Confidence bounds for the extremum determined by a quadratic regression

Joseph G. Hirschberg and Jenny N. Lye[1]

May 27, 04

Abstract

A quadratic function is frequently used in regression to infer the existence of an extremum in a relationship. Examples abound in fields such as economics, epidemiology and environmental science. However, most applications provide no formal test of the extremum. Here we compare the Delta method and the Fieller method in typical applications and perform a Monte Carlo study of the coverage of these confidence bounds. We find that unless the parameter on the squared term is estimated with great precision, the Fieller confidence interval may posses dramatically better coverage than the Delta method.

Key Words:

Inverted U-Shaped, turning point, inflexion point, Fieller method, Delta method, 1st derivative function.

---

[1] Department of Economics, University of Melbourne, Melbourne, Victoria 3010, Australia. j.hirschberg@unimelb.edu.au. The research carried out in this paper was partially funded by a grant from the Faculty of Economics and Commerce at The University of Melbourne. Comments and suggestions on an earlier version of this paper from participants at seminars held at The University of Melbourne, and Southern Methodist University are greatly appreciated. In particular we wish to thank Tom Fomby and Essie Maasoumi for their comments on this work.

1. **Introduction**

The specification of a relationship between a dependent and the independent variables in a regression are often not exactly defined by theory. In many cases it is proposed that the relationship is more complex than a linear form thus a more flexible function is hypothesised. Unfortunately, it is often beyond the scope of the study to consider the estimation of a purely nonparametric function so researchers rely on the use of a polynomial in the variable of interest to accommodate the nonlinearity. In addition, in a bid to reduce the number of new parameters introduced it is often the case that a quadratic is the polynomial chosen. An implication of the use of a quadratic is the existence of an extremum point in the relationship between the regressors and the dependent variable. This paper is concerned with the implications drawn from the estimation of the location of these points.

The estimation of relationships with polynomial forms can be found in many areas of empirical research. The function found when estimating a quadratic can be either a U-shaped or inverted U-shaped relationship. Common examples of such relationships can be found in the areas of: the Kuznets curve that proposes that the relationship between income inequality and percapita income is characterized by an inverted U-shaped curve (Kuznets 1955), the wage profile that displays the characteristics of the diminishing effect of increased experience on wages earned (Murphy and Welch, 1990), the Laffer curve which relates the level of national income to the rate of taxation (Hsing, 1996), the relationship between alcohol consumption and income that displays an increase in income until the level of consumption interferes with the ability to work (Berger and Leigh 1988, Hirschberg and Lye 2001), the economies of scale of a production technology - where it is assumed that production is subject to decreasing then increasing average costs of production (Thompson and Wolf 1993), in hedonic price models for houses in which the proximity to certain types of amenities (such as shopping facilities), can be viewed as having a detrimental impact if too close but an advantage if not close enough to generate congestion but close enough to be convenient. In these and many other cases, regressions have been run in which a continuous regressor has been fit as a quadratic. But more than just fitting the model the investigators have implied that the shape of the function defined by the parameter estimates has meaning – it shows the presence of a U-shaped or an inverted U-shaped relationship. For example, a search of the recent literature in Economics using the Econ Lit bibliography of journal articles (2002-03) indicates that 297 entries refer to U- or inverted U-shaped relationships in the abstract or title with many more cases in which these findings are only mentioned in the body of the paper (a finding that can be verified by searching a subset of journals for which the full text is searchable). However, in almost all of these applications these relationships are only referred to descriptively and rarely do they employ any form of statistical inference to formalize their assertions of the presence of the extremum point.

This paper proceeds as follows. First, we examine the origins of the quadratic specification. Then, we discuss the location and relevance of the extremum point of the quadratic specification. We then discuss the construction of confidence intervals for the maximum and minimum using the delta method and the Fieller method. Several empirical applications are presented using textbook data to highlight differences between the two methods. A simulation experiment is conducted to study the coverage of these confidence bounds for typical applications. Finally, conclusions are presented.

2. The Quadratic Regression Model

Following the discussion in Cramer (1969, pages 79-83), it can be shown that the use of a quadratic form can be justified as an approximation for a continuous function that is second order differentiable via a 2nd order Taylor series expansion. In general assuming that $y = f(x_1, x_2, \cdots, x_k)$ by applying the multivariate approximation obtained from a 2nd order Taylor Series expansion we get the specification:

$$y \approx f(\mathbf{x}_0) + \mathbf{g}\left(f(\mathbf{x}_0)\right)'\left(\mathbf{x} - \mathbf{x}_0\right) + \tfrac{1}{2}\left(\mathbf{x} - \mathbf{x}_0\right)'\mathbf{H}\left(f(\mathbf{x}_0)\right)\left(\mathbf{x} - \mathbf{x}_0\right) \qquad (1)$$

where $\mathbf{g}\left(f(\mathbf{x}_0)\right)$ is the gradient vector and $\mathbf{H}\left(f(\mathbf{x}_0)\right)$ is the Hessian matrix neither of which is known. The approximation given in (1) can be rewritten as:

$$
\begin{aligned}
y \approx & \left\{ f(\mathbf{x}_0) - \mathbf{g}\left(f(\mathbf{x}_0)\right)' \mathbf{x}_0 + \tfrac{1}{2}\mathbf{x}_0{}'\mathbf{H}\left(f(\mathbf{x}_0)\right)\mathbf{x}_0 \right\}_{1\times 1} \\
& + \left\{ \mathbf{g}\left(f(\mathbf{x}_0)\right)' + \mathbf{x}_0{}'\mathbf{H}\left(f(\mathbf{x}_0)\right) \right\}_{1\times k} \mathbf{x} + \tfrac{1}{2}\mathbf{x}' \left\{ \mathbf{H}\left(f(\mathbf{x}_0)\right) \right\}_{k\times k} \mathbf{x}
\end{aligned} \tag{2}
$$

where for an unknown function the value of the terms in the braces are unknown. This approximation can be written as a linear function of a set of unknown parameters and the levels of the ***x***'s. Making the following substitutions:

$\alpha = \left\{ f(\mathbf{x}_0) - \mathbf{g}\left(f(\mathbf{x}_0)\right)' \mathbf{x}_0 + \tfrac{1}{2}\mathbf{x}_0{}'\mathbf{H}\left(f(\mathbf{x}_0)\right)\mathbf{x}_0 \right\}_{1\times 1}$, $\mathbf{B} = \left\{ \mathbf{g}\left(f(\mathbf{x}_0)\right)' + \mathbf{x}_0{}'\mathbf{H}\left(f(\mathbf{x}_0)\right) \right\}_{1\times k}$, and

$\mathbf{\Pi} = \left\{ \tfrac{1}{2}\mathbf{H}\left(f(\mathbf{x}_0)\right) \right\}_{k\times k}$ the approximation in (2) is then rewritten as $y \approx \alpha + \mathbf{Bx} + \mathbf{x}'\mathbf{\Pi x}$ or

$y \approx \alpha + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=1}^{k} \pi_{ij} x_i x_j$. Thus we have a general specification for any function based on the generalization that the true function can be approximated by a 2nd order Taylor series expansion. This can be extended to show that any higher order Taylor series expansions are equivalent to higher order polynomial regressions. Note that the properties of regression based on the Taylor Series expansions commonly used in economics has been examined by White (1980) and Byron and Bera (1983).

3. The location and relevance of the extrema points in a polynomial specification

One characteristic of polynomial specifications is that the extrema points can be found by defining the first derivative function of the polynomial and finding those values of the regressor for which the function has a zero value. This is done by solving for the roots of a polynomial that is one degree less that the estimated polynomial which is defined by the first derivative function.

In the case of the quadratic, the regression would be of the form:

$$
y_t = \left(\gamma_0 + \sum_{j=1}^{k} \gamma_j z_{jt} \right) + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t \tag{3}
$$

where we assume that there are k linearly related regressors listed as the z_j and a single regressor entered in the quadratic form are given by x. We can define the first derivative function of y with respect to x as $\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x_t$ and we can solve this linear equation (polynomial of order 1) for the value of x at the extremum. The extremum is defined as $\theta = \frac{-\beta_1}{2\beta_2}$ where θ is either the maximum (when $\beta_2 < 0$) or the minimum (when $\beta_2 > 0$). The usual method of estimation is to define $\hat{\theta} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}$ where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimates of β_1 and β_2 in (3) respectively.

Once $\hat{\theta}$ is determined it is necessary to establish if this is a relevant value for this application. Most commonly one establishes whether $\hat{\theta}$ falls within the range of values that the regressor can take. All quadratic forms when estimated by regression will result in an extremum value, however $\hat{\theta}$ may be too distant from the range of feasible values to be meaningful in the context of the analysis. And since the estimate of the extremum point ($\hat{\theta}$) is a random variable, the question of feasibility becomes a probability question which can best be answered by construction of confidence intervals for the location of the extremum value.

Typically, when determining if a quadratic relationship is warranted one first determines if $\hat{\beta}_2$ is significantly different from zero or not. Thus the standard t-test of the null hypothesis that $\beta_2 = 0$ would indicate whether one could reject the null hypothesis of linearity. In the analysis below we demonstrate the importance of this test statistic for the determination of the location of the extremum.

4. Confidence Intervals for the Value of the Extremum

The extremum value from the quadratic regression specification is found from a ratio of the parameters. However, it is well known that when we construct an estimate of a ratio composed of terms that are random variables that we need to be wary of the finite probability that the random variable in the denominator can take values close to zero. If this occurs, the moments of the random variable defined by the ratio become undefined. The most well known example of this is the Cauchy distribution formed from the ratio of two independent standard normal random variables. In the present case the denominator is the estimate of the parameter on the squared regressor $(\hat{\beta}_2)$. Thus if this parameter is not significantly different from zero one would expect that

$\hat{\theta}$ has a distribution without finite moments. But if this was the case most certainly one would have rejected the possibility that the quadratic form was appropriate due to the inability to reject $\left[H_0 : \beta_2 = 0\right]$. For the most part in this paper we assume that the parameter on the squared term has been estimated with "sufficient precision" to assume that the random variable in the denominator has a very low probability of taking a value of zero, however it will be shown that the level of significance on this parameter may be of greater importance than usually assumed. By assuming this "sufficient precision" we imply that the ratio has a fairly well defined "pseudo mean" and "pseudo standard deviation" defined by truncating the usual integrals at some finite limit where the density of the ratio fall below some arbitrarily small level.

