The Econometric Method of Mixed Estimation

An Application to the Credibility of Trend

by Paul Brehm and Denis Guenthner

BIOGRAPHIES:

Mr. Brehm is an Actuarial Officer with the St. Paul Fire & Marine in their Personal Lines Division. Prior to joining the St. Paul he attended the University of Minnesota. He has a B.S. in Economics with a minor in Statistics. He became a fellow of the CAS in 1989.

Mr. Guenthner is a Senior Actuary in the Medical Services Division of The St. Paul Fire & Marine. He received B.S. degrees in Economics and Mathematics from the University of South Dakota. He has been a fellow of the CAS since 1987.

ABSTRACT:

Actuaries are often confronted with conflicting data and information. In trend analysis, ordinary least squares regression techniques do not allow for the introduction of any extraneous information, conflicting or not. Credibility methods have been proposed to solve this shortcoming but have not been widely accepted.

This paper discusses the use of an econometric technique, known as mixed estimation, to incorporate prior information directly into the specification of the trend model. The resultant parameter estimates are credibility weighted estimates. Mixed estimation goes one step further by generating a test statistic to test the compatibility of the data and the complement information. An effort has been made to keep theory and notation to a minimum, emphasizing practical application of the technique.

The Econometric Method of Mixed Estimation: An Application to the Credibility of Trend

INTRODUCTION

The Problem at Hand. In many aspects of actuarial analysis, the actuary must measure an underlying trend in order to make a projection. Examples include ratemaking and forecasting. In trend analysis, actuaries are often faced with data that is not sufficiently credible. Alternatively, actuaries are often faced with various sources of information (data or otherwise), each seemingly plausible and perhaps even credible, that yield different trend estimates. Further, the actuary is not well equipped when rigorous statistical or numerical estimates of trend disagree with a priori beliefs.

Ordinary least squares regression is the technique most often used in trend analysis. Standard regression models provide a wealth of statistics to measure goodness of fit. Regression models do not, however, tell the actuary how credible the resultant trend estimate is. Nor can regression models incorporate conflicting pieces of information directly into the specification of a trend model.

The Proposed Solution. In 1961, Theil and Goldberger [8] faced a similar problem in the estimation of price elasticities. Theil and Goldberger note that when a model provides estimated parameters that are counter intuitive, the model is most often changed or discarded. If intuition (or, more generally, any alternative information) is so strong, they argue that it is more logical to incorporate the alternative information into the estimation process rather than to discard the model because it contradicts the omitted knowledge.

Theil and Goldberger propose a generalized least squares technique that directly incorporates alternative information in the form of linear constraints to overcome the shortcomings of standard regression analysis. The solution they found results in parameter estimates (trend parameters in our case) that are in fact credibility weighted parameters where the credibility weights are determined based on the relative error variances from the original and complement estimates. In a later paper Theil [7] proposes a test statistic to measure the compatibility of the complement information.

This paper presents an application of Theil-Goldberger mixed estimation to the calculation of trends. After a review of current credibility techniques as applied to trends, this paper discusses the formulation of the mixed estimation model, and offers some examples of the application of the methodology in practice. Throughout this paper an attempt has been made to move the more serious mathematics to the appendix and concentrate on concepts and examples in the body of the paper.

AN OVERVIEW of CREDIBILITY as APPLIED to TREND

There are a number of examples of trend procedures utilizing credibility, both in practice and in literature. Despite the examples, credibility as applied to trend has not been widely used. Following is a brief survey of a few of the examples the authors have found.

ISO. The Insurance Services Office uses a credibility routine in its trend analysis for private passenger auto [2]. ISO starts with three years of paid data for claims costs and six years of data for frequency. Values are calculated on a rolling four quarter basis to eliminate seasonality. An exponential model is fit to the frequency and severity data on both a countrywide and a statewide basis. State trends are credibility weighted with countrywide. State credibility for a given coverage is based on the number of claims in the year ending in the most recent quarter. A classical credibility approach is used with full credibility assumed at 10,623 claims for each coverage. Partial credibility is tabular and based on the square root rule. Weighted trend estimates are capped at certain minimums and maximums.

The ISO technique is not the most theoretically appealing. It relegates credibility to a component in a series of mechanical decision rules. Furthermore, classical credibility is usually held in lower regard than Bayesian approaches. On the other hand, the ISO method has the advantage of simplicity in application.

NCCI. The National Council on Compensation Insurance uses a credibility procedure in conjunction with trend factors for adjusted loss ratios [5]. The NCCI trend procedure has undergone change over the past few years. Previously, credibility was assigned based on the Spearman D-statistic. Currently credibility is assigned based on the magnitude of the standard error of the regression as a percent of the projected point. Full credibility exists when this quantity is less than or equal to .0006. This standard presumably leads to a 90% probability that the actual loss ratio will be within 6% of the projection. A square root rule is used for partial credibility. For medical benefits, the balance of the credibility is assigned to the countrywide trend. For indemnity, the complement is unity.

The NCCI methodology retains the classical approach adopted by ISO. Here, however, full credibility is based on what amounts to the width of a confidence interval rather than an arbitrary number of claims. The above method is similar in nature to one proposed by Gary Venter.

Venter. In his paper, "Classical Partial Credibility with Application to Trend" [10], Mr. Venter rightly identifies classical partial credibilities as nothing more than the ratio of two confidence intervals. For example, if we wish to be within 5% (desired confidence interval width) of the true projection 90% of the time and our regression yields a 90% confidence interval width that is 10% of our projected point, the partial credibility is 50% (5%/10%). In this way, as Venter points out, volume is only important in as much as it affects the goodness of fit of the regression line.

Venter's technique is simple in application and has a greater theoretical appeal. It's main disadvantages stem from its classical credibility approach, disadvantages enumerated by Mr. Venter himself. In this, like all classical approaches, the greatest downfall is the degree of subjective judgement required. In addition, credibility is only meaningful in reference to a projected point, rather than the underlying trend.

Van Slyke. In an article that actually preceded Venter's, O.E. Van Slyke authored "Credibility-Weighted Trend Factors" [9]. Van Slyke's paper presents a Bayesian formula for the partial credibility of a trend factor. Partial credibility is calculated based on the relative variance between alternative estimates. In the example given the competing models are trend and no trend (average). If we let V_a be the variance of the straight average and V_t be the variance of the projected point, the partial credibility of the trend estimate is:

$$\frac{(1/V_t)}{(1/V_t) + (1/V_a)}$$

The above formula has the greatest theoretical appeal of all examples presented here. It is unfortunate that this work has not been widely used in practice. We will show that the above formula is the same as the one advanced in this paper, with the exception that we will replace V_a with a variance from an alternative trend model (Van Slyke hints at, but stops short of, formulating the above for the case where the alternative model is another estimate of trend) and V_t with the regression error variance.

Our formulation will be done in the context of, and in reference to, an econometric model rather than an actuarial one. Mixed estimation will extend credibility procedures by incorporating the complement information, statistical or otherwise, directly into the specification of the trend model. The procedure will transcend current credibility techniques by testing the compatibility of the complement information.

DEVELOPMENT of MIXED ESTIMATION

Background. In econometric analysis, it is not uncommon that a regression equation yields results which are not consistent with a priori expectations. The a priori expectations may be derived from theoretical considerations. For example, when a demand function is estimated, economic theory requires that the sum of all price and income elasticities must be zero. Also, we expect that price elasticities of demand are negative, and income elasticities are positive.

Alternatively, prior expectations may come from previous statistical analysis or another independent sample of information. For example, a prior statistical analysis may provide an estimate of the coefficients of a production function. In a demand study, we may have two samples of observations. Since both samples provide information on price elasticities, it is logical to specify the model such that both samples contribute simultaneously to the estimation process.

Conventional regression analysis does not allow alternative information to be incorporated into the model. Yet, the existence of the a priori information is precisely the reason that the regression results may be rejected, or the alternative information must be rejected in light of the new sample data.

Mixed estimation is a generalized regression technique, pioneered by Theil and Goldberger, that provides an intuitive credibility weighting, or melding, of alternative information. Theil and Goldberger's work is an extension of the work done by Durbin [1], which proposed an approach for pooling time series and cross sectional data.

Following is a brief description of the mixed estimation process. The reader is referred to Appendices A and B which summarize Ordinary and Generalized Least Squares, in matrix notation.

Model Specification. Assume we have sample of observations which satisfy the following linear relationship:

$$y = X\beta + u$$

where $y = n \times 1$ vector of the dependent variable,

X = n x k matrix of explanatory variables ($k < n$),

β = k x 1 vector of parameters to be estimated,

u = n x 1 vector of residuals

such that $E(u) = 0$ and $E(uu') = \Omega$. Note that the above assumptions are identical to the assumptions of Generalized Least Squares (GLS) model as defined in Appendix B.

