Considerations in Estimating Loss Cost Trends

Kurt S. Dickmann, FCAS, MAAA and
James R. Merz, FCAS, MAAA

Abstract

The application of loss trends has long been a fundamental part of the ratemaking process. Despite this, the actuarial literature is somewhat lacking in the description of methods by which one can estimate the proper loss trend from empirical data. Linear or exponential least squares regression is widely used in this regard. However, there are problems with the use of least squares regression when applied to insurance loss data.

In this paper, some common pitfalls of least squares regression, as it is commonly applied to insured loss data, and two alternative methods of evaluating loss trends will be illustrated. Both methods are based on simple least squares regression, but include modifications designed to account for the characteristics of insurance loss data.

The results of various methods are compared using industry loss data. Stochastic simulation is also used as a means of evaluating various trend estimation methods.

The concepts presented are not new. They are presented here in the context of analyzing insured loss data to provide actuaries with additional tools for estimating loss trends.

Introduction

This paper is organized into eight sections. The first section will describe the importance of estimating loss cost trends in Property/Casualty ratemaking. In addition, it will introduce the common industry practices used to estimate the underlying loss cost inflation rate.

The second section will provide a review of basic regression analysis since regression is commonly utilized for estimating loss trends. It will also describe other relevant statistical formulae.

The third section will describe some characteristics of insured loss data. This section will describe how insured losses violate some of the basic assumptions of the ordinary least squares model. It will also describe the complications that result because of these violations.

The fourth section will describe several methods that can be utilized along with informed judgement to identify outliers.

The fifth and sixth sections will describe two alternative methods that address the shortcomings of ordinary least squares regression on insured loss data.

The seventh section applies the common method of exponential least squares regression and the two alternative methods to industry loss data and compares the results.

In the last section, the performance of exponential least squares regression and the alternative methods will be evaluated using stochastic simulation of loss data with a known underlying trend.

While the determination and use of credibility is an essential component of loss trend determination, it is beyond the scope of this paper. However, the concepts and methods presented here apply equally to the determination of the trend assigned to the complement of credibility. The methods presented here are designed to extract as much information about the underlying trend from the available data. They are not intended to minimize the importance or use of credibility.

In addition to credibility, there are many other considerations that must be taken into account when applying loss trends, such as the effect of limits and deductibles. These issues are beyond the scope of this paper.

Section 1: Actuarial Literature and Industry Practice

In the ratemaking process, it is widely agreed that trend selection is the component that requires the most judgement.[1] According to the *Actuarial Standards of Practice,* the application of the appropriate trending procedures is essential to estimating future costs in the determination of rates.[2]

Despite the importance of trending in ratemaking and the degree of judgment required, there is little written specifically regarding the determination of loss trends. Most ratemaking papers cite trending as an integral part of the process and describe the author's selected approach. This is entirely appropriate as the subject of these papers is ratemaking and not specifically trend estimation.

The actuarial literature is sparse on the process of selecting the type of data to evaluate, preparing trend data, choosing the most appropriate model and assessing the appropriateness of the selected trends.

There are papers addressing several of the important basic issues of trending. These include the appropriate trending period and the overlap fallacy.[3] In addition, the CAS examination syllabus addresses the permissibility of using calendar year data to determine trends applied to accident year data.[4] These authors have well and fully addressed these topics and they need not be revisited.

---

[1] David R. Chernick, "Private Passenger Auto – Physical Damage Ratemaking", p. 6.

[2] ASP #13...

[3] Chernick, ibid., Charles F. Cook, "Trend and Loss Development Factors", CAS Proceedings, Vol. LVII, p. 1 and McClenahan, Foundations of Casualty Actuarial Science, 2d. Ed., Casualty Actuarial Society, Arlington, VA, 1990, Chapter 2.

[4] Cook, ibid.

In much of the syllabus material, both past and present, there are considerable differences between the types of data used for trending and the amount of discussion dedicated to the selection of the trend. Generally, each paper selects either calendar or accident year data and utilizes either the simple linear or exponential regression model with little guidance regarding which is more appropriate or discussion of the data to which the model is applied. These omissions are understandable since the subject of the articles is ratemaking, of which trend selection is only one component. There are acknowledgements of a need for better loss trending procedures contained in several papers.

A survey of rate filings was conducted to assess common industry practice. From this review, it is difficult to know definitively the amount of analysis that underlies the selection of trends. However, each company and the one rating agency examined display four-quarter-ending calendar year data with either simple linear or exponential regression results to support loss trend selections.[5]

As illustrated in both literature and practice, it is common in the Property & Casualty industry to estimate loss cost trends using either linear or exponential least squares regression. This is understandable since least squares regression is familiar to both regulators and company management. Further, least squares regression has been integrated into all commonly used electronic spreadsheet packages.

The validity of using linear or exponential least squares regression, the basic assumptions of regression analysis and the characteristics of loss data, in evaluating ratemaking trends has not been widely addressed. When selecting a model to estimate future trends, it is important to consider whether the data used violates assumptions of the model.

Loss Data

An essential consideration in evaluating loss trend involves the selection of the type of loss statistics to analyze. It is often useful to analyze both paid and incurred loss frequency and severity if available.

---

[5] Allstate, Nationwide, Progressive, State Farm and ISO

For example, paid claim counts may include claims closed without payment. Therefore, changes in claim handling procedures during the period under review may affect the trend estimate. Likewise, changes in case reserving practices and adjuster caseloads may affect incurred and/or paid severity amounts.

Analysis of both paid and incurred amounts, or amounts net versus gross of salvage and subrogation, can assist in identifying changes in claims handling. In any event, the loss statistics used should be defined consistently throughout the experience period. For example, if the paid loss amounts are recorded gross of salvage and subrogation for a portion of the time period, and net for the remaining, the amounts should be restated to a consistent basis prior to analysis.

Section 2: Least Squares Regression Basics

Least squares regression is a general term that refers to an extensive family of analytical methods. All of these methods share a common basic form.

$$Y_i = f(\vec{X}_i, \vec{\beta}_i) + \varepsilon_i$$

where,

Y_i is the i^{th} observation of the response variable.

$\vec{\beta}_i$ is a vector of model parameters to be estimated.

$\vec{X}_i$ is a vector of the the independent variables.

ε_i is the random error term.

Regression models are designed to use empirical data to measure the relationship between one or more independent variables and a dependent variable assuming some functional relationship between the variables. The functional relationship can be linear, quadratic, logarithmic, exponential or any other form.

The important point is that the functional relationship, the model, is assumed prior to calculation of the model parameters. Incorrect selection of the model is an element of parameter risk.

In addition to selection of the model, regression analysis also involves assumptions about the probability distributions of the observed data. This is essential in the development of statistical tests regarding the parameter estimates and the performance of the selected model.

Simple Linear Regression

The most common form of regression analysis is simple linear regression. The simple linear regression model has the following form.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where,

> Y_i is the i^{th} observation of the response variable.
>
> β_0 and β_1 are the model parameters to be estimated.
>
> X_i is the i^{th} value of the independent variable.
>
> ε_i is the random error term.

The parameters of the regression model are estimated from observed data using the method of least squares. This method will not be described in detail here. It is sufficient for our purpose to note that the least squares estimators, b_i, have the following characteristics:

1. They are unbiased. That is, $E[b_i] = \beta_i$.
2. They are efficient. The least squares estimators have the minimum variance among all unbiased linear estimators.
3. The least squares estimators are the same as the maximum likelihood estimators when the distributions of the error terms are assumed to be independent and normally distributed with a mean of zero and a variance of σ^2.

Because the normal distribution of the error terms is assumed, various statistical inferences can be made. Hypothesis testing can be performed. For example, the hypothesis that the trend is zero can be tested. Confidence intervals for the regression parameters can be calculated. Also, confidence intervals for $\hat{Y}$ and a confidence band for the regression line can be calculated. These very useful results make simple linear regression appealing.

