A STOCHASTIC APPROACH TO TREND AND CREDIBILITY

Joseph A. Boor

ABSTRACT

This paper contains a new approach to analyzing loss statistics which uses stochastic processes. The author views loss statistics as samples from a specific type of stochastic process. The author believes that type of process is the most consistent with the realities of insurance statistics, and he explains why. Using that mathematical framework the author develops a formula for credibility when the complement of credibility is applied to trend. The paper also contains a formula for trending data that is more consistent with the stochastic approach (and hence the realities of insurance statistics) than the trend line.

A STOCHASTIC APPROACH TO TREND AND CREDIBILITY

Joseph A. Boor

Even though insurance and econometric statistics are driven by random forces, actuaries usually treat them as deterministic. For instance, actuaries tend to assume that insurance losses follow some perfect line or exponential curve over time. Since that implies the growth in losses is a function of time alone, we are implicitly assuming that it is time alone that causes loss cost levels to change.

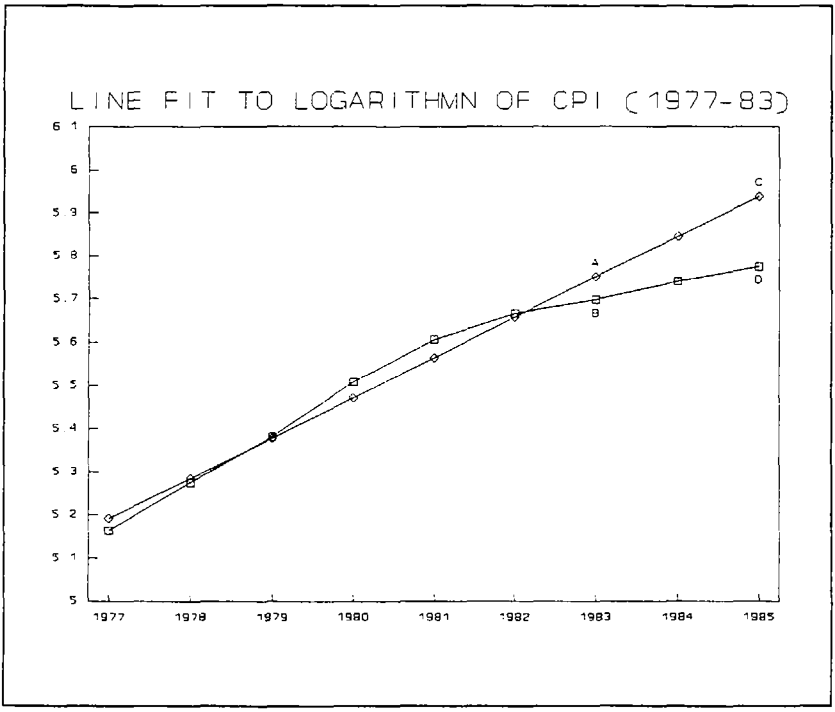
Of course, we all realize that assumption is false. But, we also recognize that we must reflect inflation and other environmental changes in ratemaking. So, in the absence of better models we use deterministic models. This paper contains a new model that reflects the randomness in econometric data.

Why the Trend Line Doesn't Work.

'I don't know where we've come from.
I don't know where we're going to.
And if all this should have a reason....
We would be the last to know.' -John Kay

Trend lines often produce unrealistic results when they are used on econometric data. Consider the United States Consumer Price Index when it began to come out of its inflationary spiral in the early 1980's. At that time a CPI prediction based on a trend line would err for two reasons: not only because the projected increase since the last actual observed point would be too high; but also because the fitted trend line value at the last observation time would be higher than the actual observed CPI at that observation time.

For example, the curve below represents that specific set of circumstances. The trend line represents a loglinear fit to the CPI during 1977-83. 'C' represents the predicted 1985 CPI log using the trend from 1977-83. 'D' is the actual recorded 1985 CPI log. The difference C-D is large because the recorded 1983-87 CPI log increase (.131) was below the trend (.374). And it is larger yet because the 1983 recorded CPI log 'B' was below the trend line value 'A'.



In this case trend line analysis works very poorly. It does so because its fundamental assumptions are contrary to the way economic systems work.

On one hand, the trend line mathematics assumes there is a straight line (or exponential curve in the case of loglinear fit) underlying the data. It assumes that the only reason the data do not lie on that straight line is that each point is imperfectly observed. In mathematic terms it assumes there is an observation error (with common variance E^2) at each point.

On the other hand, with econometric data the prediction error does not result from imperfect observation of the existing data as much as it results from year-to-year changes in the trend. There is really no logical reason for the CPI to follow a perfect exponential curve. The fact that it increased by 4% in 1984 does not mean it has to increase by exactly 4% in 1985 (although it does make it more likely). The trend line and regression have many reasonable applications in physics and chemistry; where laws of nature require that one variable be related to another by some precise formula. But at present there are no formulas that specify the behavior of econometric data. So, the author believes econometric data reflects random trend with minimal observation error rather than constant trend with significant observation errors. So, regression on econometric data may yield large errors. Some observers then conclude it is futile.

Unfortunately, the premiums and losses that are the actuary's stock in trade are econometric quantities. They inflate very much like the CPI. So actuaries need a realistic way to predict econometric quantities.

A Realistic Model

The argument above suggests we should assume that trend is random but there is no observation error. That follows from the fact that econometric data may be a series of numbers, but those numbers represent the aggregate actions of an enormous number of individuals.

For instance, the CPI is an aggregate of the buying and selling decisions of everyone in the United States. Those millions of people buy or sell independently, but their actions tend to be guided by two parameters: what others are doing (market prices) and what they see as the trend of the economy (historic inflation and other inputs). Assuming that broad econometric changes are a result of many small changes[1]; and that those changes tend to be proportional to the price level when the changes occur; results in the model below

$$y(t+\Delta)=y(t)\cdot\prod_{i=1}^{n(t,t+\Delta,\lambda)}(1+c_i(\lambda))$$

Where:

y is the econometric variable being observed (e.g. the CPI).

$(t, t+\Delta)$ is the time period over which y changes.

$n(t, t+\Delta, \lambda)$ is the number of small changes in y made between times t and $t+\Delta$. The actual number of changes, n, is random, but it is distributed around a mean of $\Delta\lambda$.

$c_i(\lambda)$ is the percentage effect on y of the 'i^{th}' change. The $c_i(\lambda)$ are random, but identically and independently distributed about some mean $C(\lambda)$.

Those bold presumptions about the pattern of y deserve further explanation. As stated earlier, econometric data represents a broad aggregate of the decisions of millions of people. If we say there are k annual exchanges between buyers and sellers; and prices agreed to by buyers and sellers change in an average of 100N% of the k exchanges; then we can expect λ=kN changes over the course of the year. So long as the k occur evenly throughout the year, $\Delta\lambda$=kN changes should occur in the interval (t,t+Δ).

Further, the changes occur with a constant frequency. And each change's occurrence is independent of the other changes. So, the number of changes $n(t, t+\Delta, \lambda)$ follows a Poisson distribution with mean $\Delta\lambda$ (see pages 21-22 in [2]).

Each time a price changes, the change only affects one of the k exchanges. So each change $c_i(\lambda)$ is very small. The size of each individual c_i is random; but the product of λ changes (the iterated product above) should average to the long-term trend of inflation 1+G. So, $E[1+c_i(\lambda)]$ should be roughly the λ'th root of 1+G. As one can see, when λ is very large and G remains fixed, $E[1+c_i(\lambda)]$ will be very close to one. So $E[c_i(\lambda)]$ will be very close to zero.

Importantly, the result of all those changes should be their product, not their sum. That is because I believe buyers and sellers consider the overall price level (y) rather than the last particular price for their exchange when the price change is determined.

Because there are so many exchanges each year, I believe λ is so large that the limit as $\lambda \to \infty$ is a close approximation to the real world. To that end, I shall define $n(t,t+\Delta,\lambda)$ to be distributed Poisson($\Delta\lambda$) (where $\lambda \to \infty$). The $c_i(\lambda)$'s should be distributed with a mean approximately equal to the λ'th root of 1+G. However, taking the Taylor's series expansion by Z of $(1+G)^Z$, $\ln(1+G)/\lambda$ is a very close approximation to the λ'th root of 1+G (at least as long as $\lambda \to \infty$, so $1/\lambda \to 0$, the Taylor's series approximation works).

Of course, that suggests that the expected value of the $c_i(\lambda)$'s will be zero as $\lambda \to \infty$. But, bear in mind that as the $c_i(\lambda)$'s go to zero, $\lambda \to \infty$. So, the product averages to $(1+G)^\Delta$.

I have deliberately failed to prescribe the distribution of the $c_i(\lambda)$'s. While I have good reasons to believe the number of changes will follow a Poisson distribution, I have no such information on the distribution of change amounts. On the other hand, the central limit law suggests that the only important characteristics of their distribution are the mean and variance.

Now the mathematic framework is set, I will use the phrases 'very small' and 'very large' for the c_i and n throughout the rest of the paper. That should be taken as the case where $\lambda \to \infty$. Further, to simplify matters, I will set $\Delta=1$ and let $c_i = c_i(\lambda)$, $n=n(t,t+1,\lambda)$.

Since the year-to-year change is the limit of an interated product, it is easier to work with the natural logarithm of $y(t)$

$$x(t+1)=\ln(y(t+1))=x(t)+\sum_{i=1}^{n}\ln(1+c_i)$$

But $\ln(1+c_i)$ is very close to zero, and each c_i is very small. So, the Taylor series expansion $\ln(1+c_i)=c_i-c_i^2+2c_i^3-3c_i^4+....$ will contain a small term c_i, and powers of c_i that are orders of magnitude smaller. That indicates $\ln(1+c_i)$ will be very close to c_i when λ is large and c_i is approximately the small quantity $\ln(1+G)/\lambda$. So, when $\lambda \to \infty$

$$x(t+1)=x(t)+\sum_{i=1}^{n} c_i .$$

The $y(t)$ curve was driven by a driving trend $(1+G)^t$. So, if $T=\ln(1+G)$ the expected value of Σc_i should be T. Since the sizes of the changes are independent of the number of changes (n), T must equal $\mu_n \cdot \mu_c$. Since $\mu_n = \lambda \to \infty$, μ_c must equal $(T/\lambda) \to 0$ as noted earlier. Further, the variance of each $x(t+1)-x(t)=\Sigma c_i$ is $\sigma_x^2=\lambda\sigma_c^2+\lambda\mu_c^2$ because of the formula for the collective variance of a count and amount distribution $(\lambda_n\sigma_s^2+\mu_s^2\sigma_n^2$ [3]$)$.

But there is another way to look at the variance. Since the variance generated by the combination of n and the c_i's should converge to the variance $\sigma_x^2 = \mathrm{Var}(x(1)|x(0))$, we should require that $\lambda(\sigma_{c,\lambda}^2+\mu_{c,\lambda}^2)=\sigma_x^2$ for each λ. So, the limit as $\lambda \to \infty$ of the variance $\lambda(\sigma_c^2+\mu_c^2)$ must clearly be the fixed variance σ_x^2. So, even though the precise distribution of the c_i's, is undetermined; $\sigma_c^2+\mu_c^2$ must implicitly be a function of λ(the mean number of changes per unit time). Specifically,

$$\sigma_c^2+\mu_c^2=\sigma_x^2/\lambda .$$

So, the only other criterior for the $c_i(\lambda)$'s is that their variance be $(\sigma_x^2/\lambda) - (\mu_c^2/\lambda)$. As stated earlier, the central limit law will ultimately suggest that all other characteristics of the distribution of the $c_i(\lambda)$'s are irrelevant.

In fact $x(t)$ is a special form of stochastic process. Since $Var(c) \le \sigma_x^2/\lambda$ is finite, the central limit law indicates

$$\frac{1}{n}\sum_{i=1}^{n} c_i$$

is approximately a normal distribution ($\sim N(\mu_c = T/n,\ \sigma_c^2/n)$) when n is very large and fixed. But practically, since $n \sim \text{Poisson}(\lambda)$ and $\lambda \to \infty$, n has an extremely small relative standard deviation $(\sigma_n/\mu_n = \sqrt{\lambda}/\lambda = 1/\sqrt{\lambda} \to 0)$. So, n may be regarded as being nearly invariant when it is large; and for all practical purposes, the total change follows a normal distribution.

$$\Sigma\, c_i \sim N(nT/n,\ n^2\sigma_c^2/n) = N(T,\ n\sigma_c^2 = \lambda\sigma_c^2)\ .$$

These produce the seemingly contradictory results that $\sigma_x^2 = \lambda(\sigma_c^2 + \mu_c^2)$ and $\sigma_x^2 = \lambda\sigma_c^2$. But noting that $E(\Sigma c_i) = T$, μ_c^2 must equal T^2/λ^2. So, as $\lambda \to \infty$, $\mu_c^2 = T^2/\lambda^2 \to 0$ and $\sigma_c^2 = \sigma_x^2/\lambda \to 0$. That means μ_c^2 goes to zero like $1/\lambda^2$ whereas σ_c^2 only decreases like $1/\lambda$. So, the σ_c^2 term predominates and the other μ_c^2 term is functionally zero. And σ_x^2 is roughly equal to $\lambda\sigma_c^2$. In fact, at the limit as $\lambda \to \infty$, σ_x^2 is equal to $\lambda\sigma_c^2$.

