THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 906

OCTOBER 2005

**INFERENCES FOR THE EXTREMUM OF
QUADRATIC REGRESSION MODELS**

by

Joseph G. Hirschberg
&
Jenny N. Lye

Department of Economics
The University of Melbourne
Melbourne   Victoria   3010
Australia.

Inferences for the extremum of quadratic regression models

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

October, 05

Abstract

Quadratic functions are often used in regression to infer the existence of an extremum in a relationship although tests of the location of the extremum are rarely performed. We investigate the construction of the following confidence intervals: Delta, Fieller, estimated first derivative, bootstrapping, Bayesian and likelihood ratio. We propose interpretations for the unbounded intervals that may be generated by some of these methods. The coverage of the confidence intervals is assessed by Monte Carlo; the Delta and *studentized* bootstrap can perform quite poorly. Of all the methods, the first derivative method is easiest to implement.

Key Words: Inverted U-Shaped, turning point, Fieller method, Delta method, 1st derivative function, Bayesian, Likelihood ratio, Bootstrap.

JEL Classification numbers: C21, C40, C52

Word Count: 5,626

---

1. Introduction

Quadratic specifications have found a wide application when it is desirable for a linear regression to allow the relationship between the regressor and the dependent variable to exhibit a limited degree of nonlinearity. It can be shown that half the negative ratio of the linear to the squared coefficients in a quadratic specification define the extremum or turning point of the relationship between the regressor and the dependent variable. In this paper, we examine various methods for constructing the confidence interval for this ratio.

Drawing inferences from the ratio of regression coefficients is elemental in a number of statistical applications. These applications include statistical calibration, slope-ratio assay, parallel-line assay, cost-effectiveness, surrogate endpoints and the extremum of a quadratic regression. Monte Carlo experiments to assess the performance of a number of different methods have been performed by Jones et al. (1996) for statistical calibration, Williams (1986) and Sitter and Wu (1993) in bioassay, Polsky et al. (1997) and Briggs et al. (1999) for cost-effectiveness ratios and Freedman (2001) for intermediate or surrogate endpoints.

Generally, the results from these Monte Carlo simulations indicate that the Fieller-based methods work reasonably well under a range of assumptions including departures from normality. The delta-based method is a consistent poor performer and often underestimates the upper limit of the intervals. These studies also found merit in alternative methods including bootstrap approaches, and for bioassay, the interval based on the likelihood ratio approach.

Examples of the estimation of quadratic relationships in the economics literature are common. The Kuznets curve proposes that the relationship between income inequality and per capita income is characterized by an inverted U-shaped relationship (Kuznets 1955) and the related environmental Kuznets curve proposes the same relationship between the level of environmental degradation and per capita income. Wage profiles display the characteristic of a diminishing effect of increased experience on wages (Murphy and Welch, 1990). The Laffer curve relates the level of national income to the rate of taxation (Hsing, 1996). The relationship between alcohol

consumption and income implies that income increases until the level of consumption interferes with the ability to work (Berger and Leigh 1988, Lye and Hirschberg 2004). The economies of scale of a production technology suggests that as the scale of production increases, average costs decrease before starting to increase (Thompson and Wolf 1993). Regressions in which a continuous regressor is included as a quadratic imply either a convex or concave relationship. However, for these results to imply that the relationship between the regressor and the dependent variable changes sign the extremum point must be feasible or finite. We propose that this can best be achieved by constructing the appropriate confidence interval for the extremum.

In this paper, we investigate the application of a number of different methods for constructing confidence intervals for ratios of regression parameters to determine the location of the extremum of a quadratic regression specification. These methods include: the Delta method; Fieller's confidence interval; a confidence interval for the 1st derivative function; the *percentile* and *studentized* bootstrap; a Bayesian method that assumes normally distributed regression disturbances with a Jeffreys' prior; and a confidence interval based on the distribution of the likelihood ratio.

This paper proceeds as follows. First, we discuss the location and relevance of the extremum point of the quadratic specification. Then, we describe alternative methods for constructing confidence intervals. To highlight differences between these methods two empirical applications based on typical data are presented. To compare the performance of the different confidence intervals, we then perform a simulation experiment to study the coverage of these confidence bounds for a typical application. Finally, conclusions are presented.

2. The location and relevance of the extrema points in a quadratic specification

The regression is of the form:

$$ y_t = \beta_1 x_t + \beta_2 x_t^2 + \left(\beta_0 + \sum_{j=4}^{k} \beta_j z_{jt} \right) + \varepsilon_t \tag{1} $$

where we assume that there is a single regressor x entered as a quadratic function, $k-4$ linearly related regressors denoted by z_j, and T observations. By setting the first derivative of $\mathrm{E}(y|x,z)$ with respect to x, $\frac{\partial \mathrm{E}(y|x,z)}{\partial x} = \beta_1 + 2\beta_2 x_t$, equal to zero, we can derive the value of x at the extremum as $\theta = \frac{-\beta_1}{2\beta_2}$. θ is either the maximum (when $\beta_2 < 0$) or the minimum (when $\beta_2 > 0$). The usual estimate of the extremum is $\hat{\theta} = \frac{-\hat{\beta}_1}{2\hat{\beta}_2}$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimates of β_1 and β_2 in (1) respectively.

Once $\hat{\theta}$ is determined it is necessary to establish if this is a relevant value in the context of the application. This depends on whether $\hat{\theta}$ falls within the range of values that the regressor can take. Although all quadratic functions, when estimated by regression, result in an extremum value (provided $\hat{\beta}_2$ is not equal to 0), $\hat{\theta}$ may be too distant from the range of the regressor to be meaningful in the context of the analysis. In addition, since the estimate of the extremum point ($\hat{\theta}$) is a random variable, confidence intervals can be used to determine if it is a feasible value for x.