Exponential Regression

While linear regression models are often satisfactory in many circumstances, there are situations where non-linear models seem more appropriate. Loss cost inflation is often assumed to be exponential. The exponential model assumes a constant percentage increase over time rather than a constant dollar increase for each time period.

The general form of the exponential regression model is given by

$$Y_t = \gamma_0 + \gamma_1 e^{\gamma_2 X_t} + \varepsilon_t$$

The parameter estimates of a non-linear regression model usually cannot be described in closed form. Therefore, numerical methods are used to determine parameter estimates using either the least squares or maximum likelihood method. Often electronic spreadsheet software will include tools to estimate the parameters for several non-linear regression models.

As with linear regression, statistical inferences such as confidence intervals for the parameter estimates, hypothesis testing and a confidence band for the fitted curve can be made.

The Exponential to Linear Transformation

In practice, the linear regression algorithm is often applied to the natural logarithm of the observed data. This transformation of the observed data simplifies the calculation of the regression parameters. However, in using this approach the analyst has, perhaps unknowingly, assumed the error terms are lognormally distributed rather than normally distributed.

The observed data is modeled using the equation,

$Ln(Y_t) = \beta_0 + \beta_1 X + \varepsilon_t$

This transformation is equivalent to the model,

$Y_t = Ke^{\beta_1 X_t} \cdot \left[e^{\varepsilon_t}\right]$, where $K = e^{\beta_0}$ and e^{ε_t} is the error term.

and the trend is obtained from the linear least squared regression estimate of β_1.

If the error term of the linear regression model, ε_i, is assumed to have a N(0,σ) distribution, it can be shown that the error term in the transformed model is lognormal with expected value $e^{\sigma^2/2}$. The error terms are positively skewed. This distribution of the error terms in the linearized model may be preferable to the normal distribution if the analyst believes it is more likely that observed values are above the mean than below the mean. This certainly may be the case with insured loss data.

Note that the lognormal distribution of the error term in the linearized model affects the calculation of confidence intervals and test statistics for the model. The familiar forms of the test statistics based on the normal distribution do not apply.

The Coefficient of Determination, R^2

Perhaps the most cited statistic derived from regression analysis is the coefficient of determination, R^2. R^2 can be interpreted as the reduction of total variation about the mean that is explained by the selected model. When R^2 is closer to one, the greater is the modeled relationship between X and Y, whether the model is linear, exponential or some other form.

The Durbin-Watson Statistic

The Durbin-Watson statistic, D, is used to test for serial correlation of the residual errors, e_t. The value of D is calculated from the observed and fitted values of Y, where $e_t = (Y_t - \hat{Y}_t)$.

$$D = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

This value is compared to critical values, d_L and d_U, calculated by Durbin and Watson. The critical values define the lower and upper bounds of a range for which the test is inconclusive. When $D > d_U$, there is no serial correlation present. When $D < d_L$, there is some degree of serial correlation present.[6]

Section 3: Insured Loss Data

There are several distinct characteristics of insured loss data that should be recognized when selecting a regression model. In broad terms, one expects data to be comprised of an underlying trend, a seasonality component, a possible cyclical nature and a random portion.[7] These traits make the estimation of the underlying trend more difficult and the rigid use of simple linear or exponential regression imprudent.

Unusual Loss Occurrences

The nature of insured losses may violate the common assumptions of simple linear or exponential least squares regression. For example, loss events that cause widespread damage can generate extraordinarily high claim frequencies in a given time period. The reverse, a time period with an extraordinarily low claim frequency, is unlikely. A similar skewness can occur in severity data for small portfolios or, almost certainly, in medium to large portfolios of liability risks due to shock losses. Examples of these characteristics are evident in trend data provided by the Insurance Services Office.

Widespread Loss Events

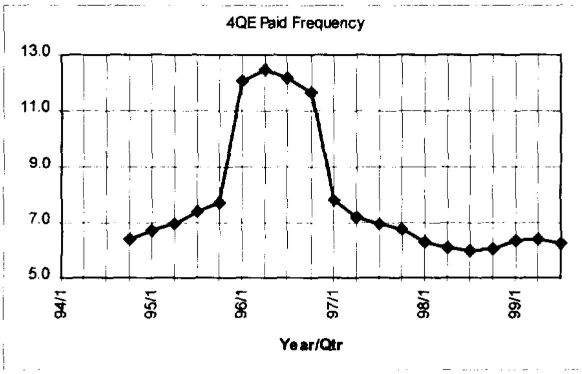
In the chart below of Homeowner claim frequencies as reported by the Insurance Services Office for the state of Oregon, there is an obviously unusual occurrence in the first quarter of 1996. The increase in claim frequency over the prior annual period is over 50%.

---

[6] Neter, et. al., Applied Linear Statistical Models, 4th ed., McGraw-Hill, Boston, 1996 p. 504.

[7] Spyros Makridakis and Steven C. Wheelwright, Forecasting Methods for Management, 5th Ed., John Wiley & Sons, New York, 1989, p. 96.

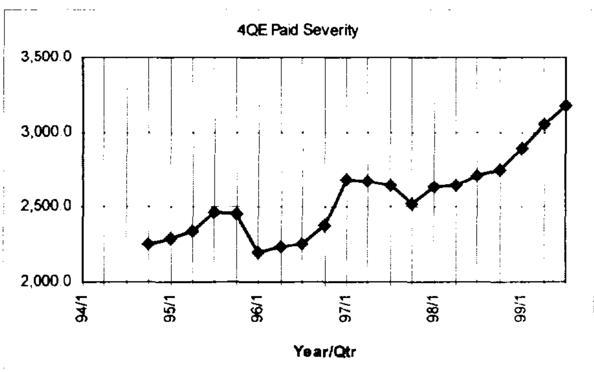
Oregon Homeowners



Because the data is twelve-month-moving, the dramatic rise in frequency that occurred in the first quarter of 1996 is transferred to the subsequent three observations. Therefore, the error terms are not independently distributed, as commonly assumed, due to the construction of the data.

A review of the severity data for the same time period shows a corresponding, though less dramatic, drop in claim severity. This is typical of a high frequency, low severity weather loss event. This drop in claim severity may go unnoticed if it were not for the associated increase in frequency. Again, due to the twelve-month moving organization of the data, the error terms are not independently distributed.

Oregon Homeowners

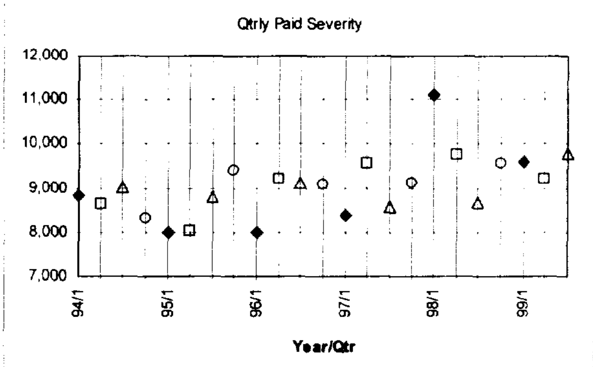


Shock Losses

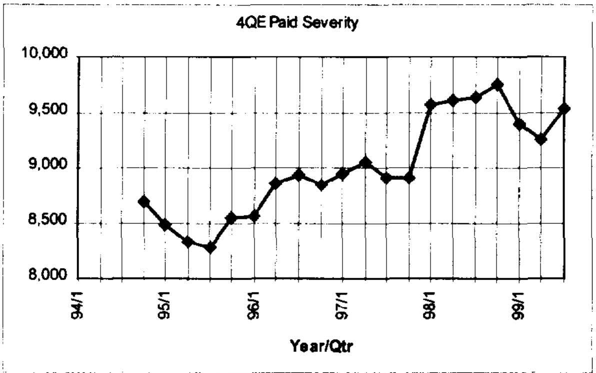
A high severity claim in a small portfolio may cause a distortion in the data and affect the trend calculated by ordinary least squares methods if no adjustments are made. A visual inspection of Nevada Private Passenger Auto Bodily Injury severity data provided by the Insurance Services Office shows an unusual occurrence in the first quarter of 1998.