4.1 The Delta Method

The Delta method is a technique for estimating the variance of a nonlinear function of random variables. It involves the application of a first order Taylor series expansion to the nonlinear function in order to linearize the relationship. A formal treatment of the Delta method, as we refer to it here, is given by Rao (1973, 385-389). He shows that when T_n is a statistic based on a sample of size n for a parameter θ where $X = \sqrt{n}\left(T_n - \theta\right)$ and $X \underset{n\to\infty}{\sim} N\left[0, \sigma^2(\theta)\right]$ if a function $g(\bullet)$ is defined with first derivative $\dot{g}(\bullet)$ then for $Y = \sqrt{n}\left(g(T_n) - g(\theta)\right)$ we find that $Y \underset{n\to\infty}{\sim} N\left[0, \left[\dot{g}(\theta)\sigma(\theta)\right]^2\right]$. The extension of this method to functions of multiple variates is straight forward. If $\mathbf{X} = \sqrt{n}\left(\mathbf{T_n} - \boldsymbol{\theta}\right)$ and is of dimension k by 1 and $\mathbf{X} \underset{n\to\infty}{\sim} N\left[\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\right]$ where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the k by k matrix covariance matrix and if $\mathbf{G}(\bullet)$ is assumed to be a multivariate first order differentiable function that maps from k to p dimensions with a p by k matrix of first derivatives defined as $\dot{\mathbf{G}}(\bullet)$, then if $\mathbf{Y} = \sqrt{n}\left(\mathbf{G}(\mathbf{T_n}) - \mathbf{G}(\boldsymbol{\theta})\right)$ then $\mathbf{Y} \underset{n\to\infty}{\sim} N\left[\mathbf{0}, \left[\dot{\mathbf{G}}(\boldsymbol{\theta})\boldsymbol{\Sigma}(\boldsymbol{\theta})\left(\dot{\mathbf{G}}(\boldsymbol{\theta})\right)'\right]\right]$.

In order to apply the Delta method for the estimation of a variance of $\left(\hat{\theta} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right)$ requires using the multivariate Delta method for two variables. The appropriate test in this case is to determine if the extremum point estimated is feasible. This can be made by constructing confidence intervals for $\left(\hat{\theta} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}\right)$ in order to define the value of the regressor at which the extremum value is most likely to occur.

The Delta method requires the vector of partial derivatives of the extremum with respect to the parameters where $\frac{\partial\hat{\theta}}{\partial\beta_1} = \left(\frac{-1}{2\hat{\beta}_2}\right)$ and $\frac{\partial\hat{\theta}}{\partial\beta_2} = \left(\frac{\hat{\beta}_1}{2\hat{\beta}_2^2}\right)$. Thus the estimated variance of $\hat{\theta}$ is given by:

$$\tfrac{1}{4}\begin{bmatrix}\left(\frac{-1}{\hat{\beta}_2}\right) & \left(\frac{\hat{\beta}_1}{\hat{\beta}_2^2}\right)\end{bmatrix}\begin{bmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2\end{bmatrix}\begin{bmatrix}\left(\frac{-1}{\hat{\beta}_2}\right) \\ \left(\frac{\hat{\beta}_1}{\hat{\beta}_2^2}\right)\end{bmatrix} = \frac{\sigma_1^2\hat{\beta}_2^2 - 2\hat{\beta}_1\hat{\beta}_2\sigma_{12} + \hat{\beta}_1^2\sigma_2^2}{4\hat{\beta}_2^4} \tag{4}$$

where σ_1^2 is the variance of $\hat{\beta}_1$, σ_2^2 is the variance of $\hat{\beta}_2$ and σ_{12} is the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$. The t-statistic for testing the null hypothesis H_0: $\hat{\theta}=0$ is then given by the expression

$t_\theta = -\hat{\beta}_1\hat{\beta}_2\left(\sqrt{\hat{\sigma}_1^2\hat{\beta}_2^2 - 2\hat{\beta}_1\hat{\beta}_2\hat{\sigma}_{12} + \hat{\beta}_1^2\hat{\sigma}_2^2}\right)^{-1}$.

4.2. A Nonlinear Least Squares Estimation of the Extremum

An alternative approach to using the Delta method is to redefine the quadratic model in a nonlinear form where the extremum value is estimated as a parameter of the model. Define the extremum in the quadratic case as the parameter θ where $\theta = \left(\frac{-\beta_1}{2\beta_2}\right)$ and rewrite (3) using the result that $\beta_1 = -2\theta\beta_2$ to derive the form of the equation as:

$$y_t = \left(\gamma_0 + \sum_{j=1}^{k}\gamma_j z_{jt}\right) + \beta_2\theta(-2x_t) + \beta_2 x_t^2 + \varepsilon_t \tag{5}$$

When the estimation of (5) is done using a Gauss-Newton algorithm (as applied in most software packages) the estimate of the covariance for the parameters employs the outer product of the gradient of the nonlinear function evaluated at each of the observations. This covariance matrix can be shown to result in an exact equivalent estimate for the standard error obtained by using the Delta method for the extremum point $\hat{\theta}$. (This is shown in Appendix A.)

4.3. The Fieller method

The Fieller method (Fieller 1932) provides a general procedure for the construction of confidence limits for statistics defined as ratios. Define a ratio $\psi = \frac{v}{w}$, where v and w are assumed to be normally distributed random variables. Instead of testing the hypothesis $H_0 : \psi = \psi_0$ versus $H_1 : \psi \neq \psi_0$ Fieller proposes rewriting the hypothesis as $H_0 : v - \psi_0 w = 0$ versus $H_1 : v - \psi_0 w \neq 0$. Thus this hypothesis does not require the formation of the ratio and the test is based on the linear combination of the normally distributed random variables.

Zerbe (1978) defines a version of Fieller's method in the regression context where the ratio is defined in terms of linear combinations of the regression parameters from the same regression $H_0 : \psi = \frac{\mathbf{K}'\mathbf{B}}{\mathbf{L}'\mathbf{B}},\ H_1 : \psi \neq \frac{\mathbf{K}'\mathbf{B}}{\mathbf{L}'\mathbf{B}}$, where in the case of the linear model defined as $\mathbf{Y}_{T\times 1} = \mathbf{X}_{T\times k}\mathbf{B}_{k\times 1} + \varepsilon_{T\times 1},\ \varepsilon \sim (\mathbf{0}_{T\times 1}, \sigma^{\mathbf{2}}\mathbf{I}_{T\times T})$, the OLS estimators for the parameters are $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$ and $\hat{\sigma}^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/(T-k)$, and the vectors $\mathbf{K}_{k\times 1}$ and $\mathbf{L}_{k\times 1}$ are known constants. Under the usual assumptions we assume that the parameters estimates are asymptotically normally distributed via $\hat{\mathbf{B}} \sim N\left(\mathbf{B}, \sigma^2 \left(X'X\right)^{-1}\right)$.

The hypothesis can be rewritten as: $H_0 : \mathbf{K}'\mathbf{B} - \psi\mathbf{L}'\mathbf{B} = 0,\ H_1 : \mathbf{K}'\mathbf{B} - \psi\mathbf{L}'\mathbf{B} \neq 0$. The *t*-statistic of this test for a particular confidence interval of $1-\alpha$ can then be given as:

$$t_0 = \frac{\mathbf{K}'\hat{\mathbf{B}} - \psi\mathbf{L}'\hat{\mathbf{B}}}{\hat{\sigma}\left[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K} - 2\psi\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L} + \psi^{\mathbf{2}}\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\right]^{½}} \tag{6}$$

distributed as a *t* distribution with *T-k* degrees of freedom. Now by squaring both sides of the equation and inserting a particular *t* value defined as t_0 (i.e. t_0 = 1.96 for $\alpha = .05$ for a two tailed test and (*n-k*) > 120) then (14) can be written as a quadratic equation $a\theta^2 + b\theta + c = 0$ where $a = (\mathbf{L}'\hat{\mathbf{B}})^{\mathbf{2}} - t_0^2\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\hat{\sigma}^2$, $b = 2\left[t_0^2\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\hat{\sigma}^2 - (\mathbf{K}'\hat{\mathbf{B}})(\mathbf{L}'\hat{\mathbf{B}})\right]$, and $c = (\mathbf{K}'\hat{\mathbf{B}})^2 - t_0^2\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}\hat{\sigma}^2$. The two roots of the quadratic equation $\left(\psi_1, \psi_2\right) = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$, define the confidence bounds of the parameter value. In the case of the extremum of the quadratic regression equation in (3) in Appendix B.1 we obtain the following values for these terms: $a = 4\left(\hat{\beta}_2^2 - t_0^2\hat{\sigma}_2^2\right)$, $b = 4\left(\hat{\beta}_1\hat{\beta}_2 - t_0^2\hat{\sigma}_{12}\right)$, and $c = \hat{\beta}_1^2 - t_0^2\hat{\sigma}_1^2$. And in order to have real roots we require the condition that $b^2 - 4ac > 0$. As Fieller(1954)) and in the regression context Dufour (1997) note, for the roots to be finite $a \neq 0$, which implies $\frac{\hat{\beta}_2^2}{\hat{\sigma}_2^2} \neq t_0^2$. But $\frac{\hat{\beta}_2^2}{\hat{\sigma}_2^2} = \hat{t}_2^2$, the square of the estimated t-statistic on $\hat{\beta}_2$. Thus $a \neq 0$ is equivalent to $\hat{t}_2^2 \neq t_0^2$. If $\hat{t}_2^2 < t_0^2$ then $a < 0$ and the bounds are not meaningful – an example of this case is shown below in section 5.4. This case would imply that the quadratic term in the equation had a *t*-statistic of lower absolute value than implied by the a priori level of type I error such as $\alpha = .05$ set for the confidence bounds. We can redefine both *a* and *c* in terms of the estimated t-statistics for the parameters:

$a = 4\hat{\sigma}_2^2\left(\hat{t}_2^2 - t_0^2\right)$, and $c = \hat{\sigma}_1^2\left(\hat{t}_1^2 - t_0^2\right)$ and define the roots as:

$$\left(\tilde{\theta}_1, \tilde{\theta}_2\right) = \frac{t_0^2\hat{\sigma}_{12} - \hat{\beta}_1\hat{\beta}_2 \pm \sqrt{\left(\hat{\beta}_1\hat{\beta}_2 - t_0^2\hat{\sigma}_{12}\right)^2 - \hat{\sigma}_2^2\hat{\sigma}_1^2\left(\hat{t}_2^2 - t_0^2\right)\left(\hat{t}_1^2 - t_0^2\right)}}{2\hat{\sigma}_2^2\left(\hat{t}_2^2 - t_0^2\right)} \tag{7}$$

4.4. The confidence bounds of the first derivative function

This approach defines an expression for the first derivative of the estimated regression with respect to the regressor *x* and finds where it crosses zero. The confidence interval around this point is then used to make a probability statement concerning the feasibility of the existence of an extremum point.