Typically, when determining if a quadratic relationship is warranted, one first establishes if the estimate for β_2 is significantly different from zero. 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 defining the confidence band of the extremum.

3.   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, as is well known, in a ratio of random variables, that the denominator can take values close to zero. If this occurs, the moments of the random variable defined by the ratio are undefined, as in the case of the Cauchy distribution, which is 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 ($\hat{\beta}_2$). In this paper, we assume that this parameter is

4

significantly different from 0 and thus 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 when deciding whether a quadratic function is warranted.

3.1   The Delta Method

The estimated variance of $\hat{\theta}$ based on the Delta method is given by (see, for example, Rao 1973, pp 385-389):

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

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 $\hat{\beta}_2$. A $100(1-\alpha)\%$ confidence interval for θ is given by:

$$\hat{\theta} \pm t_{\alpha/2}\sqrt{\frac{\hat{\sigma}_1^2\hat{\beta}_2^2 - 2\hat{\beta}_1\hat{\beta}_2\hat{\sigma}_{12} + \hat{\beta}_1^2\hat{\sigma}_2^2}{4\hat{\beta}_2^4}} \tag{3}$$

where $t_{\alpha/2}$ is the value from the t distribution with an $\left(\frac{\alpha}{2}\right)\%$ level of significance and $T-k$ degrees of freedom and $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ and $\hat{\sigma}_{12}$ are the estimated values of σ_1^2, σ_2^2 and σ_{12}, respectively.

3.2   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 in which 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 (1) by using the result that $\beta_1 = -2\theta\beta_2$ to derive the form of the equation as:

$$y_t = \beta_2\theta(-2x_t) + \beta_2 x_t^2 + \left(\beta_0 + \sum_{j=4}^{k}\beta_j z_{jt}\right) + \varepsilon_t \tag{4}$$

When (4) is estimated by using a Gauss-Newton algorithm (as applied in most software packages), the estimate of the covariance between the parameters typically employs the outer product of the

gradient of the nonlinear function evaluated at each of the observations. It can be shown that this covariance matrix is the same as the estimated standard error from the Delta method for the extremum point $\hat{\theta}$ (see, for example, Mikulich et al. 2003).

3.3   The Fieller method

The Fieller method (Fieller 1932, 1954) provides a general procedure for constructing confidence limits for statistics defined as ratios. Zerbe (1978) defines a version of Fieller's method in the regression context, where the ratio $\psi = \dfrac{\mathbf{K}'\beta}{\mathbf{L}'\beta}$ is defined in terms of linear combinations of the regression parameters from the same regression, $\mathbf{Y}_{T\times 1} = \mathbf{X}_{T\times k}\beta_{k\times 1} + \varepsilon_{T\times 1}$, $\varepsilon \sim (\mathbf{0}_{T\times 1}, \sigma^2\mathbf{I}_{T\times T})$, in which the OLS estimators for the parameters are $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\hat{\sigma}^2 = \varepsilon'\varepsilon/(T-k)$, and the vectors $\mathbf{K}_{k\times 1}$ and $\mathbf{L}_{k\times 1}$ are known constants. Under the usual assumptions, the parameter estimates are asymptotically normally distributed according to $\hat{\beta} \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$. A $100(1-\alpha)\%$ confidence interval for ψ is determined by solving the quadratic equation $a\psi^2 + b\psi + c = 0$, where

$$a = (\mathbf{L}'\hat{\beta})^2 - t_{\alpha/2}^2 \mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\hat{\sigma}^2, \ b = 2\left[t_{\alpha/2}^2 \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}\hat{\sigma}^2 - (\mathbf{K}'\hat{\beta})(\mathbf{L}'\hat{\beta})\right] \text{ and}$$

$c = (\mathbf{K}'\hat{\beta})^2 - t_{\alpha/2}^2 \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}\hat{\sigma}^2$. The two roots of the quadratic equation, $(\psi_1, \psi_2) = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$, define the confidence bounds of the parameter value. In the case of the extremum of the quadratic regression equation in (1), the values of these terms can be shown to be as follows: $a = 4\left(\hat{\beta}_2^2 - t_{\alpha/2}^2\hat{\sigma}_2^2\right)$, $b = 4\left(\hat{\beta}_1\hat{\beta}_2 - t_{\alpha/2}^2\hat{\sigma}_{12}\right)$ and $c = \hat{\beta}_1^2 - t_{\alpha/2}^2\hat{\sigma}_1^2$. We can redefine a and c as: $a = 4\hat{\sigma}_2^2\left(\hat{t}_2^2 - t_{\alpha/2}^2\right)$ and $c = \hat{\sigma}_1^2\left(\hat{t}_1^2 - t_{\alpha/2}^2\right)$, where $\hat{t}_1 = \hat{\beta}_1/\hat{\sigma}_1$ and $\hat{t}_2 = \hat{\beta}_2/\hat{\sigma}_2$. Then, the roots can be defined as:

$$(\psi_1, \psi_2) = \frac{t_{\alpha/2}^2\hat{\sigma}_{12} - \hat{\beta}_1\hat{\beta}_2 \pm \sqrt{\left(\hat{\beta}_1\hat{\beta}_2 - t_{\alpha/2}^2\hat{\sigma}_{12}\right)^2 - \hat{\sigma}_2^2\hat{\sigma}_1^2\left(\hat{t}_2^2 - t_{\alpha/2}^2\right)\left(\hat{t}_1^2 - t_{\alpha/2}^2\right)}}{2\hat{\sigma}_2^2\left(\hat{t}_2^2 - t_{\alpha/2}^2\right)} \qquad (5)$$

If $\hat{t}_2^2 > t_{\alpha/2}^2$ (5) has two real roots and one can construct a finite confidence interval. This condition corresponds to being able to reject $H_0 : \beta_2 \neq 0$ when α is the level of significance (Buonaccorsi 1979). In addition to the finite interval case, the resulting confidence interval may be the complement of a finite interval when ($b^2 - 4ac > 0$, $a < 0$) or of the whole real line when $b^2 - 4ac < 0$, $a < 0$. These conditions are discussed in Scheffé (1970) and Zerbe (1982).