The quarterly data shows the elevated severity in the first quarter of 1998 neatly as one high point while the four quarter ending data exhibits this phenomena as a four point plateau. This phenomenon occurs more often in smaller portfolios, even when utilizing basic limit data.

Nevada PPA – Bodily Injury Liability



Nevada PPA – Bodily Injury Liability



Effects of Unusual Loss Occurrences

While the cause of these events is dissimilar, the result on the data is the same. One may expect the distribution of the error term for claim frequency and severity to be positively skewed, rather than normally distributed as commonly assumed. The lognormally distributed error terms of the transformed exponential regression model may be more appropriate than the exponential model with normally distributed errors.

As demonstrated above, insured loss frequency and severity data may exhibit abnormally high random error. If these errors occur early in the time series, the resulting trend estimates from least squares regression will be understated. Conversely, if the shock value occurs late in the time series, the trend estimate will be overstated. The use of twelve-month-moving data compounds this effect since the shock is propagated to three additional data points.

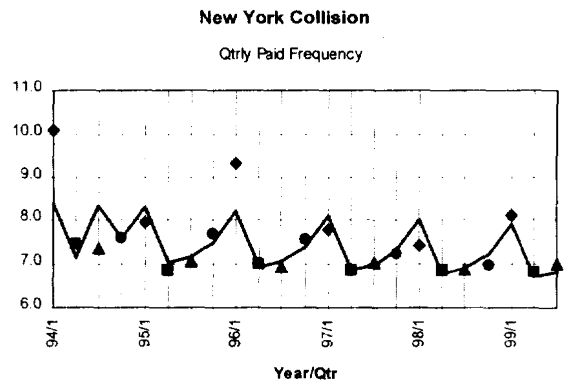
There are several methods available to identify outliers and measure their influence on the regression results. These include Studentized Deleted

Residuals, DFFITS, Cook's Distance and DFBETAS.[8] The identification of such occurrences is addressed in section four below.

Seasonality of Data

The nature of insurance coverage creates seasonal variation in claim frequency and severity. For example, winter driving conditions may cause higher Collision and Property Damage Liability claims in the first quarter. Similarly, lightning claims may be more prevalent during the summer months in certain states. The probability of severe house fires may be higher during the winter months. Auto thefts may be more frequent in summer months causing elevated severity for Comprehensive coverage.

When reviewing New York Private Passenger Auto data for Collision coverage on a quarterly basis, one can see the seasonal nature of claim frequencies. This seasonality can be illustrated by grouping like quarters together.



Generally, the use of twelve-month-moving data is a convenient method for adjusting the seasonal nature of insured losses. However, four-quarter-ending

[8] Neter, et. al., ibid., and Edmund S. Scanlon, "Residuals and Influence in Regression", CAS Proceedings, Vol. LXXXI, p. 123.

data creates serially correlated errors when used in ordinary least squares regression.

Serially Correlated Error

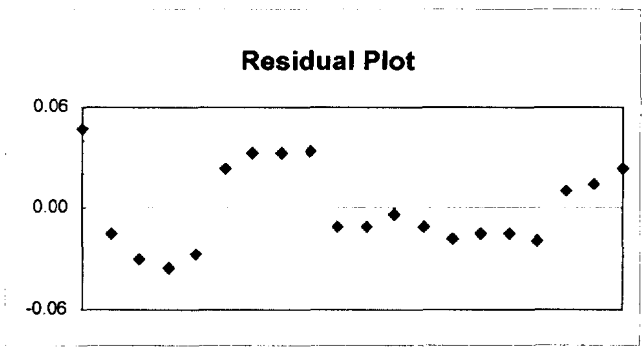
Actuarial literature shows trend data organized in a variety of ways. Some authors use twelve-month-moving calendar year data observed quarterly, others use accident year data observed annually, still others use calendar quarter data observed quarterly. Each format has advantages and disadvantages. It is important to recognize the implications of the data organization on the regression results.

Any organization of data that has overlapping time periods from one point to the next, by its construction, results in serially correlated error terms. Serial correlation of error terms occurs when the residual errors are not independent. This result is shown for twelve-month-moving calendar year data in Exhibit 2 using the Durbin-Watson statistic.

Additionally, one can plot residuals to detect serial correlation. Below the residual plot is displayed for twelve-month-moving New York Collision frequency. As one can see, the errors for adjacent points are related. As noted above, the independence of the error terms in ordinary least squares regression is generally assumed and certain conclusions about the regression statistics are based on this assumption.

New York Collision

Frequency



According to Neter, et. al., when this assumption is not met the following consequences result.

1. The estimated regression coefficients are still unbiased, but they no longer have the minimum variance property and may be quite inefficient.

2. Minimum Squared Error (MSE) may seriously underestimate the variance of the error terms.

3. The standard deviation of the coefficients calculated according to ordinary least squares procedures may seriously underestimate the true standard deviation of the estimated regression coefficient.

4. Confidence intervals and tests using the t and F distributions are not strictly applicable.

Remedial Measures

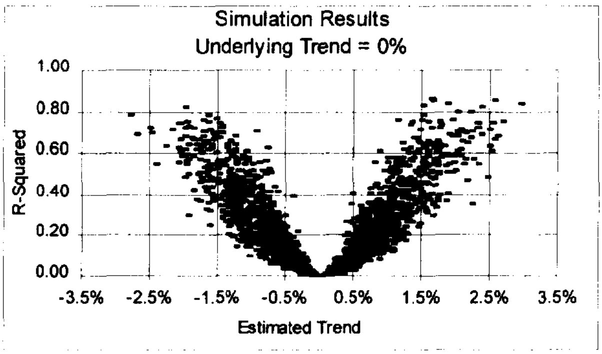
Each of the first two issues with the insured loss data, widespread loss events and extraordinary claim payments, can be resolved by removing outlying points before calculating the exponential or linear regression. The removal technique must rely on statistical tests and actuarial judgment. This will be discussed in the following section. Seasonality and serial correlation can be addressed using regression with indicator variables on quarterly data. Regression with indicator variables explicitly incorporates seasonality as a component of the model. The use of quarterly data eliminates the serial correlation resulting from the use of overlapping time periods.

Comments on Goodness-of-Fit

Estimating the underlying trend in a given dataset entails more than simply fitting a line to a set of data. During the estimation process, it is important to determine whether the underlying assumptions are met and whether the equation accurately models the observed data.[9]

Many consider R^2, the coefficient of determination, the most important statistic for evaluating the goodness-of-fit. The coefficient of determination is the proportion of the data's variability over time that is explained by the fitted curve. However, it is widely agreed that this is not sufficient.[10] The coefficient of determination, by itself, is a poor measure of goodness-of-fit.[11]

To assume that a low R^2 implies a poor fit is not appropriate. It has been shown that a low or zero trend, by its nature, has a low R^2 value.[12] Also, whenever the random variation is large compared to the underlying trend the R^2 will not be sufficient to determine whether the fitted model is appropriate. One can illustrate the low R^2 values associated with data exhibiting no trend over time. The scatter plot below was generated from a simulation with an underlying trend of zero.



---

[9] Scanlon, ibid.

[10] D. Lee Barclay, "A Statistical Note on Trend Factors: The Meaning of R-Squared", CAS Forum, Fall 1991, p. 7, and Ross Fonticella, "The Usefulness of the R^2 Statistic", CAS Forum, Winter 1998, p. 55, and Scanlon, ibid. and Neter et. al., ibid.

[11] Barclay, ibid.

[12] Barclay, ibid.

The residuals between the actual and fitted points are highly useful for studying whether a given regression model is appropriate for the data being studied.[13] It is useful to graph the fitted data against the observed data to look for patterns.[14] A random scattering of residuals occurs when the fit is proper.[15] It is important that the error term not appear systematically biased when compared to neighboring points.