For the quadratic model defined by (3) the first derivative of *Y* with respect to *x* defines a linear relationship for the slope of the quadratic function $\frac{\partial Y_t}{\partial x_t} = \beta_1 + 2\beta_2 x_t$. An estimate of the first derivative as a function of *x* can be plotted with a $1-\alpha$ confidence interval. When $\alpha = .05,\ t_{\alpha=.05} = 1.96$ bound given by:

$$\mathrm{CI}\left(\tfrac{\partial Y_t}{\partial x_t}\right) = \left(\hat{\beta}_1 + 2\hat{\beta}_2 x_t\right) \ \pm\ 1.96\sqrt{\left(\hat{\sigma}_1^2 + 4x_t\hat{\sigma}_{12} + 4x_t^2\hat{\sigma}_2^2\right)} \tag{8}$$

An estimate of the extremum value $x_t = \hat{\theta}$ is found by solving $\hat{\beta}_1 + 2\hat{\beta}_2\hat{\theta} = 0$. Similarly, the bounds defining the 95% confidence interval on $\hat{\theta}$ are found by solving

$$\left(\hat{\beta}_1 + 2\hat{\beta}_2\tilde{\theta}\right) \pm 1.96\sqrt{\left(\hat{\sigma}_1^2 + 4\tilde{\theta}\hat{\sigma}_{12} + 4\tilde{\theta}^2\hat{\sigma}_2^2\right)} = 0 \qquad (9)$$

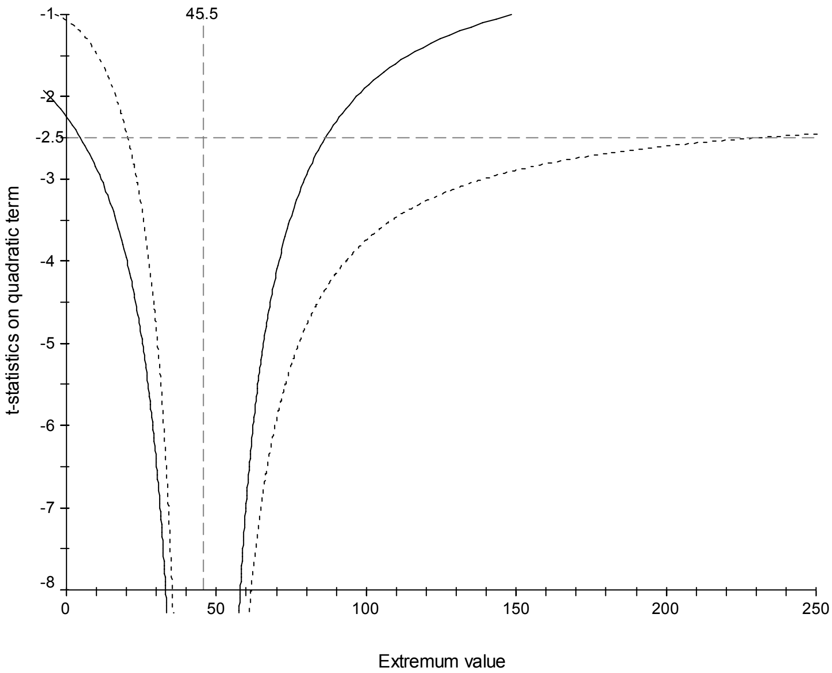
In Appendix B.2 we show that by squaring both sides of (9) we can show that the bounds on the first derivative function are equivalent to solving the roots of the same quadratic equation ($a\tilde{\theta}^2 + b\tilde{\theta} + c = 0$) we defined for the Fieller method, where *a*, *b* and *c* are defined as above in Section 4.3. Hence the confidence limits found in this way for the quadratic specification are identical to those found using the Fieller method described above. The advantage to this approach is that we can plot the first derivative function and the confidence bounds so that we can view what occurs near zero. In the examples shown in Section 5 we plot the first derivative function and the confidence bounds in every case.

This method can be generalized to any regression such as higher order polynomials, where the expression for $\frac{\partial Y_t}{\partial x_t}$ can be written as a linear function of the regression parameters. Although the solution may require searching when roots are not available from solutions to polynomials.

4.5 The relationship between the Fieller and the Delta method confidence bounds

It can be shown that the Fieller confidence bounds and the Delta confidence bounds are most different when the significance of the parameter on the squared regressor approaches the significance level set for the confidence interval (when $\left(\hat{t}_2^2 - t_0^2\right)$ is the smallest) and they converge as the significance level of the parameter on the squared term increases. In Figure 1 we plot the two sets of confidence bounds for the school test example described in Section 5.1. In this case an inverted U-shaped curve is found with an extremum. This plot shows how the 95% confidence bounds change as the standard error of the equation is varied so that the *t*-statistic on the squared regressor ($\hat{t}_2$) ranges from -2 to -8 and $t_0 = 1.96$. The parameter estimates used are those reported in Table 1 and we also use the same design matrix from this regression. The extremum in this case is estimated as 45.511 and a line is plotted at this point. For the Fieller interval the upper bound approaches ∞ as the absolute value of the *t*-statistic on the parameter estimate for the denominator parameter (the parameter on the squared term $\hat{\beta}_2$) approaches -2. In contrast, the confidence interval generated from the Delta method is finite and symmetric at all values of the t-statistic.

Figure 1 The confidence bounds on the extremum with varying values of the t-statistic on $\hat{\beta}_2$. The dotted lines indicate the Fieller interval and the solid lines show the Delta bounds.



From Figure 1 we note that the most dramatic differences between the confidence bounds in this case occur when the absolute value of the t-statistic on the squared term falls between 2 and 6 ($6 < |t_2| < 2$). The predominant feature of the difference between these two bounds is the asymmetry of the Fieller bounds versus the application of the Delta method. In addition, we note that the Delta method continues to imply upper bounds on the location of the extremum even when the *t*-statistic indicates that a significant mass of the distribution of the denominator is located at zero which would imply that the location of the extremum is indeterminate or that it does not end which would indicate the existence of a plateau instead. We also note that the Fieller lower bound would indicate that even with a absolute value of the t-statistic of less than 2 we could reject the one sided hypothesis that the extremum has a lower bound that is less than or equal to 0 at the 97.5% level (which is equivalent to the test for the existence of a plateau). This implies that if one were interested in establishing the existence of the lower bound of a plateau it may be sufficient to find a t-statistic on the quadratic term ($|t_2|$) that is much less than the usual cut off value of 2. From Figure 1 we see that the value of $|t_2|$ where the Fieller lower bound cut the zero axis is approximately 1.05.

5. Empirical Applications

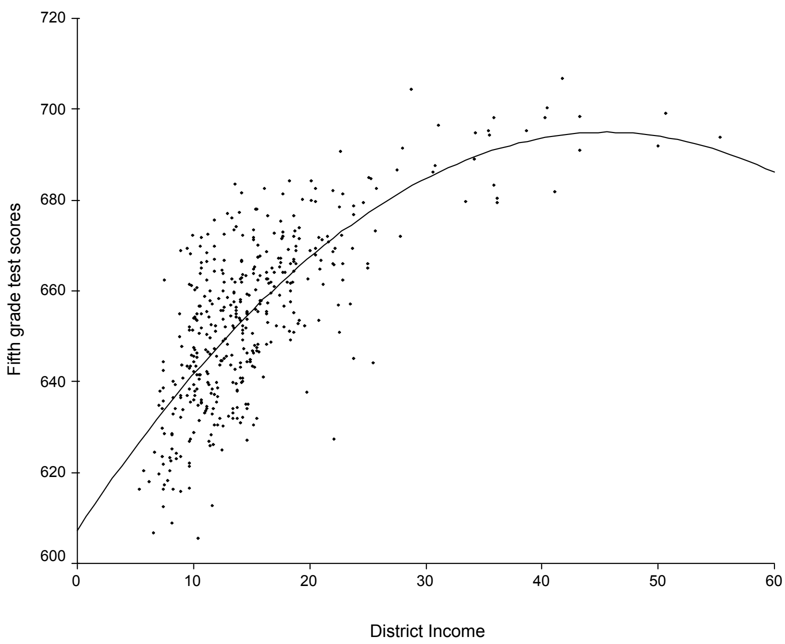
To illustrate differences between confidence limits calculated from using the Delta method and from those calculated from using the Fieller method we look at some standard textbook examples of estimating quadratic functional forms. We find the extremum value and the bounds implied by both methods as well as plotting the first derivative function and confidence bounds for each example.

5.1. The Californian Test Score Data Set

The data set taken from Stock and Watson (2003) contains data for 1998 for 420 school districts. The data set contains data on fifth grade test scores (y) and the average annual per capita income in the school district ("district income") measured in thousands of 1998 dollars (x) The median district income is 13.7 (that is $13,700 per person) and it ranges from 5.3($5,300 per person) to 55.3 ($55,300 per person).

A scatter plot of fifth-grade test scores (x) against district income (y) presented in Figure 2 indicates some curvature in the relationship between the two. To allow for this curvature test scores can be modelled as a function of income and the square of income. In this case one would be interested to know if there was an income at which the scores stopped changing. Is there an optimal income level?

Figure 2 Fifth-grade test scores (x) against district income (y)



OLS estimates of the relationship

$$y_t = \gamma_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t; \qquad t = 1, \ldots 420 \tag{10}$$

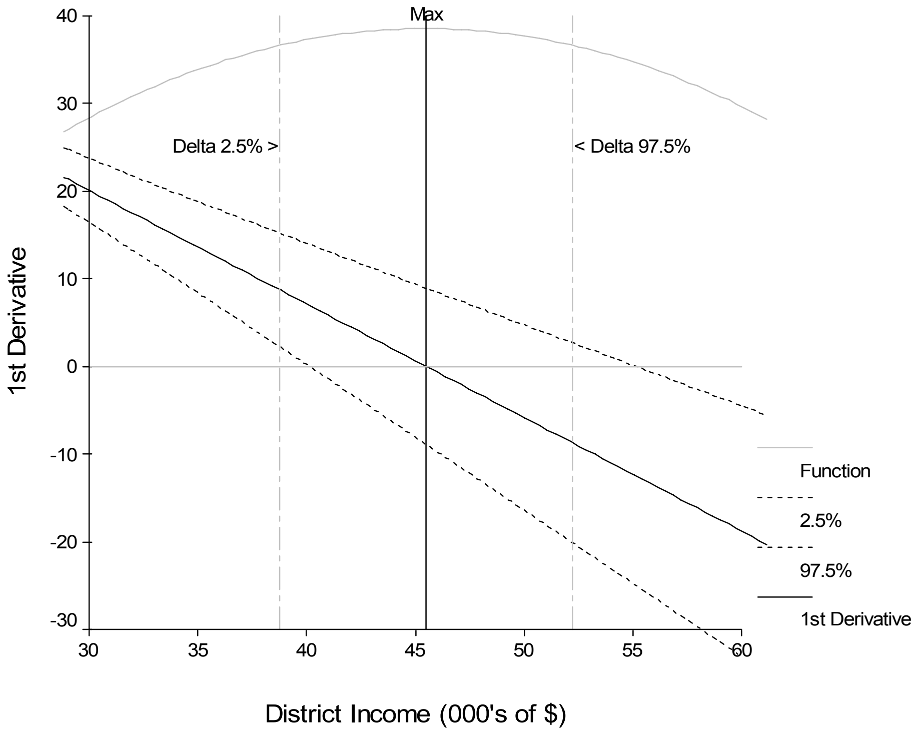
as computed by Eviews 4.1 software are given in Table 1.

Table 1 Regression of fifth grade test scores and district income.