Econometric Data as a Random Walk

As I stated earlier, $x(t)$ is actually a special form of stochastic process called a random walk. The expected increase between times t and s is $T(s-t)$. And T does not vary with s or t. Further, the changes over any two disjoint intervals ($x(a)-x(b)$ and $x(s)-x(t)$) are statistically independent with means proportional to the time difference. Mathematically, $E[x(a)-x(b)]=T(a-b)$ and $E[x(s)-x(t)]=T(s-t)$. In the language of stochastic processes, that means x has stationary, independent increments.

But what about the variance? Since the starting point $x(0)$ has not been defined, it does not yet make sense to talk about $\mathrm{Var}(x(t))$. But one can analyze $\mathrm{Var}(x(s)|x(t)=u)$. Consider the changes that affect x as it moves from $x(t)=u$ to $x(s)$. Since λ was the parameter used to denote the (very large) expected number of changes per unit time we expect very close to $n=\lambda(s-t)$ changes of size $c_1, \ldots, c_n$. The analysis of the previous section shows that the conditional distribution $x(s)|x(t)=u$ is a normal distribution with mean $E(n)\cdot T/\lambda=\lambda(s-t)\cdot T/\lambda=(s-t)T$ and variance $n\sigma_c^2=\lambda(s-t)\sigma_c^2$.

But that discrete model of economic change (each choice of λ and the distribution of the $c_i(\lambda)$'s) has an underlying assumption about the variance of the first year's trend. In fact, since the trend and variance are assumed to be independent of the starting value $x(0)$, one could define σ^2 by

$$\lim_{\lambda\to\infty} \lambda\sigma_{c,\lambda}^2=Var(x(1)|x(0)=u)=\sigma_u^2$$

Since the σ_u^2 are independent of u, they are all equal. So we may use the σ^2 they all equal as ${\sigma_x}^2$. And,

$$\mathrm{Var}(x(s)|x(t)=u)=n\sigma_c^2=\lambda(s-t)\sigma_c^2=(s-t)\sigma_x^2.$$

That result is entirely independent of the family of distributions $[c_i(\lambda)]$, as long as each $c_i(\lambda)$ distribution obeys the parameters imposed upon it. In other words, for any appropriate family of distributions $[C(\lambda)]$, the limiting variances will always be proportional solely to time. The above argument shows the resulting variances between times must be some constant variance parameter (σ_x^2) multiplied by the time difference.

That allows us to form some conclusions about this econometric 'random walk'.

1) The conditional distributions $x(s)|(x(t)=k)$ are normal distributions with mean and variance proportional solely to distance and starting point

$$[x(s)|(x(t)=k)] \sim N(k+(s-t)T, (s-t)\sigma^2) .$$

The variance is entirely independent of the starting point and is related solely to distance.

2) Since only the mean in 1) was influenced by the starting point $x(t)=k$. The distribution of $x(s)-x(t)$ is solely a function of the time difference $s-t$, i.e. it is $\sim N((s-t)T, (s-t)\sigma^2)$.

3) The process is piecewise continuous. Said another way, it produces piecewise continuous random walks. This is because $x(t+\Delta) \sim N(x(t)+\Delta T, \Delta\sigma^2)$ means that for any 'small' E

$$\lim_{\Delta \to 0} P(x(t+\Delta) \in (x(t)-E, x(t)+E)) = 1$$

4) The random functions $x(t)$ generated by the process, while continuous, will almost certainly be nondifferentiable (i.e. fractals). That is because the random nature of the process dictates that while $x(t+\Delta)-x(t)$ may show a slope of M; $x(t+\Delta/2)-x(t)$ being random, will show some different slope.

The above conclusions form the classic conditions for a random walk propelled by a constant force (T).

Insurance Data and Imperfect Observation

Of course the goal of most actuarial analyses is to find a better way to use historical insurance data to predict future losses. That requires recognizing both random change and observation error. There is an underlying propensity to loss $x(t)$ that results from a continuous random walk. But since insurance data only provides a random sample of the underlying propensity to loss, insurance data usually represents some $\hat{x}(t)$. The observed values $\hat{x}(t)$ differ from each $x(t)$ by some independent error variables $\epsilon(t) \sim N(0,E^2)$. So, insurance data is characterized by <u>both</u> random change and observation error.

With the prior analysis of econometric data switching between an exponentially trending stochastic process $y(t)$ and its linear trending cousin $x(t)=\ln(y(t))$; it is important to specify which one models insurance data. Insurance data is a reflection of a propensity to loss that is always positive and is subject to exponential inflationary pressures. So insurance data represents $y(t)$. Further, since the driving force behind the increase in $y(t)$ is severity (inflation) rather than frequency, the errors $\epsilon(t)=\hat{y}(t)-y(t)$ should be proportional to $y(t)$. Taking the log transform $x(t)=\ln(y(t))$, $\hat{x}=\ln(\hat{y}(t))$ yields an x subject to a linear random walk. And $\hat{x}$ is such that each $x(t)-\hat{x}(t)$ is from a set of independent, presumably identically distributed [4] $\epsilon(t) \sim N(0,E^2)$.

The insurance problem then reduces to:

> 'Given prior observations $\hat{x}(1)$, $\hat{x}(2)$ ..., $\hat{x}(n)$ of $\log(\hat{y})$, what is the best predictor of $\hat{y}(n+t)=\exp(\hat{x}(n+t))$?'

The Distribution of Future Losses - A Backward Approach

'Forward into the past'

-Firesign Theatre

Obviously, finding the best predictor of $\hat{x}(t+n)$ will involve finding the probability distribution of $x(n+t)$ given observed $\hat{x}(1)$, $\hat{x}(2)$, ..., $\hat{x}(n)$. That distribution will involve finding the reverse likelihood of $\hat{x}(1)$, $\hat{x}(2)$, ..., $\hat{x}(n)$ given $x(t+n)$. The process is complicated by the fact that each $\hat{x}(i)$ is derived from a compound process... first generating $x(i)$ using a random walk, and then generating $\hat{x}(i)$ by adding observation error $\epsilon(i) \sim N(0, E^2)$. Analyzing $x(t+n)|(\hat{x}(i), \hat{x}(j))$ will be especially difficult because the characteristics of a random walk dictate that all three observations will be highly interdependent. Unfortunately, the dependence is through the related variable $x(i)$, not direct.

That indirect dependence requires that parts of the analysis use x rather than $\hat{x}$. To do so requires creating a distribution of $x(i)|\hat{x}(i)$ rather than $\hat{x}(i)|x(i)$.

Determining that 'backward' distribution requires using both Bayes' Theorem and a uniform distribution on $(-\infty,+\infty)$ (a 'diffuse prior' distribution). Appendix I contains a 'reverse probability' theorem. That theorem shows that if the random variable A is a priori uniformly distributed on $(-\infty,+\infty)$ (i.e. each possible value is equally likely), then the density function $f(A=a|B=b)$ is proportional to B given A ($f(B=b|A=a)$). The constant of proportion is $1/\int f(B=b|A=x)\,dx$

That theorem involves the essence of this 'backward' analysis. To determine the likelihood of each potential x(n+t) ($f(x(n+t)|\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(n))$) I will use $f(\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(n)|x(n+t))$. Along the way, I will note that $f(x(i)|\hat{x}(i))=f(\hat{x}(i)|x(i))$ (per Appendix I).

In any event, to determine the likelihood of observing $\hat{x}(1)=\hat{x}_1,\ \hat{x}(2)=\hat{x}_2,\ \ldots$ $\ldots,\ \hat{x}(n)=\hat{x}_n$ given $x(n+t)=x_{n+t}$, it is first necessary to determine the likelihood of any $x(1)=x_1,\ x(2)=x_2$, etc. Then, going backward, while $f(x_1,\ x_2,\ \ldots,x_n|x_{n+t})$ may be complicated, $f(x_n|x_{n+t})$ is distributed $N(x_{n+1}-tT,\ t\sigma^2)$, $f(x_{n-1}|x_n)\sim N(x_n-T,\sigma^2)$, $f(x_{n-1}|x_n)\sim N(x_n-T,\sigma^2)$, $f(x_{n-2}|x_{n-1})\sim N(x_{n-1}-T,\sigma^2)$. Because the random walk has no memory those may be combined. In other words, as long as $s<u<v$, $f(x(s)=x_s|x(u)=x_u\wedge x(v)=x_v)=f(x(s)=x_s|x(u)=x_u)$, so we may multiply the adjacent conditional probabilities to obtain the overall density, $f(x_1,x_2,\ldots,x_n|x_{n+t})$.

Setting

$$f(x(1)=x_1,x(2)=x_2,\ldots x(n)=x_n|x(n+t)=x_{n+t}) = f(x_1,\ x_2,\ \ldots x_n|x_{n+t}),$$

and using the independence of the random change over time,

$$=f(x_n|x_{n+t})\cdot f(x_{n-1}|x_n)\cdot f(x_{n-2}|x_{n-1})\ \ldots\ f(x_1|x_2)$$

$$=(1/((\sqrt{2\pi}(\sqrt{t}\sigma)\ \exp\ (-(x_{n+t}-tT-x_n)^2/(2t\sigma^2))$$
$$\cdot(1/((\sqrt{2\pi}\sigma))\ \exp\ (-(x_n-T-x_{n-1})^2/(2\sigma^2))$$
$$\cdot(1/((\sqrt{2\pi}\sigma))\ \exp\ (-(x_{n-1}-T-x_{n-2})^2/(2\sigma^2))$$
$$\ldots$$
$$\cdot(1/((\sqrt{2\pi}\sigma))\ \exp\ (-(x_2-T-x_1)^2/(2\sigma^2))$$

$$=[1/((\sqrt{2\pi})^n\sigma^n\sqrt{t})]\ \exp\ [-(1/2)((x_{n+t}-tT-x_n)^2/(t\sigma^2)+(1/\sigma^2)\sum_{i=1}^{n-1}(x_i+T-x_{i+1})^2)].$$

Further, since the ϵ_i's are independent, identically distributed, and independent of the x_i's

$$f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n | x_1, \ldots, x_n)$$

$$= (1/((\sqrt{2\pi}E)) \exp(-(x_n-\hat{x}_n)^2/(2E^2))$$
$$\cdot (1/((\sqrt{2\pi}E)) \exp(-(x_{n-1}-\hat{x}_{n-1})^2/(2E^2))$$
$$\ldots$$
$$\cdot (1/((\sqrt{2\pi}E)) \exp(-(x_1-\hat{x}_1)^2/(2E^2))$$

$$= [1/(\sqrt{2\pi}E)^n] \cdot \exp[-(1/2)((1/E^2) \cdot \sum_{i=1}^{n} (x_i-\hat{x}_i)^2)] .$$

So, since the ϵ's are independent of the x's

$$f(x_1, x_2, \ldots, x_n, \hat{x}_1, \ldots, \hat{x}_n | x_{n+t})$$

$$= [1/((2\pi)^n \sigma^n \sqrt{t} E^n)] \cdot \exp[-(1/2)((x_{n+t}-tT-x_n)^2/(t\sigma^2)$$
$$+ (1/\sigma^2) \sum_{i=1}^{n-1} (x_i+T-x_{i-1})^2 + (1/E^2) \sum_{i=1}^{n} (x_i-\hat{x}_i)^2)] .$$

Then, to eliminate the reliance on $x_1, \ldots, x_n$, all that is necessary is to integrate over all possible x_i's, i.e.

$$f(\hat{x}_1, \ldots, \hat{x}_n | x_{n+t})^{[5]}$$

$$= \int_{x_1}\int_{x_2} \ldots \int_{x_n} f(\hat{x}_1, \ldots, \hat{x}_n, | x_1, x_2, \ldots, x_n, x_{n+t})\, dx_n, \ldots, dx_1$$

$$= [1/((2\pi)^n \sigma^n \sqrt{t} E^n) \cdot \int_{x_1}\int_{x_2} \ldots \int_{x_n} \exp[-(1/2)((x_{n+t}-tT-x_n)^2/(t\sigma^2)$$
$$+ (1/\sigma^2) \sum_{i=1}^{n-1} (x_i+T-x_{i+1})^2 + (1/E^2) \sum_{i=1}^{n} (x_i-\hat{x}_i)^2)] \quad dx_n \ldots dx_2\, dx_1$$

Ultimately, the best predictors of x_{n+t} will maximize that function. But since it is very unwieldy, a brief digression will illustrate what it means in concrete situations.

