The Secret Life of Trend—Including Other Algorithms for Trend and Credibility for Trend

November 4, 2020

Abstract

Actuaries use loglinear trend regularly. However, there are several aspects of trend that are not common knowledge among actuaries. For example, the loglinear model for trend is not the only model for trend. Two alternatives involving the effects of random drift on expected losses in addition to the effects of trend are presented. One including the point-by-point error associated with regression, and one without it, are presented. Further, there are alternate algorithms for computing trend. Trend estimation is discussed in all three contexts and using those alternate algorithms. Corresponding credibility formulas for trend are provided as well.

1 Introduction

The loglinear trend process is well established. But for certain situations, the standard algorithm may be unwieldy. For example, when using the linear regression approach, determining the uncertainty in the trend formula engendered by the underlying uncertainty the loss development may be challenging. Further, the regression formula for determining trend using the logarithms of the data points, involves a very specific model of trend. In effect, it assumes that there is a constant trend effecting each year to year step, but the data points are affected by error values with a common variance σ^2. That may not reflect the reality of the data. So, within this paper, optimum trend models are presented, alternate algorithms for calculating the trend under each formula are presented, and some situations in which the alternate algorithms are useful to actuaries are presented.

2 Models of Trend and the Corresponding Formulas for Estimating Trend

Many actuaries use regression-based loglinear trend. However, as mentioned earlier, it has several features that may not be obvious.

2.1 A Comment on the "Log" Part of "Loglinear" Trend

Most of the common models of loss cost inflation/trend recognize that inflationary forces are better modeled by nonlinear models. In effect, because trend is believe to compound, inflationary models with assumptions like those used in compound interest are typically used. So, the cost level of the losses at some time may be modeled by some $C(t) = \exp(a + bt)$ (where, elsewhere in this paper, a

may not be constant from time to time). Therefore, it is helpful to work with the natural logarithm of the cost level $\ln(C(t)) = a + bt$ to get data that has a simplified, linear, character.

Of course, in most cases, the actual values of $C(t)$ for various t's are not available. Rather the (often) annual historical loss severities "$f(t)$", loss frequencies "$n(t)$", pure premiums "$\pi(t)$", or annual loss ratios "$LR(t)$", present in the data are what is available. They can be expected to differ from the true underlying expected severities, frequencies, etc. by some error amount.

The next issue to resolve would be deciding which probability distribution best reflects that range of that error. The Central Limit Theorem provides a rationale for the normal distribution, since a number of claim values are added together, then divided by the number of exposures. However, the normal distribution produces negative as well as positive values, so it is not always a realistic model for a trend driven by any sort of loss cost inflation. The lognormal distribution, on the other hand, is based on a geometric rather than arithmetic version of the Central Limit Theorem. Essentially, the lognormal error scenario is presumed to involve a large number of (positive) error terms that are multiplied by, not added to, the true costs. Therefore, the lognormal approach does not produce negative values. Further, the lognormal error is more consistent with trending and loss development calculations. So it is used (via it's σ parameter) throughout the remainder of this paper. Additionally, the geometric/multiplicative lognormal error approach underlies the loglinear trend that is used so often by so many actuaries[1]. Recognizing all those considerations, the lognormal modelof the trend will be used in much of this analysis.

Since the lognormal generates random variables that are exponential functions of normal distributions, one is left with a generalized trending equation, say for pure premium, of $\pi(t) = \exp(a+bt)$, where the constant $\exp(a)$ and the growth factor $\exp(b)$ may each be subject to error and other randomness. Taking logarithms of the $\pi(t)$, etc. values produces a much-more tractable linear-type model $\ln(\pi(t)) = a+bt$. So, throughout the remainder of this paper the focus will be on the simpler, linear type, algorithms.

2.2 The Loglinear Trend Model: Constant Trend with Process Error

Of course, the loglinear trend approach is the workhorse trend estimation method used by casualty actuaries. Nevertheless, there are aspects of it that may be of interest.

2.2.1 The Basic Approach Used in Loglinear Trend

To clarify the discussion in subsection 2.1, the classic loglinear trend involves the following underlying analysis:

1. There is an underlying constant geometric trend factor $1 + T$ that causes the underlying expected pure premiums ($E[\pi(t)]$'s), expected severities, or whatever else is being analyzed to grow exponentially ($E[\pi(t + 1)] = (1 + T)E[\pi(t)]$) as t increases.

2. Of course, if the expected losses above from item 1 were known, determining the trend would be trivial. However, in practice, few trend datasets are that perfect. So, it is then assumed that even though the expected losses have a perfect pattern, the data are subject to some error[2] that acts independently, but with a common variance, on each value. So, each of the

[1] Admittedly, though, that is more of an endorsement by the crowd than the result of statistical principles.

[2] For example, this might be process or parameter error.

π's is considered to be the true data $E[\pi(t)]$ multiplied by one of a series of equal independent lognormal distributions "$M(t)$", all with mean "1" (unity) and having identical coefficient of variation "v". So, each historic loss ratio, etc. value is $\pi(t) = M(t)E[\pi(t)]$.

3. The logarithmic transfer reduces the results of number 2 to a set of $\ln(M(t)E[\pi(t)]) = \ln(\pi(0)) + t\ln(1+T) + \ln(M(t))$. That is a constant, a slope multiplied by time t, and a set of identical normally distributed process error[3] terms, each with mean zero. After estimating the optimum values of the slope and constant with regression, the projection of any future point $\pi(t+s)$ may be found using that line formula and computing the exponential function on the results.

4. That format involves fitting a constant and a slope so that the constant plus the product of a slope and time minimizes the sum of squared differences between the fitted values at the various times and the actual data points[4]. It amounts to using regression to determine $y = a + tb$ given the historical data points[5] the $y_1, y_2, ..., y_k$ and independent time variables $t_1, t_2, ..., t_k$.

5. To simplify the notation, the remainder of this section will simply focus on expressing the regression using t's and y's.