Observations: 420				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	607.3017	3.046219	199.3624	0.0000
X	3.850995	0.304262	12.65685	0.0000
X^2	-0.042308	0.006260	-6.758474	0.0000
R-squared	0.556173	Mean dependent var	654.1565	
Adjusted R-squared	0.554045	S.D. dependent var	19.05335	
S.E. of regression	12.72381			

The *p*-value on the squared income term is significant with a t-statistic value of -6.76 indicating that the null hypothesis that the population regression function is linear can be rejected against the alternative that it is quadratic. The maximum of the quadratic specification occurs at a value of x = 45.511 (that is, $45,511 per person) with a standard error estimated via the Delta method as 3.438. A 95% confidence interval using the Delta method would then be given by 38.771 to 52.251. Using the Fieller method the corresponding 95% confidence interval would be given by the bounds 40.218 to 54.881. These confidence intervals are illustrated in Figure 3 from which it is seen that the confidence interval associated with the Fieller method unlike the Delta method is not symmetric around the maximum value.

Figure 3 The plot of the 1st derivative and the 95% confidence bounds.



5.2. American Electric Generating Companies

Nerlove(1963) analysed the production costs of 145 American electric generating companies. Using the version of Nerlove's data made available by Berndt (chapter 3, 1991) on *Costs* measured in millions of 1955 dollars and *Output* measured in kilowatt hours we estimate the Log of average cost as a function of total product defined as follows:

$$y_t = \gamma_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t; \quad t = 1, \ldots 145 \tag{11}$$

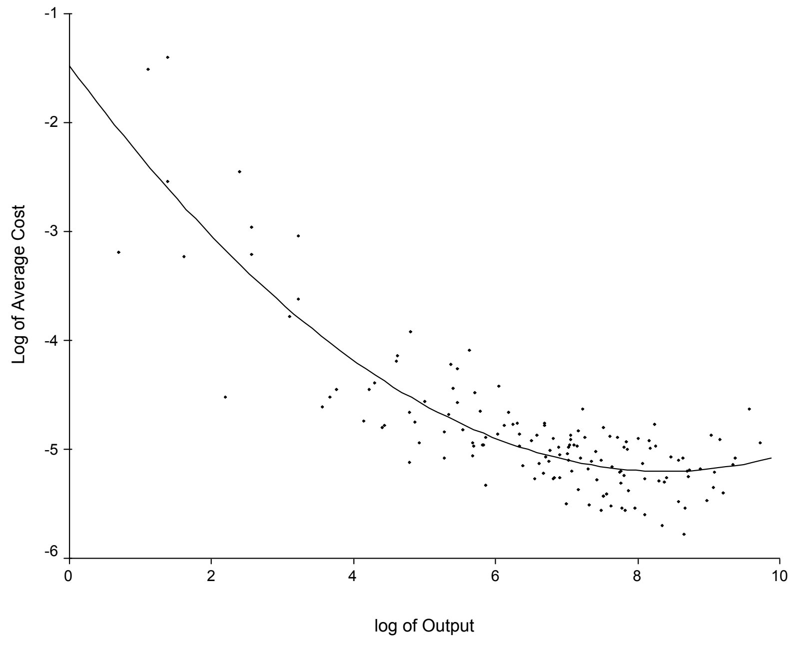
where $y = \log\left(\text{Cost}/\text{Output}\right)$ and $x = \log(\text{Output})$. Note that the interpretation of this specification is more straight-forward than the equivalent regression of the log of output in quadratic form on the log of costs.

A scatter plot of x against y is illustrated in Figure 4. The median value of x occurs at 7.011 ranging from a value of 0.693 to a value of 9.724. The curvature shown in Figure 4 indicates that the expectation in estimating (11) would be to identify a minimum value of x. In this case the location of the extreme of this quadratic indicates the size of the electricity generator that results in economies of scale in production. The confidence bound on this value can be used to determine the minimum size plant that is subject to economies of scale and the maximum plant that is not subject to diminishing returns to scale.

Table 2 Regression of log average cost as a function of log electricity output.

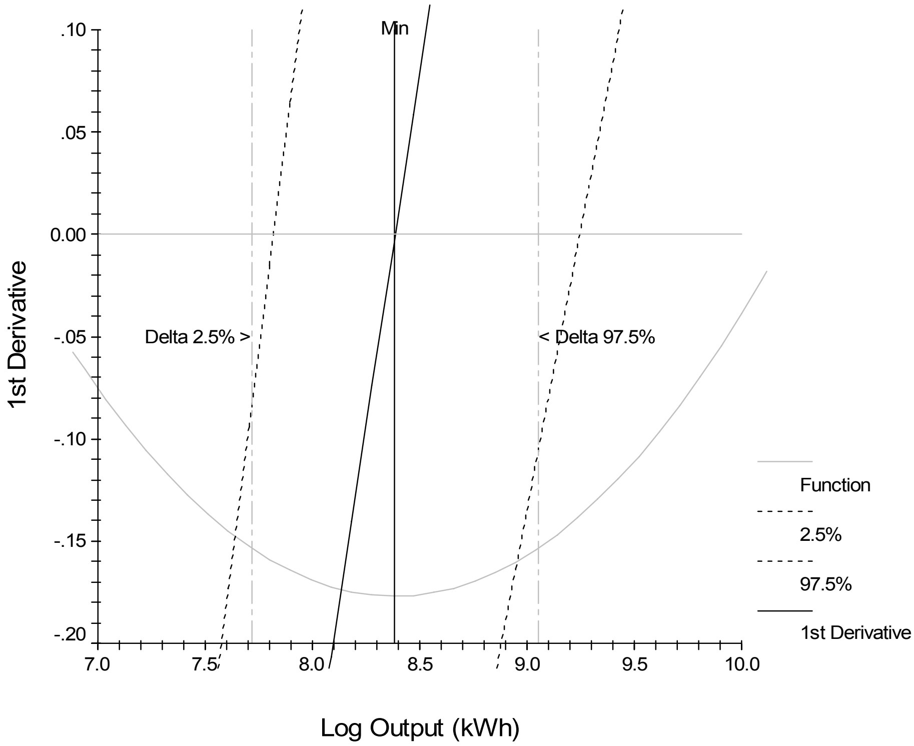
Observations: 145				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.476669	0.192806	-7.658846	0.0000
X	-0.888496	0.069078	-12.86221	0.0000
X^2	0.052983	0.005978	8.863379	0.0000
R-squared	0.762331	Mean dependent var		-4.831988
Adjusted R-squared	0.758983	S.D. dependent var		0.700324
S.E. of regression	0.343813			

Figure 4 A scatter plot of the log of average cost versus the total electricity output.



The estimation results are reported in Table 2. The p value on the squared term of x is significant with a t-statistic value equal to 8.863. The minimum value of x is equal to 8.385. Applying the Delta method to obtain a 95% confidence interval gives the range 7.717 to 9.051. Applying the Fieller method gives a 95% confidence interval ranging from 7.827 to 9.224. These confidence intervals are illustrated in Figure 5. In this example the Delta and Fieller approach give very similar bounds which is not surprising since the coefficient on the squared term of X is highly significant. However the tendency is still to estimate confidence intervals that are both higher for the Fieller approach than for the Delta approach.

Figure 5 1st derivative of the scale economies function estimated from the Nerlove data.



5.3. Birth Weights and Prenatal Visits

Using data from Wooldridge (2003) on birth weight of 1651 children and factors that potentially affect birth weight the following model was estimated:

$$\text{Log(bwght)} = \gamma_0 + \gamma_1\text{npvis} + \gamma_2\text{npvis}^2 + \gamma_3\text{cigs} + \gamma_4\text{fage} + \gamma_5\text{male} + \text{u} \qquad (12)$$

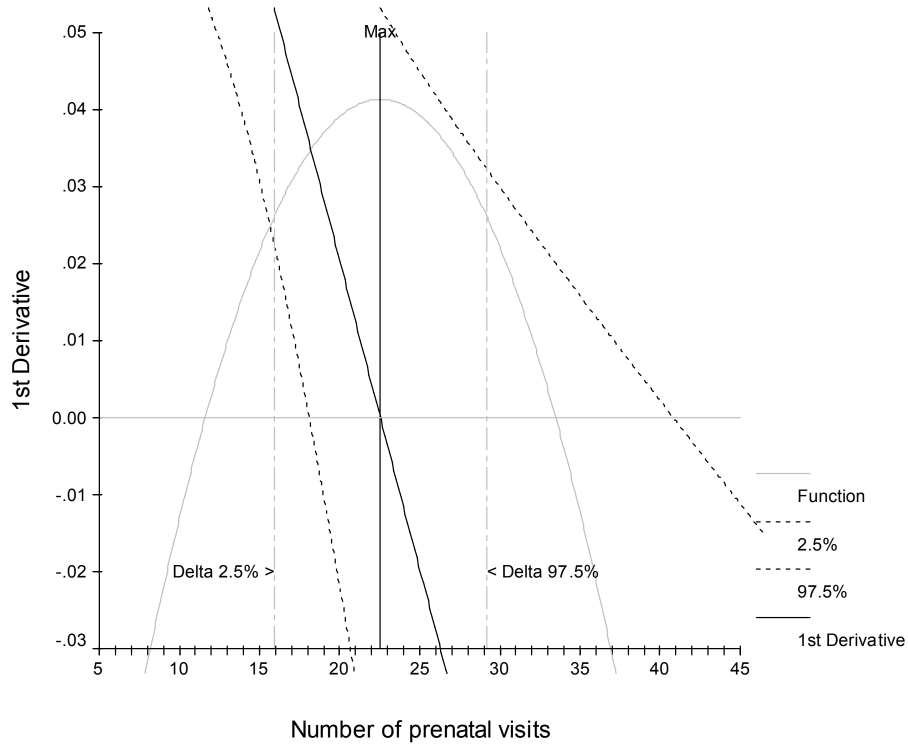
where *bwght* is the birth weight in grams; *npvis* is the total number of prenatal visits; *cigs* is the average number of cigarettes per day; *fage* is the father's age in years and *male* is a dummy variable that takes the value 1 if the baby is a male. The variable *npvis* has a median value of 12 and a minimum value of 0 and a maximum value of 40.

Table 3 Regression results for the Birth weight Example

Observations: 1651				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
NPVIS	0.015591	0.003701	4.212386	0.0000
NPVIS^2	-0.000346	0.000120	-2.887333	0.0039
CIGS	-0.002440	0.001170	-2.085741	0.0372
FAGE	0.002093	0.000867	2.414580	0.0159
MALE	0.022855	0.009772	2.338860	0.0195
C	7.911528	0.038046	207.9455	0.0000
R-squared	0.027325	Mean dependent var		8.116947
Adjusted R-squared	0.024368	S.D. dependent var		0.200724
S.E. of regression	0.198263			

The estimation results are reported in Table 3. The *t*-statistic on the squared term (*NPVIS^2*) is significant however not as significant as the previous examples. The number of prenatal visits that maximises the log of the birth weight is 22.54. The corresponding 95% confidence interval calculated using the Delta method ranges from 15.926 to 29.170 whereas the 95% confidence interval calculated using the Fieller method ranges from 18.098 to 40.822. In Figure 6 a plot of the first derivative method is illustrated where it is clearly evident that the difference between the two intervals is that the Fieller method is asymmetric around the value of the maximum with a very skewed upper bound. In this case the interpretation of these results would indicate that the minimum number of prenatal visits does not vary that much between the methods – however the number of prenatal visits that is detrimental to birth weight is either 29 or 41 depending on the limit chosen – a major difference – but still a troubling finding – do prenatal visits actually diminish health?