The use of the R^2 statistic or plots of the residuals may result in the decision that the model is an appropriate fit to the data. This conclusion applies to the historical period based on this analysis. Another consideration is the extrapolation of the trend model into the future. As McClenahan illustrates with the use of the 3rd degree polynomial, a perfect fit within the data period does not always result in the appropriate trend in the future.[16] Extrapolation beyond the data period should also be considered before the decision to proceed with the model is undertaken.

Section 4: Identification of Outliers

This section describes methods by which one can identify extraordinary values from observed loss data. These methods are designed to identify outliers from a dataset on which regression is to be performed. An excellent reference on these and other statistical methods is Applied Linear Statistical Models by Neter et. al.

Each of these methodologies cannot be applied without judgement. None of the methods is so robust as to produce reliable results in all circumstances.

[13] Neter, et. al., ibid, p. 25.

[14] Fonticello, ibid.

[15] Barclay, ibid.

[16] McClenahan, ibid.

Therefore, the selected points should always be compared to the original dataset.

The identification of the cause of the outlier is preferred. For example, if possible, the claims department should be consulted if a single large claim or if a widespread claims event, such as a catastrophe, appear to distort the data.

Visual Methods

When performing simple linear regression there are several visual methods which can result in easy identification of outlying points. Among these graphs are residual plots against the independent variable, box plots, stem-leaf plots and scatter plots[17]. While residual plots may lead to the proper inference regarding outliers, there are instances when this is more difficult. When the outlier imposes a great amount of leverage on the fitted regression line, the outlier may not be readily identifiable due to the resulting reduction of the residual.

Studentized Residuals

There are several standard methods that can be utilized to assist with the identification of outliers, each with advantages and disadvantages. The studentized residual detects outliers based on the proportional difference of the error term, e_i, and the variance of these errors. The studentized residual is defined:

$$r_i = \frac{e_i}{s\{e_i\}},$$

Where $s\{e_i\}$ is an estimate of the standard deviation of the residual. This estimate is easily calculated as $s\{e_i\} = \sqrt{MSE(1-h_{ii})}$, where h_{ii} is the diagonal element of the hat matrix $H = X(X'X)^{-1}X'$. Interestingly, $\hat{Y} = HY$ and $e = (1-H)Y$. The hat matrix will be used in future development of outlier identification for simplification of the formulae.

This method has the same disadvantage as identification of outliers using residual graphing. The variance of the errors includes the error of the i[th]

[17] Neter, et. al, ibid.

observation. In addition, there is no statistical test from which one can base a decision regarding outliers.

Studentized Deleted Residuals

A significant improvement in identifying outliers uses the studentized deleted residual. For the i^{th} observation the deleted residual, d_i, is the difference between the i^{th} observation, Y_i, and the fitted point when the fitted curve includes all but the i^{th} observation, $\hat{Y}_{i(i)}$. By excluding the i^{th} observation one can determine the influence of the observation on the fitted function. Fortunately, the deleted residual can be computed relatively easily.

$$d_i = \frac{Y_i - \hat{Y}_i}{1 - h_{ii}} = Y_i - \hat{Y}_{i(i)} \text{ where } h_{ii} \text{ is the diagonal from } H.$$

The deleted residual, d_i, when studentized (divided by the estimated standard deviation of d_i), follows the t(n-p-1) distribution. Therefore, each studentized deleted residual can be tested using $t(1-\alpha/2n; n-p-1)$. Fortunately, the studentized deleted residuals, t_i, can be computed without performing n separate regressions. It can be shown that ,

$$t_i = \frac{e_i}{\sqrt{MSE_{(i)} \cdot (1 - h_{ii})}} = e_i \left[\frac{n-p-1}{SSE(1-h_{ii}) - e_i^2} \right]$$

DFFITS

One measure of influence is the DFFITS statistic. The DFFITS is the standardized difference between the fitted regression with all points included and with the i^{th} point omitted.

$$DFFITS_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{MSE_{(i)} h_{ii}}} = t_i \cdot \left[\frac{h_{ii}}{1 - h_{ii}} \right]^{\frac{1}{2}}$$

This represents the number of standard deviations $\hat{Y}_i$ increases or decreases with inclusion of the i^{th} observation. Note that the DFFITS statistic is a function of the studentized deleted residual and can be computed without performing multiple regressions. Observations are considered outliers if the DFFITS is greater than one for medium datasets and $2\sqrt{p/n}$ for large datasets.

Cook's D

Another measure of influence is Cook's Distance measure, D_i. Scanlon utilizes Cook's D statistic to identify outliers.[18] Cook's D measures the influence of the i^{th} case on all fitted values.

$$D_i = \frac{\sum_{j=1}^{n} (\hat{Y}_j - \hat{Y}_{j(i)})^2}{p \cdot MSE}$$

The denominator standardizes the squared difference measure of the numerator. Evaluation of Cook's D is accomplished by utilizing the F(p, n-p) distribution. A percentile value less than 10-20% shows little influence on the fitted values, while a percentile value of 50% or more indicates significant influence.

Fortunately, Cook's D can be calculated for each observation from a single regression using the following relationship.

$$D_i = \frac{e_i^2}{p \cdot MSE} \cdot \left[\frac{h_{ii}}{(1-h_{ii})^2} \right]$$

As with all models good judgement is imperative and comparison to the original data is advised. In addition to the methods described above, one can calculate a confidence band around the fitted curve. Observations outside the confidence band are candidates for removal.

Each of these methods is designed to identify a single outlier from the remaining data. These techniques may not be sufficient to distinguish outliers when other outliers are adjacent or nearby. Each of these methods is extendable to identify multiple outliers from the remaining data. However, a discussion of these extensions is beyond the scope of this paper.

[18] Scanlon, ibid.

Section 5: Manual Intervention - Deletion/Smoothing of Outliers

Manual Intervention

The identification of extraordinary values is certainly a matter of judgement. In the analysis that follows, the determination of outliers is completed by use of visual inspection.

In many cases a visual review of the twelve-month-moving data can identify outliers. However, the occurrence of two ouliers within four quarters of each other can be difficult to detect using twelve-month-moving data. For this analysis the data is decomposed into the quarterly loss data shown below.

Table 1 – Quarterly Frequency – Oregon Homeowners

	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
1994	6.167	5.778	6.194	7.319
1995	7.573	6.665	8.076	8.613
1996	24.861	8.456	7.006	6.555
1997	9.303	6.053	5.906	5.778
1998	7.300	5.301	5.592	5.986
1999	8.539	5.463	4.965	

The observed frequency in the first quarter of 1996 is identified as an outlier.

Treatment of Outliers

Once the outliers have been identified, one can proceed in several ways. First, the analyst may simply remove the outlying point from consideration and complete the analysis as if the observation did not occur. While this alternative may seem appealing, it does not allow for the reconstruction of twelve-month-moving data.

The second approach is to replace the outlier with the fitted point from the regression after removal of the outlier. This removes the outlier from the regression entirely, but allows reconstruction of the four-quarter-ending data.

The final approach is to replace the outlying point with the fitted point plus or minus the width of a confidence interval, as appropriate. This choice mitigates the extent to which the outlier affects the regression results, without removing the point entirely.

For simplicity, the authors have selected the first approach for comparison purposes but acknowledge that the other two procedures may be appropriate in other circumstances.

Parameter Estimation
Estimation of the underlying trend in the data is completed through exponential regression on the quarterly data, excluding the outliers, with indicator variables to recognize any seasonality.

Section 6: Qualitative Predictor Variables for Seasonality

This method of least squares regression recognizes the seasonal nature of insured losses through the use of qualitative predictor variables, or indicator variables. Indicator variables are often used when regression analysis is applied to time series data. Also, since the data used in this method is quarterly rather than twelve-month-moving, first-order autocorrelation of the error terms is not present. Hence, the issues that arise from such autocorrelation are eliminated.