Additional information on β is available. This alternative, or extraneous information may be of the form of statistical data, or simply an a priori estimate of β and may be stated in the form of linear restrictions as:

$$r = R\beta + v$$

where r is a g x 1 vector, R is a g x k matrix, and v is a g x 1 vector such that $E(v) = 0$ and $E(vv') = \Sigma$.

To estimate β, we can use the sample information alone, in which case the GLS estimator (Appendix B) is the best linear unbiased estimate. However, the sample and extraneous information may be combined from the beginning by specifying the model as:

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} u \\ v \end{bmatrix} \qquad \text{where}$$

$$E \begin{bmatrix} u \\ v \end{bmatrix} = 0 \quad \text{and} \quad E \begin{bmatrix} u \\ v \end{bmatrix} [\, u' \quad v' \,] = \begin{bmatrix} \Omega & 0 \\ 0 & \Sigma \end{bmatrix} .$$

The mixed estimate of β is obtained by applying GLS to the above model and provides the following best linear unbiased estimate:

$$\tilde{\beta} = [\ X'\Omega^{-1}X + R'\Sigma^{-1}R]^{-1}[X'\Omega^{-1}y + R'\Sigma^{-1}r] \quad . \tag{1}$$

The formula for the mixed estimate can be simplified if we make some additional assumptions. If we assume that the u's are mutually independent with a common variance σ_u^2 and the v's are mutually independent with a common variance σ_v^2, then $E(uu') = \Omega = \sigma_u^2 I$ and $E(vv') = \Sigma = \sigma_v^2 I$. The mixed estimator, $\tilde{\beta}$, then simplifies to:

$$\tilde{\beta} = [(1/\sigma_u^2)(X'X)+(1/\sigma_v^2)(R'R)]^{-1}[(1/\sigma_u^2)(X'y)+(1/\sigma_v^2)(R'r)] \ . \tag{2}$$

In formula (2) it can be seen that the mixed estimator is weighting together the components of two regressions with weights equal to the inverse of their respective error variances.

If there is only a single piece of extraneous information, such as an a priori opinion of β_1, the first element of β, then the mixed estimate becomes

$$\tilde{\beta} = [\ X'X + \frac{\sigma_u^2}{\sigma_v^2} \begin{bmatrix} 1 & 0 \ldots 0 \\ 0 & 0 \ldots 0 \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ 0 & 0 \ldots 0 \end{bmatrix}]^{-1}[\ X'y + \frac{\sigma_u^2}{\sigma_v^2} \begin{bmatrix} \beta_1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}] \quad . \tag{3}$$

Formula (3) highlights the fact that alternative information need not be statistical data. A priori estimates of β_1 and σ_v^2 could be derived from any source. In an example of estimating price elasticities in their paper, Theil and Goldberger [8] use the point estimates of price elasticities and associated variances from a prior study as the alternative information.

Confidence Intervals. The Theil-Goldberger mixed estimator has an intuitively appealing graphical depiction. But first, a digression on regression parameter confidence intervals is in order.

In a simple regression of y = a + bx + e, we can calculate the confidence intervals around a and b by:

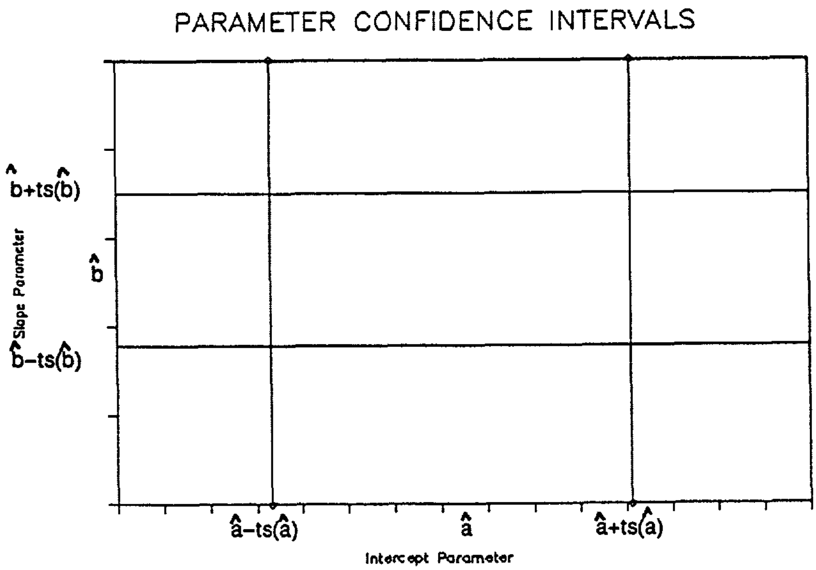
$$a \pm t_{n-2,(1-\alpha/2)} s_a \text{, and}$$
$$b \pm t_{n-2,(1-\alpha/2)} s_b \text{,}$$

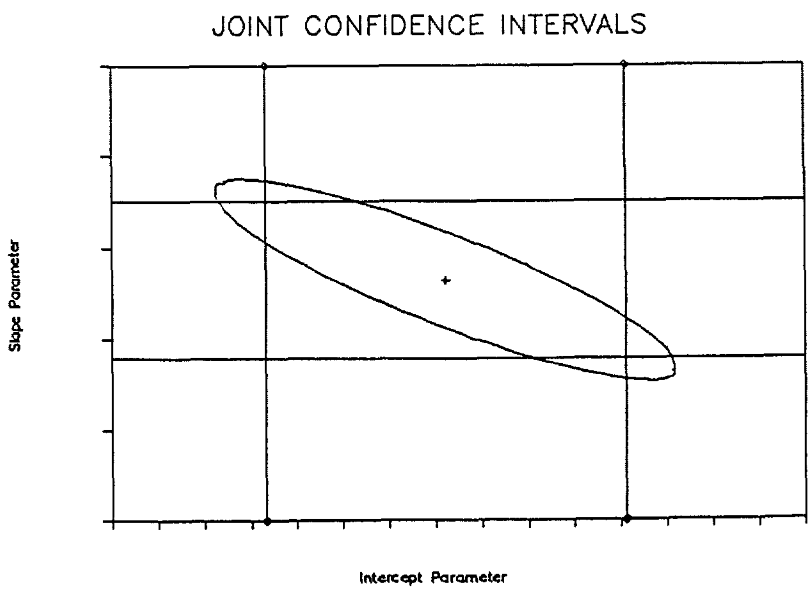
$$\text{where } s_a^2 = s_u^2[(1/n) + x^2/\Sigma(x-\bar{x})^2] \text{, and}$$
$$s_b^2 = s_u^2/\Sigma(x-\bar{x})^2 \text{ .}$$

Graphed in parameter space this would appear as a rectangle as shown on the top of Exhibit 1.

Statisticians and actuaries are trained to know that, because of the covariance between a and b, the joint confidence region is more efficient than the above rectangle. The joint confidence region is actually an ellipse as shown on the bottom of Exhibit 1. The size of the ellipse is determined by the desired confidence level. The set of all joint confidence regions is a

Exhibit 1
Regression Confidence Intervals





series of concentric ellipses. For further details on joint confidence regions, c.f., [7].

In the case of mixed estimation, if we perform two separate regressions on the sample data and the alternative information, we would have two sets of concentric ellipses. The solution for $\tilde{\beta}$, the mixed estimate, turns out to be a point on the locus of tangencies between the joint confidence ellipses from the sample and alternative regressions. The specific point on the locus of tangencies depends on the relative variances between the regressions.

The graphical solution described above is demonstrated through examples later in this paper.

Credibility. Under the assumptions presented in this section, the mixed estimator is equivalent to giving weight to the sample and extraneous information in proportion to the inverse of the variances. The credibility of the sample information is

$$z = \frac{[1/\sigma_u^2]}{[(1/\sigma_u^2) + (1/\sigma_v^2)]} \quad . \tag{4}$$

Formula (4) is equivalent to Van Slyke's formulation except it 1) substitutes the variance of the extraneous information for the variance of the straight average, and 2) is expressed in terms of error variances rather than a variance of a projected point.

In the formulations of $\tilde{\beta}$ and z, we rely on σ_u^2 and σ_v^2. However, in practice these quantities are rarely known. Where the true variance parameters are unknown, the unbiased estimators s_u^2 and s_v^2 (respectively) are substituted.