Then, the linear model with process error assumes that the logarithms of the expected cost level, the $E[y_i]$'s, indeed lie on a line and follow

$$E[y_i] = a + bt_i. \quad (1)$$

But, the available historical data is different from the true expected cost levels, and each regression data point y_i differs from the underlying $E[y_i]$ by some normally distributed error, $err(i) = E[y_i] - y_i$. The $err(i)$s, per the model, all are expected to be independent of one another and have the same variance, σ^2. So, $y_i = E[y_i] - err(i)$ for all i, and

$$y_i = a + bt_i - err(i); \text{ for all } i. \quad (2)$$

Of course, much of the basics of linear regression are part of the basic education of casualty actuaries. But the material above is presented in order to stress complete clarity on the exact assumptions underlying loglinear regression. This will also set up the approach to be used for other models of trend.

2.2.2 The "Weights" Assigned to the Regression Data Points

The next step is to look at the "objective" function that the trend estimate is designed to minimize. The underlying likelihood that a set of points $y_1, ..., y_k$ are generated from a given a and b, is a constant, times the exponential function, of the negative of

[3] In the mathematics,this is referred to as "'observation error". This is potentially a broad definition of process error that would include any independent error variables sharing a common variance and a common mean of zero.

[4] Minimizing the sum of squared differences would, per the mathematics of the normal distribution, amount to the maximum likelihood estimator of the constant $\ln(\pi(0))$ and the slope $\ln(1+T)$

[5] Normally, n would be used for the number of points, but it was already used to denote the number of points, but it has already been used to represent the frequency. So, "k" will denote the the number of historical data points used in the regression underlying the trend.

$$\sum_{i=1}^{k} \frac{(y_i - a - bt_i)^2}{\sigma^2}. \tag{3}$$

So, minimizing that sum of squared differences maximizes the probability that the historical y's could arise from expected costs that follow the trend line.

Unsurprisingly, this devolves to finding an a and b that minimize a sum of squared errors. But, surprisingly, it does not depend on σ^2. One may begin with the covariance formula for the slope, and not yet specify that the times used in the regression are regular annual, quarterly, etc. Then, focusing on the averages, $\bar{y}$ of the y_i's and $\bar{t}$ of the t_i's, the slope may be written as

$$b = \frac{\sum_{i=1}^{k}(t_i - \bar{t})(y_i - \bar{y})}{\sum_{i=1}^{k}(t_i - \bar{t})^2} = \frac{\sum_{i=1}^{k}(t_i - \bar{t})y_i}{\sum_{i=1}^{k}(t_i - \bar{t})^2} = \sum_{i=1}^{k} \frac{t_i - \bar{t}}{\sum_{j=1}^{k}(t_j - \bar{t})^2} y_i \tag{4}$$

(noting that the constant $\bar{y}$ in the first term is multiplied by some values that add to zero.)

Therefore, the slope b is just a linear combination of the values in the regression data. Further, theory of sums of series indicates that the denominator is equal to $\frac{k^3-k}{12}$. The values from below the mean $\bar{t}$ of time are negative, the others are positive. The simpler expression for the slope is

$$b = \sum_{i=1}^{k} \frac{t_i - \bar{t}}{\frac{k^3-k}{12}} y_i. \tag{5}$$

Of course, in practice the t_i's are consecutive years, consecutive quarters, or something similar. Hence, it makes sense to focus on examples using consecutive and evenly spaced times.

As an example, lets say the values from 2011, 2012, 2013, 2014, and 2015 are to be used to estimate the slope. $\bar{t}$ is clearly 2013=year 3. The value of the denominator is $\frac{k^3-k}{12}$ is $(5^3 - 5)/12 = 10$,treating 2011 as year one. Further, the "k" weights starting from that of 2011, are -2/10, -1/10, 0, 1/10, 2/10 = -.2, -.1, 0 ,.1, .2. Note the linear progression stemming from equation (4), and the symmetry up to a minus sign. Both are general characteristics when the times are year to year, quarter, to quarter, etc. without breaks.

It is easy to see that the midpoint $\bar{t}$ of the numbers $t_i = i = 1, 2, ..., k$ is $\frac{k+1}{2}$. So, the weights for computing the slope from annual data may simply be stated as

$$12\frac{i - \frac{k+1}{2}}{k^3 - k} = 6\frac{2i - k - 1}{k^3 - k}. \tag{6}$$

Thus, the slope is really just a difference of weighted sums[6] average of the y values. One may also note that the values $12\frac{i-\frac{k+1}{2}}{k^3-k}$, $12\frac{j-\frac{k+1}{2}}{k^3-k}$ have constant denominators, so the points further from the center, where $|i - \frac{k+1}{2}|$ is larger, have greater influence. Thus, one may readily see that the endpoints receive the largest weight.

[6]The slope is not a weighted average of the values. It rather a difference between weighted sums. One may see that the "weights" sum to zero, since the value $i - (k+1)/2$ at i is always the negative of the value $(k+1)/2 - i$ on the opposite side at $(k+1) - ii - (k+1)/2$ at the time on the opposite side $(k+1) - i - (k+1)/2$ of the center at $(k+1)/2$.

2.2.3 The Impact of Loss Development Uncertainty

This provides insight into the impact of loss development uncertainty on trend. Consider a regression slope computed using five points. The middle point has zero weight, the first two have negative weight, and points four and five have positive weight. Arguably, either all the negative or all the positive values can be thought of as determining the slope. Between points four and five, the last, and the most developed[7] data point is point five, which has two-thirds of the weight. For a ten point slope, the last point has 36% of the weight.

If one has perspective on the development uncertainty in the data to be trended, and the uncertainties in the various points are statistically independent, one may estimate the variance of the slope due to development uncertainty. If there is a standard deviation in an ultimate loss of ϕ, then the the standard deviation of its logarithm may be estimated by $\gamma = \phi \times \frac{dlog}{dx}(\mu) = \frac{\phi}{\mu}$, or the coefficient of variation of the distribution of possible ultimate losses. Given that the various points' uncertainties are independent, one need only multiply the resulting values (squared, for variance) by the squares of the weights in equation (9) to get the consequential variance of the trend estimate across possible actual values of the ultimate losses.