The linearized form of the exponential regression model is given as

$$Ln(Y_t) = \beta_0 + \beta_1 X_t + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 + \varepsilon_t$$

Where,

Y_t is the dependent variable
X_t is the independent variable (time)
D_2 = 1, if second quarter, 0 otherwise
D_3 = 1, if third quarter, 0 otherwise
D_4 = 1, if fourth quarter, 0 otherwise
ε_t is the random error term

The model above can be viewed as four regression models, one for each set of quarterly data.

The exponential equivalents, without error terms, are

First Quarter: $Y_t = [e^{\beta_0}] \cdot e^{\beta_1 X_t}$
Second Quarter: $Y_t = e^{\beta_2} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$
Third Quarter: $Y_t = e^{\beta_3} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$
Fourth Quarter: $Y_t = e^{\beta_4} \cdot [e^{\beta_0}] \cdot e^{\beta_1 X_t}$

One can think of e^{β_1} as the trend component of the model and e^{β_2}, e^{β_3} and e^{β_4} as the seasonal adjustments to e^{β_0}.

Essentially, the assumption is that the rate of change in frequency or severity over time is constant for all quarters, but the level of frequency or severity differs by quarter. This differs from multiple regression models, which assume separate trends for each quarter. A single trend, rather than four different trends, is intuitively appealing for ratemaking applications.

Section 7: Comparison of Results

This section compares trend estimates derived from five estimation methods applied to industry data provided by The Insurance Services Office. The data is displayed in Exhibit 1. Exponential least squares regression on twelve-month-moving data, quarterly data and annual data are used as examples of common industry practice. The results from the exponential regressions will be compared to results derived from the alternative methods described above.

Detailed calculations using the Oregon Homeowners data are shown in the attached exhibits. The results in the tables below show the annual trend derived from each method and the associated R^2 value in parentheses.

Table 1 - Oregon Homeowners Frequency

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	-1.5% (.06)	-13.9% (.53)	-17.0% (.62)	-6.9% (.17)
Quarterly	-15.6% (.32)	-26.7% (.45)	-13.2% (.21)	-3.9% (.03)
Annual	--	-5.3% (.50)	-19.2% (.72)	-10.1% (.34)
Manual Adjustment	--	-6.8% (.79)	-8.4% (.58)	-2.6% (.20)
Indicator Variables	-9.4% (.91)	-22.2% (.75)	-10.9% (.48)	-2.6% (.27)

Table 2 – New York PPA Collision Frequency

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	0.3% (.04)	-1.7% (.43)	-2.2% (.61)	-1.9% (.58)
Quarterly	-0.6% (.00)	-1.6% (.07)	-2.8% (.17)	-1.7% (.10)
Annual	--	-0.6% (.14)	-2.3% (.66)	-1.2% (.37)
Manual Adjustment	--	--	-1.0% (.80)	-0.8% (.84)
Indicator Variables	1.7% (.83)	-0.6% (.80)	-2.2% (.76)	-1.2% (.74)

Table 3 – Nevada PPA Bodily Injury Severity

	# Years of Observations			
Method	2 yr.	3 yr.	4 yr.	5 yr.
12 MM	1.2% (.06)	3.0% (.52)	3.1% (.72)	3.1% (.78)
Quarterly	4.9% (.10)	4.3% (.20)	4.1% (.31)	2.7% (.25)
Annual	--	3.5% (.63)	2.8% (.71)	3.7% (.85)
Manual Adjustment	--	1.2% (.85)	1.9% (.65)	1.4% (.41)
Indicator Variables	9.4% (.57)	4.9% (.36)	4.0% (.37)	2.7% (.27)

The manual adjustment method and regression using indicator variables provide additional estimates of the underlying loss trend to assist the actuary in selecting appropriate adjustment for ratemaking.

Section 8: Evaluation of Methods Using Stochastic Simulation

In this section, a simulation is constructed to test the accuracy of each estimation method. Each of the five methods above is applied to the simulated data.

Personal auto severity data is simulated using a known underlying trend, a normally distributed random error term, a seasonal adjustment for each quarter and a shock variable to simulate a single large claim payment.

Simulation Parameter Estimation

Based on the Nevada PPA Bodily Injury severity analysis from the previous section the following simulation parameters were selected.

Table 5 – PPA Bodily Injury Severity Simulation Parameters

Trend	3.5%	$e^{\beta_1} - 1$
Severity Variance	5.048 10^{-2}	$MSE / \hat{Y}_0^2$
Base Severity	$8,700	$e^{\beta_0} = E[Y_0]$

Quarter	Seasonal Adjustment		Shock Probability	Shock Magnitude
First	1.000	--	1/23	20%
Second	1.013	e^{β_2}	1/23	20%
Third	0.987	e^{β_3}	1/23	20%
Fourth	1.03	e^{β_4}	1/23	20%

The shock probability and magnitude were chosen based on the observed data. Of the 23 observations, only one observation appeared to have an extraordinarily high severity. The magnitude of the shock is fixed at 20%. The simulation could be further modified to include a stochastic variable for the shock magnitude. Simulations for other states and lines of business would incorporate other parameter values based on observed data.

The simulation function is given by,

$$Ln(Y_i) = [\beta_0 + \beta_1 X_i + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4] \cdot \hat{W}_i + \hat{\varepsilon}$$

where,

$$\Pr[\hat{W}_i = (1 + \delta_i)] = 1/23,$$
$$\Pr[\hat{W}_i = 1.00] = 22/23,$$

and

$\hat{\varepsilon}_i$ is $N(0, \sigma^2)$

The shock value of the natural logarithm of the severity, $1+\delta_i$, corresponding to the shock value of the severity must be calculated. It can be shown that the value of δ_i is given by

$\dfrac{\ln(1+\alpha)}{\beta_0+\beta_1 X_i+\beta_2}$, where α is the shock value for Y.

Likewise, the error variance, σ^2, for $\ln(y_i)$ is derived from the estimated variance of $Y_i = {MSE}/{\hat{Y}_0^2}$ according to the following relationship.

$$\frac{MSE}{\hat{Y}_0^2} = e^{\sigma^2}\left(e^{\sigma^2}-1\right)$$

Simulation Results

Ten thousand simulated data sets were generated. The five estimation methods were applied to each data set.

It is important to note that the application of the manual intervention method assumed correct identification of the extraordinary observations in every simulation. In practice, identification of extraordinary values depends on judgement and statistical methods as described previously. Therefore, the comparison that follows may overstate the accuracy of the manual intervention method.

The table below summarizes the results of each regression method based on 10,000 simulations of twenty observations. Since the underlying trend in the simulation is known, accuracy is measured using the absolute difference between the estimated trend and the actual trend. The percentage of estimates above the actual trends is also shown in order to detect upward bias in the estimation method. Also, the percent of estimates within various neighborhoods of the actual trend are calculated.

The simulation was constructed with a seasonal component and outliers. Therefore, it is not surprising that the manual intervention method that excludes the outliers and includes quarterly indicator variables produces good results.

Table 5 – Comparison of Methods (based on 10,000 simulations)							
			Percentage of Estimates				
Method	Average Trend Estimate	Average Absolute Difference	Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.52%	0.82%	50.7%	37.7%	54.1%	66.9%	.74
Quarterly	3.33%	0.91%	44.1%	34.5%	49.0%	62.5%	.34
Annual	3.51%	0.93%	50.2%	33.6%	48.6%	61.3%	.75
Indicator Variables	3.51%	0.92%	50.4%	34.4%	49.3%	62.0%	.48
Manual Adjustment	3.50%	0.81%	49.4%	37.7%	53.9%	67.6%	.54

A similar process can be used to simulate frequency data which include the probability of loss events that produce large numbers of claims.

Other Simulation Results

Four other simulations were performed. The first compares results when no shocks are present. The second simulation included only data when shock values were present. The third simulation included shocks early in the time series only. The final simulation included shocks only late in the time series.