Two Extreme Examples

To gain some insight into the structure underlying the 'best' predictor of x_{n+t}, I will analyze two extreme examples. One is the case of 'total determinism' ($\sigma^2=0$). The other is 'perfect observation' ($E^2=0$)

'Total determinism' ($\sigma^2=0$) fulfills all the criteria needed for regression: 1) The underlying exposure $x(t)$ is a straight line; and 2) The only reason the observed data $x(t)$ do not fall on a straight line is the presence of independent, identically distributed, observation errors $\epsilon(t)$.

The fitted line $x(t)=\bar{x}+m(t-\bar{t})$ represents the regression estimate. Further, since the vectors $a_1=[1, 1, \ldots,]$ and $a_2=[-(n-1)/2, -(n-3)/2, \ldots, (n-1)/2]$ are independent, we can use them to produce the regression. Since a_1 is a 'pure constant', $\bar{x}=a_1\cdot[\hat{x}_i]/\|a_1\|^2$. And, since a_2 is pure slope, $m=a_2\cdot[\hat{x}_i]/\|a_2\|^2$. But, after some algebra, $a_2\cdot[\hat{x}_i]$ may be rewritten

$$a_2\cdot[\hat{x}_i]=[\sum_{i=1}^{n}[i-((n+1)/2)]\hat{x}_i]/\sum_{i=1}^{n}[i-((n+1)/2)].$$

Which, after some series algebra become

$$=K\sum_{i=1}^{n-1}\frac{(in-i^2)}{2}(\hat{x}_{i+1}-\hat{x}_i)$$

(where K is constant with respect to $[\hat{x}_i]$ and the $in-i^2$ are the weights used on the differences $(\hat{x}_{i+1}-\hat{x}_i)$).

So, regression is based on averaging over the observation period. The prediction keys off an average value of x - roughly its predicted value at the middle of the observation times. It adds a slope multiplied by the time elapsed since the middle of the observation times. The slope is computed by using a weighted average of year-to-year changes in $\hat{x}$. Just as the mean keys off the middle of the observation times, the weights applied to year-to-year changes place heavier weight near the middle of the observation period (Consider the shape of $in-i^2$. It is a quadratic with a maximum at $n/2$). In short, regression is oriented toward the middle of the observation period.

The 'perfect observation' case ($E^2=0$) produces estimates based largely on the latest point. Since the series has no memory, (i.e. $u<v<t$ implies $f(x(t)=x_t|x(v)=x_v) = f(x(t)=x_t|x(v)=\hat{x}_v \wedge \hat{x}(u)=x_u)$ the points prior to $\hat{x}_n=x_n$ are irrelevant except for estimating trend. In other words, the best estimate of $x(n+t)$ will be x_n+tT.

To estimate T, note that the perfect observation of the $\hat{x}_i$'s means there is no ϵ_i influencing either $\hat{x}_i-\hat{x}_{i-1}$ or $\hat{x}_{i+1}-\hat{x}_i$. Consequently, each $\hat{x}_{i+1}-\hat{x}_i$ is independent. So, each $\hat{x}_{i+1}-\hat{x}_i$ is an independent, identically distributed estimate of T. Thus, the best estimate of T is their average $T'=(1/(n-1))\sum_{i=1}^{n}\hat{x}_{i+1}-\hat{x}_i$.

Telescoping the differences produces $T'=(\hat{x}_n-\hat{x}_1)/(n-1)$. Combining the two results yields the optimum estimate for x_{n+t}

$$x_{n+t}=x_n+T(\hat{x}_n-\hat{x}_1)/(n-1) .$$

(To verify the above verbal argument, set $\hat{x}_i=x_i$ in the integral shown previously and maximize. The E^2 as a constant is superfluous.) So, the 'perfect observation' case dictates that the constant be the last observed point and the trend be an equal weighting of the observed differences.

Summarizing, the two extreme cases both key off a fixed point and a trend from the fixed point. In the case where $\sigma^2=0$ the fixed point is the mean of the observed points and the trend is a weighted average of the annual change (alternately, one could view the trended mean $\overline{x}+(n/2)T$ as the fixed point). In the perfect observation case ($E^2=0$) the trend is a straight average of the annual changes. From another perspective, when $E^2=0$ the fixed point applies 100% weight to the last observed point, and when $\sigma^2=0$ the fixed point equation applies equal weight to all the observed points.

In the typical case both E^2 and σ^2 will be non-zero. The key question is 'Where will the fixed point and trend lie between those extremes?'

The General Solution

'The only solution... isn't it amazing'

Jim Morrison

Appendix III shows the best estimator of x_{t+n} given observed $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$, a predetermined trend T, and a predetermined ratio E^2/σ^2. It uses a weighted average of the trended observed points for the fixed point and the trend T beyond the fixed point. The weights do not lend themselves to a closed-form formula readily, but they are easy to compute.

First, you compute the recursive values, F_i. To start, set

$$F_1=1,\ F_2=E^2+\sigma^2.$$

Then, you calculate each succeeding F_i using

$$F_{i+1}=(2E^2+\sigma^2)F_i-E^4F_{i-1}.$$

And then the best estimator of x_{n+c} is

$$x_{n+c} = cT + [\sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i)T)] / [\sum_{i=1}^{n} F_i E^{2(n-i)}] .$$

(i.e. the weights for the fixed point are $F_i E^{2(n-i)}$).

Unfortunately, that estimator depends on first choosing the average trend T and the variance relationship E^2/σ^2. Appendix IV contains an estimating formula for the trend, T. The author has not yet determined the best estimator for E^2 and σ^2, but the estimating process used in appendices III and IV could be extended to produce an estimate for them as well.

In any event, the formula provides a means of assigning weights for each of the last five available years of fire experience, or each of the last three years of workers compensation class experience, etc. That alone makes it useful.

Credibility Against Straight Trend - Exponential Smoothing and Ratemaking

A useful by-product of the previous formula is a credibility formula to use when the complement of credibility is applied to straight trend.

Specifically, when the ratemaking formula is

$$ZL+(1-Z)(R+T)=R' .$$

Where L represents the rate based on raw experience, R is the existing rate, T is trend, and R' is the result of credibility. Then, the best credibility (Z) is

$$Z = [\sigma^2 + \sigma\sqrt{4E^2+\sigma^2}] / [2E^2 + \sigma^2 + \sigma\sqrt{4E^2+\sigma^2}]$$

(where E^2 and σ^2 are as defined previously).

To prove this, first note that

$$R(i+1)=ZL(i+1)+(1-Z)(T+R(i)).$$

So,

$$R(i+1)=ZL(i+1)+(1-Z)(T+ZL(i)+(1-Z)R(i-1))$$

$$=ZL(i+1)+Z(1-Z)(L(i)+T)+(1-Z)^2(R(i-1)+T).$$

And, extending the expansion

$$R(i+1)=Z\sum_{j=0}^{i}(L(i-j)+jT)(1-Z)^j.$$

so, R is really an exponentially smoothed estimate of the loss level with smoothing parameter $(1-Z)$.

Next, I will show that the $F_iE^{2(n-i)}$ weights are also exponential in character. A theorem from numerical analysis states that the results of a recursion relation $ax_{n+1}=bx_n+cx_{n-2}$ will be $K_1r_1^n+K_2r_2^n$; where r_1 and r_2 are the roots of $ax^2-bx-c=0$.

In the case of the F_i's this means a linear combination of the form

$$F_i=K_1[(2E^2+\sigma^2+\sigma\sqrt{4E^2+\sigma^2})/2]^i + K_2[(2E^2+\sigma^2-\sigma\sqrt{4E^2+\sigma^2})/2]^i$$

But, as i gets very large, the larger root's power will grow much faster than the smaller root's. So, for large i

$$F_i\approx K_1[(2E^2+\sigma^2+\sigma\sqrt{4E^2+\sigma^2})/2]^i.$$

Now, in the estimating formula for x_{t+n}, the weights are $F_i E^{2(n-i)}$. So the smoothing parameter for successively older observed points is roughly

$$F_{i-1}E^{2(n-i+1)}/(F_i E^{2(n-i)})=E^2 F_{i-1}/F_i\,,$$

or

$$2E^2/(2E^2+\sigma^2+\sigma\sqrt{4E^2+\sigma^2})\;.$$

Since $(1-Z)$ is the smoothing parameter,

$$Z = 1-[2E^2/(2E^2+\sigma^2+\sqrt{4E^2+\sigma^2})] = (\sigma^2+\sigma\sqrt{4E^2+\sigma^2})/(2E^2+\sigma^2+\sigma\sqrt{4E^2+\sigma^2})\;.$$

which is the result we seek.

Parenthetically, note that since trend is usually exponential rather than linear a logarithmic transform produces the formula $L(i)^Z\cdot(R(i)(1+T))^{1-Z}$ rather than the linear sum formula $ZL(i)+(1+Z)R(i)(1+T)$.

Summary

The random nature of most economic forces creates random behavior in econometric data, especially insurance data. So, the most effective way to project econometric series involves viewing them as a random walks. Within the general framework that imposes, the projection becomes a compromise between: 1) formula trend and random observation; and 2) random trend and error-free observation. Two of the formulas presented in this paper illustrate the 'most accurate' estimators for random walk data. The author believes those formulas to be merely the beginning. Viewing insurance data as a random walk will give actuaries many opportunities to refine our formulas and thereby make better predictions.

The author wishes to acknowledge the assistance of Greg DeCroix. Greg's assistance in finding the recursion formula for the F_i's was invaluable. The author also wishes to thank Darlene Hodges, who typed many revisions of this paper.

DIFFUSE PRIORS AND BAYES THEOREM

Many problems seek an unknown quantity (such as the best rate to charge) which could, a priori, be any number. They can be solved through uniform distributions on infinite intervals. Those are called diffuse priors. For example, a basic problem in statistics involves the following scenario: Observed data from a normal distribution $x_1, x_2, \ldots, x_n$ are available. There are sufficient data points to give an acceptable estimate of the mean (x) and variance (σ^2), but the distribution of the true mean μ is desired. A priori, all the potential $\mu\epsilon(-\infty,\infty)$ are equally likely candidates, but obviously the μ close to $\bar{x}$ deserve greater probability.

If μ and the x_i were restricted to some finite interval (a, b) then Bayes' theorem would yield

$$f(\mu|[x_i])=f([x_i]|\mu)\cdot f(\mu)/f([x_i])=f([x_i]|\mu)(b-a)^{n-1}$$

In other words, since b-a is constant, Bayes theorem indicates the likelihood of μ given $[x_i]$ is proportional to the likelihood of those $[x_i]$ given μ.

The problem lies when the $[x_i]$ and μ, a priori, take any value in $(-\infty,\infty)$ with equal likelihood (i.e. they are uniformly distributed on $(-\infty,\infty)$). The solution involves the use of 'diffuse priors' (uniform distributions on infinite sets). The author is not familiar with whatever approaches to diffuse priors are currently used by others, but I hope to convey enough of my thinking to solve the practical problems underlying this paper.