Since $s^2 = \Sigma(y-\hat{y})^2/(n-k) = SSE/(n-k)$, we can simplify the above expression for z to look more like the Bayesian formula actuaries are used to. Let the subscripts "u" denote values from the original regression and "v" denote values from the alternative. Then

$$z = \frac{(n-k_u)/SSE_u}{(n-k_u)/SSE_u + (g-k_v)/SSE_v}$$

If the numerator and denominator are both multiplied by SSE_u,

$$z = \frac{(n-k_u)}{(n-k_u) + (g-k_v)[SSE_u/SSE_v]} \qquad (5)$$

Here we have an expression for credibility in the form N/N+K, where "N" is equal to the number of degrees of freedom in the original regression $(n-k_u)$, and "K" is equal to the ratio of the sums of squared errors from the two regressions weighted by the degrees of freedom from the alternative.

Formula (5) is consistent with known properties of credibility: 1) z is between zero and one, 2) z increases with N and SSE_v (variance of the complement), and 3) z decreases with SSE_u (variance of data).

Manipulating formula (5) yields some insight into practical considerations when using mixed estimation. Consider the case where $k_u = k_v = k$, then by dividing numerator and denominator by $(n-k)$,

$$z = \frac{1}{1 + [(g-k)/(n-k)][SSE_u/SSE_v]}$$

Note that the relative sample sizes, n and g, will affect the assigned credibility directly as well as through the relative goodness of fit. For example, if n is substantially larger than g, z will be relatively larger, ceterus paribus. However, if g is selected equal to n, credibility is solely a function of the relative error variances.

TEST of COMPATIBILITY

A major advantage of the mixed estimation procedure is that not only does it implicitly meld together alternative sets of information, but is also provides a test of whether the two sets of information are compatible. To test the null hypothesis that the sample and extraneous information are compatible, Theil [8] proposed the following test statistic:

$$\tau = (r-R\hat{\beta})'[s_u^2 R(X'X)^{-1}R' + \Sigma]^{-1}(r-R\hat{\beta}) \ , \qquad (6)$$

where $\hat{\beta}$ is the OLS estimate of the sample information, and s_u^2 is the estimate of σ_u^2. Theil showed that τ is distributed as Chi-square with g degrees of freedom. If τ exceeds the selected Chi-square critical value, the sample and extraneous information are not compatible, and the mixed estimate should not be computed.

In the simplified case of formula (2),

$$\tau = (r-R\hat{\beta})'[s_u^2 R(X'X)^{-1}R' + \sigma_v^2 I]^{-1}(r-R\hat{\beta}). \qquad (7)$$

The test statistic equips the actuary, for the first time, with a tool for testing what amounts to "the goodness of fit" between the observed data and the complement of credibility.

EXAMPLE 1: ALTERNATIVE STATISTICAL DATA

In this example we are seeking to identify the underlying annual average trend in average severity in State X for a homeowners rate indication. We have State X severity data available to us as shown in Appendix C, Sheet 1. Since State X severity has a quarterly seasonality, we fit the following regression equation using ordinary least squares:

$$\text{Average Severity} = \alpha e^{\{\beta_1(\text{time}) + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4\}},$$

where D_2, D_3, and D_4 are dummy variables for second, third, and fourth quarters respectively. The equation is linearized by taking the logs of both sides.

The average annual trend is calculated to be 9.5% ($e^{4\hat{\beta}_1}-1$) for State X, a trend we believe high based on knowledge of countrywide trends. A similar analysis of countrywide trends yielded a annual rate of change of 3.7% (Appendix C, Sheet 2). Rather than disregard the high trend figure as a fluke or presume instead that it the only possible answer, Theil-Goldberger mixed estimation is used to incorporate the available countrywide data into the specification of our state trend model. To keep this example simple, we will assume that for both the state and the countrywide regression, $E[uu']=\sigma_u^2 I$ and $E[vv']=\sigma_v^2 I$, respectively.

Since we are primarily interested in the trend parameter calculated by the mixed estimator, we have chosen to neutralize the seasonality and differences in intercepts (magnitude) between State X and countrywide by first deseasonalizing the data and then indexing the deseasonalized severities to the first point. Deseasonalizing is accomplished using factors based on $e^{\hat{\beta}_2}$, $e^{\hat{\beta}_3}$, and $e^{\hat{\beta}_4}$ calculated in Appendix C, Sheets 1 and 2. Regressions are now fit to the adjusted severity indices for State X (Sheet 3) and the countrywide (Sheet 4). The model in each case was:

$$\text{Severity Index} = \alpha e^{\beta(\text{time})}$$

which is linearized by taking the logs of both sides. Note that the trend estimates $(e^{4\hat{\beta}}-1)$ from both regressions are the same as the original regressions, but we can now focus on fewer parameters in our mixed estimation. The mixed estimator is then calculated using formula (2).

The matrix manipulations can look intimidating, but formula (2) can actually be calculated in LOTUS using /RANGE TRANSPOSE, /DATA MATRIX MULTIPLY, and /DATA MATRIX INVERT. $\tilde{\beta}$ can also be calculated as

$$\tilde{\beta} = z(\hat{\beta} \text{ from original regression}) + (1-z)(\hat{\beta} \text{ from alternative regression})$$

using the credibility formula (4).

Appendix C, Sheet 5 shows the mixed estimation parameters. The estimated trend is found to be 4.7%. Sheet 5 also shows the graphical interpretation of the solution. The graph shows two sets of two concentric ellipses, one corresponding to the original regression, one to the alternative. The mixed estimate lies on the locus of tangencies between all such concentric ellipses.

The precise point on the locus is determined by the relative regression error variances, hence, credibility.

Note that credibility was never explicitly calculated in this example. No values of p or k needed to be defined. The only need for actuarial judgement was in the selection of the complement information. Credibility can be calculated with formula (4):

$$\frac{(1/s_u^2)}{(1/s_u^2) + (1/s_v^2)} = \frac{(1/0.003)}{(1/0.003) + (1/0.0006)} = 16.8\% .$$

Credibility can also be backed out given the above parameter estimates.

$$(9.5\%)(Z) + (3.7\%)(1-Z) = 4.7\%$$

$$\text{or} \qquad Z = 16.8\%.$$

Actuarial judgement is always required in the selection of a complement of credibility. This is inherent in credibility procedures. The advantage of mixed estimation over other procedures is the ability to statistically test the judgement employed. By calculating Theil's test statistic, the actuary can rigorously test the compatibility of the complement information. Using formula (7), in this example the value of $\tau = 12.6$, g is equal to 15. The critical value of Chi-squared at the 95% confidence level is 25. Since $12.6 < 25$, we cannot reject the hypothesis that the alternative information, (countrywide severity) is a viable complement to the data (state severity).

This example makes an interesting point as regards the compatibility test. The test is one of **compatibility** not **equality.** A statistician would look at the joint confidence intervals shown on Sheet 5 (cont.) of Appendix C and conclude that the two sets of regression parameters are different at the appropriate level of confidence (since the ellipses do not intersect). We, as actuaries, conclude only that we cannot reject the possibility that the alternative information is a viable complement.

EXAMPLE 2: ALTERNATIVE is an ECONOMIC INDEX

The importance of the test of compatibility is highlighted by this example. Recall that if we reject the hypothesis that our sample data and the alternative information are compatible, the mixed estimation procedure should not be employed.

Our goal in this example is to determine the long term trend in the severity of claims for physicians and surgeons professional liability. The data sample consists of countrywide severity at basic limits, developed to an ultimate basis, for notice years 1979 to 1988. As shown on Appendix D, Sheet 1, the severities are indexed to 1.000 for 1979, and the following regression line is estimated:

$$\ln(\text{severity index}) = \alpha + \beta(\text{time})$$

where ln is the natural (base e) logarithm. The resulting annual trend is 10.2%.

For the alternative information, we consider related Consumer Price Indices. The All Medical Care Items index was selected, which reflects changes in price levels for doctors fees, hospital room charges, and drug and prescription costs. The CPI index has been adjusted to set 1979 equal to 1.000. Both the data sample and the alternative information have been indexed at 1979 so that the resulting parameter estimates of the intercept are of similar magnitudes. The following model is then estimated:

$$\ln(\text{CPI index}) = \alpha + \beta(\text{time}) \ ,$$

and the resulting estimated annual trend is 8.2% as shown on Appendix D, Sheet 2.

Before moving on to the mixed estimator, we first consider the test of compatibility of the two sources of information. After fitting the two individual regression lines, we have all the data needed to compute the compatibility statistic, τ (formula (7)). Given the trend estimates are similar, 10.2% and 8.2%, we might expect the test to indicate compatibility. However, as shown on the bottom of Appendix D, Sheet 3, τ is calculated to be 30.0, which is greater than the Chi-square critical value of 18.3 at a 95% confidence level. In this example, the hypothesis that the CPI index is an acceptable alternative source of information for estimating long term trends for doctors' medical malpractice severity is rejected. The mixed estimate should not be calculated, i.e., the CPI index is not a statistically appropriate complement.