3.4 The confidence bounds of the first derivative function

For the quadratic model defined by (1), the first derivative of y with respect to x defines a linear relationship for the slope of the quadratic function given by $\frac{\partial E(y|x,z)}{\partial x} = \beta_1 + 2\beta_2 x$. An estimate of the first derivative as a function of x can be plotted with a $100(1-\alpha)\%$ confidence interval defined as:

$$\mathrm{CI}\left(\tfrac{\partial E(y|x,z)}{\partial x}\right) = \left(\hat{\beta}_1 + 2\hat{\beta}_2 x\right) \pm t_{\alpha/2}\sqrt{\left(\hat{\sigma}_1^2 + 4x\hat{\sigma}_{12} + 4x^2\hat{\sigma}_2^2\right)} \tag{6}$$

An estimate of the extremum value, $x = \hat{\theta}$, is found by solving $\hat{\beta}_1 + 2\hat{\beta}_2\hat{\theta} = 0$. Similarly, the bounds that define a $100(1-\alpha)\%$ confidence interval on θ are found by solving for x in the relationship:

$$\left(\hat{\beta}_1 + 2\hat{\beta}_2 x\right) \pm t_{\alpha/2}\sqrt{\left(\hat{\sigma}_1^2 + 4x\hat{\sigma}_{12} + 4x^2\hat{\sigma}_2^2\right)} = 0 \tag{7}$$

which is equivalent to solving the roots of the equation:

$$\left(\hat{\beta}_1 + 2\hat{\beta}_2 x\right)^2 - t_{\alpha/2}^2\left(\hat{\sigma}_1^2 + 4x\hat{\sigma}_{12} + 4x^2\hat{\sigma}_2^2\right) = 0. \tag{8}$$

By rearranging the terms in (8), this can be written as the quadratic equation, $ax^2 + bx + c = 0$, where a, b and c are defined as in the Fieller method described in Section 3.3. This result follows analogously from Miller (1981 pp. 117-120). In this case, we are considering only one possible value of the dependent variable. Hence, the confidence limits found in this way for the quadratic specification are identical to those found by using the Fieller method. An advantage of this approach is that it can provide a graphical representation of the various possible Fieller-type confidence intervals (see also Williams 1959, pp. 110-113).

3.5   Bayesian Inference

A number of authors have applied Bayesian techniques to draw inferences about the ratio of regression parameters (see, for example, Zellner 1978, Buonaccorsi and Gatsonis 1988 and Ghosh et al. 2003).

Following Buonaccorsi and Gatsonis (1988), if θ is the extremum of a quadratic regression, then a locally flat prior on θ seems an appropriate expression of the investigator's lack of information. A Jeffreys' prior is a reference prior when all regression parameters in the model are treated as being equally important. We assume a Jeffreys' prior given by:

$$\pi^J(\theta) \propto \sigma^{-((k+1)+1)}\left(\frac{1}{Q(\theta)}\right) \tag{9}$$

where $Q(\theta) = c_{11}\theta^2 + 2c_{12}\theta + c_{22}$, $C = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix} = A_{11} - A_{12}A_{22}^{-1}A_{21}$,

$$A_{11} = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}, \; A_{12} = A_{21}' = \begin{pmatrix} s_{13} & \cdots & s_{1k} \\ s_{23} & \cdots & s_{2k} \end{pmatrix}, \; A_{22} = \begin{pmatrix} s_{33} & \cdots & s_{3k} \\ \vdots & \ddots & \vdots \\ s_{k3} & \cdots & s_{kk} \end{pmatrix}, \; X'X = T\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and }$$

$s_{jl} = T^{-1}\sum_{i=1}^{T} x_{ij}x_{il}, \; j,l = 1\cdots k$. For the quadratic model defined in (1), A_{11} is the 2 by 2 sub-matrix of the cross products of the regressor (x) and the regressor squared. Ghosh et al. (2003) show that the marginal posterior distribution of θ, assuming the regression model in (1), where ε_t are *i.i.d.* $N(0, \sigma^2)$, and given a Jeffreys' prior, is:

$$\pi^J(\theta / y) \propto \left(\frac{1}{Q(\theta)}\right)\left\{SSE + \frac{T|C|\left(2\hat{\beta}_2\theta + \hat{\beta}_1\right)^2}{Q(\theta)}\right\}^{\frac{T+1}{2}} \tag{10}$$

where SSE is the sum of squared residuals from regression (1). In our applications below, we numerically integrate the density implied by (10) to determine the confidence intervals.

3.6   The Bootstrap Approach

The bootstrap method estimates the sampling distribution of a statistic based on resampling with replacement from the original data (see Efron, 1979). The empirical estimate of the sampling distribution can then be used to construct confidence intervals. While this approach does not rely on parametric assumptions concerning the underlying distribution, it can be computationally intensive and may require special purpose programs. Using a conditional bootstrap for regression parameters (Freedman, 1981) we use both the *percentile* and the *studentized* methods for constructing confidence intervals.