In many cases, the variances of the logarithms of the ultimate losses, etc. may be directly estimated (and be statistically independent), perhaps by using the approach in Hayne 1985. Then the total variance of the error in the slope estimate due to loss development uncertainty is simply the sum of the variances due to development at the various years, etc., each multiplied by the square of the corresponding "weight"

$$36\frac{\sum_{i=1}^{k}\gamma_i^2(2i-k-1)^2}{(k^3-k)^2} \tag{7}$$

(where each γ_i^2 is the variance of the logarithm of the i^{th} data point.)

As one may see, this could be quite substantial sometimes. Consider though, that the result above does not represent the entire error variance associated with the resulting slope. The regression result is also an estimate, thus it is also part of the total error variance of the resulting slope.

If one has a fairly good handle on the variance of the process error, the results may be improved some by switching from the standard regression to "weighted regression". Weighted regression in this case maybe illustrated by the goal it seeks. Standard regression minimizes the squared differences between the points on the line and the data values. Weighted regression, weights are assigned to each of the squared differences. They correspond to the total variance (process and loss development) affecting each point. Thus one would seek[8]

$$\sum_{i=1}^{k}\frac{(a+bi-y_i)^2}{\sigma^2+\gamma_i^2} = min. \tag{8}$$

One alternative is to use calendar year trend, which requires no loss development. However, one must weigh that against its susceptibility to say, a claims department's decision to close a large percentage of their inventory in one of the calendar years (and how much that might distort the trend) and the fact that the data is from somewhat older accident/report years.

[7] Exactly how much development is involved of course depends on the line of business, perhaps the class of business, etc.

[8] As information, extensions to some popular spreadsheet software packages that perform this calculation are available at present if one does not wish to use a goal seek solution routine to compute this.

2.2.4 The Weights (Yes Weights) of the Year-to-Year Differences in the Regression Data

Section 2.2.2 provided the "weights" for determining the slope as a linear combination of the data points. The next step is to show that the estimated slope from consecutive times $t = 1, 2, ..., k$ is a weighted average of the year-to-year increases $y_{i+1} - y_i$. The first step in computing the weights involves noting that, when k is odd, the weight w_1 for $y_2 - y_1$ must match the subsection 2.2.2 point-by-point type weight for y_1 of

$$-6\frac{k-1}{k^3-k}. \tag{9}$$

A moment's review of the sums will show that the weight w_i applied to $y_i - y_{i-1}$ must equal the weight for just the point y_i $\left(\text{i.e. } 6\frac{2i-k+1}{k^3-k}\right)$ less w_{i-1}. The provides a point-by-point formula for the w_i's.

That formula for the w_i's may be solved, and the resulting weights for the one year slopes are

$$w_i = 6\frac{i(k-i)}{k^3-k} \text{ for each } y_{i+1} - y_i. \tag{10}$$

These weights (w_i's) have a very important property—they sum to unity. Thus, the projected linear trend (slope) b is really just a weighted average of the year-to-year slopes in the data. So the original external projected trend ratio $T+1 = \exp(b)$ will be a geometric average of the year-to-year growth values in that data. The same weights will be used, but they will represent exponents for the various year-to-year growth values within the geometric average. Further, one should note that although the weights for individual points y_i are larger as one moves away from the center of the experience period, the weights for the $y_{i+1} - y_i$'s are larger near the center of the period.

2.2.5 Summary of the Results for Regression

In conclusion, the regression slope (logarithm of the trend value) may be expressed as a difference between weighted sums of the loss, etc. values, or as a weighted average of the year-to-year changes. That leads to an estimate of the effect of loss development uncertainty on the fitted slope. Further, it is well known that the optimum prediction under regression to some period $k + j$ is a straight average (identical weights of $1/k$) of the y values, plus the calculated slope times the number of years from the mean time $\bar{t}$ to the future period. As shown above, for this annual data, both the beginning point and the slope to future periods are weighted values of the y's. So any linear projection to some future period $k + j$ may be expressed as a weighted sum/linear function of the y's.

$$est(y_{k+j}) = \sum_{i=1}^{k}\left[\frac{1}{k} + \left(\frac{k+1}{2} + j\right)\frac{6(2i-k+1)}{k^3-k}\right]\hat{y}_i. \tag{11}$$

(Essentially, the $\frac{y_i}{k}$ sum to the mean of the y_i's. The $\frac{k+1}{2}$ trends from the mean time associated with the mean of the y_i's to the time associated with the last data point. Lastly, the j term moves it to the future time period desired for the projection.)

Overall, one may see that the standard loglinear trend algorithm is based on computing a straight average for the starting point and a weighted average for the trend.

2.3 The Trend with Random Drift Model: Varying Trend but no Process Error

The loglinear approach deals with situations where the true underlying expected loss values for each year are not completely known, but the underlying trend is constant. The trend with random drift case involves perfectly known expected loss values. However, in addition to the trend, those expected losses"'drift" in a random way.

This is really another use for a model that is widely employed by some other financial service providers. Use of that model, geometric Brownian motion,to reflect changing costs is not new. It is commonly used in the investment community as a model of risk-adjusted stock price evolution. It has already been used in the actuarial world in Boor 1993 and McNichols and Rizzo 2012.

Basically, it assumes that the cost, etc. levels $C(t)$ are affected by a constant trend (T as always). However, the cost level is also buffeted by constant but random changes, so that from year to year its logarithm is changed by a random selection from a normal distribution (in addition to the slope). The effects of these changes are cumulative in that all the prior changes are embedded in each value. In the transformed distribution $\ln(C(t))$ values at time s and time t differ not only by the logarithm of the trend, but also by some value from a normal distribution with some variance parameter δ^2. It may be written

$$\ln(C(t)) - \ln(C(s)) = (t-s)\ln(1+T) + (t-s)\delta N(0,1) \tag{12}$$

(where $N(0,1)$, in a slight abuse of notation (for clarity), represents a sample from the standard normal distribution). One may describe $\delta N(0,1)$ as random drift. Since it is linear now, it is associated with the slope of a line rather than with the compounding trend in the original trend data. So, this is "slope with random drift" rather than "trend with random drift". Notably in this linear case, in each set of intervals (s,t) and (u,v) that do not overlap (other than at the endpoints), the samples from the normal distribution are independent.