NO SHOCKS

Table 6 – Comparison of Methods (based on 10,000 simulations)							
			Percentage of Estimates				
Method	Average Trend Estimate	Average Absolute Difference	Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.50%	0.69%	50.0%	43.2%	60.7%	75.0%	.80
Quarterly	3.33%	0.78%	43.2%	39.2%	55.6%	68.9%	.40
Annual	3.51%	0.78%	50.7%	39.0%	55.4%	68.8%	.81
Indicator Variables	3.51%	0.78%	50.8%	39.2%	55.4%	68.9%	.53
Manual Adjustment	3.51%	0.78%	50.8%	39.2%	55.4%	68.9%	.53

The results of this simulation show that there is little difference between traditional regression techniques and regression using qualitative predictor variables for seasonality.

ALL SHOCKED

Table 7 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates: Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	3.52%	0.89%	49.9%	35.2%	50.0%	63.0%	.70
Quarterly	3.35%	0.97%	44.7%	31.9%	46.3%	58.9%	.31
Annual	3.53%	1.01%	50.5%	30.5%	44.6%	57.3%	.72
Indicator Variables	3.53%	0.98%	50.6%	31.4%	45.5%	58.6%	.45
Manual Adjustment	3.51%	0.81%	49.7%	37.9%	54.0%	67.2%	.54

The results of the simulation using only data with shocks illustrate the increased accuracy of the manual adjustment method described previously under these circumstances.

SHOCKED EARLY

Table 8 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates: Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	1.68%	1.88%	6.7%	12.6%	19.3%	26.4%	.35
Quarterly	1.87%	1.78%	11.9%	15.4%	23.2%	30.5%	.16
Annual	1.93%	1.77%	14.1%	15.8%	23.5%	31.6%	.46
Indicator Variables	2.05%	1.66%	15.1%	17.4%	25.5%	34.9%	.33
Manual Adjustment	3.50%	0.84%	49.9%	36.6%	52.1%	65.6%	.53

This simulation illustrates the understatement of trend estimates by traditional methods when shock values occur early in the time series. While proper elimination of the shocks may be difficult, this simulation shows the value of the proper identification.

SHOCKED LATE

Table 9 – Comparison of Methods (based on 10,000 simulations)

Method	Average Trend Estimate	Average Absolute Difference	Percentage of Estimates Above Actual	Within .5% of Actual	Within .75% of Actual	Within 1% of Actual	Average R^2
12 MM	5.37%	1.93%	93.3%	12.4%	19.1%	26.6%	.79
Quarterly	5.23%	1.85%	89.5%	15.5%	23.1%	31.3%	.39
Annual	5.66%	2.22%	93.5%	10.6%	16.8%	22.7%	.80
Indicator Variables	5.50%	2.07%	92.5%	12.1%	18.7%	25.5%	.52
Manual Adjustment	3.52%	0.85%	50.3%	37.1%	52.3%	65.1%	.54

This simulation illustrates the overstatement of trend estimates by traditional regression techniques when shocks occur late in the time series.

Conclusion

The regression concepts discussed here are not new to actuaries. Nor are the characteristics of insured loss data. Actuaries are familiar with the stochastic nature of claim frequency and severity. Actuaries are also keenly aware of the potential for loss events, be they weather events that generate an extraordinary number of "normal" sized claims, or single claims with extraordinary severity, that do not fit the assumptions of basic regression analysis.

While outlier identification techniques are described in section four, they have not been applied to the industry data. The evaluation of these techniques is a subject worthy of further research. In addition, the authors would welcome development of techniques to discriminate between random noise and seasonality, to identify turning points in the trend and to distinguish between outliers and discrete but "jumps" in the level of frequency and severity.

Hopefully, the authors have presented some additional tools for ratemaking and stimulated interest in developing trend estimation techniques that recognize the unique characteristics of insured losses.

Acknowledgements

The authors would like to thank the Insurance Services Office for their generosity in supplying the industry data used in this analysis, the ratemaking call paper committee for their guidance and our families for their understanding and support throughout the process of drafting and editing this paper.

Index to Exhibits

Industry Loss Data	Exhibit 1
Nevada Bodily Injury Liability	Page 1
New York Collision	Page 2
Oregon Homeowners	Page 3
Oregon Homeowners Frequency Regression Results	Exhibit 2
Quarterly Data with Seasonal Indicator Variables	Page 1
Quarterly Data	Page 2
Four-Quarter-Ending Data	Page 3
Annual Data	Page 4
Manually Adjusted Qrtly Data w/ Indicator Variables	Page 5

Nevada Bodily Injury
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	2.018	8,836.39
94/2	2.042	8,634.60
94/3	2.100	9,021.44
94/4	2.186	8,310.44
95/1	2.108	8,000.58
95/2	2.140	8,040.02
95/3	1.967	8,786.99
95/4	2.064	9,415.44
96/1	1.954	7,993.37
96/2	1.842	9,213.77
96/3	1.751	9,124.03
96/4	1.757	9,084.54
97/1	1.739	8,371.74
97/2	1.861	9,572.92
97/3	1.837	8,560.24
97/4	1.831	9,103.45
98/1	1.770	11,106.61
98/2	1.999	9,743.20
98/3	1.778	8,651.21
98/4	1.749	9,552.60
99/1	1.799	9,594.95
99/2	1.830	9,205.35
99/3	1.755	9,799.76

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	2.087	8,694.21
95/1	2.110	8,486.04
95/2	2.134	8,338.22
95/3	2.100	8,277.70
95/4	2.070	8,557.51
96/1	2.031	8,560.77
96/2	1.956	8,855.56
96/3	1.901	8,935.94
96/4	1.825	8,840.39
97/1	1.772	8,950.26
97/2	1.778	9,049.11
97/3	1.799	8,904.46
97/4	1.817	8,912.80
98/1	1.824	9,574.94
98/2	1.859	9,621.17
98/3	1.844	9,642.72
98/4	1.823	9,753.45
99/1	1.830	9,396.03
99/2	1.789	9,252.20
99/3	1.783	9,535.72

New York Collision
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	10.085	1,969.88
94/2	7.458	1,753.67
94/3	7.359	1,946.69
94/4	7.586	2,073.42
95/1	7.951	2,150.86
95/2	6.858	2,022.18
95/3	7.067	2,106.83
95/4	7.692	2,214.01
96/1	9.326	2,230.18
96/2	6.993	2,037.11
96/3	6.948	2,113.95
96/4	7.575	2,275.75
97/1	7.792	2,460.61
97/2	6.860	2,185.90
97/3	7.023	2,226.98
97/4	7.235	2,301.27
98/1	7.423	2,349.35
98/2	6.835	2,112.72
98/3	6.889	2,225.14
98/4	6.982	2,268.54
99/1	8.103	2,392.75
99/2	6.821	2,196.11
99/3	7.000	2,291.83

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	8.117	1,939.10
95/1	7.588	1,984.33
95/2	7.437	2,050.69
95/3	7.363	2,090.35
95/4	7.391	2,126.97
96/1	7.738	2,152.11
96/2	7.768	2,154.42
96/3	7.734	2,155.96
96/4	7.704	2,171.35
97/1	7.328	2,230.23
97/2	7.292	2,265.26
97/3	7.309	2,291.61
97/4	7.225	2,298.13
98/1	7.137	2,268.21
98/2	7.128	2,249.98
98/3	7.094	2,249.49
98/4	7.031	2,241.10
99/1	7.200	2,255.85
99/2	7.198	2,275.84
99/3	7.226	2,291.96