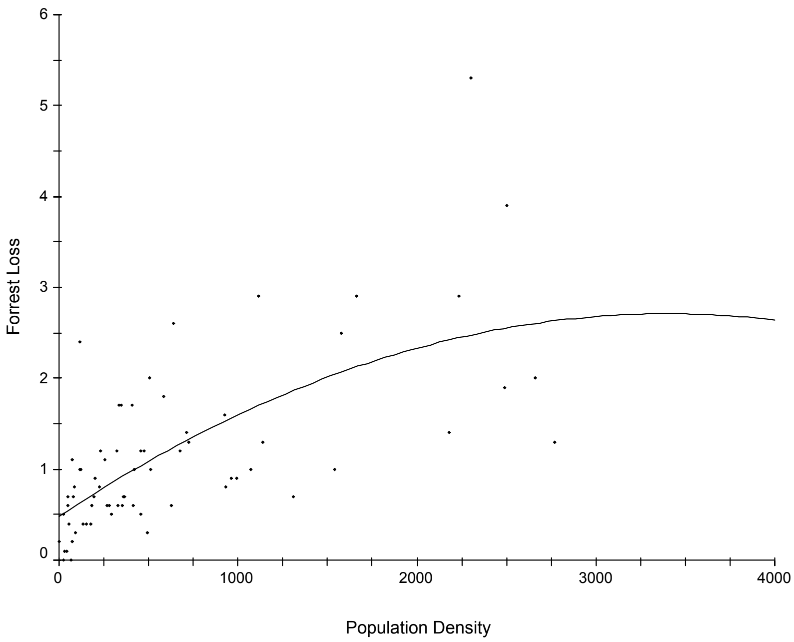
Figure 6 The Plot of the 1st derivative function and a 95% confidence bound.



5.4. Deforestation Example

Figure 7 is a scatter plot of population density (x) and deforestation (y) for 70 tropical countries using data from Koop(2000). The scatter plot illustrates that there is no obvious nonlinear relationship between the two variables although there are a number of outlying observations, the minimum value of x is 0.89 with a maximum value of 2769 and a medium value of 354. In this case the search for a maximum would be evidence that once a country reached a certain density the level of deforestation would start to abate. Given the nature of the data this plot would indicate that there are probably other factors that influence the level of deforestation other that population density.

Figure 7 Population density (x) and deforestation (y)



The OLS results of estimating a model that allows for a quadratic in x

$$y_t = \gamma_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t; \qquad t = 1, \ldots 70 \qquad (13)$$

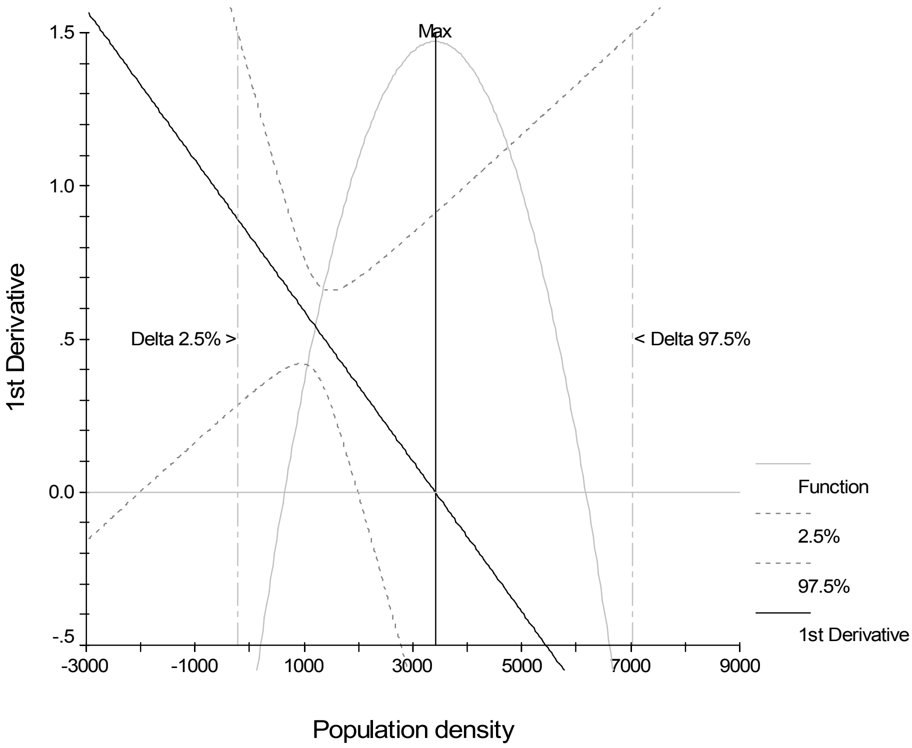
are given in Table 4 from which it can be concluded that the t-statistic on the squared term is not significant. The value of the population density that maximizes deforestation is 3406. A 95% confidence interval for this maximum using the Delta method can be constructed with bounds -226. to 7038. While this confidence limit is wide it however is still defined by finite bounds.

Table 4 Regression results for Deforestation example.

Observations: 1 70				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.480017	0.150160	3.196715	0.0021
X	0.001309	0.000406	3.221703	0.0020
X^2	-1.92E-07	1.60E-07	-1.198694	0.2349
R-squared	0.446352	Mean dependent var		1.138571
Adjusted R-squared	0.429826	S.D. dependent var		0.928189
S.E. of regression	0.700874			

Using the Fieller method to construct the 95% confidence limit for the maximum value we find a case where the significance level estimated for the squared parameter is less than the confidence level used to construct the Fieller interval thus $(\hat{t}_2^2 - t_0^2) < 0$. In this case the plot of the confidence bounds on the 1st derivative function makes the situation quite clear. When using the formula for the upper bound one actually obtains the point where the confidence bound on the lower bound of the 1st derivative function cuts the zero axis again at the lower level of *x* and the lower bound is the only bound for the extremum that can be determined. In this case the upper bound is not defined. This is illustrated in Figure 8.

Figure 8 The 1st derivative function and the Fieller and Delta method confidence bounds for a case when $(\hat{t}_2^2 - t_0^2) < 0$.



In this case we find that the first derivative result shows quite plainly that the data do not support the hypothesis of a finite extremum in this regression. Although we cannot reject the hypothesis that the quadratic term in the regression is zero at any level less than 23.49%, the Fieller interval indicates that we can reject the hypothesis that the lower bound of a plateau is less than 2000 at the 5% level. This would be consistent with the hypothesis that the deforestation is monotonically related to population density and that at a certain point this relationship stops. Thus the existence of an extremum point at which the positive relationship between deforestation and population density turns to become a negative relationship between deforestation and population density is not sustained by the data.

6. Monte Carlo Experiments to compare the coverage of the Fieller and Delta Confidence bounds.

Two series of Monte Carlo experiments were performed to investigate the properties of the confidence intervals generated by the Fieller and the Delta method for constructing the confidence bounds of the extremum value. The model used in both experiments is a quadratic in a single regressor with an intercept: $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$, where $\varepsilon_t \sim N\left(0,\sigma^2\right)$. For the purposes of this experiment we use two series based on the actual data for x from examples 5.1 and 5.2. In these experiments we change the location of the extremum by varying the values of β_0 and β_2 such that the mean of the dependent variable remains constant and set σ^2 so that the t-statistic on the estimate of β_2 would remain at 1.96. However for any particular simulation we may find cases were $(\hat{t}_2^2 - t_0^2) < 0$ thus we removed these from our sample of simulations this means that the average value for $\hat{t}_2^2$ is greater than $\left(1.96\right)^2$.

To change the location of the extremum point (θ) we vary the value of β_2 which we referred to as $\tilde{\beta}_2$ using the relationship $\tilde{\beta}_2 = \frac{-\hat{\beta}_1}{2\tilde{\theta}}$ where $\tilde{\theta}$ is the new location of the extremum. The value of the mean of the dependant variable ($\bar{y}$) is kept constant by defining the intercept (β_0) via the relationship $\tilde{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} - \tilde{\beta}_2\overline{x^2}$. To insure that the regressions simulated would not result in significance levels for β_2 that systematically vary with the value of θ we set the t-statistic to be approximately 1.96 by setting the variance of the pseudo-random errors generated where $\tilde{\sigma}^2 = \left[\tilde{\beta}_2 \middle/ 1.96\sqrt{\{\mathbf{X}'\mathbf{X}\}_{33}^{-1}}\right]^2$, $\{\mathbf{X}'\mathbf{X}\}_{33}^{-1}$ is the diagonal element of the $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ matrix corresponding to x^2 where $\mathbf{X}_t = \left[1 \quad x_t \quad x_t^2\right]$.

Once the values of $\tilde{\beta}_2$ and $\tilde{\sigma}$ are determined by the value of $\tilde{\theta}$ we only included those simulated regressions where the absolute value of the t-statistics on the estimated coefficient $\hat{\beta}_2$ was greater than 1.96. As shown in Section 4.5 above, a t-statistic of which implies a lower significance than the Fieller confidence bound ($t_2^2 \le t_\alpha^2$) – will result in an infinite bound. Using this selection rule the average value for $|t_2|$ in these simulations was approximately 2.7.

For each value of the extremum point 20,000 replications were computed using $\tilde{y}_t = \tilde{\beta}_0 + \hat{\beta}_1 x_t + \tilde{\beta}_2 x_t^2 + \tilde{\varepsilon}_t$, where $\tilde{\varepsilon}_t \sim N\left(0, \tilde{\sigma}^2\right)$. Once the simulated set of $\tilde{y}_t$ were calculated the regressions were run and the estimated extremum value was generated along with confidence bounds based on the Delta and Fieller 95% methods. Then the proportion of the total number of replications that resulted in the confidence intervals which include the true extremum value (θ) was determined. If the bounds were exact in 95%, or 19,000 replications, the confidence bounds would include the true value of θ.

6.1 Experiment based on the California Test Score data.

In the first experiment the regression design is based on the Californian test score data set used in Section 5.1. The values of *x* range from 6 to 65 and we set values of θ over the range of *x* so that θ takes on values that correspond to values of *x* in the beginning of the sample, middle or at the limit of the *x* values. Table 6.1 reports the true percentage of the confidence limits for the estimated values of θ.