That means that the yearly slopes $y_2 - y_1, y_3 - y_2, ..., y_k - y_{k-1}, (y_i = \ln(C(i))$ are all independent samples of the slope $b = \ln(1+T)$. Since there are $k-1$ samples, the slope estimate using the year-to-year changes is clearly

$$est(b) = \frac{1}{k-1}\sum_{i=2}^{k} y_i - y_{i-1} \tag{13}$$

One may note that any i not on the top or bottom of the range, is included in both $y_{i+1} - y_i$ and $y_i - y_{i-1}$. So most of the terms cancel, leaving a slope estimate for these historical points under slope with random drift of

$$est(b) = \frac{1}{k-1}(y_k - y_1). \tag{14}$$

In effect, the weights for the points are $-\frac{1}{k-1}$ for y_1, $\frac{1}{k-1}$ for y_k, and zero for the other points.

To finish the linear portion of this analysis, one may note that slope with random drift is well-known to be "Markov" or "memoryless". That means that for times after the last data point at k, y_k alone is the best point for making future predictions. $y_1, y_2, ...y_{k-1}$ would only be useful when y_k and perhaps additional values are not known. So, whereas predicting future values using regression in equation (11) involved using middle of the time values at the starting point, slope with random drift predictions begin with the most recent data point y_k. So the linear slope with random drift estimate for y at time $k+j$ is

$$est(y_{k+j}) = y_k + \frac{j}{k-1}(y_k - y_1) = \frac{j+k-1}{k-1}y_k - \frac{j}{k-1}y_1 \tag{15}$$

Unlike the regression system, translating the slope with random drift formulas back to the needed trend with random drift values ($C(i)$'s) is easy. Equation (15), based on the difference between the last point and the first point becomes a ratio, And the multiplier $\frac{1}{k-1}$ becomes an exponent, generating the $k-1^{st}$ root of $1+T = \left(\frac{C(k)}{C(1)}\right)^{\frac{1}{k-1}}$. And per the Markov property $est\left(C(t+j)\right) = C(k) \times \left(\frac{C(k)}{C(1)}\right)^{\frac{j}{k-1}}$.

2.4 Trend with Both Random Drift and Process Error

Contrary to the assumptions of the last two sections, sometimes trend is influenced both by volatility in the trend (or slope) and process error. This section presents a model to use in such a situation.

2.4.1 Explanation of the Model

The previous models each include a core assumption that could sometimes be an unrealistic . Often trend data from very large datasets that would seem to be susceptible to the trend with random drift view of subsection 2.3 appear to be a little different from what the theory would suggest. For example, consider the consumer price index data in Table 1, where one would expect little process error. Process error would create a situation where very large increases or decreases in the trend could come from more extreme errors. In that case, a large decrease in the annual trend would be followed by a large increase and vice versa. One cannot determine conclusively from the data, but the changes in the trend from 2008 to 2009, 2009 to 2010, and 2010 to 2011 do suggest that some process error (perhaps arising from the data collection process) is present.

Table 1: Year-to Year Trend Rates in Consumer Price Index (All Urban Consumers

Value at 12/31 of Year	CPI	Change in CPI
2006	210.800	
2007	210.036	-0.36 %
2008	210.228	0.09 %
2009	215.949	2.72 %
2010	219.179	1.50 %
2011	225.612	2.94 %
2012	229.601	1.77 %
2013	233.049	1.50 %
2014	234.812	0.76 %
2015	236.565	0.75 %

As one may see, the trend rates in the CPI data are fairly volatile. Further, it is possible that the dynamics of the consumer price index involve even more complexity. Therefore it is reasonable to question whether or not the loglinear trend model really captures the structure of the data it is applied to. The concern with trend with random drift is more direct. One may also see that the trend with random drift assumption that none of the points contain process error of any sort, is suboptimal for datasets subject to process risk.

Therefore, an approach that recognizes both the process, etc. risk that makes the data points imperfect representations of the underlying costs and also accommodates drift-type volatility from year to year is needed. The approach begins with the trend with random drift process consistent with

$$E\left[\ln\left(C(i+1)\right)\right] = E\left[\ln\left(C(i)\right)\right] + \ln(1+T) + \delta N(0,1) \tag{16}$$

(with the notation abusing $N(0,1)$ terms representing independent standard normal samples for each of the various intervals $(i, i+1)$). Thus, $E\left[\ln\left(C(i+1)\right)\right]$ follows the slope with random drift paradigm. But, rather than the pure random drift involved in subsection 2.3, this scenario also includes the process error included in subsection 2.2. Specifically, one may state that

$$y_i = E\left[\ln\left(C(i)\right)\right] + \sigma N(0,1) \tag{17}$$

(with the $N(0,1)$ terms independent among the various indices $i = 1, 2, ..., k$).

2.4.2 Finding the Key Variances

Considering the presence of both types of volatility, the goal is to find the a and b that are most consistent with the data. To do that, one must first define an error function or objective function to minimize. The most obvious approach would be to take a page from the playbook of the other two situations and seek the slope that is consistent with the lowest possible variance. However, in this case there are actually two variances, both the process variance from subsection 2.2 and the drift variance from subsection 2.3. So one must consider what combination of those should be minimized.