The *percentile* method uses the $100\left(\%_{2}\right)$ and $100\left(1-\%_{2}\right)$ percentile values of the bootstrap sampling distribution as the upper and lower confidence limits. Although implementation is straight-forward, the *percentile* method assumes that the bootstrap replicates are unbiased, whereas in this application, the ratio estimates are biased but consistent (Briggs et al. 1999).

The *studentized* bootstrap or bootstrap-*t* (see, for example, Efron and Tibshirani 1993, p. 160) is a widely proposed alternative to the percentile method and has been employed in a study of the calibration problem by Jones et al (1996). In this approach, the bootstrap estimates of the extremum, $\hat{\theta}_{i}$, and its estimated standard error, $\widehat{se(\hat{\theta}_{i})}$, are used to create a new vector that has elements defined as:

$$\hat{t}_{i} = \frac{\left(\hat{\theta}_{i} - \hat{\theta}\right)}{\widehat{se(\hat{\theta}_{i})}} \qquad (11)$$

The $100\left(\%_{2}\right)$ and $100\left(1-\%_{2}\right)$ percentile values of the vector $\hat{t}_{i}$ are used as the '*t*-statistics' for defining the upper and lower confidence limits. A crucial assumption underlying this method is that an appropriate estimate of the standard error exists; in most cases, the delta method would be used to obtain this.

3.7   The Likelihood Ratio Interval

The likelihood ratio interval for θ consists of all values of $\tilde{\theta}$ for which the null hypothesis $H_0 : \theta = \tilde{\theta}$ is not rejected in favour of the alternative $H_1 : \theta \neq \tilde{\theta}$ by means of a likelihood ratio test. We define the likelihood ratio statistic as $\lambda = -2 \left(LogL_{H_1} - LogL_{H_0} \right)$, which has a chi-squared distribution with one degree of freedom. $LogL_{H_0}$ is the value of the log likelihood from the reparameterized model defined by (4), and $LogL_{H_1}$ is the value of the log likelihood obtained after setting the value of $\tilde{\theta}$ in (4) to a range of alternative values. A $100(1-\alpha)$% likelihood ratio confidence interval is defined by all values of $\tilde{\theta}$ that satisfy $\lambda < \chi_1^2(\alpha)$, where $\chi_1^2(\alpha)$ is the upper α value of the chi-squared distribution with one degree of freedom. Following a procedure similar to Fieller's, as outlined by Morgan (1992, p. 63), one can define a confidence interval by finding those values of $\tilde{\theta}$ where $\lambda = \chi_1^2(\alpha)$. In our applications, we assume that ε_t in (4) is normally distributed.

## 4.   Empirical Applications

To illustrate differences between confidence limits based on the different methods, we look at two standard textbook examples that estimate quadratic functional forms.

4.1   The Californian Test Score Data Set

The data set taken from Stock and Watson (2003, p. 134) 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 and ranges from 5.3 to 55.3.

A scatter plot of y against x (presented in Figure 1) indicates a curved relationship between the two. To allow for this curvature, test scores can be modelled as a function of income and its square. In this case, one would be interested to know if scores decline from a particular income level. Estimation results for a quadratic regression obtained by using Eviews 5 software are presented in Table 1.

Figure 1 Scatter plot district income (y) against fifth-grade test scores (x)
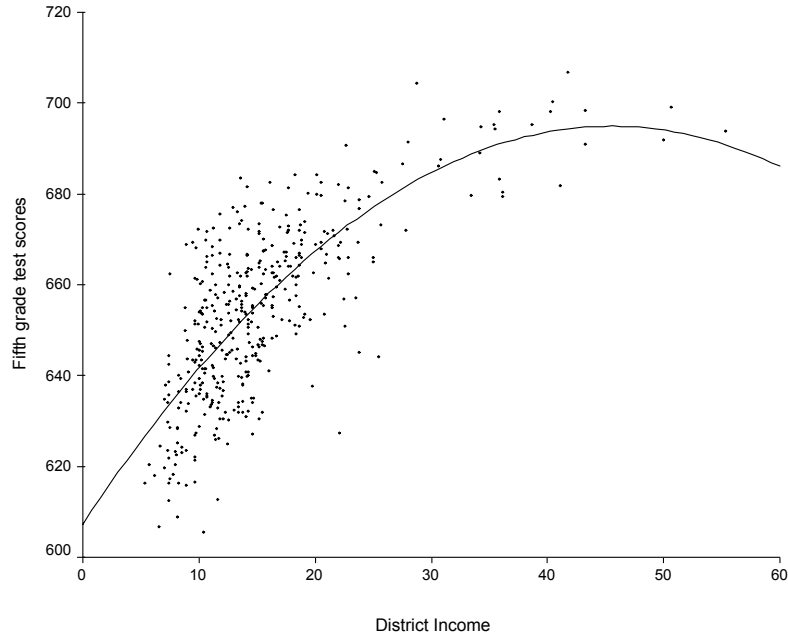
**Table 1** Regression results for Californian Test Score Data Set

Observations: 420				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	607.302	3.046	199.362	.000
X	3.851	0.304	12.657	.000
X^2	-0.042	0.006	-6.758	.000
R-squared	0.556	S.E. of regression		12.724

The *p*-value on the squared income term is significant given a *t*-statistic of -6.76, which indicates that the null hypothesis that the population regression function is linear can be rejected against the alternative that it is quadratic. The extremum of the quadratic function is estimated as 45.511; its estimated standard error based on the Delta method is 3.438. The 95% confidence interval implied by the Delta method is from 38.771 to 52.251. The corresponding 95% confidence interval based on the Fieller method is from 40.218 to 54.881.

