random effects	80.30	Negative weight on 2012
Arbitrary random effect correlations	92.14	Unusual pattern of credibility weights
Arbitrary correlations with geometrically decreasing weights	87.81	Follows a smooth pattern without restricting the correlation structure

The most reasonable method may be the final one presented. Decreasing positive weights places more emphasis on recent performance and weighting the result with the overall mean of 81 allows for the often observed decline after a successful season or run of seasons. Thus a forecast of 88 wins is a reasonable conclusion. However, the range of estimates indicates that results may vary. As a final comment, while this note was being prepared, Tampa Bay decided to retain its best pitcher for 2014, signed a free-agent relief pitcher, and executed some trades. Thus, while the forecast of 88 wins is a good starting point, this additional knowledge may be relevant when setting a final forecast.

References

Frees, E., Young, V., and Luo, Y., 1999, A Longitudinal Data Analysis Interpretation of Credibility Models, *Insurance: Mathematics and Economics*, Vol. 24:3, 229-247.

Frees, E., Young, V., and Luo, Y., 2001, Case Studies Using Panel Data Models, *North American Actuarial Journal*, Vol. 5:4, 24-42.

James, B., 1983, *The Bill James Baseball Abstract 1983*, Ballantine.

Keri, J., 2011, *The Extra 2%: How Wall Street Strategies Took a Major League Baseball Team from Worst to First*, ESPN Books and Ballantine Books.

Klinker, F., 2011, Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting, *Casualty Actuarial Society E-Forum*, Winter 2011, 1-25.

Klugman, S., Panjer, H., and Willmot, G., 2012, *Loss Models: From Data to Decisions*, 4th edition, John Wiley and Sons.

Mahler, H., 1990, An Example of Credibility and Shifting Risk Parameters, *Proceedings of the Casualty Actuarial Society*, Vol. LXXVII, 225-308.

Data Sets

The fifty point time series is:

112.82	105.82	104.42	98.25	110.60	107.51	107.03	107.78	101.46	103.28
108.72	99.14	104.47	105.21	108.13	111.96	112.26	114.10	114.20	111.83
103.83	107.32	105.00	107.72	104.99	105.98	95.31	89.62	92.83	101.44
93.66	93.56	90.98	94.79	92.15	94.48	104.47	100.11	102.73	97.22
98.26	100.63	99.75	99.00	95.50	98.69	101.75	106.86	110.98	111.64

The baseball data was obtained from baseball-reference.com.

Year	ARI	ATL	BAL	BOS	CHC	CHW	CIN	CLE	COL	DET	HOU	KCR	ANA	LAD	FLA
1998	65	106	79	92	90	80	77	89	77	65	102	72	85	83	54
1999	100	103	78	94	67	75	96	97	72	69	97	64	70	77	64
2000	85	95	74	85	65	95	85	90	82	79	72	77	82	86	79
2001	92	88	63	82	88	83	66	91	73	66	93	65	75	86	76
2002	98	101	67	93	67	81	78	74	73	55	84	62	99	92	79
2003	84	101	71	95	88	86	69	68	74	43	87	83	77	85	91
2004	51	96	78	98	89	83	76	80	68	72	92	58	92	93	83
2005	77	90	74	95	79	99	73	93	67	71	89	56	95	71	83
2006	76	79	70	86	66	90	80	78	76	95	82	62	89	88	78
2007	90	84	69	96	85	72	72	96	90	88	73	69	94	82	71
2008	82	72	68	95	97	89	74	81	74	74	86	75	100	84	84
2009	70	86	64	95	83	79	78	65	92	86	74	65	97	95	87
2010	65	91	66	89	75	88	91	69	83	81	76	67	80	80	80
2011	94	89	69	90	71	79	79	80	73	95	56	71	86	82	72
2012	81	94	93	69	61	85	97	68	64	88	55	72	89	86	69
2013	81	96	85	97	66	63	90	92	74	93	51	86	78	92	62

Year	MIL	MIN	NYM	NYY	OAK	PHI	PIT	SDP	SFG	SEA	STL	TBR	TEX	TOR	WSN
1998	74	70	88	114	74	75	69	98	89	76	83	63	88	88	65
1999	74	63	97	98	87	77	78	74	86	79	75	69	95	84	68
2000	73	69	94	87	91	65	69	76	97	91	95	69	71	83	67
2001	68	85	82	95	102	86	62	79	90	116	93	62	73	80	68
2002	56	94	75	103	103	80	72	66	95	93	97	55	72	78	83
2003	68	90	66	101	96	86	75	64	100	93	85	63	71	86	83
2004	67	92	71	101	91	86	72	87	91	63	105	70	89	67	67
2005	81	83	83	95	88	88	67	82	75	69	100	67	79	80	81
2006	75	96	97	97	93	85	67	88	76	78	83	61	80	87	71
2007	83	79	88	94	76	89	68	89	71	88	78	66	75	83	73
2008	90	88	89	89	75	92	67	63	72	61	86	97	79	86	59
2009	80	87	70	103	75	93	62	75	88	85	91	84	87	75	59
2010	77	94	79	95	81	97	57	90	92	61	86	96	90	85	69
2011	96	63	77	97	74	102	72	71	86	67	90	91	96	81	80
2012	83	66	74	95	94	81	79	76	94	75	88	90	93	73	98
2013	74	66	74	85	96	73	94	76	76	71	97	92	91	74	86