Conceptually, one could use the infinitesimal, I, sometimes used in mathematical logic. I is a (entirely theoretical) constant that is infinitely close to zero, but non-zero. So

$$\int_{-\infty}^{\infty} \mathrm{I}\,dt=1$$

Thus, if we use the a priori distribution

$$f(u)=\mathrm{I},\quad f([x_n])=\mathrm{I}';$$

then

$$f(\mu|[x_i])=f([x_i]|\mu)\,\mathrm{I}/\mathrm{I}'.$$

So, the probability of μ given the observed $[x_i]$ is proportional to the probability those $[x_i]$ would be observed when μ is the underlying mean.
In the event the $[x_i]$ come from a normal $N(\mu,\sigma^2)$ distribution, σ^2 may be determined fairly accurately from the observed x_i's. So,

$$f(\mu|[x_i])=(1/(\sigma\sqrt{2\pi}))^n\exp[(-(1/2\sigma^2)\Sigma(x_i-\mu)^2]\cdot(\mathrm{I}/\mathrm{I}')$$

which probability formulas[6] reduce to a normal distribution for the mean

$$(\sqrt{n}/(\sigma\sqrt{2\pi}))\exp[-(n/(2\sigma^2))(\overline{x}-\mu)^2]\cdot K.$$

But, since

$$\int\exp[(-(n/(2\sigma^2))(\overline{x}-\mu)^2)]\,d\mu=\sigma\sqrt{2\pi}/\sqrt{n}$$

we conclude that $K=1$, and

$$\mu\sim N(\overline{x},\sigma^2/n).$$

In general, if A and B have uniform diffuse prior distributions, then $P(A=a|B=b)=P(B=b|A=a)\cdot K$. In other words, the probability of A given B is proportional to the probability of B given A.

Mathematical Niceties

At least one article [7] suggests that Bayes' original concept of a uniform distribution on $(-\infty, \infty)$ consisted of a normal distribution with infinite variance, e.g.

$$\lim_{\sigma\to\infty} N(\mu, \sigma^2).$$

Of course, that inevitably produces a specific mean and mode for the prior distribution of μ. According some specific μ that favored status makes the distribution somewhat less than uniform. But, if one were seeking to prove some $G(x)=0$ for a uniform distribution on $(-\infty, \infty)$; one could say: If

$$\lim_{\sigma\to\infty} G(x|N(\mu,\sigma^2))=0.$$

For <u>all</u> μ, $G(x)=0$ holds for the uniform distribution on $(-\infty, \infty)$.

The author has two alternate, but potentially mathematically equivalent, approaches. The first one involves a limit of uniform distributions. In this case the requirement is that

$$\lim_{n\to\infty} G(x|U(a_n, b_n))=0$$

($U(a_n, b_n)$ representing the uniform distribution on the interval (a_n, b_n)).

More important, that result must hold for all sequences $[a_n]$ and $[b_n]$ such that $a_n\to-\infty$ and $b_n\to\infty$.

More generally, one could require that $G(x|f_n) \to 0$ for all sequences of density functions $\{f_n\}$ with an infinite, flat limit. Specifically,

$$\lim \text{(non-zero domain of } f_n) = (-\infty, \infty)$$

and

$$\lim_{n\to\infty}[\max(f_n(x))/\min(f_n(x))]=1^{[8]}$$

Whichever definition you choose, it is clear that the formulas earlier in this paper, which use I, hold.

Pitfalls

The typical problem with diffuse priors is actually a problem with finite uniform distributions, too. There may be uncertainty over what is to be uniformly distributed. For example, when developing a prior distribution for the mean, μ, of a normal distribution it is fairly clear that μ should be uniformly distributed on $(-\infty,\infty)$. But what about the variance, σ^2? Should σ^2 be uniformly distributed on $[0,\infty)$, or should σ be uniformly distributed on $(-\infty,\infty)$? Making σ^2 uniformly distributed inherently makes 'small' σ^2 more likely than making σ uniformly distributed. So, when it is not clear what should be uniformly distributed, diffuse prior distributions are inappropriate.

Fortunately, in this paper the author has used diffuse priors solely for estimating means. So, the variance issue is moot. But, there are other situations, outside the scope of this paper, where problems may arise.

INTRODUCTORY LEMMAS

Before proceeding to prove that the $F_i E^{2(n-i)}$'s are the best weights for historical experience, it will be helpful to prove two lemmas.

Lemma 1: Weighted Squared Difference Theorem.

The weighted sum of squared differences equals the squared difference from the weighted mean plus the squared differences. Mathematically,

$$\sum_{i=1}^{n} W_i (a_i - x)^2 = (\sum_{i=1}^{n} W_i)(x - [(\Sigma W_i a_i)/\Sigma W_i])^2 + (1/\Sigma W_i) \sum_{i=1}^{n} \sum_{j \le i} W_i W_j (a_i - a_j)^2$$

Practically, this means that the estimate x which minimizes the weighted squared differences from the observed points $\{a_i\}$ is the weighted average of the a_i's. Further, the residual error after choosing that best estimate consists of squared differences between the a_i's. Each such difference is weighted by the weights of the two a_i's in the difference.

The most straightforward way to prove this involves placing the weighted mean inside the sum and using brute force.

$$\sum_{i=1}^{n} W_i (a_i - x)^2 = \sum_{i=1}^{n} W_i ([(\Sigma W_j a_j / \Sigma W_j) - x] + [a_i - (\Sigma W_j a_j / \Sigma W_j)])^2$$

Expanding the square,

$$= \sum_{i=1}^{n} W_i ([\Sigma W_j a_j / \Sigma W_j) - x]^2 + 2[(\Sigma W_j a_j / \Sigma W_j) - x][a_i - (\Sigma W_j a_j / \Sigma W_j)] + [a_i - (\Sigma W_j a_j / \Sigma W_j)]^2).$$

Then, distributing the summation across the three sums,

$$= (\Sigma W_i)\,[x-(\Sigma W_j a_j/\Sigma W_j)]^2 + 2\,[(\Sigma W_j a_j/\Sigma W_j)-x]\,[\Sigma W_i a_i - \Sigma W_j a_j] + \Sigma W_i\,[a_i-(\Sigma W_j a_j/\Sigma W_j)]^2.$$

Noting that $\Sigma W_i a_i = \Sigma W_j a_j$, the polynomial equals

1) $$= (\Sigma W_i)\,[x-(\Sigma W_i a_i/\Sigma W_i)]^2 + (\Sigma W_i)\,[a_i-(\Sigma W_j a_j/\Sigma W_j)]^2.$$

Computing the square in the last term, note that

$$\Sigma W_i\,[a_i-(\Sigma W_j a_j/\Sigma W_j)]^2$$

$$= \Sigma W_i a_i^2 - 2(\Sigma W_i a_i)(\Sigma W_i a_i)/(\Sigma W_i) + (\Sigma W_i)(\Sigma W_j a_j)^2/(\Sigma W_j)^2,$$

$$= (1/\Sigma W_i)\,[(\Sigma W_i a_i^2)(\Sigma w_i) - 2(\sum_i\sum_j W_i W_j a_i a_j) + (\Sigma W_i a_i)^2],$$

$$= (1/\Sigma W_i)\,[(\Sigma W_i a_i^2)(\Sigma W_i) + \sum_i\sum_j W_i W_j (a_i a_j - 2a_i a_j)],$$

$$= (1/\Sigma W_i)\,[\sum_i\sum_j W_i W_j a_i^2 - \sum_i\sum_j W_i W_j a_i a_j].$$

Splitting the sums up into the cases where j is less than, equal to, or greater than i.

$$= (1/\Sigma W_i)\,[\sum_{i=2}^{n}\sum_{j=1}^{i-1} W_i W_j a_i^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_i W_j a_i^2$$

$$+ \Sigma W_i^2 a_i^2 - \sum_{i=2}^{n}\sum_{j=1}^{i-1} W_i W_j a_i a_j - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_i W_j a_i a_j - \Sigma W_i^2 a_i^2].$$

Subtracting the $\Sigma W_i^2 a_i^2$ terms that cancel, and interchanging i and j in two of the indices

$$=(1/\Sigma W_i)\,[\sum_{i=2}^{n}\sum_{j=1}^{i-1} W_iW_ja_i^2+\sum_{i=2}^{n}\sum_{j=1}^{i-1} W_iW_ja_j^2-\sum_{i=2}^{n}\sum_{j=1}^{i-1} W_iW_ja_ia_j-\sum_{i=2}^{n}\sum_{j=1}^{i-1} W_iW_ja_ia_j]\,.$$

Collecting terms

$$=(1/\Sigma W_i)\sum_{i=2}^{n}\sum_{j=1}^{i-1}(W_iW_ja_i^2+W_iW_ja_j^2-2W_iW_ja_ia_j)\,,$$

$$=(1/\Sigma W_i)\sum_{i=2}^{n}\sum_{j=1}^{i-1}W_iW_j(a_i-a_j)^2\,.$$

Adding the case where $i=j$; $(a_i-a_j)=0$

$$=(1/\Sigma W_i)\sum_{i=1}^{n}\sum_{j\le i}W_iW_j(a_i-a_j)^2$$

Now, substituting that result back into 1) yields the lemma:

$$\Sigma W_i(a_i-x)^2=(\Sigma W_i)\,(x-[\Sigma W_ia_i/\Sigma W_i])^2+(1/\Sigma W_i)\sum_{i}\sum_{j\le i}W_iW_j(a_i-a_j)^2\,.$$

Exponential Integral Theorem

A textbook theorem used to analyze multivariate normal distributions states

$$\int_{-\infty}^{\infty}\exp(-(1/2)\,[(x-G)^2/\sigma^2+H])\,dx=\sigma\sqrt{2\pi}\exp(-H/2)\,.$$

The proof is comparatively simple. Exp$(-H/2)$ is constant with respect to the variable of integration (x). So

$$\int_{-\infty}^{\infty} \exp(-(1/2)[(x-G)^2/\sigma^2+H])\,dx=\exp(-H/2)\int_{-\infty}^{\infty} \exp((x-G)^2/(2\sigma^2))\,dx.$$

But up to the constant $1/(\sigma\sqrt{2\pi})$ the integral is simply the density of a normal $N(G,\sigma^2)$ distribution. So its integral is $\sigma\sqrt{2\pi}$. Thus, the theorem holds:

$$=\exp(-H/2)\cdot\sigma\sqrt{2\pi}=\sigma\sqrt{2\pi}\,\exp(-H/2).$$

Lemma 2) Integral of Weighted Squared Differences

$$\int_{-\infty}^{\infty} \exp(-(1/2)\Sigma W_i(x-a_i)^2)\,dx$$

$$=\sqrt{2\pi/\Sigma W_i}\,\exp[-(1/2)(\sum_i \sum_{j\leq i} W_i W_j(a_i-a_j)^2)/(\Sigma W_i)]$$

This lemma is a straightforward combination of Lemma 1 and the exponential integral theorem.

PROOF OF THE FIXED POINT ESTIMATOR FORMULA

To prove that

$$1)\quad e_{n+t} = tT + [\sum_{1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i)T)] / [(\sum_{1}^{n} F_i E^{2(n-i)})]$$

is the best estimator for x_{n+t}, I need to first integrate the x_{n+t} density function. Then, the formula will result from some simple algebra which proves the recursion relation.

Using a diffuse prior argument

$$2)\quad f(x_{n+t}|\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$$

$$= Kf([\hat{x}_i]_1^n | x_{n+t}),$$

$$= K(1/[(2\pi)^n \sigma^n \sqrt{t} E^n]) \int_{x_n} \ldots \int_{x_1} \exp[-(1/2)((x_{n+t} - tT - x_n)^2/(t\sigma^2)$$

$$+ (1/\sigma^2)[\sum_{i=1}^{n-1} (x_i + T - x_{i+1})^2] + (1/E^2)[\sum_{i=1}^{n} (x_i - \hat{x}_i)^2])] \cdot dx_1 \ldots dx_n.$$

Combining the K into K′ (a function independent of x_{n+t}) multiplied by an exponent of squared differences

$$3)\quad = K'([x_i]_1^n, E^2, \sigma^2, t, T) \cdot \exp[-(1/2) K''(E^2, \sigma^2, t) \sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i+t)T - x_{n+t})^2]$$

Showing that the estimator from 1) minimizes that sum of squared differences will then suffice to show it is the best estimator of x_{n+t}

To solve the multiple integral from 2) I need to first prove a theorem

Multiple Integral Theorem

Given:

1) observed points $[\hat{x}_i]_1^n$ distributed around unknown means $[x_i]_1^n$;

2) generated by a normal stochastic process with mean increase T and variance parameter σ^2;

3) where each of the $[\hat{x}_i]$ differ from the $[x_i]$ by an independent $N(0, E^2)$ distribution;

4) and the times between valuation are t_i (so $f(x_{i+1}|x_i) \sim N(t_iT, t_i\sigma^2)$));

the integral

4) $$I(x_{n+1}) = \int_{x_n} \ldots \int_{x_1} \exp(-(1/2)\{(1/E^2)[\sum_{i=1}^{n}(\hat{x}_i - x_i)^2]$$

$$+[\sum_{i=1}^{n}(x_i + t_iT - x_{i+1})/(t_i\sigma^2)]\})\,dx_1, \ \ldots \ dx_n$$

$$= K([\hat{x}_i]_1^n, [t_i]_1^n, E^2, \sigma^2, T)\cdot\exp(-(1/2)[(1/F_{n+1})\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i + T(\sum_{j=i}^{n}t_j) - x_{n+1})^2]).$$

Where,

5) $F_1 = 1$

$$F_{n+1} = t_n\sigma^2(\sum_{i=1}^{n}F_iE^{2(n-i)}) + E^2F_n$$

I will prove it using mathematical induction. The proof for $n=1$ is trivial. Next, I must show the result holds for $I(x_{n+2})$ when it holds for $I(x_{n+1})$.