The major factor leading to the conclusion to reject compatibility is the small error variances in both models. Even though the estimated trends are similar, the two 95% confidence ellipses are sufficiently small such that we conclude that the CPI index will not provide a trend estimate compatible with our sample data.

Rejection of compatibility in this example is consistent with the non-intuitive result provided by the mixed estimate. Using mixed estimation would imply credibility of only

$$\frac{(1/s_u^2)}{(1/s_u^2) + (1/s_{v^2})} = 37.5\%$$

for our sample data of doctors' severity and a credibility weighted trend estimate of 8.9%, even though the model provides a good fit for our sample data. We could conclude that is appropriate to give our sample data full credibility and disregard the alternative information, or search for another complement.

Mixed estimation should not be used without computing the compatibility statistic, because mixed estimation will produce a credibility weighted trend estimate from two sources of information, regardless of whether the sources are related.

SUMMARY

Mixed estimation was developed by Theil and Goldberger as a generalized regression technique to allow alternative sources of information to be incorporated simultaneously into the model specification and estimation process. Mixed estimation can be used anytime that regression analysis is the indicated approach to the analysis. Because it is a very general procedure the formula for trend credibility can be generalized beyond the regression context for use with any type of prior information.

This econometric tool is recommended for use in the actuarial task of estimating trend and the credibility of trend. In its application to trend, mixed estimation provides advantages over some of the procedures currently in use. Unlike the ISO, NCCI, and other classical credibility approaches, mixed estimation does not require an arbitrary full credibility standard to be specified. Instead, the credibility implied by mixed estimation is proportional to the inverse of the variances of the regression lines, which is intuitively appealing and analogous to an empirical Bayesian approach.

An integral part of the mixed estimation process, the test of compatibility, is perhaps the most significant advantage of the procedure. Mixed estimation, or any other procedure, will never replace the actuarial judgement required in selecting the complement of the credibility. However, mixed estimation is the only method developed to date that provides the actuary with the ability to test whether the information selected as the complement of the credibility is appropriate to the problem at hand.

REFERENCES

[1] Durbin, J. "A Note on Regression When There is Extraneous Information About One of the Coefficients," Journal of the American Statistical Association, Vol. 48, 1953, p. 799.

[2] Insurance Services Office. Actuarial Service Circular AS-PA-89-20, New York, NY, ISO, 1989.

[3] Johnston, J. Econometric Methods (2nd Edition). New York, NY, McGraw-Hill, Inc. 1972.

[4] Maddala, G. S. Econometrics. New York, NY, McGraw-Hill, Inc. 1977.

[5] National Council on Compensation Insurance. Memorandum - Trending Procedure, New York, NY, NCCI, 1985.

[6] Neter, J. and Wasserman, W. Applied Linear Statistical Models. Homewood, Ill, Richard D. Irwin, Inc. 1974.

[7] Theil, H. "On the Use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, Vol. 58, 1963, p.401.

[8] Theil, H. and Goldberger, A. S. "On Pure and Mixed Statistical Estimation in Economics," International Economic Review, Vol. 2, 1961, p. 65.

[9] Van Slyke, O.E. "Credibility Weighted Trend Factors," PCAS LXVIII, p. 160.

[10] Venter, G. "Classical Partial Credibility with Application to Trend," PCAS LXXIII, p. 27.

APPENDIX A

ORDINARY LEAST SQUARES

This appendix provides a summary, in matrix notation, of the assumptions, parameter estimates, and properties of ordinary least squares (OLS) estimators. The purpose is to provide an outline of significant results. For derivation or mathematical proof of a particular formula, the reader is directed to one of the references [3], [4], or [5].

Assumptions. A linear relationship exists between a variable y and k explanatory variables x_1, x_2, ... x_k, with a disturbance term u. We have a random sample of n observations of the y and x's. Our linear model can be compactly stated in matrix notation as:

$$y = X\beta + u$$

where y = an n x 1 vector of the dependent variable

X = an n x k matrix of the explanatory variables

β = a k x 1 vector of the parameters to be estimated

u = an n x 1 vector of residuals.

To estimate the vector β, we need some assumptions about how the observations in the sample have been generated. Assume that:

1. The residuals are independently and identically distributed with mean 0 and variance σ^2_u:

$$E[u] = 0 \quad \text{and}$$

$$Var[u] = E[uu'] = \sigma_u^2 I$$

where I is an n x n identity matrix. This means the residuals are homoscedastic (have constant variance) and are pairwise uncorrelated.

2. The x's are nonstochastic and hence are independent of the residuals:

$$E(X'u) = 0$$

In repeated sampling, the sole source of variation in the y vector is the variation in the u vector, and the properties of our estimators and tests are conditional upon X.

3. The x's are linearly independent:

$$\text{rank } (X'X) = \text{rank } X = k \quad \text{and}$$

$$(X'X)^{-1} \text{ exists.}$$

The number of observations exceeds the number of parameters to be estimated and no exact linear relations exist between any of the variables in X.

OLS Estimation. Let β denote a k x 1 vector of estimates of the elements of β. Given our random sample of n observations,

$$y = X\beta + u \quad ,$$

where u is an n x 1 vector of residuals. The principle of least squares estimates β by minimizing u'u (the sum of the squared residuals) and produces

$$\hat{\beta} = (X'X)^{-1}X'y \,.$$

Properties of OLS Estimators. Under the above assumptions, $\hat{\beta}$ is the best linear unbiased estimator of β.

1. By best estimator, we mean $\hat{\beta}$ is the minimum variance estimator of β. The variance of $\hat{\beta}$ is given by

$$\mathrm{Var}[\hat{\beta}] = \sigma_u^2 (X'X)^{-1} \,.$$

It can be shown that this quantity is smaller than the variance of any other linear unbiased estimator of β.

2. Since $(X'X)^{-1}X'$ is a matrix of constants, the elements of $\hat{\beta}$ are linear functions of y.

3. $\hat{\beta}$ is unbiased because $E[\hat{\beta}] = \beta$.

Significance Tests and Confidence Intervals. To derive tests of significance and confidence intervals for β, we assume that the residuals, u, are normally distributed. With this additional assumption, it can be shown that the maximum likelihood estimate of β is equivalent to the OLS estimate.

1. Individual regression coefficient. To test the significance of an individual parameter estimate, the following test statistic

$$t = \frac{(\beta-\hat{\beta})}{(X'X)^{-1}(e'e)/(n-k)}$$

has the t distribution with (n-k) degrees of freedom. The 100(1-α) per cent confidence interval for B is given by

$$\hat{\beta} +/- t_{n-k,(1-\alpha/2)}(X'X)^{-1}(e'e)/(n-k) \quad .$$

2. More than one regression coefficient. When we consider all parameter estimates together, the following test statistic

$$F = \frac{(\beta-\hat{\beta})'(\beta-\hat{\beta})/k}{(X'X)^{-1}(e'e)/(n-k)}$$

has the F distribution with k and (n-k) degrees of freedom. The 100(1-α) per cent confidence region for β is given by

$$\hat{\beta} +/- F_{k,n-k,(1-\alpha/2)}(X'X)^{-1}(e'e)/(n-k)$$

For example, in a two parameter model, the above equation provides a 100(1-α) per cent elliptical confidence region.

Violations of Assumptions.

1. Heteroscedasticity. The residuals, u, do not have constant variance. The regression parameter estimates, $\hat{\beta}$, are still unbiased, but two problems arise. First, $\hat{\beta}$ is no longer an efficient (minimum variance) estimator. Second, the estimates of the variances of $\hat{\beta}_i$ are biased. Generalized least squares (see Appendix B) can provide more efficient estimators when heteroscedasticity occurs.

2. Serial Correlation. The residuals, u, are not independently distributed. The consequences of serial correlation are the same as for heteroscedasticity: $\hat{\beta}$ is still unbiased, but it is not efficient, and the estimated variance of β is biased and likely to be greatly understated. This will result in R^2, t, and F statistics that tend to be exaggerated. A commonly used test for serial correlation is the Durbin-Watson test.

3. $E(X'u) = 0$. The residuals and the independent variables are correlated. The OLS estimator $\hat{\beta}$ is no longer an unbiased estimator of β.

4. Multicollinearity. The independent variables, X, are not linearly independent. If an exact linear dependence exists among two or more of the x_i's, the OLS estimator $\hat{\beta}$ is impossible to determine because $(X'X)^{-1}$ does not exist. When some of the x's are very highly correlated, the precision of the individual parameter estimates deteriorates because the variance of $\hat{\beta}$ is very large. A regression line with high R^2 but insignificant t statistics for $\hat{\beta}$ is a common result due to multicollinearity.

5. Residuals are not normally distributed. OLS produces $\hat{\beta}$ that still is the best linear unbiased estimator of β, but all the tests of significance are not valid.