Oregon Homeowner
Insurance Industry Loss Data

YY/Q	Qtrly Paid Frequency	Qtrly Paid Severity
94/1	6.167	2,365.18
94/2	5.778	2,228.65
94/3	6.194	2,224.27
94/4	7.319	2,227.28
95/1	7.573	2,477.43
95/2	6.665	2,436.19
95/3	8.076	2,700.68
95/4	8.613	2,209.83
96/1	24.861	1,973.35
96/2	8.456	2,620.94
96/3	7.006	2,832.13
96/4	6.555	3,070.97
97/1	9.303	2,353.67
97/2	6.053	2,535.58
97/3	5.906	2,747.17
97/4	5.778	2,556.34
98/1	7.300	2,689.08
98/2	5.301	2,569.03
98/3	5.592	3,034.36
98/4	5.986	2,730.10
99/1	8.539	3,126.60
99/2	5.463	3,313.96
99/3	4.965	3,625.18

YY/Q	Four Qtr Ending Paid Frequency	Four Qtr Ending Paid Severity
94/1		
94/2		
94/3		
94/4	6.366	2,259.55
95/1	6.715	2,297.02
95/2	6.935	2,344.92
95/3	7.409	2,468.28
95/4	7.734	2,452.25
96/1	12.069	2,200.97
96/2	12.493	2,241.76
96/3	12.196	2,252.52
96/4	11.656	2,376.12
97/1	7.827	2,683.58
97/2	7.222	2,670.32
97/3	6.942	2,646.38
97/4	6.744	2,525.25
98/1	6.258	2,635.14
98/2	6.066	2,645.05
98/3	5.984	2,713.38
98/4	6.035	2,754.39
99/1	6.353	2,897.10
99/2	6.384	3,054.73
99/3	6.220	3,175.16

Oregon Homeowners

Exponential Regression with Indicator Variables on Quarterly Frequency

Observation	YY/Q	Qtrly Freq.	X_i	D2	D3	D4	Ln(Freq)	ln(Fitted Freq)	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq.
1	94/4	6.17	0.00	0	0	1	1.820	2.307	-0.4875	0.24	-	10.05
2	95/1	5.78	0.25	0	0	0	1.754	1.903	-0.1485	0.02	0.11	6.71
3	95/2	6.19	0.50	1	0	0	1.823	1.924	-0.1007	0.01	0.00	6.85
4	95/3	7.32	0.75	0	1	0	1.991	1.966	0.0249	0.00	0.02	7.14
5	95/4	7.57	1.00	0	0	1	2.024	2.281	-0.2569	0.07	0.08	9.79
6	96/1	6.66	1.25	0	0	0	1.896	1.877	0.0193	0.00	0.08	6.53
7	96/2	8.08	1.50	1	0	0	2.089	1.898	0.1918	0.04	0.03	6.67
8	96/3	8.61	1.75	0	1	0	2.153	1.940	0.2133	0.05	0.00	6.96
9	96/4	24.86	2.00	0	0	1	3.213	2.255	0.9583	0.92	0.55	9.54
10	97/1	8.46	2.25	0	0	0	2.135	1.851	0.2846	0.08	0.45	6.36
11	97/2	7.01	2.50	1	0	0	1.947	1.871	0.0759	0.01	0.04	6.50
12	97/3	6.55	2.75	0	1	0	1.879	1.913	-0.0340	0.00	0.01	6.78
13	97/4	9.30	3.00	0	0	1	2.230	2.229	0.0011	0.00	0.00	9.29
14	98/1	6.05	3.25	0	0	0	1.800	1.825	-0.0246	0.00	0.00	6.20
15	98/2	5.91	3.50	1	0	0	1.777	1.845	-0.0687	0.00	0.00	6.33
16	98/3	5.78	3.75	0	1	0	1.754	1.887	-0.1330	0.02	0.00	6.60
17	98/4	7.30	4.00	0	0	1	1.988	2.203	-0.2150	0.05	0.01	9.05
18	99/1	5.30	4.25	0	0	0	1.668	1.799	-0.1308	0.02	0.01	6.04
19	99/2	5.59	4.50	1	0	0	1.721	1.819	-0.0983	0.01	0.00	6.17
20	99/3	5.99	4.75	0	1	0	1.790	1.861	-0.0712	0.01	0.00	6.43

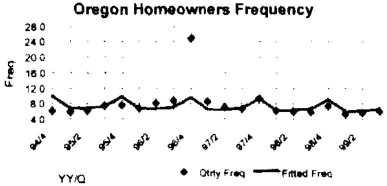
Sum	1.53	1.41
D	**0.92**	
Number of X	4.00	
Observations	20.00	
d_U at .05	1.83	
d_L at .05	0.90	

Test is Inconclusive

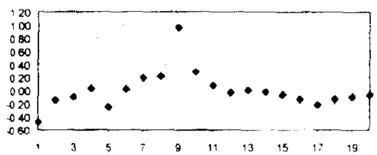
d_U at .01	1.57
d_L at .01	0.68

Test is Inconclusive

Oregon Homeowners Frequency



Residual Plot



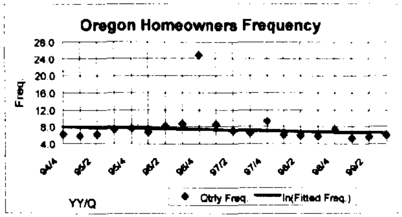
Regression Output

Trend	-2.58%
R^2	0.27
Obs.	20

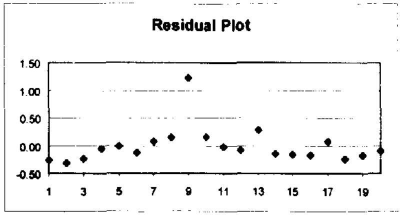
Seasonality Factors	
1st Qtr	1.000
2nd Qtr	1.028
3rd Qtr	1.079
4th Qtr	1.488

Oregon Homeowners
Exponential Regression on Quarterly Frequency

Observatior	YY/Q	Qtrly Freq.	Xi	Ln(Freq)	ln(Fitted Frec	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq.
1	94/4	6.17	0.00	1.820	2.068	-0.2485	0.06	-	7.91
2	95/1	5.78	0.25	1.754	2.058	-0.3037	0.09	0.00	7.83
3	95/2	6.19	0.50	1.823	2.048	-0.2251	0.05	0.01	7.75
4	95/3	7.32	0.75	1.991	2.038	-0.0474	0.00	0.03	7.68
5	95/4	7.57	1.00	2.024	2.028	-0.0038	0.00	0.00	7.60
6	96/1	6.66	1.25	1.896	2.018	-0.1218	0.01	0.01	7.52
7	96/2	8.08	1.50	2.089	2.008	0.0815	0.01	0.04	7.45
8	96/3	8.61	1.75	2.153	1.998	0.1551	0.02	0.01	7.37
9	96/4	24.86	2.00	3.213	1.988	1.2255	1.50	1.15	7.30
10	97/1	8.46	2.25	2.135	1.978	0.1577	0.02	1.14	7.23
11	97/2	7.01	2.50	1.947	1.968	-0.0203	0.00	0.03	7.15
12	97/3	6.55	2.75	1.879	1.958	-0.0781	0.01	0.00	7.08
13	97/4	9.30	3.00	2.230	1.948	0.2825	0.08	0.13	7.01
14	98/1	6.05	3.25	1.800	1.937	-0.1374	0.02	0.18	6.94
15	98/2	5.91	3.50	1.777	1.927	-0.1508	0.02	0.00	6.87
16	98/3	5.78	3.75	1.754	1.917	-0.1630	0.03	0.00	6.80
17	98/4	7.30	4.00	1.988	1.907	0.0805	0.01	0.06	6.74
18	99/1	5.30	4.25	1.668	1.897	-0.2296	0.05	0.10	6.67
19	99/2	5.59	4.50	1.721	1.887	-0.1663	0.03	0.00	6.60
20	99/3	5.99	4.75	1.790	1.877	-0.0871	0.01	0.01	6.54



Sum	2.03	2.90
D	**1.43**	
Number of X	1.00	
Observations	20.00	
d_U at .05	1.41	
d_L at .05	1.20	
Uncorrelated		
d_U at .01	1.15	
d_L at .01	0.95	
Uncorrelated		