Note that

$$I(x_{n+2}) = \int_{x_{n+1}} \ldots \int_{x_1} \exp(-(1/2)[(1/E^2)[\sum_{i=1}^{n+1}(\hat{x}_i - x_i)^2]$$

$$+[\sum_{i=1}^{n+1}(x_i + t_i T - x_{i+1})^2/(t_i\sigma^2)]])\ dx_1,\ \ldots,\ dx_{n+1}$$

So, pulling out the terms that are constant with respect to $x_1, \ldots, x_n$

$$= \int_{x_{n+1}} \exp(-(1/2)[(\hat{x}_{n+1} - x_{n+1})^2/E^2 + (x_{n+1} + t_{n+1}T - x_{n+2})/(t_{n+1}\sigma^2)])$$

$$\cdot \int_{x_n} \ldots \int_{x_1} \exp(-(1/2)\big[[(1/E^2)[\sum_{i=1}^{n}(\hat{x}_i - x_i)^2]$$

$$+[\sum_{i=1}^{n}(x_i + t_i T - x_{i+1})^2/(t_i\sigma^2)]\big])\,dx_1,\ \ldots,\ dx_{n+1}.$$

Then the inner 'n' integrals may use the induction hypothesis

$$= \int_{x_{n+1}} \exp(-(1/2)[(\hat{x}_{n+1} - x_{n+1})^2/E^2 + (x_{n+1} + t_{n+1}T - x_{n+2})^2/(t_{n+1}\sigma^2)]) \cdot I(x_{n+1})\,dx_{n+1},$$

$$= \int_{x_{n+1}} \exp(-(1/2)[(\hat{x}_{n+1} - x_{n+1})^2/E^2 + (x_{n+1} + t_{n+1}T - x_{n+2})^2/(t_{n+1}\sigma^2)])$$

$$\cdot K([\hat{x}_i]_1^n, [t_i]_1^n, E^2, \sigma^2, T)$$

$$\cdot \exp(-(1/2)(1/F_{n+1})[\sum_{i=1}^{n} F_i E^{2(n-i)}(\hat{x}_i + T(\sum_{j=i}^{n} t_j) - x_{n+1})^2]) \cdot dx_{n+1}.$$

6) $= K([\hat{x}_i]_1^n, [t_i]_1^n, E^2, \sigma^2, T) \int_{x_{n+1}} \exp(-\frac{A}{2})\, dx_{n+1}$.

Where

$$A = (\hat{x}_{n+1} - x_{n+1})^2/E^2 + (x_{n+1} + t_{n+1}T - x_{n+2})^2/(t_{n+1}\sigma^2) + (1/F_{n+1}) \sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + T(\sum_{j=i}^{n} t_j) - x_{n+1})^2 .$$

Now to evaluate A, the first step is to apply the integral of weighted squared differences lemma (lemma from the previous appendix) using x_{n+1} as x. Specifically,

7) $\int_{x_{n+1}} \exp(-A/2)\, dx_{n+1}$

$$= (2\pi/[(1/E^2) + (1/(t_{n+1}\sigma^2)) + (1/F_{n+1}) \sum_{i=1}^{n} F_i E^{2(n-i)}])^{1/2}$$

$$\cdot \exp(-(1/2)\Big[[(\hat{x}_{n+1} + t_{n+1}T - x_{n+2})^2/(t_{n+1}E^2\sigma^2)]$$

$$+ [(1/(F_{n+1}E^2)) \sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + T(\sum_{j=i}^{n} t_j) - \hat{x}_{n+1})^2]$$

$$+ [(1/F_{n+1}^2) \sum_{i=1}^{n} \sum_{j \le i} F_i F_j E^{2(2n-i-j)} (\hat{x}_j + T(\sum_{k=j}^{i-1} t_k) - \hat{x}_i)^2]$$

$$+ [(1/(t_{n+1}\sigma^2 F_{n+1})) \sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + T(\sum_{j=i}^{n+1} t_j) - x_{n+2})^2]\Big]$$

$$/ [(1/E^2) + (1/(t_{n+1}\sigma^2)) + (1/F_{n+1}) \sum_{i=1}^{n} F_i E^{2(n-i)}])$$

That produces quite a long expression. But, noting that the long 'sum of the weights' term

$$(1/E)^2+(1/(t_{n+1}\sigma^2))+(1/F_{n+1})\sum_{i=1}^{n}F_iE^{2(n-i)}$$

$$=(1/(E^2t_{n+1}\sigma^2F_{n+1}))\,[t_{n+1}\sigma^2F_{n+1}+E^2F_{n+1}+t_{n+1}\sigma^2E^2\sum_{i=1}^{n}F_iE^{2(n-i)}]\ ;$$

and combining the $t\sigma^2$ terms

$$=(1/(t_{n+1}E^2\sigma^2F_{n+1}))\,[t_{n+1}\sigma^2\sum_{i=1}^{n+1}F_iE^{2(n+1-i)})+E^2F_{n+1}]$$

$$=F_{n+2}/(F_{n+1}t_{n+1}E^2\sigma^2)\ ;$$

Then, plugging that back in 7)

$$\int_{x_{n+1}}\exp(-A/2)\,dx_{n+1}$$

$$=(\sqrt{2\pi t_{n+1}F_{n+1}/F_{n+2}})\,E\sigma\cdot\exp(-(F_{n+1}t_{n+1}E^2\sigma^2/(2F_{n+2}))$$

$$\Big[[(\hat{x}_{n+1}+t_{n+1}T-x_{n+2})^2/(t_{n+1}E^2\sigma^2)]+[(1/(F_{n+1}E^2))\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i+T(\sum_{j=i}^{n})-\hat{x}_{n+1})^2]$$

$$+[(1/F_{n+1}^2)\sum_{i=1}^{n}\sum_{j\le i}F_iF_jE^{2(2n-i-j)}(\hat{x}_j+T(\sum_{k=j}^{i-1}t_k)-\hat{x}_i)^2]$$

$$+[(1/(t_{n+1}\sigma^2F_{n+1}))\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1}t_j)-x_{n+2})^2]\Big])$$

That is still quite a lengthy expression. But, part of it may be reduced immediately. Since the multiplier in front of the function and the middle two terms in the sum are constant with respect to x_{n+2},

$$=K'([\hat{x}_i]_1^{n+1}, [t_i]_1^{n+1}, E^2, \sigma^2, T) \cdot \exp\Big\{-(F_{n+1}t_{n+1}E^2\sigma^2/(2F_{n+2}))$$

$$\cdot\Big[[(\hat{x}_{n+1}+t_{n+1}T-x_{n+2})^2/(t_{n+1}E^2\sigma^2)]+[(1/(t_{n+1}\sigma^2F_{n+1}))\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1}t_j)-x_{n+2})^2]\Big]\Big\}$$

That is reduced, but still lengthy. Applying the top of the quotient to the sums

$$=K' \ \exp\Big\{(-(1/2)(1/F_{n+2})\Big[[F_{n+1}(\hat{x}_{n_1}+t_{n+1}T-x_{n+2})^2]$$

$$+[E^2\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1}t_j)-x_{n+2})^2]\Big])\Big\}$$

Adding the $n+1$ term to the sum

$$=K' \ \exp(-(1/2)(1/F_{n+2})\sum_{i=1}^{n+1}F_iE^{2(n+1-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1})-x_{n+2})^2) .$$

Then plugging that back in the original formula in 6)

$$I(n+2)=K\cdot K' \ \exp \ (-(1/2)(1/F_{n+2})\sum_{i=1}^{n+1}F_iE^{2(n+1-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1}t_j)-x_{n+2})^2) ,$$

$$=K([x_i]_1^{n+1}, [t_i]_1^{n+1}, E^2, \sigma^2, T)\cdot\exp\Big\{-(1/2)(1/F_{n+2})\sum_{i=1}^{n+1}F_iE^{2(n+1-i)}(\hat{x}_i+T(\sum_{j=i}^{n+1}t_j)-x_{n+2})^2\Big\}.$$

So the induction hypothesis is proven and the integral evaluation theorem holds.

The Best Estimator

Now that we know the density function $f(x_{n+t}|[\hat{x}_i]_1^n, E^2, \sigma^2, T)$, the next step is to show that the estimator

$$e_{n+t} = tT + [\sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i)T)] / (\sum_{i=1}^{n} F_i E^{2(n-i)})]$$

is the optimum estimator for x_{n+t}. The key is to show that the true x_{n+t} is normally distributed around e_{n+t}

$$f(x_{n+t}|[\hat{x}_i]_1^n, E^2, \sigma^2, T) \sim N(e_{n+t}, \delta^2) .$$

Then, since e_{n+t} is both the mean and the mode of the distribution, it must be the best estimator.

Plugging the results of the integration theorem into the earlier formula for $f(x_{n+t})$,

$$f(x_{n+t}|[\hat{x}_i]_1^n, t, E^2, \sigma^2, T)$$

$$= K\exp(-(1/2)(1/F'_{n+1}) \cdot \sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i+t)T - x_{n+t})^2)$$

Using the weighted sum of squares lemma (Lemma 1) from appendix II, (note $F'_{n+1} = t\sigma^2(\sum_{i=1}^{n} F_i E^{2(n-i)}) + E^2 F_n$ instead of $\sigma^2(\sum F_i E^{2(n-i)}) + E_2 F_n$ because of the $t\sigma^2$ in the last term)

$$= K \exp\left\{(-(1/2)(1/F'_{n+1})\left[[(x_{n+t} - [\sum_{i=1}^{n} F_i E^{2(n-i)} (\hat{x}_i + (n-i+t)T)]) / (\sum_{i=1}^{n} F_i E^{2(n-i)}))^2\right.\right.$$

$$\left.\left. \cdot (\sum_{i=1}^{n} F_i E^{2(n-i)}) + (1/\sum_{i=1}^{n} F_i E^{2(n-i)}) \sum_{i=1}^{n} \sum_{j<i} F_i F_j E^{2(2n-i-j)} (\hat{x}_i - (i-j)T - \hat{x}_j)^2]\right]) \right\}$$

Noting that the second term in the sum is constant with respect to x_{n+t}, and using the definition of the F_i's in 5).

$$=K\ \exp\{(-(1/2)((F'_{n+1}-E^2F_n)/(t\sigma^2F'_{n+1}))$$

$$\cdot(x_{n+t}-[\sum_{i=1}^{n}F_iE^{2(n-i)}(\hat{x}_i+(n-i+t)T)]/[\sum_{i=1}^{n}F_iE^{2(n-i)}])^2)\}$$

$$=K\ \exp\{(-(1/2)[1/(t\sigma^2F'_{n+1}/(F'_{n+1}-E^2F_n))]/(x_{n+t}-e_{n+t})^2\}.$$

Since the K is merely a constant which will be adjusted to make the distribution integrate to 1.

$$f(x_{n+t})\sim N(e_{n+t},[t\sigma^2F'_{n+1}/(F'_{n+1}-E^2F_n)])$$

Which completes the proof as soon as I show that the F_i's produced by 5) follow the recursion rule

$$F_1=1$$

$$F_2=E^2+\sigma^2$$

$$F_{k+1}=(2E^2+\sigma^2)F_k-E^4F_{k-1}$$

The proof involves fairly straightforward algebra.

$$F_{k+1}=\sigma^2\sum_{i=1}^{k}F_iE^{2(k-i)}+E^2F_k$$

$$=\sigma^2F_k+E^2\sigma^2\sum_{i=1}^{k-1}F_iE^{2(k-1-i)}+E^2F_k$$

$$=(\sigma^2+E^2)F_k+E^2(\sigma^2\sum_{i=1}^{k-1}F_iE^{2(k-1-i)})\ .$$

Applying the definition of the F_i's to the sum,

$$=(\sigma^2+E^2)F_k+E^2(F_k-E^2F_{k-1})$$

$$=(\sigma^2+2E^2)F_k-E^4F_{k-1}\ .$$

So, the F_i's fulfill the recursion rule, and thus, e_{n+t} is the best estimate.