Figure 2 plots both the estimated relationship between income and test scores and the straight line that defines the first derivative function and its associated 95% confidence bounds. In Section 3.4, we showed that the points at which the confidence bounds of the first derivative function are equal to zero define an interval that is equivalent to the confidence interval associated

11

with the Fieller method. We have also plotted the estimated confidence interval based on the Delta method.

Figure 2 The plot of the first derivative function and the 95% confidence bounds.

District Income (000's of $)
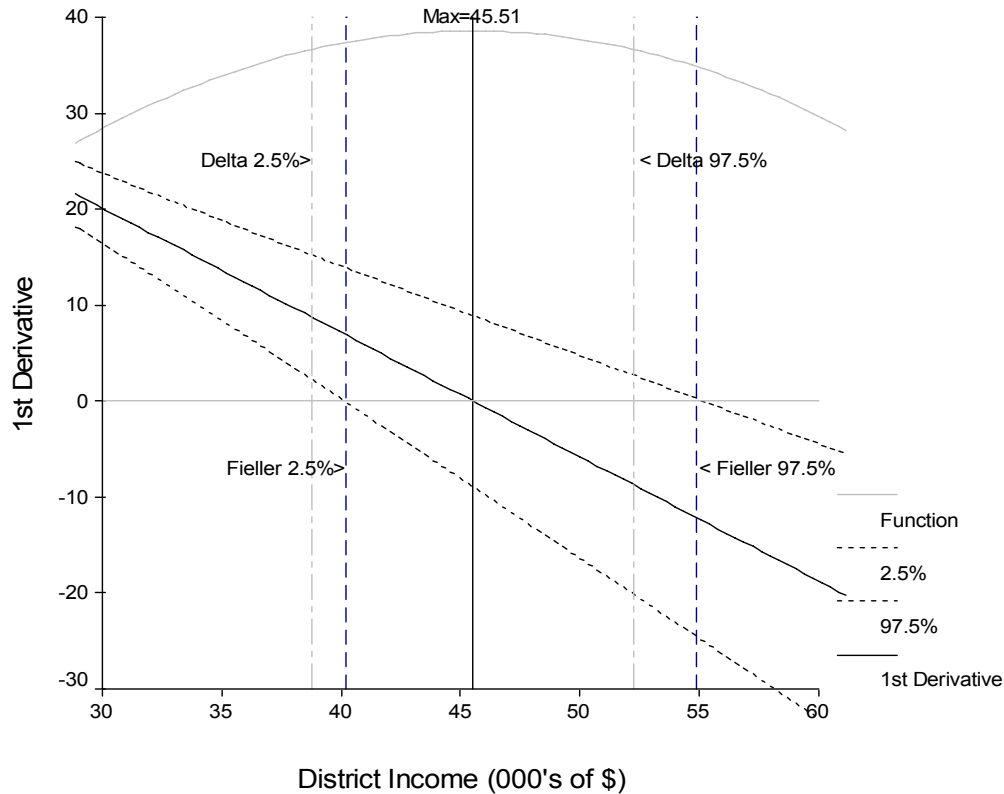
Using a conditional bootstrap with a pseudo sample size of 1000 and a first-order balanced draw that ensures all observations have an equal probability of being drawn, as proposed by Davison et al. (1986), we obtain an interval of 40.022 to 55.290 by using the *percentile* method. Using the Delta method to approximate the standard errors for the *studentized* bootstrap for the same draws yields a 95% interval of 37.036 to 50.478. The Bayesian method, based on assuming a Jeffreys' prior and normally distributed disturbances yields an interval of 40.210 to 54.886. The likelihood ratio interval is computed as 40.224 to 54.861. Table 2 provides a summary of the bounds associated with the different methods for constructing the confidence intervals. From Table 2, the Fieller, the Bayesian, the likelihood ratio and the *percentile* bootstrap intervals are quite similar. While the Delta and the *studentized* bootstrap intervals, both of which rely on the same estimate of the standard error, yield lower bounds.

12

Table 2 Summary of Confidence Interval Bounds for Californian Test Score Data

Method	Lower Bound	Upper Bound
Delta	38.771	52.251
Fieller	40.218	54.881
Studentized	37.036	50.240
Percentile	40.022	55.240
Bayesian	40.040	54.886
Likelihood	40.224	54.861

.

4.2   Deforestation Example

As a counter example, we look at a case in which the evidence is less strong for the existence of an extremum point. Figure 3 is a scatter plot of deforestation (y) and population density (x) for 70 tropical countries based on data from Koop (2000, p. 24). The scatter plot indicates no obvious nonlinear relationship between the two variables although there are a number of outlying observations; the minimum value of x is 0.89, the maximum value is 2,769 and the median is 354. In this case, a maximum would imply that once a country reaches a certain density, deforestation declines.

Figure 3 Scatter plot of deforestation (y) against population density (x)
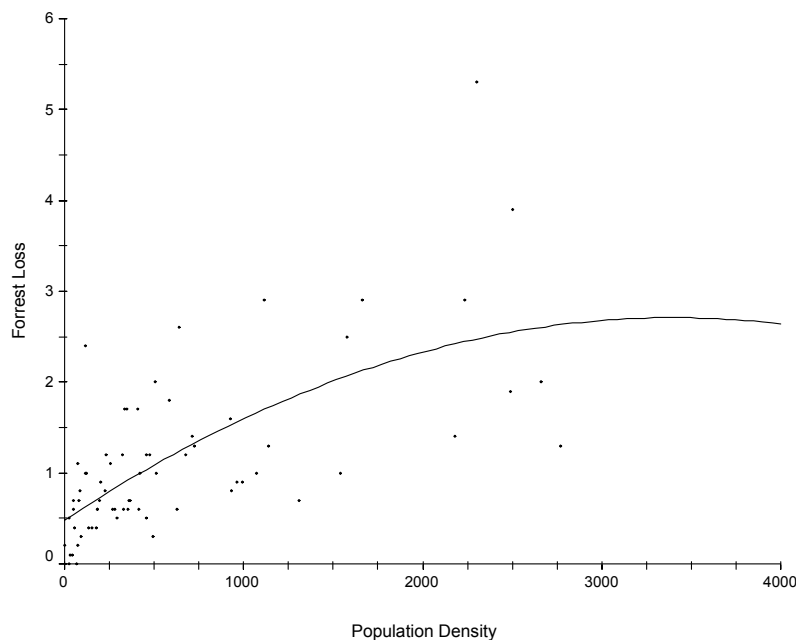