For illustration, note that σ and δ (actually σ^2 and δ^2) may be estimated using historical data. Per Boor 2015[9], once the slope is removed σ^2 and δ^2 may be computed using

$$\frac{E\left[\sum_{i=1}^{k-1}(y_{i+1}-y_i)^2 - (y_k-y_1)^2\right]}{2(k-2)} = \sigma^2, \tag{18}$$

and

$$\frac{E\left[(k-1)(y_k-y_1)^2 - \sum_{i=1}^{k-1}(y_{i+1}-y_i)^2 - (y_k-y_1)^2\right]}{(k-1)(k-2)} = \delta^2. \tag{19}$$

where each y_j is the value of the linear (generally, log-transformed) value for the j^{th} year, month, etc.

However, when one attempts to simultaneously estimate the slope, process error variance σ^2, and drift variance δ^2, the problem tends to become too unwieldy to perform reliably, at least per per a few methods employed by the author. Therefore, one may suggest using the trend with random drift approach when the values appear to be fairly compact around a curve, the approach of this section when there is a similar non-linear appearance, but the values are not compact around the curve, and the regression approach otherwise, at least as a starting point.

[9] To facilitate its use, be aware that in the referenced paper the "y_i's' were labeled as "S_i"s to limit conflicts among variable names.

2.4.3 Estimating the Slope

On the other hand, if reasonable estimates of σ^2 and δ^2, may be made, then it as least possible to provide an estimate for the slope. The idea involves estimating the underlying "expected loss"[10]/slope with random drift/no process error path, then use the standard slope with random drift estimate of the slope of the expected values, [(most recent point) − (first point)]/$(k-1)$, of the slope from subsection 2.3.

To do so, it is helpful to define the best approximation point-by-point. For example, at the first data value y_1, the only information available is the value y_1, so that would be the estimate e_1 of the first point. Its variance around the true value[11] on the underlying path of the expected values of losses would be σ^2, which can be set as the initial value τ_1^2.

For the second value along the path, we have two estimators, $e_1 + b$ and y_2. The expected prediction variance of y_2, with respect to the value of the underlying expected losses would logically be it's variance from the those losses, or σ^2. The expected squared prediction error generated by $e_1 + b$ would be the squared error inherent in e_1, or the process variance σ^2, plus the inherent volatility as one moves from year to year along the path, δ^2. Since the two may generally be thought to be independent, the variance of the error between the estimate e_1+b and the true value along the path is $\delta^2 + \sigma^2$.

That begins the iteration. Since the expected squared error e_1 makes predicting the expected value is $\tau_1^2 = \sigma^2$, and the drift along the path is independent of the process error, the error $e_1 + b$ makes (where b is the currently unknown slope) in predicting the second true point on the path adds one year of random drift to make $\tau_1^2 + \delta^2$ the error variance of $e_1 + b$ in predicting the second point on the path. The error variance of y_2 would be σ^2. A formula from best estimate credibility (see Boor 1992) indicates that for these two independent estimators, the best estimate results from weighting each one by the expected squared prediction error of the other[12]. Therefore, the (best) estimated value of the second point on the path the expected losses underlying the data actually followed is

$$e_2 = \frac{\sigma^2(e_1 + b) + (\tau_1^2 + \delta^2)y_2}{\tau_1^2 + \delta^2 + \sigma^2}. \tag{20}$$

Since the two components are (clearly) independent, the variance of the result above is just the result of multiplying the variances of the two by the scalar multipliers (like credibilities). A little algebra results in in a formula for the error variance of e_2 of

$$\tau_2 = \frac{\sigma^2 \times (\tau_1^2 + \delta^2)}{\tau_1^2 + \delta^2 + \sigma^2}. \tag{21}$$

Those may be generalized into recursive formulas for the e's and τ's

$$e_{i+1} = \frac{\sigma^2(e_i + b) + (\tau_1^2 + \delta^2)y_{i+1}}{\tau_i^2 + \delta^2 + \sigma^2}. \tag{22}$$

[10]This could also be expected frequency, severity, etc

[11]The language is key here. In this case, the mean of e_1 is equal to the actual underlying first point on the path the expected losses follow as they drift. However, if they did not match, the expected squared error predicting the initial point on the path would have to include the squared difference between the mean of e_1 and the mean of that initial point along with the variance of e_1. Considering that all the distributions used in this section are presumed to be unbiased and independent, per Boor 1993, it should not be an issue. However, it is mentioned for completeness and clarity.

[12]If the reader is so inclined, one may verify that the formula works in this instance by using Bayesian methods.

$$\tau_{i+1} = \frac{\sigma^2 \times (\tau_i^2 + \delta^2)}{\tau_i^2 + \delta^2 + \sigma^2}. \tag{23}$$

Of course, since that formula assumes that one already knows the slope, it is not directly useful for estimating the slope. However, it may be used indirectly. In the regression model, the slope is set so that squared residuals between the fitted line and the actual points are minimized. So, it would be logical to seek the value of b for which the iterations of equations (22) and (23) generate the least squared residuals between the best estimate e_i's and the actual points (y_i's). Table 2 illustrates the process when a 10% exponential trend is accompanied by process error corresponding to $\sigma^2 = .005$ after the logarithmic transform and drift variance $\delta^2 = .002$, also in the logarithmic transformed data. Of course, the exponential/trend with random drift data is first converted to a linear system using logarithms so that all the variables are on a linear basis. Then the calculations mentioned earlier are carried out in Table 2. Lastly, the spreadsheet software searches for the value in medium gray of the slope that minimizes the sum of squared differences between the data points and the best estimate of the points along the underlying path. That, after conversion to loglinear trend, forms the trend estimate.