ESTIMATING THE TREND

The best estimate of the trend is a weighted average of differences between adjacent points

$$T = \left[\sum_{i=1}^{n-1} W_i (\hat{x}_{i+1} - \hat{x}_i) \right] / \sum_{i=1}^{n-1} W_i .$$

The weights are somewhat complicated, but not overly difficult to compute.

1) $W_i = E^2 - (2E^4 F_i / F_{i+1}) - (E^{2(i+1)} / F_{i+1}) + E^{-2i} [F_{i+1} - E^2 F_i] G_i$

where the G_i are recursively calculated from n down, e.g.

2) $G_n = (E^{4n} + E^{2n+2} F_n) / (F_n (F_{n+1} - E^2 F_n))$

$G_i = G_{i+1} + [(E^{4i} + 2E^{2i+2} F_i) / (F_i F_{i+1})]$.

To prove that is the best estimate of the trend T, I will follow several steps. First, I will isolate the terms that involve T from the probability function for $\hat{x}_1, \hat{x}_x, \ldots, \hat{x}_n$. That will represent the function I must maximize. Maximizing it will involve minimizing a sum of squared differences between T and the differences between adjacent points $(\hat{x}_{i+1} - \hat{x}_i)$.

Before minimizing that function, I must show it is independent of the time (t) since the last observation. Then, I will convert it from functions of T and differences between faraway points $\hat{x}_i - \hat{x}_j$ into differences between T and differences between adjacent points $\hat{x}_{i+1} - \hat{x}_i$. That will produce a complicated set of weights for each difference $\hat{x}_{i+1} - \hat{x}_i$. Next, I will simplify those weights to show they are the weights in equations 1) and 2).

The Function to Minimize - The New Distribution of Observed Points

The previous appendix showed that the distribution of the potential observed points $\hat{x}_1, \ldots, \hat{x}_n$ given a future value x_{n+c} was proportional to a term involving x_{n+c} and a constant, e.g.

$$f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n | x_{n+c}, T, \sigma^2, E^2) = K_1 \exp(-K_2(e_{n+c}-x_{n+c})^2+K_3),$$

$(K_1, K_2, K_3$ constant w.r.t. $x_{n+c})$

That made e_{n+c} the best estimator of x_{n+c}. I would like to isolate T the way I isolated x_{n+c} to produce a formula

$$f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n | x_{n+c}, T, \sigma^2, E^2) = K_1 \exp(-K_2(T'-T)^2-K_3(e_{n+c}-x_{n+c})^2+K_4)$$

$(K_1, K_2, K_3, K_4$ constant w.r.t. both x_{n+c} and T).

Then, the expression T′ will represent the best (maximum likelihood) estimator of T.

The first step is to combine the terms involving T[9], e.g. to find $f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n | x_{n+c}, T, \sigma^2, E^2) = K_1 \exp(-g(T)-K_3(e_{n+c}-x_{n+c})^2+K_4)$

Thankfully, finding g(T) is fairly easy. Simple inspection of the multiplier of $(e_{n+c}-x_{n+c})^2$ shows it is independent of T as well as x_{n+c}. The function g(T) then simply represents the terms 'cast off' as constant when integrating over the x_i's plus the x_i terms cast off when the weighted squared differences between many individual terms and x_{n+c} were combined (at the end of appendix III).

First, let me discuss out the terms cast off when integrating over the x_l's. The terms cast off into the constant when evaluating the multiple integral over x_l were

$$\exp\Big\{-(1/2)\,[(t_l\sigma^2/F'_{l+1})\sum_{j=1}^{l-1}F_jE^{2(l-1-j)}(\hat{x}_j+(l-j)T-\hat{x}_l)^2$$

$$+(t_lE^2\sigma^2/(F_lF'_{l+1}))\sum_{j=1}^{l-1}\sum_{k\le j}F_jF_kE^{2(2l-2-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2]\Big\}$$

($t_l=1$, except for $t_n=t$, and $F'_{l+1}=F_{l+1}$, except for F'_{n+1} which is $t\sigma^2\sum F_lE^{2(n-l)}+E^2F_n$.)

Which, after moving some E^2 terms outside the sums,

$$3)\quad =\exp\Big\{-(1/2)\,[(t_l\sigma^2/(E^2F'_{l+1}))\sum_{j=1}^{l-1}F_jE^{2(l-j)}(\hat{x}_j+(l-j)T-\hat{x}_l)^2$$

$$+(t_l\sigma^2/(E^2F_lF'_{l+1}))\sum_{j=1}^{l-1}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2]\Big\}$$

For simplicity, let me call the first term A_l and the second B_l to get

$$4)\quad =\exp\{-(1/2)\,[A_l+B_l]\}$$

But there is another T term to add. When the final individual terms $(\hat{x}_i+(t+n-i)T-x_{n+t})^2$ were combined by the weighted sum of squares theorem in appendix III (to get $(e_{n+t}-x_{n+t})^2$), the following terms were 'cast off'.

$$5)\quad \exp\{-(1/2)\,[(1/F'_{n+1})(1/\sum_{i=1}^{n}F_iE^{2(n-i)})\cdot\sum_{l=1}^{n}\sum_{j\le l}F_lF_jE^{2(2n-l-j)}(\hat{x}_j+(l-j)T-\hat{x}_l)^2]\}$$

$$=\exp(-(C_n/2))$$

Combining all the terms involving T, I get

$$6)\quad g(T) = -(1/2)\,[C_n + \sum_{I=1}^{n} A_I + B_I]$$

Looking back at the pieces of g(T), it is much more difficult to work with than it needs to be. First, it uses t_I and F_I'-two clumsy expressions. But, as we will see later, the sum g(T) is actually independent of t.

Before proving that, I need to prove several lemmas. One will be used to prove the independence from t. The others will be used later to simplify g(T).

Lemmas

Before showing g(T) is independent of t, I need to make a brief digression. I will need several lemmas to complete the analysis. Since I need one of them to prove g(T) is independent of t, I should prove them before discussing g(T) further.

Interchange of Sum Indices Lemma.

$$7)\quad \sum_{a=1}^{n} \sum_{b=1}^{a-1} h(a,b) = \sum_{b=1}^{n-1} \sum_{a=b+1}^{n} h(a,b)$$

Proof: the indices on either side describe the case where $b<a\le n$.

An alternate version, where $b\le a\le n$, is

$$8)\quad \sum_{a=1}^{n} \sum_{b=1}^{a} h(a,b) = \sum_{b=1}^{n} \sum_{a=b}^{n} h(a,b)$$

Sum of the F_I's

$$9)\quad \sum_{a=1}^{n} E^{2(n-a)} F_a = (F_{n+1} - E^2 F_n)/\sigma^2$$

Proof: Using the summation definition of the F_i's from appendix III

$$F_{n+1}=\sigma^2\left[\sum_{a=1}^{n} E^{2(n-a)}F_a\right]+E^2F_n.$$

Simple algebra produces the result.

Partial Sum of the F_i's Lemma.

$$10)\quad \sum_{a=b}^{n} E^{2(n-a)}F_a=\frac{[F_{n+1}-E^2F_n-E^{2(n-(b-1))}F_b+E^{2(n-(b-2))}F_{b-1}]}{\sigma^2}$$

Proof:

$$\sum_{a=b}^{n} E^{2(n-a)}F_a=\sum_{a=1}^{n} E^{2(n-a)}F_a-\sum_{a=1}^{b-1} E^{2(n-a)}F_a = \sum_{a=1}^{n} E^{2(n-a)}F_a-E^{2(n-(b-1))}\sum_{a=1}^{b-1} E^{2(b-1-a)}F_a$$

Using equation 9) twice produces the result.

Sum of the iF_i's Lemma.

$$11)\quad \sum_{a=1}^{n} aE^{2(n-a)}F_a=\frac{nF_{n+1}-(n-1)E^2F_n+E^{2n}}{\sigma^2}$$

Proof: Noting that $a=\sum_{b=1}^{a}$

$$\sum_{a=1}^{n} aE^{2(n-a)}F_a=\sum_{a=1}^{n}\ \sum_{b=1}^{a} E^{2(n-a)}F_a$$

Using the interchange of sum indices lemma 8)

$$=\sum_{b=1}^{n}\ \sum_{a=b}^{n} E^{2(n-a)}F_a$$

Using the formula for the partial sum (equation 10))

$$=\sum_{b=1}^{n}\frac{F_{n+1}-E^2F_n-E^{2(n-(b-1))}F_b+E^{2(n-(b-2))}F_{b-1}}{\sigma^2}$$

Distributing the sum across the addition and pulling terms constant relative to b outside the sum.

$$=\frac{nF_{n+1}-nE^2F_n-E^2\left[\sum_{b=1}^{n}E^{2(n-b)}F_b\right]+E^2\left[\sum_{b=0}^{n-1}E^{2(n-b)}F_b\right]}{\sigma^2}$$

Removing one term from the first sum

$$=\frac{nF_{n+1}-nE^2F_n-E^2F_n-E^2\left[\sum_{b=1}^{n-1}E^{2(n-b)}F_b\right]+E^2\left[\sum_{b=0}^{n-1}E^{2(n-b)}F_b\right]}{\sigma^2}$$

Now, the problem summing from $b=0$ to $n-1$ is that F_0 is undefined. Since it occurs where $b=1, F_1-E^2F_0=0$, it appears $F_0=1/E^2$ (Note that then $F_2=E^2+\sigma^2=(\sigma^2+2E^2)F_1-E^4F_0$). And the equation is

$$=\frac{nF_{n+1}-nE^2F_n-E^2F_n-E^2\sum_{b=1}^{n-1}E^{2(n-b)}F_b+E^2\sum_{b=1}^{n-1}E^{2(n-b)}F_b+E^{2n}}{\sigma^2}$$

$$=\frac{nF_{n+1}-(n-1)E^2F_n+E^{2n}}{\sigma^2}$$

Partial Sum of the F_i's Lemma

$$12)\quad \sum_{a=b}^{n}aE^{2(n-a)}F_a=(1/\sigma^2)\{nF_{n+1}-(n-1)E^2F_n-(b-1)E^{2(n-(b-1))}F_b+(b-2)E^{2(n-(b-2))}F_{b-1}\}$$

Proof: same basic argument as equation 10).

Telescoping Sum Lemma

$$13)\quad (\hat{x}_k+(j-k)T-\hat{x}_j)^2=(j-k)\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2-\sum_{i=k}^{j-1}\sum_{m\leq i}(\hat{x}_{i+1}-\hat{x}_i-(\hat{x}_{m+1}-\hat{x}_m))^2$$

Proof: set

$$(\hat{x}_k+(j-k)T-\hat{x}_j)^2=(\sum_{i=k}^{j-i}(\hat{x}_{i+1}-\hat{x}_i-T))^2=(j-k)^2(T-(1/(j-k))\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i))^2$$

and then use the weighted sum of squares theorem from appendix II.

g(T) is Independent of t

Now that those lemmas are proven, I must show the 't' in g(T) may be replaced with '1'.

Since the trend is something reflected in the observed points $\hat{x}_1, \ldots, \hat{x}_n$, rather than something instrinsic to the length of the projection period (t), it seems that estimated trend (T') should be independent of t. That will follow from the independence of g(T) from t.

To prove g(T) is independent of t, all that is necessary is to show that the few terms in g(T) that contain a t are actually constant with respect to t. Reviewing equations, 3), 5), and 6), those are $C_n+A_n+B_n$. E.g.

$$g(T) = -(1/2)(K+C_n+A_n+B_n),$$

where K is the terms that are obviously constant with respect to t.