APPENDIX B

GENERALIZED LEAST SQUARES

In the development of the OLS estimator $\hat{\beta}$ in Appendix A, one of the assumptions was that the residuals, u, have constant variance and are not auto correlated:

$$E(uu') = \sigma_u^2 I \ .$$

Let us now relax this assumption and instead assume that

$$E(uu') = \Omega$$

where Ω is a symmetric positive definite matrix of order n, which defines the variances and covariances of the residuals. The Generalized Least Squares (GLS) estimator of β is given by

$$\hat{\beta}_{(GLS)} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \quad .$$

The variance of this estimator is given by

$$Var[\hat{\beta}_{(GLS)}] = \sigma^2 (X'\Omega^{-1}X) \quad .$$

As mentioned in Appendix A, it can be shown that

$$Var[\hat{\beta}_{(GLS)}] \leq Var[\hat{\beta}_{(OLS)}] \quad .$$

when $E(uu') = \Omega$ instead of $\sigma_u^2 I$. Given Ω, the usual significance tests and confidence intervals for $\hat{\beta}$ can be computed as well.

APPENDIX C

Data and Statistical Analysis for Example 1

Sheet 1: Average Homeowners Severity for State X

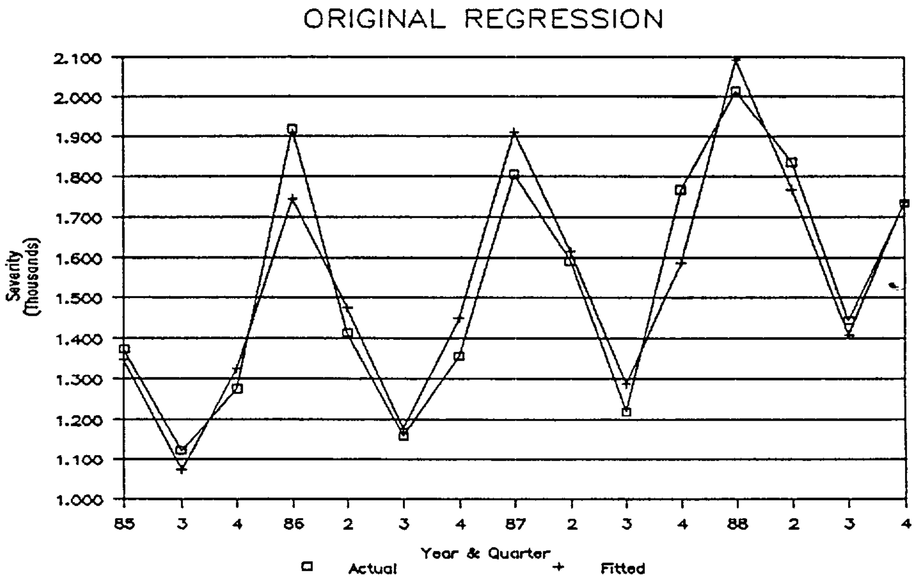
	[X] Matrix				[y] Vector				
	Time	Q2	Q3	Q4	Actual	ln[Sev.]	Fitted	Fitted	
Date	Index	Dummy	Dummy	Dummy	Sev.	[y]	[y]	Actual	[e]
85	1	1	0	0	1,374	7.225	7.206	1,347	0.019
3	2	0	1	0	1,122	7.023	6.979	1,074	0.044
4	3	0	0	1	1,274	7.150	7.189	1,324	(0.039)
86	4	0	0	0	1,918	7.559	7.465	1,745	0.095
2	5	1	0	0	1,414	7.254	7.296	1,475	(0.042)
3	6	0	1	0	1,158	7.055	7.069	1,175	(0.015)
4	7	0	0	1	1,357	7.213	7.279	1,450	(0.066)
87	8	0	0	0	1,805	7.498	7.555	1,910	(0.057)
2	9	1	0	0	1,590	7.371	7.387	1,614	(0.015)
3	10	0	1	0	1,219	7.106	7.160	1,287	(0.054)
4	11	0	0	1	1,767	7.477	7.370	1,587	0.108
88	12	0	0	0	2,014	7.608	7.645	2,091	(0.038)
2	13	1	0	0	1,835	7.515	7.477	1,767	0.038
3	14	0	1	0	1,444	7.275	7.250	1,409	0.025
4	15	0	0	1	1,733	7.457	7.460	1,737	(0.003)
								ave=	0.000
								s.d.=	0.004
								t=	0.000

Model: Severity = exp{Xb} + u', or ln[Severity] = y = Xb + u

Regression Output:

Constant		7.374		
Std Err of Y Est		0.064		
R Squared		92.1%		
No. of Observations		15		
Degrees of Freedom		10		
X Coefficient(s)	0.023	(0.191)	(0.440)	(0.253)
Std Err of Coef.	0.004	0.049	0.049	0.049
	[trend]	[Q2]	[Q3]	[Q4]
t-statistics	5.853	(3.914)	(9.049)	(5.188)
R-bar-squared		89.0%		
F-statistic		29.273		
Durbin-Watson [d]		2.578		
Trend exp{[trend]*4}-1:		9.5%		

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 1: Average Homeowners Severity for State X (cont)



Calculation of Seasonal Parameters

	Regression Estimates	exp{B}	off balanced
B1	0.000	1.000	1.232
B2	(0.191)	0.826	1.018
B3	(0.440)	0.644	0.793
B4	(0.253)	0.776	0.957
	average:	0.812	1.000

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 2: Average Homeowners Severity for Total United States

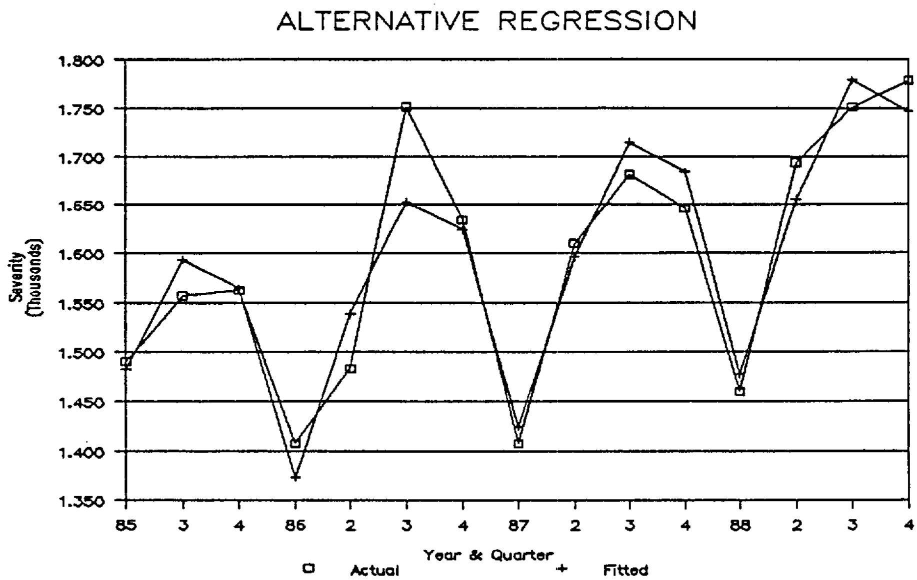
	[R] Matrix				[r] Vector				
	Time	Q2	Q3	Q4	Actual	ln[Sev.]	Fitted	Fitted	
Date	Index	Dummy	Dummy	Dummy	Sev.	[r]	[r]	Actual	[e]
85	1	1	0	0	1,490	7.307	7.302	1,483	0.005
3	2	0	1	0	1,557	7.350	7.374	1,593	(0.023)
4	3	0	0	1	1,563	7.354	7.356	1,565	(0.001)
86	4	0	0	0	1,407	7.249	7.225	1,373	0.024
2	5	1	0	0	1,484	7.302	7.339	1,538	(0.036)
3	6	0	1	0	1,751	7.468	7.410	1,653	0.058
4	7	0	0	1	1,634	7.399	7.392	1,623	0.006
87	8	0	0	0	1,407	7.250	7.262	1,425	(0.012)
2	9	1	0	0	1,610	7.384	7.375	1,596	0.009
3	10	0	1	0	1,682	7.427	7.447	1,715	(0.019)
4	11	0	0	1	1,646	7.406	7.429	1,684	(0.023)
88	12	0	0	0	1,460	7.286	7.298	1,478	(0.012)
2	13	1	0	0	1,694	7.435	7.412	1,656	0.023
3	14	0	1	0	1,752	7.468	7.484	1,779	(0.015)
4	15	0	0	1	1,779	7.484	7.466	1,747	0.018
								ave=	0.000
								s.d.=	0.002
								t=	0.000

Model: Severity = exp{Rb} + v', or ln[Severity] = r = Rb + v

Regression Output:

Constant		7.188		
Std Err of Y Est		0.029		
R Squared		90.6%		
No. of Observations		15		
Degrees of Freedom		10		
X Coefficient(s)	0.009	0.104	0.167	0.140
Std Err of Coef.	0.002	0.022	0.022	0.022
	[trend]	[Q2]	[Q3]	[Q4]
t-statistics	5.288	4.760	7.636	6.375
R-bar-squared		86.9%		
F-statistic		24.145		
Durbin-Watson [d]		2.751		
Trend exp{[trend]*4}-1:		3.7%		

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 2: Average Homeowners Severity for Total United States (cont)



Calculation of Seasonal Parameters

	Regression Estimates	exp{B}	off balanced
B1	0.000	1.000	0.901
B2	0.104	1.110	1.000
B3	0.167	1.182	1.064
B4	0.140	1.150	1.036
	average:	1.110	1.000

APPENDIX C

Data and Statistical Analysis for Example 1

Sheet 3: Average Homeowners Severity for State X, Deseasonalized and Indexed

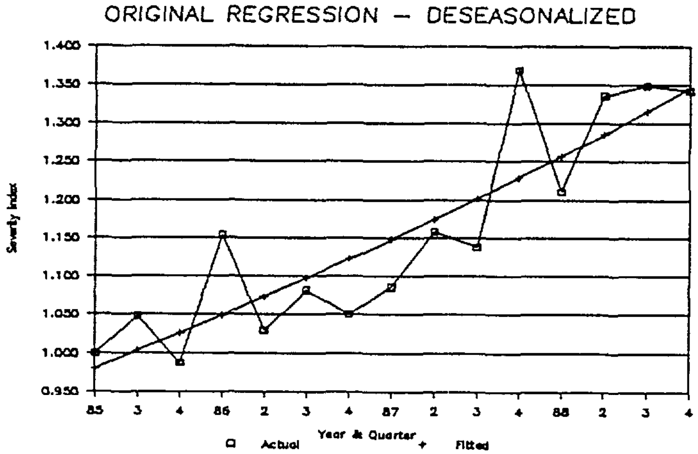
Date	[X] Time	Index	[y] Vector ln[Index] [y]	Fitted [y]	Fitted Index	[e]
85	1	1.000	0.000	(0.019)	0.981	0.019
3	2	1.048	0.047	0.003	1.003	0.044
4	3	0.987	(0.013)	0.026	1.026	(0.039)
86	4	1.154	0.143	0.048	1.050	0.095
2	5	1.029	0.029	0.071	1.074	(0.042)
3	6	1.082	0.079	0.094	1.098	(0.015)
4	7	1.051	0.050	0.116	1.123	(0.066)
87	8	1.085	0.082	0.139	1.149	(0.057)
2	9	1.158	0.146	0.161	1.175	(0.015)
3	10	1.139	0.130	0.184	1.202	(0.054)
4	11	1.369	0.314	0.207	1.230	0.108
88	12	1.211	0.191	0.229	1.258	(0.038)
2	13	1.336	0.290	0.252	1.286	0.038
3	14	1.349	0.299	0.274	1.316	0.025
4	15	1.342	0.295	0.297	1.346	0.000
					ave=	0.000
					s.d.=	0.004
					t=	0.047

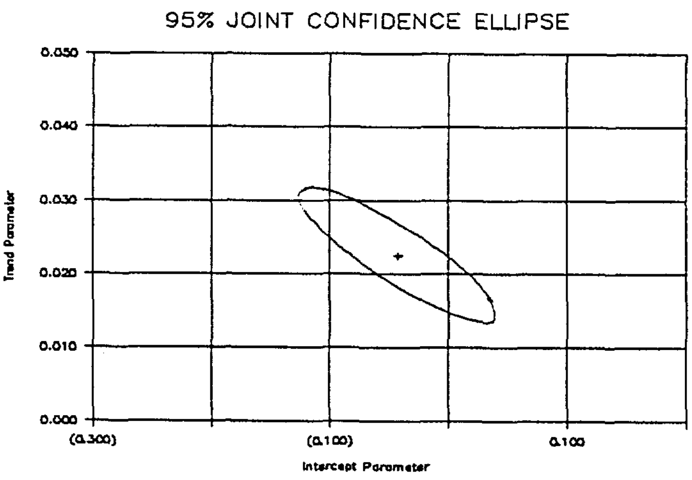
Model: Severity = exp{Xb} + u', or ln[Severity] = y = Xb + u

Regression Output:

Constant		(0.042)
Std Err of Y Est		0.056
R Squared		77.9%
No. of Observations		15
Degrees of Freedom		13
X Coefficient(s)	0.023	
Std Err of Coef.	0.003	
	[trend]	
t-statistics	6.771	
F-statistic		45.849
Durbin-Watson [d]		2.575
Trend exp{[trend]*4}-1:		9.5%

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 3: Average Homeowners Severity for State X, Deseasonalized and Indexed (cont)





APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 4: Average Homeowners Severity for Total United States, Deseasonalized and Indexed

Date	[R] Time	Index	[r] Vector ln[Index] [r]	Fitted [r]	Fitted Actual	[e]
85	1	1.000	0.000	(0.005)	0.995	0.005
3	2	0.981	(0.019)	0.004	1.004	(0.023)
4	3	1.012	0.012	0.014	1.014	(0.001)
86	4	1.048	0.047	0.023	1.023	0.024
2	5	0.996	(0.004)	0.032	1.032	(0.036)
3	6	1.104	0.099	0.041	1.042	0.058
4	7	1.058	0.056	0.050	1.051	0.006
87	8	1.048	0.047	0.059	1.061	(0.012)
2	9	1.080	0.077	0.069	1.071	0.009
3	10	1.060	0.058	0.078	1.081	(0.019)
4	11	1.066	0.064	0.087	1.091	(0.023)
88	12	1.087	0.084	0.096	1.101	(0.012)
2	13	1.136	0.128	0.105	1.111	0.023
3	14	1.104	0.099	0.114	1.121	(0.015)
4	15	1.152	0.141	0.124	1.132	0.018
					ave=	(0.000)
					s.d.=	0.002
					t=	(0.000)

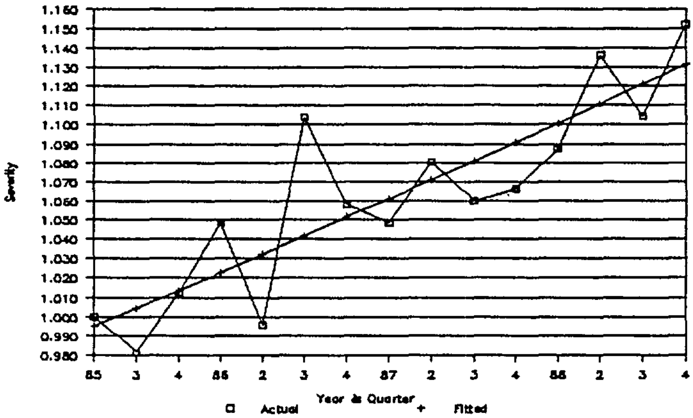
Model: Severity = exp{Rb} + v', or ln[Severity] = r = Rb + v

Regression Output:

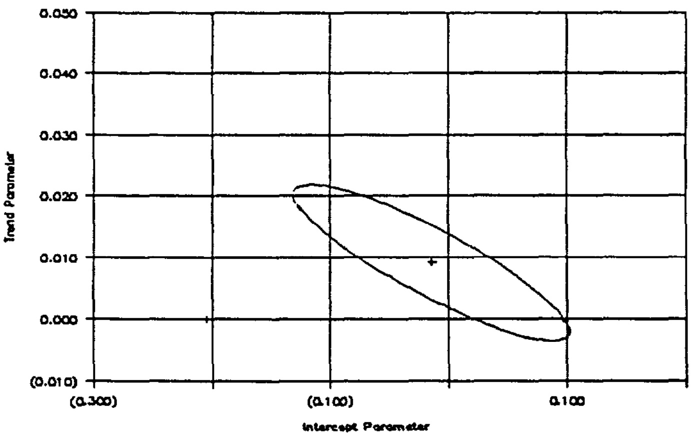
Constant		(0.014)
Std Err of Y Est		0.025
R Squared		74.2%
No. of Observations		15
Degrees of Freedom		13
X Coefficient(s)	0.009	
Std Err of Coef.	0.001	
	[trend]	
t-statistics	6.117	
F-statistic		18.707
Durbin-Watson [d]		2.751
Trend exp{[trend]*4}-1:		3.7%

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 4: Average Homeowners Severity for Total United States, Deseasonalized and Indexed