Table 2: Estimation of Underlying Slope With Both Random Drift and Process Error When Actual Value (10%) is Larger than the Standard Deviations

Constants:

A. σ^2 =	0.0050,	σ =	.0706	(considered known)
B. δ^2 =	0.0020,	δ =	.0447	(considered known)
Exponential Trend =	10.00%,	Slope of Logs =	9.53%	(both to be found)

Year	(1) Simulated Loss Ratio	(2) "S" Natural Log of Loss Ratio	(3)=[Prior (6)] "e" Current Point Est of Expected Loss Level	(4)=[Prior (7) Incoming Variance "τ"	(5)=(4)+B. Drift Variance to Next e	(6)={[(3)+b]A.+(1)(5)} /{A. +(5)} Next e	(7) = (5)×A. /(A.+(5) Outgoing Value of τ	(8)=((2)-(3))^2 Difference Data Point S and e
1	101.3 %	0.0128	0.0128	0.0050	0.0070	0.1034	0.0029	0.00000
2	110.4 %	0.0987	0.1034	0.0029	0.0049	0.1925	0.0025	0.00002
3	120.6 %	0.1876	0.1925	0.0025	0.0045	0.3128	0.0024	0.00002
4	140.0 %	0.3365	0.3128	0.0024	0.0044	0.4383	0.0023	0.00056
5	159.3 %	0.4657	0.4383	0.0023	0.0043	0.4822	0.0023	0.00075
6	155.1 %	0.4389	0.4822	0.0023	0.0043	0.6342	0.0023	0.00188
7	198.2 %	0.6843	0.6342	0.0023	0.0043	0.6619	0.0023	0.00251
8	183.1 %	0.6047	0.6619	0.0023	0.0043	0.7690	0.0023	0.00328
9	218.2 %	0.7803	0.7690	0.0023	0.0043	0.8587	0.0023	0.00013
10	235.2 %	0.8551	0.8587	0.0023	0.0043			0.00001

b=Estimated Slope of Logs= 9.40% Sum of Differences Between Loss Level Path and Data Points= .00916
T=Est Loglinear Trend= 9.85%

One could question whether the fairly high accuracy (estimate of 9.85% vs. an actual 10.0%) of this method is due to the larger slope predominating over the two variances. Therefore, the same calculations were done for a trend rate of 3% in Table 3.

The sums of squared differences in Table 3 match Table 2 because the data to be analyzed was the same, up to the slope, in both examples. However, notice that even the estimate of the lower 3% trend was very, very good. In the experience of the author, if the data looks like it complies with the regression assumptions, typically this only provides a marginal improvement in accuracy over

Table 3: Estimation of Underlying Slope With Both Random Drift and Process Error When Actual Value (3%) is Smaller than the Standard Deviations

Constants:				
A. $\sigma^2 =$	0.0050,	$\sigma =$	.0706	(considered known)
B. $\delta^2 =$	0.0020,	$\delta =$	.0447	(considered known)
Exponential Trend =	3.00%,	Slope of Logs =	2.96%	(both to be found)

Year	(1) Simulated Loss Ratio	(2) "S" Natural Log of Loss Ratio	(3)=[Prior (6)] "e" Current Point Est of Expected Level	(4)=[Prior (7) Incoming Variance "τ"	(5)=(4)+B. Drift Variance to Next e	(6)={[(3)+b]A.+(1)(5)} /{A. +(5)} Next e	(7) = (5)×A. /(A.+(5) Outgoing Value of τ	(8)=((2)-(3))^2 Difference Data Point S and e
1	1.013 %	0.0128	0.0128	0.0050	0.0070	0.0363	0.0029	0.00000
2	1.034 %	0.0330	0.0363	0.0029	0.0049	0.0603	0.0025	0.00001
3	1.058 %	0.0561	0.0603	0.0025	0.0045	0.1124	0.0024	0.00002
4	1.149 %	0.1393	0.1124	0.0024	0.0044	0.1695	0.0023	0.00072
5	1.225 %	0.2027	0.1695	0.0023	0.0043	0.1570	0.0023	0.00110
6	1.116 %	0.1101	0.1570	0.0023	0.0043	0.2336	0.0023	0.00220
7	1.336 %	0.2898	0.2336	0.0023	0.0043	0.2074	0.0023	0.00315
8	1.155 %	0.1445	0.2074	0.0023	0.0043	0.2442	0.0023	0.00397
9	1.290 %	0.2543	0.2442	0.0023	0.0043	0.2682	0.0023	0.00010
10	1.301 %	0.2633	0.2682	0.0023	0.0043			0.00002

b=Estimated Slope of Logs=	2.82%	Sum of Differences Between Between Loss Level Path and Data Points=	.00916
T=Est Loglinear Trend=	2.86%		

the regression estimate. These alternate views are better suited to situations where one expects that the underlying trend changed significantly during the time period of the data.

In conclusion, while there is a workable formula for estimating the variance structure given knowledge of the trend, there is also a workable formula for estimating the trend given the variance structure. However, the author is not aware of any good approach to estimate both simultaneously. Nevertheless, in certain situations, investing the time needed to execute this method can yield better accuracy in the trend calculation.

3 Limited Fluctuation Credibility for Trend

Now that formulas that relate the trend calculations to individual points are available (at least for regression and pure trend with random drift), it is possible to develop credibility formulas that are designed specifically for trend. The two versions for limited fluctuation credibility follow.

3.1 What Would a (Limited Fluctuation) Credibility Formula for Trend Look Like?

When considering credibility for trend, it is relevant to begin with the core goal of the given credibility process. Although actuaries typically think of limited fluctuation credibility in terms of claim counts, it is really about the rate, trend, etc. not changing too much unless the data clearly indicate that a given change is needed. For example, for the common 1082 standard, the objective is to not allow pure randomness in the data to arbitrarily change rates by more than 5% (up or down), unless the data indicate such a change is needed. Further, since any amount of loss can conceivably happen, "the data indicate such a change is needed" is defined as "There is a 10% (or less) chance that the credibility-adjusted result will randomly create more than a 5% change." Those requirements are not based on claim counts, claim counts are merely a convenient way to compute the credibility.

So, in the case of limited fluctuation credibility, the stated goal is that probability that random chance causes losses to exceed some threshold ± R is limited to some suitably low probability p. So, the goal is to find (or estimate) the p^{th} and $1-\mathrm{p}^{th}$ percentiles of the distribution of possible trends. Then one may appropriately throttle the distribution with a credibility factor Z so that $Z \times F_{trend}^{-1}(\mathrm{p}) \geq -\mathrm{R}$ and $Z \times F_{trend}^{-1}(1-\mathrm{p}) \leq \mathrm{R}$, where $F_{trend}^{-1}(p)$ is the pth percentile of the distribution of possible changes in trend.