First, rewrite C_n by replacing l and j with j and k to get

$$14)\quad A_n+B_n+C_n=(t\sigma^2/(E^2F'_{n+1}))\sum_{j=1}^{n-1} F_jE^{2(n-j)}(\hat{x}_j+(n-j)T-\hat{x}_n)^2$$

$$+(t\sigma^2/(E^2F_nF'_{n+1}))\sum_{j=1}^{n-1}\sum_{k\le j} F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2$$

$$+(1/F'_{n+1})(1/\sum_{i=1}^{n} F_iE^{2(n-i)})\cdot\sum_{j=1}^{n}\sum_{k\le j} F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2.$$

Then, the strategy is to convert the expression above into an expression in t times a double sum constant relative to t. Then I will show the expression in t is actually constant relative to t. The first step is to note that the first term is the case where $j=n$ for the second term (with j playing the role of k).

$$= (t\sigma^2/(E^2F_nF'_{n+1}))\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2$$

$$+(1/(F'_{n+1})(1/\sum_{j=1}^{n}F_iE^{2(n-i)})\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2$$

Now the double sums in each term are identical and independent of t. So, we may set

$$=[(t\sigma^2/(E^2F_nF'_{n+1}))+(1/F'_{n+1})(1/\sum_{i=1}^{n}F_iE^{2(n-i)})]\cdot K$$

Now, all that remains is to show that is independent of t. Using the 'Sum of the F_i's Lemma' 9) (and correcting for the difference betwen the definition of F'_{n+1} and F_{n+1})

$$=[(t\sigma^2/(E^2F_nF'_{n+1})+(1/F'_{n+1})(t\sigma^2/(F'_{n+1}-E^2F_n))]\cdot K$$

$$= K\cdot\{(t\sigma^2/F'_{n+1})/[(1/(E^2F_n))+(1/(F'_{n+1}-E^2F_n))]\}$$

Performing more algebra

$$=K\cdot\{(t\sigma^2/F'_{n+1})[F'_{n+1}/(E^2F_n(F'_{n+1}-E^2F_n))]\}$$

$$=K\cdot\{t\sigma^2E^2F_n/(F'_{n+1}-E^2F_n)\}$$

$$=K\,E^2F_n/\sum_{i=1}^{n}E^{2(n-i)}F_i$$

Which is independent of t. So, in equation 3), 5) and 6) we may treat the t's as 1's and the F'_{n+1}'s as F_{n+1}.

The next step is to convert the expression involving the differences between faraway (j and k) terms to differences of adjacent terms (i and $i+1$).

g(T) as Differences Between Adjacent Points

g(T) can be converted to the following expression involving differences between adjacent terms.

15) $$g(T)=[C_n+\sum_{l=1}^{n} A_l+B_l]/2=-(\sigma^2/2E^2)(U(T)+V(T))+K;$$

where K is constant with respect to T; and

16) $$U(T)=\sum_{l=2}^{n-1}(1/(F_lF_{l+2}))\sum_{j=2}^{l}\sum_{k<j}E^{2(2l-j-k)}F_jF_k\cdot(j-k)\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2;$$

and,

17) $$V(T)=(1/(F_n(F_{n+1}-E^2F_n)))\sum_{j=2}^{n}\sum_{k<j}E^{2(2l-j-k)}F_jF_k\cdot(j-k)\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2.$$

(Notice that U and V are identical except for the terms to the left of the double sum. If the $F_{n+1}-E^2F_n$ in V were simply F_{n+1}, V could be combined into the sum over the l's in U).

To prove that, I must state equations 3), 5) and 6) without t; perform some algebra to simplify the sums; then use the Telescoping Sum Lemma 13).

First, let me point out that when 't' is replaced by '1',

$$g(T)=-[C_n+\sum_{l=1}^{n}A_l+B_l]/2$$

$$18)\quad =-(1/2)\Big[\{\sum_{l=1}^{n}(\sigma^2/(E^2F_{l+1}))\sum_{j=1}^{l-1}F_jE^{2(l-j)}(\hat{x}_j+(l-j)T-\hat{x}_l)^2\}$$

$$+\{\sum_{l=1}^{n}(\sigma^2/(E^2F_lF_{l+1}))\sum_{j=1}^{l-1}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}$$

$$+\{(1/F_{n+1})(1/\sum_{l=1}^{n}F_lE^{2(n-l)})\sum_{l=1}^{n}\sum_{j\le l}F_lF_jE^{2(2n-l-j)}(\hat{x}_j+(l-j)T-\hat{x}_l)^2\}\Big].$$

That unwieldly expression can be simplified considerably. The first step is to note that in the first term the sum over j and the expression to the right form the case where j=l in the second term, so

$$=-(1/2)\Big[\{\sum_{l=1}^{n}(\sigma^2/(E^2F_lF_{l+1}))\sum_{j=1}^{l}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}$$

$$+\{(1/F_{n+1})(1/\sum_{l=1}^{n}F_lE^{2(n-l)})\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}\Big].$$

Then, the sum in the second multiplier in the second term can receive the benefit of the 'sum of the F_l's lemma 9).

$$=-(1/2)\Big[\{\sum_{l=1}^{n}(\sigma^2/(E^2F_lF_{l+1}))\sum_{j=1}^{l}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}$$

$$+\{(\sigma^2/(F_{n+1}(F_{n+1}-E^2F_n)))\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}\Big].$$

Then, the second term may be combined with the case where l=n in the first term to get

$$=-(1/2)\Big[\{\sum_{l=1}^{n-1}(\sigma^2/(E^2F_lF_{l+1}))\sum_{j=1}^{l}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}$$

$$+\{[(\sigma^2/(E^2F_nF_{n+1}))+(\sigma^2/(F_{n+1}(F_{n+1}-E^2F_n)))]\cdot\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}\Big]$$

Using some algebra to simplify the multiplier in the second term

19) $$=-(1/2)\Big[\{\sum_{l=1}^{n-1}(\sigma^2/(E^2F_lF_{l+1}))\sum_{j=1}^{l}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}$$

$$+\{[(\sigma^2/(E^2F_n(F_{n+1}-E^2F_n)))\cdot\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}(\hat{x}_k+(j-k)T-\hat{x}_j)^2\}\Big].$$

Then, all that remains is to use the telescoping sum lemma and cast off the $\hat{x}_{i+1}-\hat{x}_i-(\hat{x}-\hat{x}_m)$ terms (since they are constant with respect to t.

$$=-(\sigma^2/2E^2)\{\sum_{l=1}^{n-1}(1/(F_lF_{l+1}))\sum_{j=1}^{l}\sum_{k\le j}F_jF_kE^{2(2l-j-k)}\cdot(j-k)\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\}$$

$$-(\sigma^2/2E^2)\{(1/(F_nF_{n+1}))\sum_{j=1}^{n}\sum_{k\le j}F_jF_kE^{2(2n-j-k)}\cdot(j-k)\sum_{i=k}^{j-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\}$$

Noting that $j-k=0$ when $j=k$; $\sum_{j=1}^{n}\sum_{k\le j}K(j-k)=\sum_{j=2}^{n}\sum_{k<j}K(j-k)$, so

$$g(T)=-(\sigma^2/(2E^2))(U(T)+V(T))+K$$

So, g(T) may be described as weighted squared differences between T and the differences between adjacent points.

U(T) and V(T) as Sums Over Differences Between Adjacent Points

The next step is to simplify 16) by repeatedly using the 'interchange of sum indices' lemma., e.g.

$$20)\quad U(T)=\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\sum_{l=i+1}^{n-1}\ \sum_{j=i+1}^{l}\ \sum_{k=1}^{i}(j-k)E^{2(2l-j-k)}F_jF_k/(F_lF_{l+1})\ ;\ \text{and}$$

$$21)\quad V(T)=\sum_{i=1}^{n-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\sum_{j=i+1}^{n}\ \sum_{k=1}^{i}(j-k)E^{2(2n-j-k)}F_jF_k/(F_n(F_{n+1}-E^2F_n))\ .$$

The proof of each involves repeated and straightforward use of the two interchange of sums lemmas.

Summing the Weights Over j and k

To make the expressions for U(T) and V(T) more tractable, the last two sums should be simplified. Their sum is

$$22)\quad \sum_{j=i+1}^{l}\ \sum_{k=1}^{i}(j-k)E^{2(2l-j-k)}F_jF_k$$

$$=(E^{2(l-i)}/\sigma^4)\{F_{i+1}[(l-i)F_{l+1}-(l-i-1)E^2F_l+E^{2l}]$$

$$-E^2F_i[(l-i+1)F_{l+1}-(l-i)E^2F_l+E^{2l}]-E^{2i}[F_{l+1}-E^2F_l]\}$$

The proof involves using the lemmas proved earlier for the sum of the F_l's (equations 9) and 10)) and the sum of the iF_l's lemmas (equations 11) and 12)). The first step is to split the j-k term and pull the constants across the 'k' sum.

$$\sum_{j=i+1}^{l}\ \sum_{k=1}^{i}(j-k)E^{2(2l-j-k)}F_jF_k=\sum_{j=i+1}^{l}E^{2(2l-i-j)}F_j\{[j\sum_{k=1}^{i}E^{2(i-k)}F_k]-\sum_{k=1}^{i}kE^{2(i-k)}F_k\}$$

Using equations 9) and 11) on the two sums,

$$= \sum_{j=i+1}^{l} E^{2(2l-j-i)} F_j \{ [j(F_{l+1}-E^2F_l)/\sigma^2 - (iF_{l+1}-(i-1)E^2F_l+E^{2i})/\sigma^2 \}$$

Pulling out the terms that are constant with respect to j,

$$= (1/\sigma^2) E^{2(l-i)} \{ (F_{l+1}-E^2F_l) [\sum_{j=i+1}^{l} jE^{2(l-j)}F_j]$$

$$- (iF_{l+1}-(i-1)E^2F_l+E^{2i}) [\sum_{j=i+1}^{l} E^{2(l-j)}F_j] \}$$

Summing the 'j' sums using equations 10) and 12)

$$= (E^{2(l-i)}/\sigma^2 \Big\{ \frac{(F_{l+1}-E^2F_l)[lF_{l+1}-(l-1)E^2F_l-iE^{2(l-i)}F_{l+1}+(i-1)E^{2(l-(i-1))}F_i]}{\sigma^2}$$

$$- \frac{(iF_{l+1}-(i-1)E^2F_i+E^{2i})\cdot(F_{l+1}-E^2F_l-E^{2(l-i)}F_{l+1}+E^{2(l-(i-1))}F_i)}{\sigma^2} \Big\}$$

Multiplying those polynomials in the F's and collecting and cancelling terms produces

$$= \frac{E^{2(l-i)}}{\sigma^4} \{ F_{i+1}[(l-i)F_{l+1}-(l-i-1)E^2F_l+E^{2l}]$$

$$-E^2F_i[(l-i+1)F_{l+1}-(l-i)E^2F_l+E^{2l}]-E^{2i}[F_{l+1}-E^2F_l] \}.$$

Which is exactly equation 22).

Summing U(T) Over l

The sum over "l" in U(T) may be computed to produce

$$23)\quad U(T)=\left(\frac{1}{\sigma^4}\right)\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\cdot\Big\{E^2-(n-i)E^{2(n-i)}F_{i+1}/F_n-2E^4F_i/F_{i+1}$$

$$\cdot\Big\{E^2-(n-i)E^{2(n-i)}F_{i+1}/F_n-2E^4F_i/F_{i+1}+(n-i+1)E^{2(n-i+1)}F_i/F_n-E^{2(i+1)}/F_{i+1}+E^{2n}/F_n\Big\}$$

$$+E^{-2i}(F_{i+1}-E^2F_i)\Big[\sum_{l=i+1}^{n-1}(E^{4l}+2E^{2(l+1)}F_l)/(F_lF_{l+1})\Big]\Big\}$$

Before I show that, let me note that U has become too long to be tractable. So, let me break it up into three terms. Using equations 20) and 22)

$$U(T)=\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\cdot\sum_{l=i+1}^{n-1}(E^{2(l-i)}/\sigma^4)\Big\{F_{i+1}[(l-i)F_{l+1}-(l-i-1)E^2F_l+E^{2l}]$$

$$-E^2F_i[(l-i+1)F_{l+1}-(l-i)E^2F_l+E^{2l}]-E^{2i}[F_{l+1}-E^2F_l]\Big\}/(F_lF_{l+1})\,.$$

Pulling out the constant terms and collecting coefficients produces

$$24)\quad U(T)=(1/\sigma^4)\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2[A_i-B_i-C_i]\,,$$

where

$$25)\quad A_i=E^{-2i}F_{i+1}\sum_{l=i+1}^{n-1}[((l-i)E^{2l}/F_l)-((l-i-1)E^{2(l+1)}/F_{l+1})+(E^{4l}/(F_lF_{l+1}))]\,,$$

$$26)\quad B_i=E^{-2(i-1)}F_i\sum_{l=i+1}^{n-1}[((l-i+1)E^{2l}/F_l)-((l-i)E^{2(l+1)}/F_{l+1})+(E^{4l}/(F_lF_{l+1}))]\,,$$

$$27)\quad C_i=\sum_{l=i+1}^{n-1}[(E^{2l}/F_l)-(E^{2(l+1)}/F_{l+1})]\,,$$

Next, I must simplify each expression. Note that the second term within the sum of A is nearly the first term evaluated at a higher index. E.g.