Table 5 The Coverage of the confidence intervals from the Delta and Fieller methods

θ	Delta	Fieller
6	74.96%	93.99%
7	75.00%	93.92%
8	75.99%	94.19%
9	75.61%	93.85%
10	76.09%	93.95%
11	77.16%	94.17%
12	77.88%	94.45%
13	78.23%	94.18%
14	79.83%	93.99%
15	81.18%	93.98%
16	83.12%	94.07%
17	85.06%	94.18%
18	87.91%	94.46%
19	91.01%	94.58%
20	93.78%	94.44%
21	96.59%	94.87%

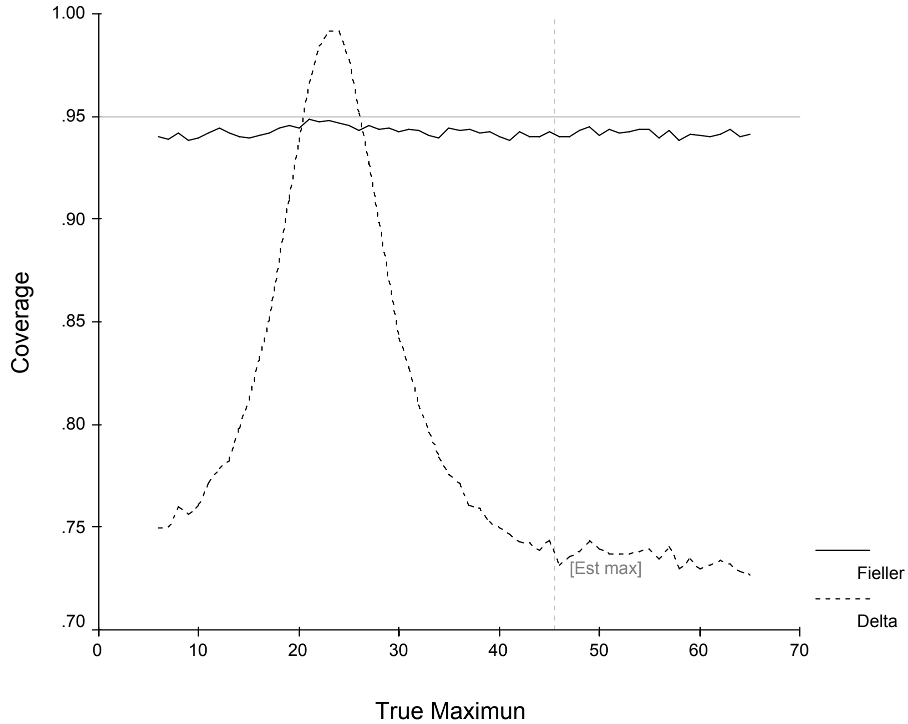
θ	Delta	Fieller
22	98.41%	94.74%
23	99.17%	94.81%
24	99.13%	94.65%
25	97.75%	94.55%
26	95.18%	94.34%
27	92.71%	94.58%
28	89.83%	94.38%
29	86.98%	94.43%
30	84.23%	94.25%
31	82.74%	94.37%
32	80.94%	94.30%
33	79.61%	94.06%
34	78.41%	93.95%
35	77.54%	94.44%
36	77.11%	94.32%
37	76.05%	94.35%

θ	Delta	Fieller
38	75.95%	94.20%
39	75.25%	94.27%
40	74.99%	94.03%
41	74.68%	93.84%
42	74.27%	94.25%
43	74.24%	94.03%
44	73.87%	94.04%
45	74.37%	94.27%
46	73.12%	94.04%
47	73.57%	93.99%
48	73.83%	94.29%
49	74.34%	94.48%
50	73.95%	94.06%
51	73.69%	94.36%
52	73.69%	94.19%

θ	Delta	Fieller
53	73.71%	94.28%
54	73.84%	94.39%
55	73.93%	94.39%
56	73.44%	93.97%
57	74.06%	94.29%
58	72.96%	93.82%
59	73.43%	94.16%
60	72.99%	94.09%
61	73.15%	94.04%
62	73.37%	94.16%
63	73.21%	94.35%
64	72.85%	94.04%
65	72.69%	94.12%
average	79.54%	94.23%

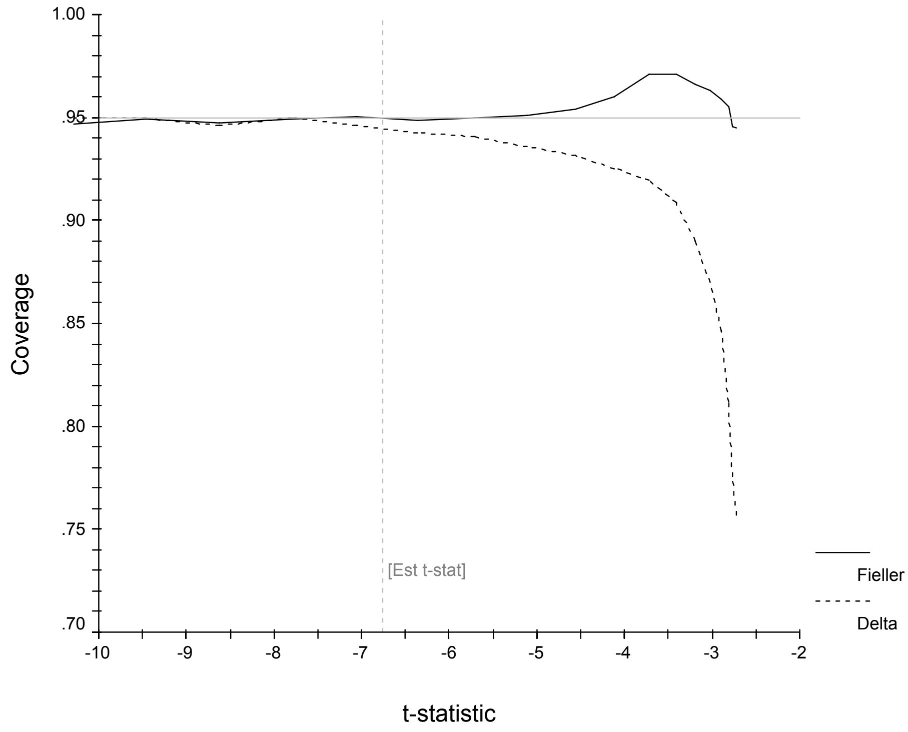
From these results we conclude that the coverage of the Fieller method is consistently close to 95% with an average of 94.23% while the Delta method gives confidence intervals that vary considerably on the value of θ and in some circumstances the true percentage is much less than 95% whereas in other cases it is higher than 95%. Figure 9 provides a plot of the coverage from this experiment. From this plot we can determine that, if the t-statistic for the actual estimation was -2.7, then the coverage of the Delta method would be closer to 75% than the intended coverage of 95%. And from Figure 1 it can be seen that for cases like the example in Section 5.2 it is in the definition of the upper bound where the Fieller and the Delta methods differ the most. Note also that the values of the extremum point where the Delta method appears to change in coverage most dramatically (99.17%) appears at an extremum value of 23. This occurs in the area of the central tendency of the regressor with a median of 13.7, mean of 15.3 and midpoint of 31.8.

Figure 9 The coverage of the Delta and Fieller confidence bounds when the average value of the t-statistic for β_2 is approximately -2.7.



The actual estimated maximum from the true y variable is 45.511, however the t-statistic on the $\hat{\beta}_2$ is much greater than the average value in this simulation (-2.7) it is closer in value to -6.7. As shown in Figure 2 as the absolute value of the t-statistic becomes as large as 6.7 there is very little difference in the two confidence bounds. In order to demonstrate the influence of the t-statistic for the quadratic term, $(\hat{\beta}_2)$ as defined in equation (3), on the coverage of these two methods for determining the confidence interval, we performed a modified simulation where we changed the t-statistic but kept the location of the maximum at 45.551. Figure 10 displays the coverage estimated from this simulation.

Figure 10 The coverage of the Delta and Fieller confidence bounds when the maximum of the quadratic is fixed at 45.511 and the average value of the t-statistic for β_2 $\left(\hat{t}_2\right)$ is varied.



Again in Figure 10 it can be seen that the Fieller interval has a much closer coverage to 95% than the Delta method over all ranges of the t-statistic. Note that as the absolute value of the t-statistic becomes larger than 4 the Fieller method remains much more stable than the Delta method that slowly approaches 95% from below. As shown in Figure 1, this is not surprising since as the t-statistic becomes greater the two confidence bounds begin to coincide. When the t-statistic is equal to -6.758 as estimated in Table 1 we note that the coverage of both the Delta and Fieller is quite close to 95% however the Delta still appears to be systematically low. The similarity of these intervals can be noted in Figure 3. This finding is similar to the findings in Sitter and Wu (1993) where they report that the Fieller coverage can be found to be markedly superior to the Delta method when the sample sizes in are small – which would be analogous to smaller t-statistics as in the present case.

6.1 Experiment based on the American Electric Generating Companies

In the second experiment the data on *x* was taken from the American Electric Generating Companies example in Section 5.2. In this case the extremum point establishes the location of the minimum of the log average cost curve. Again we used the actual regressors and set the values of β_0 and β_1. Since the values of *x* in this case range from 0.693 to 9.724 we assume values of the true minimum θ is bounded by these values of *x*. Table 6 reports the coverage of the confidence limits for these values of θ.

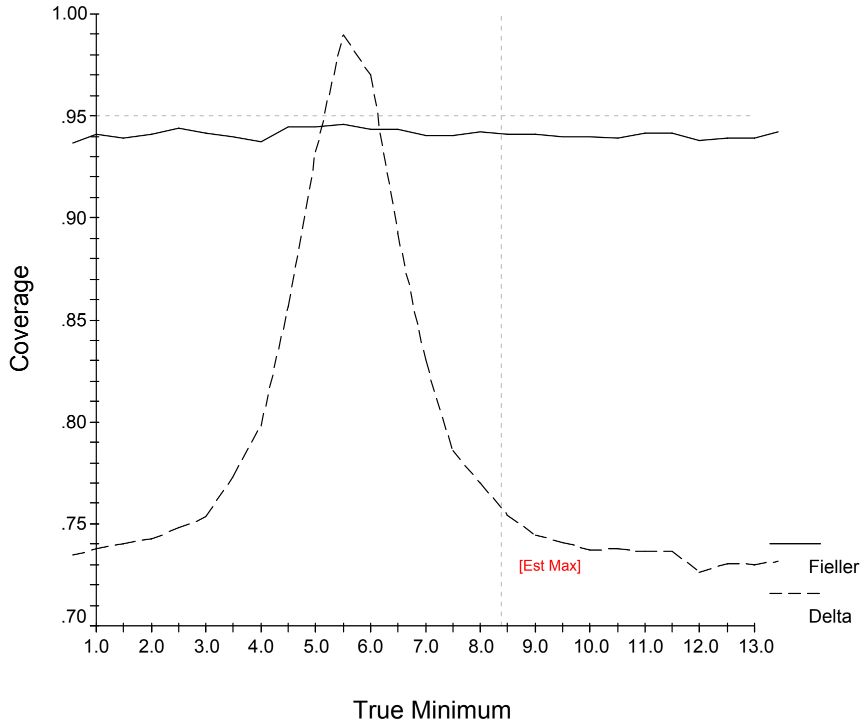
Table 6 The simulated coverage of the Fieller and Delta methods when applied to the model for Electricity Generators.