Table 3 reports the results for the quadratic regression. In Table 3 the t-statistic on the squared term is not significant at the 5% level. The estimated extremum of the quadratic function is

3,406. The Delta method 95% confidence interval is from -226 to 7,038; the upper bound is more than twice the maximum.

Table 3 Regression results for Deforestation example.

Observations: 70				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.480	0.150	3.197	0.002
X	0.001	0.000	3.222	0.002
X^2	-1.92E-07	1.60E-07	-1.199	0.235
R-squared	0.446	S.E. of regression		0.701

The bounds defined by a *percentile* bootstrap are given by -11,373 and 17,915. The latter bound is well above the maximum data point. However, the 95% confidence interval based on the *studentized* bootstrap, which is from -5,923 to 3,998 is much narrower. The estimated standard errors of the extremum values in the bootstrap pseudo samples are larger the greater the estimates of coefficients, which results in a very compressed distribution for the $\hat{t}_i$. The Bayesian interval based on the assumption of a Jeffreys' prior and normally distributed disturbances is from -6,227 to 12,377, which is wider than the Delta interval. These methods yield finite confidence intervals even if they are inappropriate.

When the Fieller method is used to construct the 95% confidence limit for the maximum value, we have a case where $(\hat{t}_2^2 - t_{\alpha/2}^2) < 0$ (see equation 5). However, the quadratic equation that defines the Fieller interval still has two real roots. This occurs when the interval is the complement of a finite interval. In this case the complement of a finite interval is the entire real line excluding the interval from -2,036 to 1,987. A plot of the confidence bounds on the first derivative function clarifies the situation.

The likelihood ratio based confidence interval is similar to that of the Fieller, given a lower bound of 2,001.42 and no upper bound. In this case, the first derivative result clearly shows 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 significance level below 23.5%, the Fieller interval indicates that we can reject the hypothesis that the lower bound of the

plateau is less than 1,987 at the 5% level. This is consistent with the hypothesis that deforestation is

monotonically related to population density up to a certain point. Using both of these methods, the

data indicate that there is no extremum point at which the positive relationship between

deforestation and population density ceases and goes into reverse. Table 4 lists the confidence

bounds estimated for the deforestation case.

Table 4 Summary of Confidence Interval Bounds for Deforestation Example

Method	Lower Bound	Upper Bound
Delta	-226	7,038
Fieller	1,987	∞
Studentized	-5,923	3,998
Percentile	-11,373	17,915
Bayesian	-6,227	12,377
Likelihood	2,001	∞

Figure 4 shows that the lower confidence bound for the first derivative function cuts the zero

axis twice while the upper confidence bound for the first derivative function does not cross the zero

axis. When the confidence interval cuts the zero axis at -2,036, since the first derivative function

does not approach zero, the range from -2,036 and below is not relevant for defining the lower

bound. Hence, the bound at 1,986 can be viewed as the relevant lower bound for the first derivative

function and the upper bound is infinite.

Figure 4 The first derivative function and the Fieller and Delta method confidence bounds when $(\hat{t}_2^2 - t_{\alpha/2}^2) < 0$

Population density
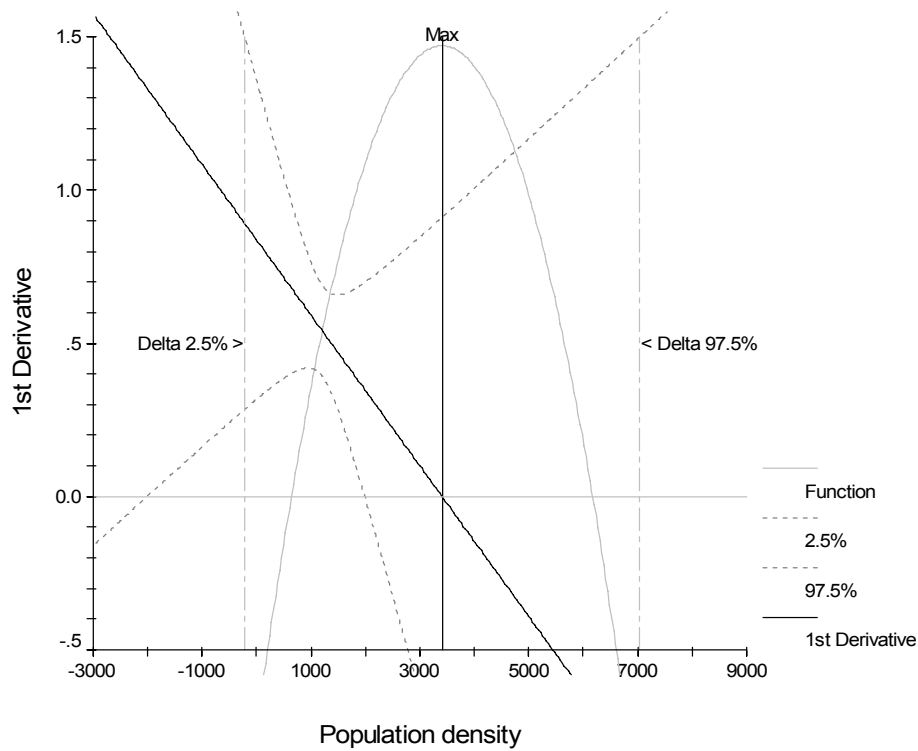
5.   Monte Carlo Experiments to compare the coverage of Confidence bounds.