Regression Output

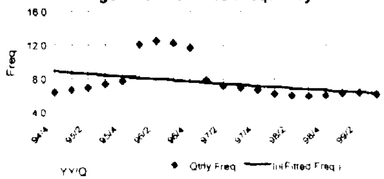
Trend	-3.94%
R^2	0.03
Obs.	20

Oregon Homeowners
Exponential Regression on 4QE Frequency

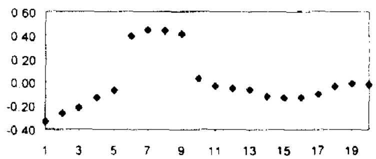
							Durbin-Watson		
Observation	YY/Q	4QE Freq	X_i	Ln(Freq)	ln(Fitted Freq	Residuals	e_t^2	$(e_t-e_{t-1})^2$	Fitted Freq
1	94/4	6 37	0 00	1 852	2 187	-0 3355	0 11	-	8 91
2	95/1	6 72	0 25	1 905	2 169	-0 2641	0 07	0 01	8 75
3	95/2	6 93	0 50	1 936	2 151	-0 2155	0 05	0 00	8 60
4	95/3	7 41	0 75	2 003	2 133	-0 1306	0 02	0 01	8 44
5	95/4	7 73	1 00	2 045	2 116	-0 0705	0 00	0 00	8 29
6	96/1	12 07	1 25	2 491	2 098	0 3930	0 15	0 21	8 15
7	96/2	12 49	1 50	2 525	2 080	0 4450	0 20	0 00	8 00
8	96/3	12 20	1 75	2 501	2 062	0 4394	0 19	0 00	7 86
9	96/4	11 66	2 00	2 456	2 044	0 4120	0 17	0 00	7 72
10	97/1	7 83	2 25	2 058	2 026	0 0316	0 00	0 14	7 59
11	97/2	7 22	2 50	1 977	2 008	-0 0316	0 00	0 00	7 45
12	97/3	6 94	2 75	1 937	1 991	-0 0533	0 00	0 00	7 32
13	97/4	6 74	3 00	1 908	1 973	-0 0647	0 00	0 00	7 19
14	98/1	6 26	3 25	1 834	1 955	-0 1207	0 01	0 00	7 06
15	98/2	6 07	3 50	1 803	1 937	-0 1337	0 02	0 00	6 94
16	98/3	5 98	3 75	1 788	1 919	-0 1307	0 02	0 00	6 82
17	98/4	6 04	4 00	1 798	1 901	-0 1029	0 01	0 00	6 69
18	99/1	6 35	4 25	1 848	1 883	-0 0350	0 00	0 00	6 58
19	99/2	6 38	4 50	1 853	1 866	-0 0124	0 00	0 00	6 46
20	99/3	6 22	4 75	1 828	1 848	-0 0199	0 00	0 00	6 35
						Sum	1 04	0 40	
						D	**0.38**		
						Number of X	1 00		
						Observations	20 00		

d_U at 05	1 41
d_L at 05	1 20
First Order Auto-Correlated	
d_U at 01	1 15
d_L at 01	0 95
First Order Auto-Correlated	

Oregon Homeowners Frequency



Residual Plot



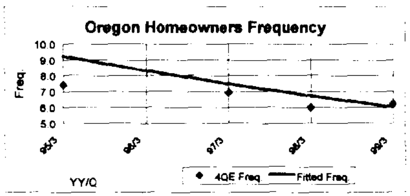
Regression Output

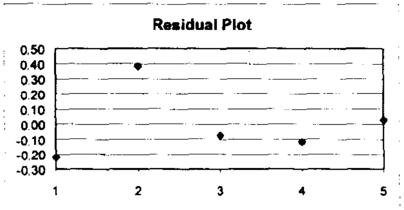
Trend	-6 89%
R^2	0 17
Obs	20

Oregon Homeowners
Exponential Regression 4QE Frequency, Annual Observations

Observatior	YY/Q	4QE Freq	Xi	Ln(Freq)	ln(Fitted Frec	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq
1	95/3	7.41	0.00	2.003	2.224	-0.2213	0.05	-	9.25
2	96/3	12.20	1.00	2.501	2.118	0.3836	0.15	0.37	8.31
3	97/3	6.94	2.00	1.937	2.012	-0.0742	0.01	0.21	7.47
4	98/3	5.98	3.00	1.788	1.905	-0.1168	0.01	0.00	6.72
5	99/3	6.22	4.00	1.828	1.799	0.0288	0.00	0.02	6.04

Sum	0.22	0.60
D	2.77	
Number of X	1.00	
Observations	5.00	
dU at .05	na	
dL at .05	na	
dU at .01	na	
dL at .01	na	





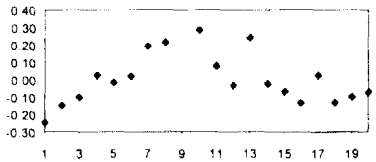
Oregon Homeowners
Manually Adjusted Exponential Regression with Indicator Variables on Quarterly Frequency

Observatior	YY/Q	Qtrly Freq	Xi	D2	D3	D4	Ln(Freq)	Fitted Freq	Residuals	Durbin-Watson e_t^2	Durbin-Watson $(e_t-e_{t-1})^2$	Fitted Freq
1	94/4	6.17	0.00	0	0	1	1.820	2.068	-0.2479	0.06	-	7.91
2	95/1	5.78	0.25	0	0	0	1.754	1.903	-0.1485	0.02	0.01	6.71
3	95/2	6.19	0.50	1	0	0	1.823	1.924	-0.1007	0.01	0.00	6.85
4	95/3	7.32	0.75	0	1	0	1.991	1.966	0.0249	0.00	0.02	7.14
5	95/4	7.57	1.00	0	0	1	2.024	2.042	-0.0173	0.00	0.00	7.70
6	96/1	6.66	1.25	0	0	0	1.896	1.877	0.0193	0.00	0.00	6.53
7	96/2	8.08	1.50	1	0	0	2.089	1.898	0.1918	0.04	0.03	6.67
8	96/3	8.61	1.75	0	1	0	2.153	1.940	0.2133	0.05	0.00	6.96
10	97/1	8.46	2.25	0	0	0	2.135	1.851	0.2846	0.08	0.01	0.00
11	97/2	7.01	2.50	1	0	0	1.947	1.871	0.0759	0.01	0.04	6.36
12	97/3	6.55	2.75	0	1	0	1.879	1.913	-0.0340	0.00	0.01	6.50
13	97/4	9.30	3.00	0	0	1	2.230	1.989	0.2407	0.06	0.08	6.78
14	98/1	6.05	3.25	0	0	0	1.800	1.825	-0.0246	0.00	0.07	7.31
15	98/2	5.91	3.50	1	0	0	1.777	1.845	-0.0687	0.00	0.00	6.20
16	98/3	5.78	3.75	0	1	0	1.754	1.887	-0.1330	0.02	0.00	6.33
17	98/4	7.30	4.00	0	0	1	1.988	1.963	0.0246	0.00	0.02	6.60
18	99/1	5.30	4.25	0	0	0	1.668	1.799	-0.1308	0.02	0.02	7.12
19	99/2	5.59	4.50	1	0	0	1.721	1.819	-0.0983	0.01	0.00	6.04
20	99/3	5.99	4.75	0	1	0	1.790	1.861	-0.0712	0.01	0.00	6.17
												6.43
									Sum	0.38	0.32	
									D	**0.86**		
									Number of X	4.00		
									Observations	19.00		

Oregon Homeowners Frequency

d_U at .05	1.85
d_L at .05	0.86
First Order Auto-Correlated	
d_U at .01	1.58
d_L at .01	0.65
Test is Inconclusive	

Residual Plot



Regression Output

Trend	-2.58%
R^2	0.20
Obs	19

Seasonality Factors	
1st Qtr	1.000
2nd Qtr	1.028
3rd Qtr	1.079
4th Qtr	1.171