θ	**Delta**	**Fieller**
1.0	73.76%	94.08%
1.5	74.04%	93.94%
2.0	74.25%	94.08%
2.5	74.80%	94.40%
3.0	75.33%	94.13%
3.5	77.33%	93.98%
4.0	79.80%	93.74%
4.5	85.65%	94.44%
5.0	93.23%	94.47%
5.5	98.98%	94.56%
6.0	97.03%	94.33%
6.5	89.20%	94.32%
7.0	83.00%	94.05%

θ	**Delta**	**Fieller**
7.5	78.61%	94.04%
8.0	76.97%	94.23%
8.5	75.42%	94.12%
9.0	74.44%	94.08%
9.5	74.10%	93.97%
10.0	73.71%	93.95%
10.5	73.76%	93.93%
11.0	73.63%	94.15%
11.5	73.65%	94.16%
12.0	72.64%	93.78%
12.5	73.06%	93.90%
13.0	73.01%	93.89%

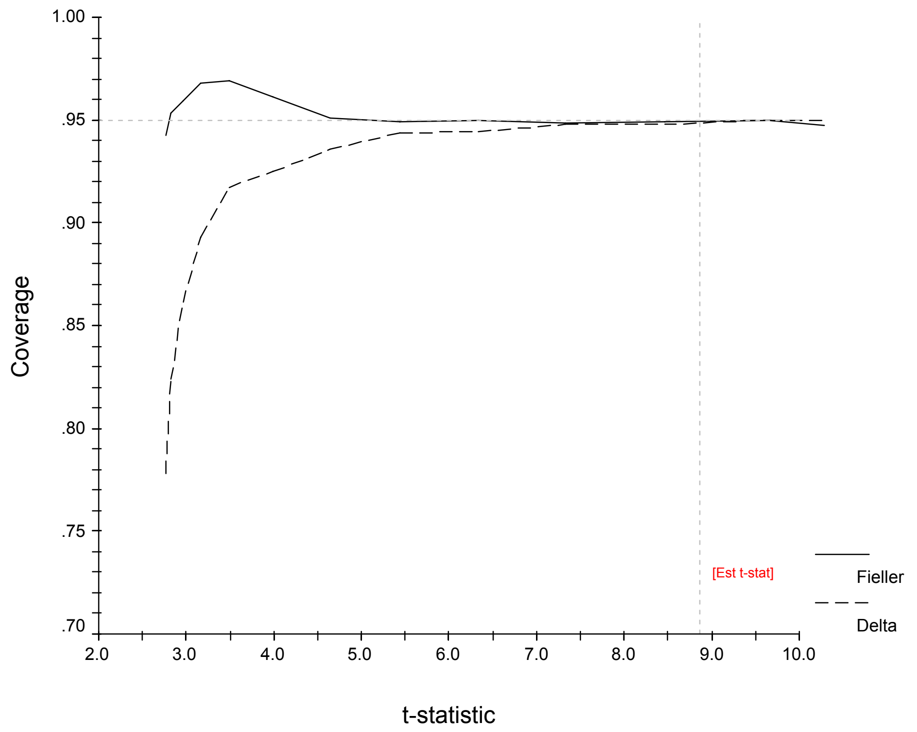
Figure 11 illustrates the coverage of the American electric generator economies of scale example. Note that again at the extremum point (a minimum at 8.383) we find that the coverage for the Delta method is much lower than for the Fieller method and that it varies with the value of the true minimum. Note again that the Fieller method results in coverages that are much closer to the 95% and that the Delta method intervals peak in coverage at 98.96% at a minimum value of 5.5.

Figure 11 The coverage of the Delta and Fieller confidence bounds when the average absolute value of the t-statistic for β_2 is 2.7.



As in the case with the first simulation we also examined the coverage at the estimated extremum point as the t-statistic on the quadratic term was varied. Figure 12 shows the same pattern found in Figure 10. This implies that for relatively low precision on the quadratic parameter the coverage of the confidence interval generated via the delta method is much lower than the equivalent Fieller interval.

Figure 12 The coverage of the Delta and Fieller confidence bounds when the minium of the quadratic is fixed at 8.385 for the American electric generators and the average value of the t-statistic for β_2 $\left(\hat{t}_2\right)$ is varied.



In sum, both simulations demonstrate that when the coefficient on the squared regressor is just significant at just over the 95% level the Delta method will have a tendency to underestimate the size of the confidence bound and imply a more precise estimate of the extremum point – where the standard error is under estimated. In addition, the Delta method exhibits a rather alarming tendency to vary in the degree of coverage at certain values of the extremum value. Note that we replicated the main results of these experiments when we assumed non-normally identically and independently distributed disturbances. These experiments relied on the usual assumption of asymptotic normality of the regression parameters estimated from observations from a data generating process of this type.

7. Conclusion

This paper has shown that the extremum value of a U- or inverted U- shaped relationship found by the application of a quadratic regression can have a confidence interval determined in either of two analytic ways. The most common approach is to use the formula for the extremum value and apply the Delta method which approximates a linear relationship for the extremum point which is defined as negative one half times the ratio of two regression parameters. A procedure which we show is equivalent to applying the Delta method is to redefine the regression as a nonlinear specification. The second method involves the application of the Fieller method for the construction of fiducial confidence bound. We show this to be equivalent to finding the values of the regressor where the confidence interval on the function defined by the first derivative is set equal to zero.

When the parameter on the square of the regressor ($\hat{\beta}_2$ in equation 3) is estimated with a high degree of precision, *p*-values (probability of making a type I error) of .001 or less, we find that both the Delta and the Fieller bounds are very similar. However, from our simulations we find that if the precision of the estimate of $\hat{\beta}_2$ is lower than this, though not so low as to fail to reject the null hypothesis $\left[H_0 : \beta_2 = 0\right]$, we may encounter severe difficulties with the Delta method for estimating finite bounds on the extremum. We find that even when the *p*-value for this parameter is .005 the coverage of the 95% Delta confidence interval is closer to being 80% while the alternative Fieller based method is within 1% of 95%.

In conclusion, we find that the usual confidence intervals based on the application of the Delta method may not suffice for ratios of regression parameters used in the estimation of the extremum of a quadratic unless the parameter on the squared term has a t-statistic in excess of 6. The usual rule of thumb that t-statistics of 2 or more may well not be sufficient when estimating confidence bounds for the extremum. In the case that the quadratic is warranted by the estimation of a t-statistic on the squared regressor between 2 and 6 it is strongly recommended that the alternative method based on Fieller's approach be employed for the construction of confidence intervals.

This paper has been limited only to the consideration of the quadratic specification and focuses on the existence of extremum points. A potential difficulty with the quadratic is that it results in a very limited functional relationship. All non trivial quadratic functions have a single extremum. A more complex functional form that allows for multiple extrema and inflexion points may be warranted. As we have shown in Appendix B, the Fieller approach is equivalent to the determination of the point where the confidence bounds of the first derivative function crosses the zero axis. Future research will help to establish how the first derivative function can be used to locate the extremum point for any function in which one can define the first derivative as a linear function of the estimated regressors. Such functions are found when employing higher order polynomials and fractional polynomials as suggested by Barnett (1983) and Royston and Altman (1994) or even trigonometric specifications as proposed by Gallant (1981) and Eubank and Speckman (1990).

In addition, another hypothesis that is excluded by the traditional hypothesis testing based on the application of the Delta method is the presence of a plateau in the relationship. In the case of a positively valued regressor this would imply that the lower bound is non- zero for the extremum point but that there is no finite upper bound. The application of the Delta method implies that the probability of such a hypothesis is zero due to the imposition of symmetry on the test and the assumption that all upper bounds are finite. However, the Fieller method allows the consideration of an asymmetric interval that can be used to establish bounds on the t-statistic of the quadratic term which would indicate the region in which one would conclude that the regression implies the existence of a plateau – a value of the regressor past which the relationship to the dependent variable ceases and does not recommence. This would imply that there is no extremum which acts as a turning point where the relationship between the regressor and dependent variable reverses.

References

Barnett, W. A., 1983, "New Indices of Money Supply and the Flexible Laurent Demand System", *Journal of Business and Economic Statistics*, 1, 7-23.

Berger, M. C. and J. P. Leigh, 1988, "The effects of alcohol use on wages", *Applied Economics*, 20, 1343-51.

Berndt, E. R., 1991, *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley.

Byron, R. P. and A. K. Bera, 1983, "Least Squares Approximations to Unknown Regression Functions: A Comment", *International Economic Review*, 24, 255-260.

Cramer, J. S., 1969, *Empirical Econometrics*, North Holland.

Dufour, J., 1997, "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models", *Econometrica,* 65, 1365-1387.

Eubank, R. L. and Paul Speckman, 1990,"Curve fitting by polynomial-trigonometric regression", *Biometrika*, 77, 1-9.

Fieller, E. C., 1932, "The Distribution of the Index in a Normal Bivariate Population", *Biometrika*, 24, 428-440.

Fieller, E. C., 1954, "Some Problems in Interval Estimation", *Journal of the Royal Statistical Society. Series B*, 16, 175-185.

Gallant, A. R., 1981, 'On the Bias in Flexible Functional Forms and an Essentially Unbiased Form", *Journal of Econometrics*, 15, 211-245.

Hsing, Y., 1996, "Estimating the Laffer Curve and Policy Implications", *Journal of Socio-Economics*, 25, 395-401.

Hirschberg, J. G., and J. N. Lye, 2001, "The influence of smoking on the labour market consequences of drinking", *Health Promotion Journal of Australia*, 12, 48-51.

Koop, G., (2000), *Analysis of Economic Data*, Wiley USA.

Kuznets, S., 1955, "Economic growth and income inequality", *American Economic Review*, 45, 1-28.

Murphy, K. M. and F. Welch, 1990, "Empirical Age-Earnings Profiles", *Journal of Labor Economics*, 107, 285-326.

Nerlove, M., 1963, "Returns to Scale in Electricity Supply", Chapter 7 in Carl F Christ, ed. *Measurement in Economics: Studies in Honor of Yehuda Grundfeld*, Stanford University Press, Stanford, Cal., 167-198.

Rao, C. R., 1973, *Linear Inference and its Applications*, 2nd edition, John Wiley and Sons, New York.

Royston, Patrick and Douglas G. Altman, 1994, “Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling”, *Applied Statistics,* 43, 429-467.

Sitter, R. A. and C. F. J. Wu, 1993, “On the Accuracy of the Fieller Intervals for Binary Response Data”, *Journal of the American Statistical Association*, 88, 1021-1025.

Stock, J. and M. Watson, (2003), *Introduction to Econometrics*, Pearson Education, Inc. USA.