Monte Carlo experiments were performed to investigate the properties of the confidence

intervals generated by the methods outlined in Section 3. The model used in the experiments is

quadratic in a single regressor with an intercept: $y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$ and

the Californian test score data (from Section 4.1) is used.   In each experiment, β_1 is set to the value

estimated in Section 4.1; that is, 3.851. The values of x range from 10 to 65 and we set values of θ

over the range of x. That is, we modify the location of the extremum by varying the values of β_2. β_0

is varied so that the mean of the dependent variable has the same value in each experiment. We set

σ^2 so that $\beta_2 / \sqrt{\sigma_2^2}$, the '*true*' t-statistic for β_2, equals -1.96, where σ_2^2 is the variance of β_2 .

For each value of the extremum point, 10,000 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 $\tilde{\varepsilon}_t \sim N(0, \tilde{\sigma}^2)$. Once the simulated set of $\tilde{y}_t$ were calculated, the

regressions were run and the estimated extremum value was generated along with 95% confidence

16

bounds based on the Delta method, the Fieller method, the *percentile* bootstrap, the *studentized* bootstrap (with 1000 pseudo samples), the Bayesian method with a Jeffreys' prior and the likelihood ratio method based on a regression with normal errors. The coverage, which is the percentage of replications that resulted in confidence intervals that included the true extremum value (θ), was determined. Table 5 reports the coverage of each confidence interval and the average t-statistic on the quadratic term obtained from each Monte Carlo experiment.

Table 5 The Coverage of the confidence intervals from various methods when the location of the extremum varies.

True Max	Avg t-stat in simulations	Delta	Fieller	BS %	BS t-stat	Bayes	L Ratio
10	-2.71	0.76	0.94	0.94	0.45	0.95	0.94
15	-2.71	0.81	0.94	0.94	0.58	0.96	0.94
20	-2.70	0.94	0.95	0.96	0.92	0.97	0.95
25	-2.70	0.98	0.95	0.96	0.99	0.98	0.95
30	-2.72	0.85	0.94	0.94	0.65	0.96	0.94
35	-2.71	0.77	0.94	0.94	0.48	0.94	0.94
40	-2.70	0.75	0.94	0.94	0.39	0.94	0.94
45	-2.71	0.74	0.94	0.94	0.34	0.94	0.94
50	-2.70	0.74	0.94	0.94	0.32	0.94	0.94
55	-2.70	0.73	0.94	0.94	0.31	0.94	0.94
60	-2.70	0.74	0.94	0.94	0.30	0.94	0.94
65	-2.71	0.73	0.95	0.94	0.28	0.94	0.94

These results indicate that the coverage of the Fieller, *percentile* bootstrap, Bayesian and likelihood ratio methods are consistently close to 95% while the Delta method and the *studentized* bootstrap method, which uses the Delta method to estimate the standard error of the estimated θ, yield confidence intervals that vary considerably with the value of θ; in some cases the true percentage is well below 95%. Figure 5 provides a plot of the coverage from this experiment.

Figure 5 The coverage of the various confidence bounds when the location of the extreme varies.
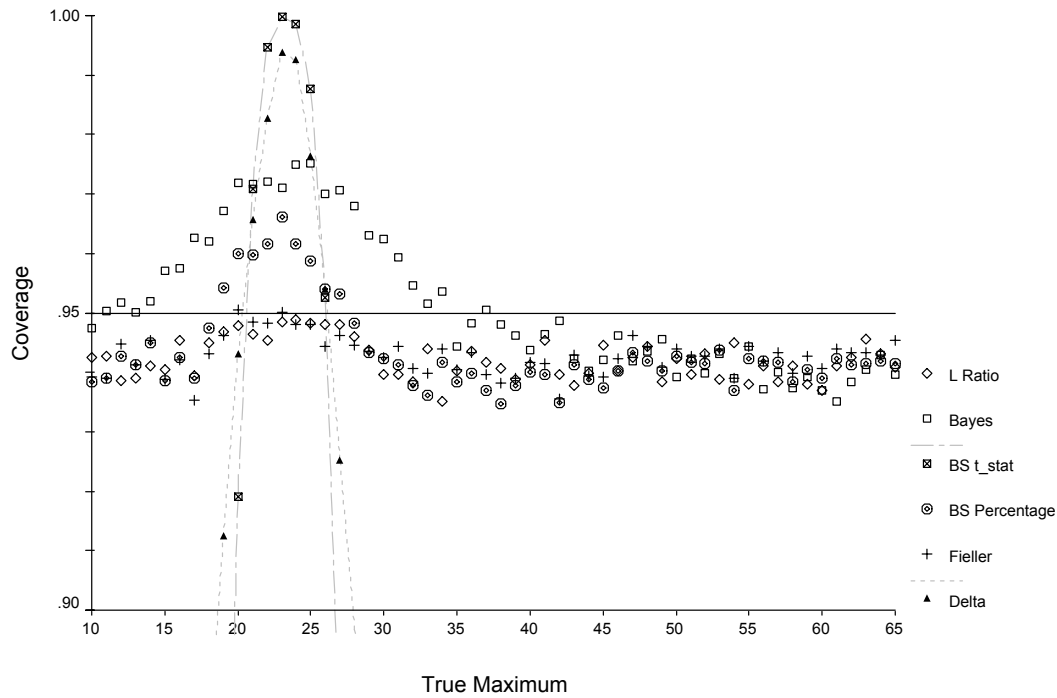
In the next set of Monte Carlo experiments we vary the value of σ^2. The location of the

extremum is set to the value obtained when the actual data for *y* and *x* was used in Section 4.1. That

is, we set the location of the maximum at the value 45.551, by setting β_1 =3.851 and β_2 =-0.042.