$$A_i=E^{-2i}\Big\{[\sum_{l=i+1}^{n-1}(l-i)E^{2l}/F_l]$$

$$-\sum_{l=i+1}^{n-1}(l-i-1)E^{2(l+1)}/F_{l+1}]+[\sum_{l=i+1}^{n-1}E^{4l}/(F_lF_{l+1})]\Big\}$$

$$=E^{-2i}\Big\{[\sum_{l=i+1}^{n-1}(l-i)E^{2l}/F_l]-\sum_{l=i+2}^{n}(l-i)E^{2l}/F_l]$$

$$+[\sum_{l=i+1}^{n-1}2E^{2(l+1)}/(F_{l+1}]+[\sum_{l=i+1}^{n-1}E^{4l}/F_lF_{l+1})]\Big\}.$$

Then, the second and third term telescope to produce

$$=E^{-2i}F_{i+1}\Big\{(E^{2(i+1)}/F_{i+1})-(n-i)E^{2n}/F_n$$

$$+[\sum_{l=i+1}^{n-1}2E^{2(l+1)}/F_{l+l}]+[\sum_{l=i+1}^{n-1}E^{4l}/(F_lF_{l+1})]\Big\}.$$

Then, combining the last two terms, and distributing the multiplier

$$28)\quad A_i=E^2-(n-i)E^{2(n-i)}F_{i+1}/F_n$$

$$+E^{-2i}F_{i+1}\cdot[\sum_{l=i+1}^{n-1}(E^4+2E^{2(l+1)}F_l)/(F_lF_{l+1})]\ .$$

Simplifying B_i in a similar fashion produces

$$29)\quad B_i = \frac{2E^4 F_i}{F_{i+1}} - \frac{(n-i+1)\,E^{2(n-i+1)}F_i}{F_n}$$

$$+E^{-2(i-1)}F_i\,[\sum_{l=i+1}^{n-1} \frac{(E^{4l}+2E^{2(l+1)}F_l)}{(F_l F_{l+1})}]$$

Simplifying C is simpler. The sums telescope to produce

$$30)\quad C_i = [\sum_{l=i+1}^{n-1} \frac{E^{2l}}{F_l}] - [\sum_{l=i+1}^{n-1} \frac{E^{2(l+1)}}{F_{l+1}}]$$

$$= \frac{E^{2(i+1)}}{F_{i+1}} - \frac{E^{2n}}{F_n}$$

Then, combining equations 28) for A_i, 29) for B_i, and 30) for C_i into equation 24)

$$U(T) = (1/\sigma^4)\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\Big\{E^2 - \frac{(n-i)\,E^{2(n-i)}F_{i+1}}{F_n} - \frac{2E^4F_i}{F_{i+1}} + \frac{(n-i-1)\,E^{2(n-i-1)}F_i}{F_n}$$

$$-\frac{E^{2(i+1)}}{F_{i+1}} + \frac{E^{2n}}{F_n} + E^{-2i}[F_{i+1}-E^2F_i]\,[\sum_{l+i+1}^{n-1}\frac{(E^{4l}+2E^{2(l+1)}F_l)}{(F_lF_{l+1})}]\Big\}.$$

Which is exactly equation 23).

Summing V(T)

V(T) may also be summed to produce

$$31)\quad V(T)=(1/\sigma^4)\sum_{i=1}^{n-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\Big\{(n-i)E^{2(n-i)}F_{i+1}/F_n$$

$$+\frac{E^{2(n-i-1)}F_{i+1}}{(F_{n+1}-E^2F_n)}+\frac{E^{2(2n-i)}(F_{i+1}-E^2F_i)}{(F_n(F_{n+1}-E^2F_n))}$$

$$-\ (n-i-1)E^{2(n-i+1)}\left(\frac{F_i}{F_n}\right)-\frac{E^{2(n-i+2)}F_i}{(F_{n+1}-E^2F_n)}-\frac{E^{2n}}{F_n}\Big\}$$

The proof requires using the equation for the sum over j and k (22) on equation 21). Then, simple algebra produces the result.

Combining U(T) and V(T)

Now that the sums in U(T) and V(T) have been simplified, the next step is to combine them to produce the complete weights

$$32)\quad g(T)=-\left(\frac{\sigma^2}{(2E^2)}\right)(U(T)+V(T))+K$$

$$=K-\left(\frac{1}{(2E^2\sigma^2)}\right)\Big\{\sum_{i=1}^{n}(\hat{x}_{i+1}-\hat{x}_i-T)^2$$

$$\cdot[E^2-\left(\frac{2E^4F^i}{F_{i+1}}\right)-(E^{2(i+1)}/F_{i+1})+E^{-2i}(F_{i+1}-E^2F_i)G_i]\Big\}$$

To prove it, we need to combine equation 23) for U(T) and equation 31) for V(T) and simplify the result. Combining the two equations produces

$$g(T)=K-(1/(2E^2\sigma^2))\Big\{\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\Big[E^2-(n-i)E^{2(n-i)}F_{i+1}/F_n$$

$$-2E^4F_i/F_{i+1}+(n-i+1)E^{2(n-i+1)}F_i/F_n$$

$$-E^{2(i+1)}/F_{i+1}+E^{2n}/F_n+E^{-2i}(F_{i+1}-E^2F_i)\Big[\sum_{l+i+1}^{n-1}(E^{4l}+2E^{2(l+1)}F_l)/(F_lF_{l+1})\Big]\Big]\Big\}$$

$$-(1/(2E^2\sigma^2))\Big\{\sum_{i=1}^{n-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\Big[(n-i)E^{2(n-i)}F_{i+1}/F_n+E^{2(n-i+1)}F_{i+1}/(F_{n+1}-E^2F_n)$$

$$+E^{2(2n-i)}(F_{i+1}-E^2F_i)/(F_n(F_{n+1}-E^2F_n))$$

$$-(n-i+1)E^{2(n-i+1)}F_i/F_n-E^{2(n-i+2)}F_i/(F_{n+1}-E^2F_n)-E^{2n}/F_n\Big]\Big\}$$

That is an incredibly long expression. But thankfully, many of the U and V terms cancel or combine (at least for i between 1 and n-2) to produce

$$g(T)=K-(1/(2E^2\sigma^2))\Big\{\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\Big[E^2-2E^4F_i/F_{i+1}-E^{2(i+1)}/F_{i+1}$$

$$+E^{-2i}(F_{i+1}-E^2F_i)\cdot\Big[(E^{4n}+E^{2(n+1)}F_n)/(F_n(F_{n+1}-E^2F_n))$$

$$+\sum_{l=i+1}^{n-1}(E^{4l}+2E^{2(l+1)}F_l)/(F_lF_{l+1})\Big]\Big]\Big\}$$

$$-(1/(2E^2\sigma^2))(\hat{x}_n-\hat{x}_{n-1}-T)^2\Big\{E^2+E^4F_n/(F_{n+1}-E^2F_n)+E^{2(n+1)}(F_n-E^2F_{n-1})/(F_n(F_{n+1}-E^2F_n$$

$$-2E^4F_{n-1}/F_n-E^6F_{n-1}/(F_{n+1}-E^2F_n)-E^{2n}/F_n\Big\}$$

Then, noting that the definition of the G_i from equation 2), and combining some of the terms in the second product

$$g(T)=K-(1/(2E^2\sigma^2))\Big\{\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\big[E^2-2E^4F_i/F_{i+1}-E^{2(i+1)}/F_{i+1}$$

$$+E^{-2i}(F_{i+1}-E^2F_i)G_{i+1}-(1/2E^2\sigma^2))(\hat{x}_n-\hat{x}_{n-1}-T)^2\cdot\big\{E^2+E^4(F_n-E^2F_{n-1})/(F_{n+1}-E^2F_n)$$

$$-E^{2(n+1)}(F_n-E^2F_{n-1})/(F_n(F_{n+1}-E^2F_n))-2E^4F_{n-1}/F_n-E^{2n}/F_n)$$

Then, combining some of the terms applied to $(\hat{x}_n-\hat{x}_{n-1}-T)^2$

$$g(T)=K-(1/2E^2\sigma^2))\Big\{\sum_{i=1}^{n-2}(\hat{x}_{i+1}-\hat{x}_i-T)^2\big[E^2-2E^4F_i/F_{i+1}-E^{2(i+1)}/F_{i+1}+E^{-2i}(F_{i+1}-E^2F_i)G_{i+1}\big]\Big\}$$

$$-(1/2E^2\sigma^2)(\hat{x}_n-\hat{x}_{n-1}-T)^2\cdot\big\{E^2-2E^4F_{n-1}/F_n-E^{2n}/F_n+E^{-2(n-1)}(F_n-E^2F_{n-1})G_n\big\}$$

Which yields the result in 32).

$$g(T)=K-(1/(2E^2\sigma^2))\sum_{i=1}^{n-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2\big[E^2-2E^4F_i/F_{i+1}-E^{i+1}/F_{i+1}+E^{-2i}(F_{i+1}-E^2F_i)G_i\big]$$

Which could be restated as

$$33)\quad g(T)=K-(1/(2E^2\sigma^2))\sum_{i=1}^{n-1}(\hat{x}_{i+1}-\hat{x}_i-T)^2W_i,$$

Where the W_i are the weights from 1) that should be the weights used to average the $(\hat{x}_{i+1}-\hat{x}_i)$'s to produce T'.

The Final Formula

Producing the final estimate is now a fairly straightforward process of using the weighted sum of squares theorem from appendix II to restate g(T), and then showing T' minimizes it.

Applying the weighted sum of squares theorem to equation 33) produces

$$g(T)=K-(1/(2E^2\sigma^2))\,(\sum_{i=1}^{n-1} W_i)\cdot(T-[(\sum_{i=1}^{n-1} W\text{-}i(\hat{x}_{i+1}-\hat{x}))/\sum_{i=1}^{n-1} W_i])^2$$

+ other terms that do not involve T.

Combining the first and last terms into the constant

$$g(T)=K_1+K_2(T-[(\sum W_i(\hat{x}_{i+1}-\hat{x}_i))/\sum W_i])^2$$

Which is clearly maximized by setting

$$34)\quad T=T'=[\sum_{i=1}^{n-1} W_i(\hat{x}_{i+1}-\hat{x}_i)]/\sum_{i=1}^{n-1} W_i$$

So, T' is the best estimator

REFERENCES/FOOTNOTES

[1] This model assumes that the average size of the changes is proportional solely to the current price level, and is independent of historical price increases. One could make an argument that sellers try to follow the current trend of the price level. That of course would lead to inflation creating more inflation. The author believes the model in this paper is a good first step, and that reflecting the current trend of the price level is a logical enhancement.

[2] Robert V. Hogg, Stuart A. Klugman, *Loss Distributions* (New York, New York: John Wiley & Sons, 1984), pp. 21-22.

[3] Casualty Actuarial Society, *Exposure Draft, Casualty Contingencies, Chapter Fourteen, Reinsurance, Retentions and Surplus*, (Atlanta, Georgia: Educational Foundation, Inc., 1975) pp. 24-25.

[4] In practice, each $\epsilon(t)$ may have a different variance, corresponding to the number of customers insured. Realistically, they are usually close to constant. Further, the model may be refined for the case where the E's vary over time.

[5] Per the definition of a marginal distribution. See, for instance pp. 66 in *Introduction to Mathematical Statistics* by Robert V. Hogg and Allen T. Craig, Fourth Edition, (New York, New York: Macmillan, 1978)

[6] This follows from the fact that when you have n samples from a $N(\mu,\sigma^2)$ distribution their mean is distributed $N(\mu,\sigma^2/n)$.

[7] Bradley Efron, 'Controversies in the Foundation of Statistics', originally printed in the *American Mathematical Monthly*, Volume 85, Number 4, April 1978, pp. 231-238. Reprinted in the *Casualty Actuarial Society Forum*, Fall 1991, pp. 259-275. See pp. 266 in the latter.

[8] Further, one could create non-uniform diffuse prior distributions by setting $\int Ih(x)\,dx=1$ when $\int h(x)\,dx=\infty$. Then $x\sim Ih(x)$ has a non-uniform diffuse prior distribution.

[9] A careful reader might point out that the function e_{n+t} is dependent on T, so I cannot isolate all the T-terms into G(T). But that is irrelevant. If I express the probability function as $K_1\exp(-K_2(T^*-T)^2-K_3(e_{n+t}-x_{n+t})^2+K_4)$ and then set $T=T^*$ and $x_{n+t}=e_{n+t}(T)$; the resulting $K_1\exp(K_4)$ is still the largest attainable probability for $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$.