Thompson Jr., H. G. and L. L. Wolf, (Regional Differences in Nuclear and Fossil-Fuel Generation of Electricity”, *Land Economics*, 69, 234-248.

White, H., 1980, “Using Least Squares to Approximate Unknown Regression Functions”, *International Economic Review,* 21, 149-170.

Wooldridge, J., (2003), *Introductory Econometrics A Modern Approach*, 2nd edition, Thomson South-Western, USA.

Zerbe, G. O., 1978, “On Fieller’s Theorem and the General Linear Model”, *The American Statistician*, 32, 103-105.

Appendix A The equivalence of the Delta method and the nonlinear least squares covariance matrix.

First we define the variance as estimated from the application of the delta method to the estimate of the extremum as defined by the ratio of two parameters estimated via a linear regression. Then we use the nonlinear least squares equivalent specification to estimate the extremum as shown in Section 4.2 and determine the estimated variance on this parameter to show the equivalence.

A.1 The Delta Method Variance Estimate

Consider the regression model containing a single regressor entered in the quadratic form given by x

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t \ , \ \ \varepsilon_t \text{ i.i.d.} \sim (0,\sigma^2)$$

where the extremum is given by $\theta = \dfrac{-\beta_1}{2\beta_2}$. The OLS estimates $\hat{\beta}_0,\ \hat{\beta}_2,\ \hat{\theta} = \dfrac{-\hat{\beta}_1}{2\hat{\beta}_2}$ by application of the Delta method the corresponding variance covariance matrix of these 3 parameters is given by

$$\widehat{\text{cov}}\left(\hat{\beta}_0,\ \hat{\beta}_2, \hat{\theta}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & {}^{-1}\!/\!_{2\hat{\beta}_2} & {}^{\hat{\beta}_1}\!/\!_{2\hat{\beta}_2^2} \end{bmatrix} \widehat{\text{cov}}(\hat{\beta}_0,\ \hat{\beta}_1\ \hat{\beta}_2) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & {}^{-1}\!/\!_{2\hat{\beta}_2} \\ 0 & 1 & {}^{\hat{\beta}_1}\!/\!_{2\hat{\beta}_2^2} \end{bmatrix}$$

$$= \hat{\sigma}^2 \mathbf{Q}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{Q}'$$

where $\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & {}^{-1}\!/\!_{2\hat{\beta}_2} & {}^{\hat{\beta}_1}\!/\!_{2\hat{\beta}_2^2} \end{bmatrix}$ and $= \widehat{\text{cov}}(\hat{\beta}_0,\ \hat{\beta}_1\ \hat{\beta}_2) = \hat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.

A.2 The Nonlinear Least Squares Variance Estimate

Now consider estimating the parameters $\hat{\beta}_0,\ \hat{\beta}_2,\ \hat{\theta} = \dfrac{-\hat{\beta}_1}{2\hat{\beta}_2}$ using the nonlinear estimation approach given by $y_t = \beta_0 + \beta_2\theta\left(-2x_t\right) + \beta_2 x_t^2 + \varepsilon_t$ where the errors are identically and independently distributed with mean zero and a variance of σ^2 $\left(\varepsilon_t \sim (0,\sigma^2)\right)$. When this equation is estimated using an algorithm method in which the estimation of the covariance of the parameter vector $\left(\widehat{\text{cov}}(\hat{\beta}_0,\ \hat{\beta}_2, \hat{\theta})\right)$ of the nonlinear function employs the outer product of the gradient of the nonlinear function evaluated at each of the observations the estimated variance-covariance matrix of the parameters is given by $\hat{\sigma}^2\left(\mathbf{F}'\mathbf{F}\right)^{-1}$ where **F** is the *T* by *k* matrix of derivatives of the parameter vector evaluated at the regression parameter estimates at each observation. That is a row of **F** is defined as:

$$f_t = \left(\left(\frac{\partial Y}{\partial \beta_0}\right) \ \left(\frac{\partial Y}{\partial \beta_2}\right) \ \left(\frac{\partial Y}{\partial \theta}\right)\right)$$

$$= \left(1 \quad -2\hat{\theta}\mathrm{x}_t + \mathrm{x}_t^2 \quad -2\hat{\beta}_2\mathrm{x}_t\right)$$

Or $f_t = \mathbf{x}_t\mathbf{H}'$ where

$$\mathbf{H}=\begin{bmatrix}1 & 0 & 0\\ 0 & -2\hat{\theta} & 1\\ 0 & -2\hat{\beta}_2 & 0\end{bmatrix}, \text{ or } \mathbf{H}=\begin{bmatrix}1 & 0 & 0\\ 0 & \hat{\beta}_1/\hat{\beta}_2 & 1\\ 0 & -2\hat{\beta}_2 & 0\end{bmatrix}$$

Then:

$$\begin{aligned}\hat{\sigma}^2\left(\mathbf{F}'\mathbf{F}\right)^{-1} &= \hat{\sigma}^2\left((\mathbf{H}\mathbf{X}')(\mathbf{X}\mathbf{H}')\right)^{-1}\\ &= \hat{\sigma}^2\left(\mathbf{H}'\right)^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}^{-1}\end{aligned}$$

and

$$\begin{aligned}\left(\mathbf{H}'\right)^{-1} &= \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & -1/2\hat{\beta}_2 & \hat{\beta}_1/2\hat{\beta}_2^2\end{bmatrix}\\ &= \mathbf{Q}\end{aligned}$$

The estimates of the covariance matrix are identical as long as the parameter estimates from OLS are the same as those obtained from the nonlinear method. This should be the case unless the nonlinear method is not started at appropriate starting values.

Appendix B. The equivalence between the Fieller confidence bounds and the confidence bounds on the first derivative.

First we define the nature of the confidence bounds for the ratio of two linear combinations of regression parameters as shown by Zerbe(1978). Then we show that the confidence bounds found for the first derivative function result in the same interval.

B.1 The Fieller bounds as given by Zerbe

Where the regression estimated was of the form:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \beta_3 z_t + \varepsilon_t$$

Following the notation in Zerbe(1978) we can define the confidence bounds on a ratio of linear combinations of regression parameters as:

$$\theta = \frac{K'B}{L'B}$$

When the linear combination is the extremum point defined for the implied parabola given as: $\theta = \frac{-\beta_1}{2\beta_2}$ we find that the linear combinations are given as:

$$K'B = \begin{pmatrix}0 & -1 & 0 & 0\end{pmatrix}\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{pmatrix},\ K'B = -\beta_1$$

$$L'B = \begin{pmatrix} 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \; L'B = 2\beta_2$$

If we define the estimated covariance of the regression estimates of the β's as : $cov = \hat{\sigma}_\varepsilon^2 \left(X'X \right)^{-1}$ then we have that the solutions to the quadratic equation defined as: $a\theta^2 + b\theta + c = 0$ define the confidence bounds when :

$$a = (L'B)^2 - t^2 L'covL$$
$$b = 2\left[t^2 K'covL - (K'B)(L'B) \right]$$
$$c = (K'B)^2 - t^2 K'\text{cov}\, K$$

By our definition of the linear combinations we have that :

$$L'covL = \begin{pmatrix} 0 & 0 & 2 & 0 \end{pmatrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{03} & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \; L'covL = 4\sigma_2^2,$$

$$K'covK = \begin{pmatrix} 0 & -1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{03} & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \; K'covK = \sigma_1^2$$

$$K'covL = \begin{pmatrix} 0 & -1 & 0 & 0 \end{pmatrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{03} & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix}, \; K'covL = -2\sigma_{12}$$

Thus when $K'B = -\beta_1$, $L'B = 2\beta_2$ $L'covL = 4\sigma_2^2$, $K'covK = \sigma_1^2$, and $K'covL = -2\sigma_{12}$. The estimates of the terms in the quadratic equation are defined as $a\theta^2 + b\theta + c = 0$ where: $a = 4\hat{\beta}_2^2 - 4t^2\hat{\sigma}_2^2$, $b = -4t^2\hat{\sigma}_{12} + 4\hat{\beta}_1\hat{\beta}_2$, and $c = \hat{\beta}_1^2 - t^2\hat{\sigma}_1^2$. And we find two solutions to the equation given by $\theta = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ as:

$$\tfrac{1}{2}\left(\hat{\beta}_2^2 - t^2\hat{\sigma}_2^2 \right)^{-1} \left(t^2\hat{\sigma}_{12} - \hat{\beta}_1\hat{\beta}_2 \pm \sqrt{\left(t^4\hat{\sigma}_{12}^2 - 2t^2\hat{\sigma}_{12}\hat{\beta}_1\hat{\beta}_2 + \hat{\beta}_2^2 t^2\hat{\sigma}_1^2 + t^2\hat{\sigma}_2^2\hat{\beta}_1^2 - t^4\hat{\sigma}_2^2\hat{\sigma}_1^2 \right)} \right)$$

B.2 The first derivative method and the Fieller method

The first derivative method employs the confidence bounds of the first derivative function of the estimated regression. The confidence bounds of the first derivative function are plotted to establish where they equal to zero.

We assume that the first derivative of the regression function can be written as a linear combination of the parameter set: $\dfrac{\partial y}{\partial x} = D'B$ where the matrix D is a function of x and constants. Thus the confidence bounds for the first derivative can be written as: $D'B \pm t\sqrt{D'covD}$. If we wish to equate these bounds to zero we can then specify

$D'B \pm t\sqrt{D'covD} = 0$. Thus we can find those values of x for which this is true. To do this we can subtract $D'B$ from both sides of the equation and then square both sides to get $\left(D'B\right)^2 = t^2 D'covD$ and moving both terms to the same side of the equals sign we obtain $\left(D'B\right)^2 - t^2 D'covD = 0$.

Using the regression equation given as (3) we can define the first derivative as a linear function of the parameters as:

$$D'B = \begin{pmatrix} 0 & 1 & 2x & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \beta_1 + 2x\beta_2$$

And the variance of the 1st derivative function is defined as:

$$D'covD = \begin{pmatrix} 0 & 1 & 2x & 0 \end{pmatrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} & \sigma_{03} \\ \sigma_{01} & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{02} & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{03} & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \\ 2x \\ 0 \end{pmatrix} = \sigma_1^2 + 4x\sigma_{12} + 4x^2\sigma_2^2$$

By setting the value of x where the confidence bounds cross the zero axis to ρ we find the expression of the form:

$$0 = \left(\beta_1 + 2\rho\beta_2\right)^2 - t^2\left(\sigma_1^2 + 4\rho\sigma_{12} + 4\rho^2\sigma_2^2\right)$$

and this is equivalent to :

$$0 = \left(4\beta_2^2 - 4t^2\sigma_2^2\right)\rho^2 + \left(-4t^2\sigma_{12} + 4\beta_1\beta_2\right)\rho + \beta_1^2 - t^2\sigma_1^2$$

Which provides the identical quadratic equation which when solved results in the confidence bounds as the Fieller method applied to regression parameter ratios as shown in Section B.1.