3.2 Limited Fluctuation Credibility with Regression-Based Trend

Subsection 3.1 allows one the opportunity to define a general credibility formula for the regression slope. Such a formula would depend on the variances, not on claim counts. Specifically, it is well known that approximately 90% of the probability in the normal distribution is within two standard deviations (up or down) of the mean (note that for purposes of considering pure randomness, the mean would be "no change"). So if $Z = 5\%/(90\% \times CV(linear\ slope))$(where CV denotes the standard coefficient of variation, the standard deviation divided by the mean), the criteria underlying the 1082 standard will be fulfilled by credibility weighting the slope with some very reliable ancillary data. However, that does require one to know the variance of the fitted slope around the true slope.

It should then be clear that the main challenge in determining credibility for the regression slope that underlines the trend is finding the percentiles of the distribution of possible slopes. Thankfully, a standard statistic is available to help. For example, using the special additional regression option available in the som common spreadsheet software, one may compute the key variance statistic needed. After first taking logarithms of the CPI data in Table 1, the spreadsheet option produces (as approximately excerpted from the regression output).

Table 4: Excerpt from Supplementary Information Spreadsheet Software Provided in Loglinear Regression of Table 1 Data

ANOVA				
	df	*SS*,	*MS*	*F*
Regression	1	0.01965	0.01965	173.0653602
Residual	8	0.00091	0.00011	
Total	9	0.02056		
	Coefficients	Standard Error,	t Statistic	p value
Intercept	-25.62411	2.35860	-10.86411	4.55693E-06
X Variable	0.01543	0.00117	13.15543	1.06096E-06

The values in gray are the the mean (fitted) slope and the standard error (error standard deviation) of the regression slope. Thus, the 1082-equivalent credibility calculation for the linear regression would be $5\%/(90\% * (.00117/.01543)) = 73\%$

However, the credibility of the final trend T of the actual CPI values is slightly different. To estimate the trend in the original CPI data, one must invert the logarithm and convert the trend factor $1 + T = \exp(slope)$ to the trend rate $T = \exp(slope) - 1$. That affects the relative error (which is supposed to, under the 1082 standard, be 5% or less) through the magnification or shrinkage generated by the exponential function, as well as the denominator used in computing the relative error. The magnification or shrinkage multiplier would be generated by the derivative of $\exp(x) - 1$, or $\exp(x)$ at $x = .01543$. The value is 1.01555. Then, since $T = \exp(.01543) - 1 = .01555$, the 5% relative error allowed in determining the slope translates to relative error of $5\% \times 1.0155 \times .01543/.01555 \approx 5.04\%$. So the 5% threshold still essentially holds in this example. However, it is important to complete this final step to be certain that the credibility fulfills its function appropriately.

3.3 Limited Fluctuation Credibility with Trend with Random Drift

Just as in loglinear trend, the key to limited fluctuation credibility for trend with random drift trend lies in computing the variance. In this case, the process is much more straightforward. In say, ten years, of data there are nine year-to-year slopes. It is not difficult to calculate the variance of those slopes. Then, since the estimate of the trend with random drift is simply the average of those nine slopes, all one need do is divide the variance of the individual slopes by nine. That estimates the variance of the error in the slope estimate, and the remaining process mirrors that used in subsection 3.2.

3.4 Usefulness of Limited Fluctuation Credibility for Trend

Of course, once it is computed, limited fluctuation credibility can be used in a wide variety of situations. This can be used when the complement of credibility benchmark is countrywide trend for the line of business, or when it is the tend in last year's rate filing. The flexibility and (comparative) ease of computing this are offset somewhat by the fact that it does not result in the most accurate estimate of trend. The methods of the next section will focus on accuracy, but consequently the formulas are less robust..

4 Best Estimate Credibility for Loglinear Regression

As noted in Boor 1992, best estimate credibility depends not just on how well the data predicts the loss costs, it also depends on how well (or poorly) the complement of credibility benchmark predicts the loss costs. It further depends on whether or not the prediction errors generated by the data and the benchmark are correlated. If one is using countrywide trend as a benchmark, one might expect the errors to be uncorrelated. However, if one is using the trend generated last year as a benchmark, one might expect substantial correlation. Therefore, the two approaches are analyzed in separate sections.

4.1 Best Estimate Credibility with External Benchmark Data

Per Boor 1992, best estimate credibility is a function of the error each statistic (dataset) makes in predicting the underlying quantity being estimated (in this case, the slope underlying the trend).

For two predictors that make uncorrelated errors, the credibility weight of one statistic is proportional to the squared error generated by the other statistic. One might expect that, say, trend computed from countrywide data would make prediction errors that are largely uncorrelated with those generated by the trend data of a small state. In general, most benchmarks tend to make errors that are unrelated to the trend of a dataset with lesser volume. So that simpler, no covariance, formula would apply.

The next step that is required is to estimate the two variances. However, as discussed in subsection 3.2, that may be begun, for both the target data and the benchmark, by using the standard errors of the two slopes obtained in the regression process.

For the subject data, that suffices to produce an estimate of the squared error. However, the squared error the benchmark makes in estimating the subject slope solely because the benchmark and the subject data simply have different underlying slopes requires another term. Therefore, the squared error the benchmark makes is more than just variance from the regression fit. That amount, at least in this characterization is a constant bias[13] rather than variance. It does contribute to the squared error, though. It is not hard to see that the expected squared error is equal to the variance of the predictor plus the square of that bias.