ALTERNATIVE REGRESSION — DESEASONALIZED

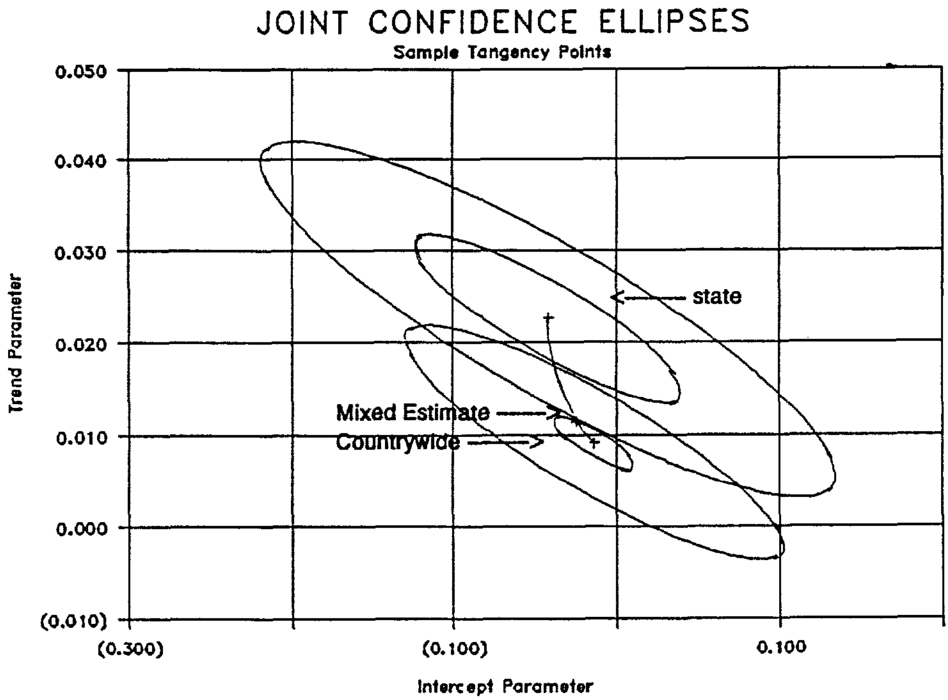


95% JOINT CONFIDENCE ELLIPSE



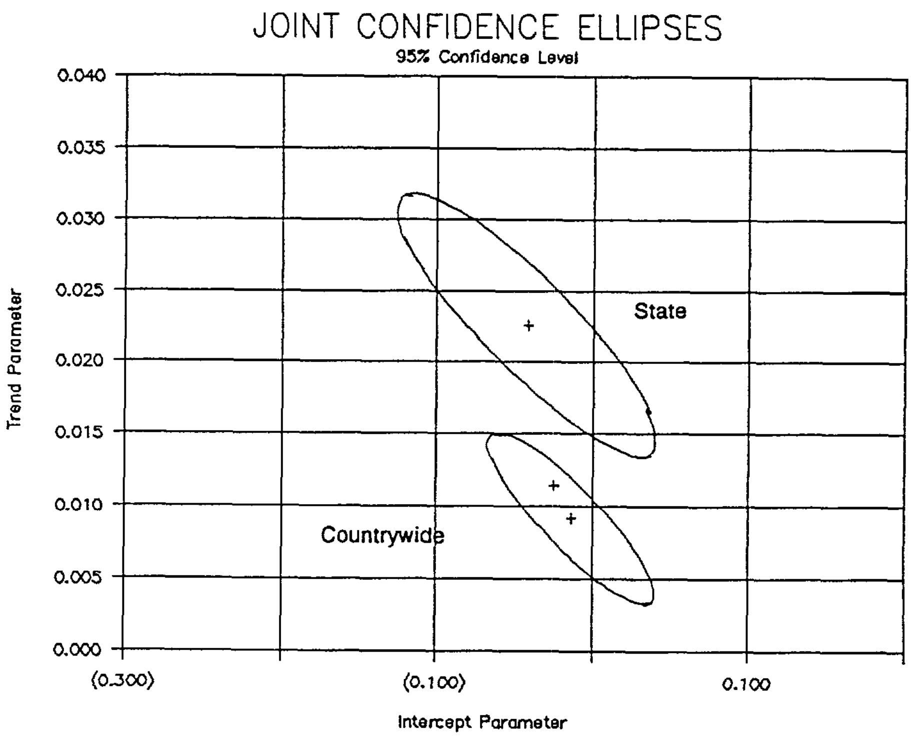
APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 5: Mixed Estimation

Mixed Estimation Parameters [b]:		constant =	(0.019)
		trend =	0.011
Credibility:	16.8%		
Estimated Average Annual Trend: exp(4*trend)-1 =			4.7%



Chi-Squared Statistic for Compatibility:	12.6
Degrees of Freedom:	15
Critical Value:	25.0
Disposition of Hypothesis:	Accept

APPENDIX C
Data and Statistical Analysis for Example 1
Sheet 5: Mixed Estimation *(cont)*



APPENDIX D
Data and Statistical Analysis for Example 2
Sheet 1: Countrywide Doctor Professional Liability Severity, Indexed

Date	[X] Time	Actual	[y] Vector Index	[y]	Fitted [y]	Fitted Index		[e]
1979	1	16,429	1.000	0.000	0.008	1.008		(0.008)
1980	2	18,145	1.104	0.099	0.106	1.111		(0.006)
1981	3	21,301	1.297	0.260	0.203	1.225		0.057
1982	4	22,312	1.358	0.306	0.300	1.350		0.006
1983	5	24,089	1.466	0.383	0.397	1.487		(0.014)
1984	6	25,091	1.527	0.423	0.494	1.639		(0.071)
1985	7	28,962	1.763	0.567	0.591	1.806		(0.024)
1986	8	34,150	2.079	0.732	0.688	1.990		0.043
1987	9	37,019	2.253	0.812	0.785	2.193		0.027
1988	10	39,298	2.392	0.872	0.882	2.417		(0.010)
							ave=	0.000
							s.d.=	0.004
							t=	0.000

Model : Severity = exp{Xb} + u', or ln[Severity] = y = Xb + u

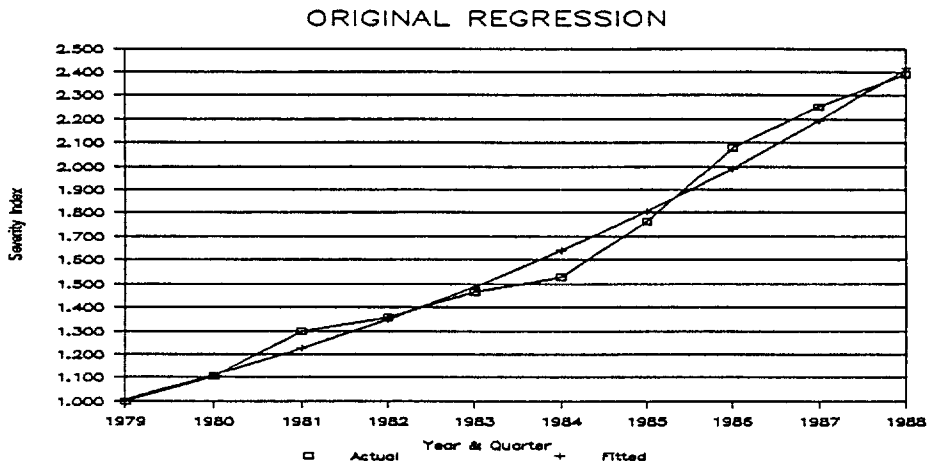
Regression Output:

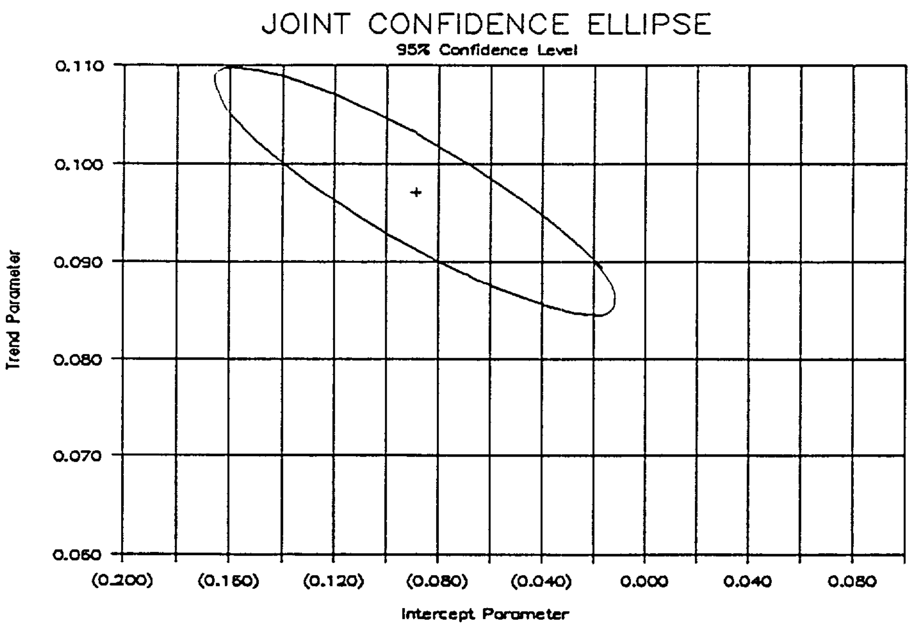
Constant		(0.089)
Std Err of Y Est		0.039
R Squared		98.5%
No. of Observations		10
Degrees of Freedom		8
X Coefficient(s)	0.097	
Std Err of Coef.	0.004	
	[trend]	
t-statistics	22.876	
F-statistic		130.8
Durbin-Watson [d]		1.561
Trend exp{[trend]}-1:		10.2%