Now, one may only know the actual bias by knowing the underlying slopes that are being estimated. However, one could use the difference between the slope of the subject data and the slope of the benchmark data to estimate the bias. Then the credibility of the slope in the subject data is:

$$
\begin{aligned}
&Z(\text{subject data}) = \\
&\qquad \left[\text{standard error}^2(\text{benchmark}) + (\text{difference of slopes})^2\right] \\
&\quad / \left[\text{standard error}^2(\text{subject data}) + \text{standard error}^2(\text{benchmark}) + (\text{difference of slopes})^2\right].
\end{aligned}
$$

For example, When the Table 1 data is loglinearly regressed, the slope is 0.01543 and the standard error is 0.005. If one could then identify a benchmark to supplement this (CPI) data, and it had a slope of .017 and standard error of .003, the the credibility of the CPI data would be $\frac{.003^2+(.01543-.17)^2}{.005^2+.003^2+(.01543-.17)^2} = 31\%$. Thus, given the regression output, this is not a challenging calculation.

It should be apparent that this general approach will also work with the slope with random drift using the variance of year-to-year changes formula from subsection 3.3. Details are not provided as the formula should also be apparent.

Of course, at this point, with either trend scenario, a credibility weighted estimate of the slope is produced, but what is actually needed is the exponential-based trend. The procedure for converting the expected squared error in the slope to that in the exponential trend has already been discussed in subsection 3.2. Alternately, one may simply perform the credibility calculation on the regression slopes, and then convert the result to a trend by applying the exponential function to the slope and subtracting unity.

[13]Technically "bias", although it does not have the strong negative connotation associated with bias that distorts the results—as this improves the results.

4.2 Best Estimate Credibility When Updating Loglinear Regression

In updating trend, the complement of credibility is either last year's trend or last ear's slope It seems logical to use the same credibility for each. In contrast to the uncorrelated nature of the errors in external benchmarks with internal data, last year's slope uses much of the same actual data as the subject data. So, then Boor 1992 indicates (after some algebra) that the optimum credibility for the slope is

$$\begin{aligned} &Z(\text{subject data}) = \\ &\Big[\text{standard error}^2(\text{old slope}) + (\text{difference of slopes})^2 - \text{covariance}(\text{new slope,old})\Big] \\ &\Big/\Big[\text{standard error}^2(\text{new slope}) + \text{standard error}^2(\text{old slope}) + (\text{difference of slopes})^2 \\ &-2 \times \text{covariance}(\text{new slope,old})\Big]. \end{aligned} \tag{24}$$

The new issue to be resolved, then, is that of estimating the covariance between this year's slope and last year's slope. The first tool lies in some additional output of the regression routine in the spreadsheet software. The data in Table 5 also comes from the regression on the Table 1 CPI data.

Table 5: Additional Excerpt from Supplementary Information Spreadsheet Software Provides in Loglinear Regression of Table 1 Data

Multiple R	0.97766
R Square	0.95582
Adjusted R Square	0.95029
Standard Error	0.01066
Observations	10

The standard error in gray is the standard deviation of the "residuals", or differences between the fitted curve and the data points. Combining this with the weights from equation (9) in subsection 2.2, that are effectively used in computing the slope, one may see that the variance of each (independent) point/residual and weight combination is equal to the square of the weight times the standard error squared. To show that this works note this alternate computation of the previously provided (Table 4) standard error of the slope.

$$\sqrt{.01066^2 \times \sum_{i=1}^{k} \left(6\frac{2i-k-1}{k^3-k}\right)^2} = .00117. \tag{25}$$

The goal, though, is to compute the covariance of this year's slope with last year's slope. That may be done in a similar fashion. It is not hard to see that the first year in the current trend period was the second year in last year's trend period, and so on. Then, if we further define r_{new} to be the standard error in the new regression and r_{old} to be that of the prior regression, one may obtain (after some algebra) the following formula for the covariance

$$
\begin{aligned}
Cov(\text{new slope,old slope}) &= \sum_{i=1}^{k-1} r_{new} r_{old} \times 6\frac{2i-k-1}{k^3-k} \times 6\frac{2i-k+1}{k^3-k}. \\
&= r_{new} r_{old} \times 12\frac{k-3}{k(k^3-k)}.
\end{aligned}
\tag{26}
$$

Note that this is essentially a constant, identical across all updates and data used in k period slope, times the product of the two standard errors.

One may also note that this does not completely resolve the updating credibility problem. It works when the complement of credibility term is last year's slope, but not when it is the credibility weighted average of several prior years that updating would have generated the last year. However, that problem is fairly complex. Per H. Gerber and D. Jones 1975, when these sort of covariances exist, successive updates often require changes in the credibility mix of older years to be truly optimal. Likely, there is still some view of what is optimal that would accommodate this particular situation. Hopefully, this provides a first step. Further, in context the formulas of this subsection could conceivably be used to provide a proxy for the full updating credibility.

5 Summary

A detailed analysis of the trend, and associated linear slope, calculations was presented. Three alternate scenarios for the underlying process driving the trend were presented, along with some guidance for estimating the trend in each case. How often they produce materially different trends is not known at present, but might represent an opportunity for further analysis. Lastly, using details of the analyses of the trend process in this article, credibility formulas on both a limited fluctuation and best estimate basis were provided. Those formulas focused primarily or conventional regression trend, but form a template for the trend with random drift as well.

References

J. Boor. Credibility based on accuracy. *Proceedings of the Casualty Actuarial Society*, 79:165–185, 1992.

J. Boor. A stochastic approach to trend and credibility. *Casualty Actuarial Society Forum,Special Ratemaking Edition*, pages 341–400, 1993.

J. Boor. The credibility of the overall rate indication: making the theory work. *Variance*, 9:167–186, 2015.

H. Gerber and D. Jones. Credibility formulas of the updating type. *Transactions of the Society of Actuaries*, 27:31–46, 1975.

R. Hayne. An estimate of statistical variation in development factor methods. *Proceedings of the Casualty Actuarial Society*, 72:25–43, 1985.

J. McNichols and J. Rizzo. Stochastic gbm methods for modeling market prices. *Casualty Actuarial Society Forum,Summer Edition*, pages 21–18, 2012.