APPENDIX D

Data and Statistical Analysis for Example 2

Sheet 1: Countrywide Doctor Professional Liability Severity, Indexed (cont)





APPENDIX D
Data and Statistical Analysis for Example 2
Sheet 2: Consumer Price Index, All Medical Care Items, Indexed

Date	[R] Time	[r] Vector Index	[r]	Fitted [r]	Fitted Actual	[e]
1979	1	1.000	0.000	0.044	1.045	(0.044)
1980	2	1.109	0.103	0.123	1.130	(0.019)
1981	3	1.229	0.206	0.201	1.223	0.005
1982	4	1.371	0.316	0.279	1.322	0.036
1983	5	1.491	0.399	0.358	1.430	0.042
1984	6	1.583	0.459	0.436	1.547	0.023
1985	7	1.681	0.519	0.515	1.673	0.005
1986	8	1.807	0.592	0.593	1.809	(0.001)
1987	9	1.927	0.656	0.671	1.957	(0.015)
1988	10	2.052	0.719	0.750	2.117	(0.031)
					ave=	0.000
					s.d.=	0.003
					t=	0.000

Model : Index = exp{Rb} + v', or ln[Index] = r = Rb + v

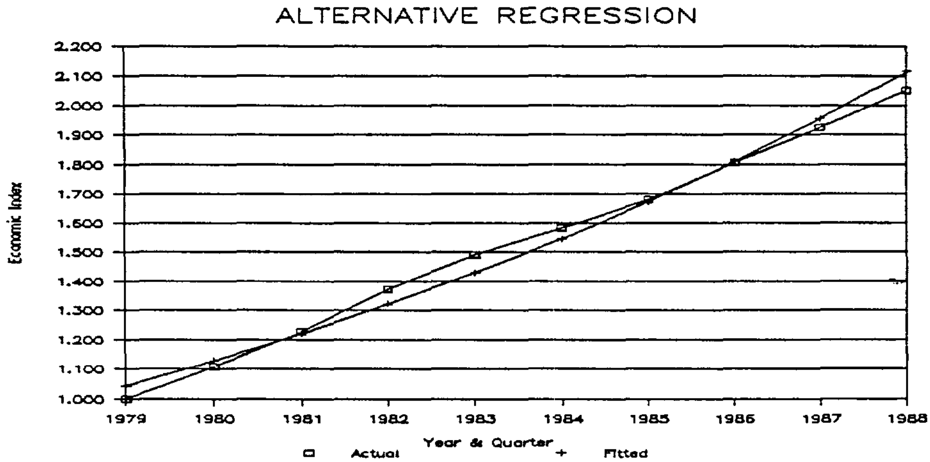
Regression Output:

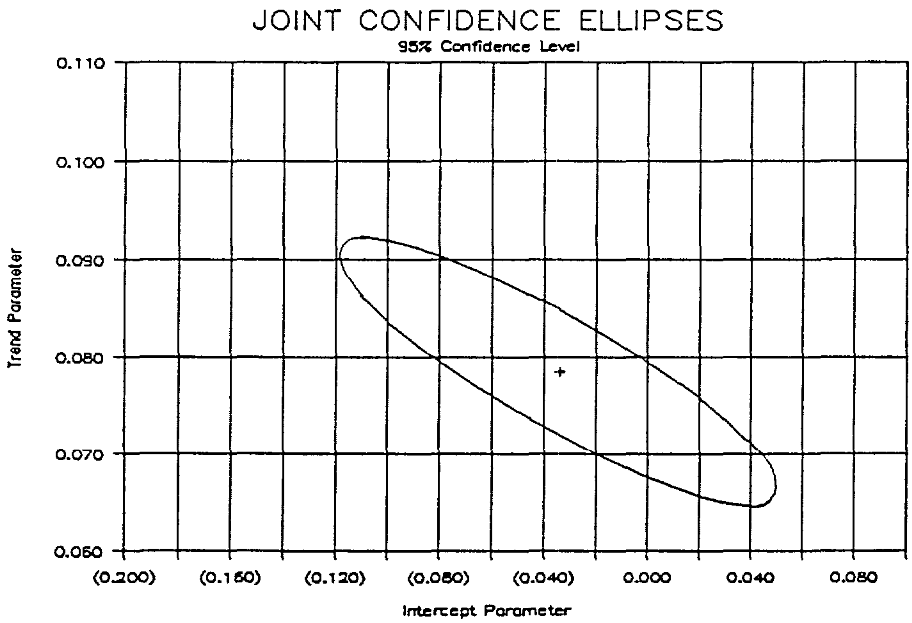
Constant		(0.034)
Std Err of Y Est		0.030
R Squared		98.6%
No. of Observations		10
Degrees of Freedom		8
X Coefficient(s)	0.078	
Std Err of Coef.	0.003	
	[trend]	
t-statistics	23.827	
F-statistic		141.9
Durbin-Watson [d]		0.487
Trend exp{[trend]}-1:		8.2%

APPENDIX D

Data and Statistical Analysis for Example 2

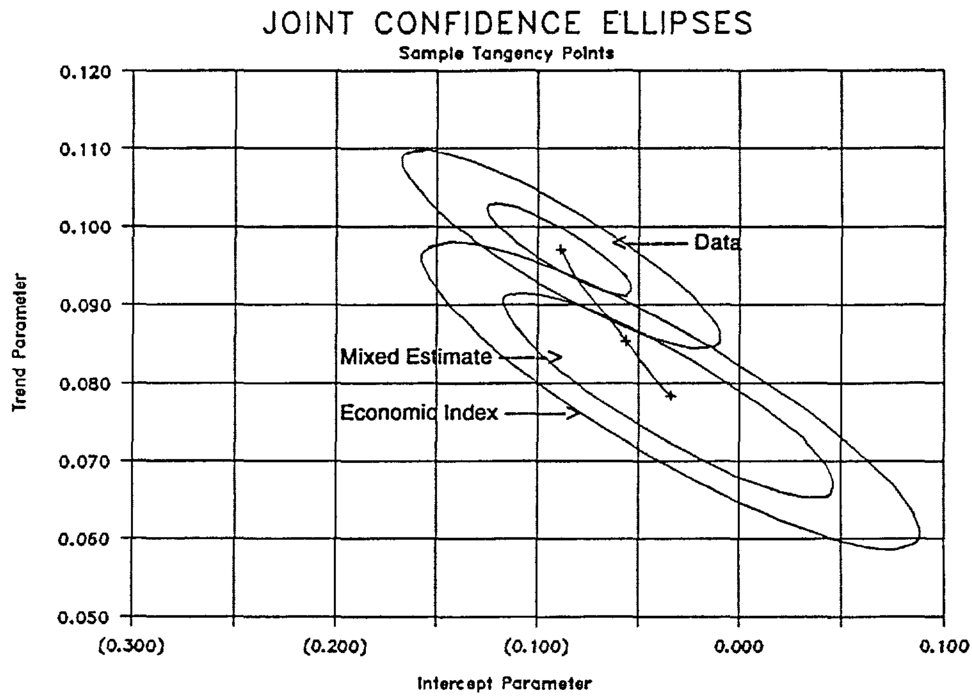
Sheet 2: Consumer Price Index, All Medical Care Items, Indexed (cont)





APPENDIX D
Data and Statistical Analysis for Example 2
Sheet 3: Mixed Estimation

Mixed Estimation Parameters:		constant =	(0.055)
		trend =	0.085
Credibility:	37.5%		
Estimated Average Annual Trend: exp(trend)-1 =			8.9%



Chi-Squared Statistic for Compatibility:	30.0
Degrees of Freedom:	10
Critical Value @ 95% Confidence Level:	18.3
Disposition of Hypothesis:	Reject

APPENDIX D

Data and Statistical Analysis for Example 2

Sheet 3: Mixed Estimation (cont)