β_0 is varied so that the mean of the dependent variable has the same value in each experiment. In

each experiment the value of σ^2 is set to yield a value for the *'true'* t-statistic for β_2,

$\left(\text{that is,} \frac{\beta_2}{\sqrt{\sigma_2^2}}\right)$, that varies from -3 to -12. The coverage estimated from this simulation is displayed

in Table 6 and Figure 6. For values of $\frac{\beta_2}{\sqrt{\sigma_2^2}}$ equal to -6.758 or less, there is very little difference in

the confidence bounds. However, the *studentized* bootstrap continues to provide coverage that is

well below that of the other methods. As the values of $\frac{\beta_2}{\sqrt{\sigma_2^2}}$ goes down to -4, the coverage of the

confidence intervals based on the Delta and *studentized* bootstrap methods tend to be much less

than 95%, while the coverage of other methods slightly exceeds 95%. [2]

---

[2] Another set of Monte Carlo experiments was performed by using data on production costs in which the location of a minimum was of interest. Similar results to those reported here were obtained relating to the coverage of the confidence intervals for each of the intervals.

Table 6 The Coverage of the confidence intervals from various methods when the '*true*' *t*-statistic on β_2 is varied.

"true" t-stat for β_2	*Avg t-stat in simulations*	*Delta*	*Fieller*	*BS* %	*BS* t-stat	*Bayes*	*L Ratio*
-3.0	-3.27	0.90	0.97	0.97	0.74	0.97	0.97
-4.0	-4.06	0.92	0.96	0.97	0.82	0.97	0.96
-5.0	-5.02	0.94	0.95	0.95	0.86	0.95	0.95
-6.0	-6.00	0.94	0.95	0.95	0.89	0.95	0.95
-7.0	-7.00	0.95	0.95	0.95	0.90	0.95	0.95
-8.0	-8.01	0.95	0.95	0.95	0.90	0.95	0.95
-9.0	-9.02	0.95	0.95	0.95	0.91	0.95	0.95
-10.0	-10.03	0.95	0.95	0.95	0.92	0.95	0.95
-11.0	-11.03	0.95	0.95	0.95	0.93	0.95	0.95
-12.0	-12.04	0.95	0.95	0.95	0.93	0.95	0.95

Figure 6 The coverage of the various confidence bounds when the '*true*' t-statistic on β_2 is varied.
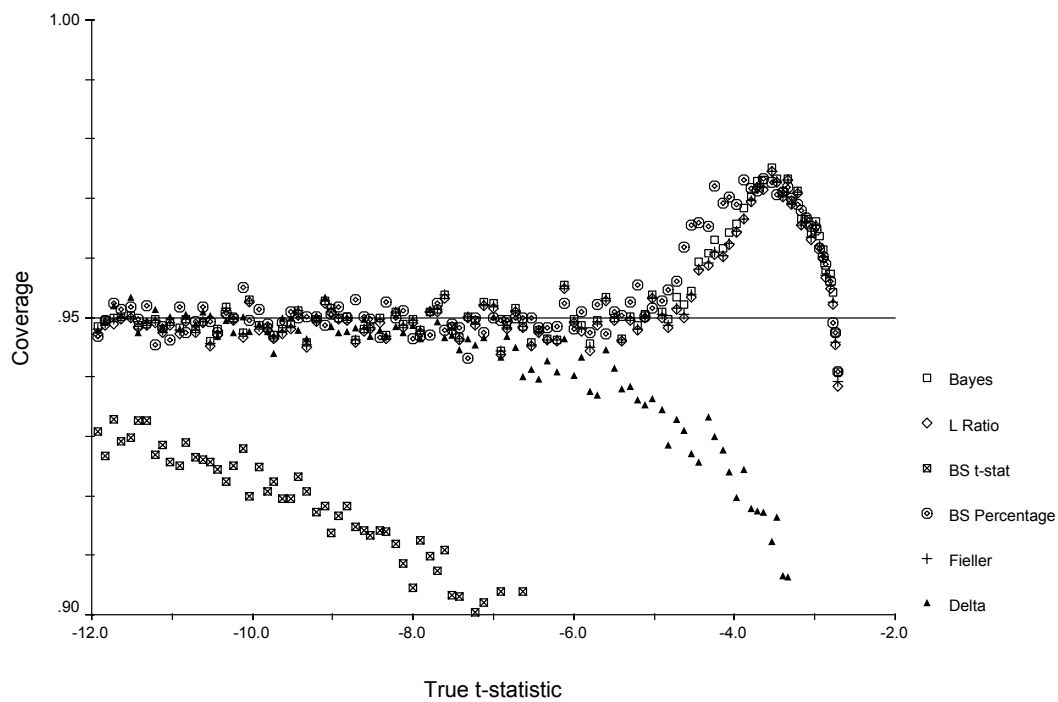
6.   Conclusions

In this paper we propose that inferences about the extremum value of a U- shaped or inverted U- shaped relationship based on the application of a quadratic regression can be made by

19

examining its confidence interval. We investigated the properties of the following methods for this confidence interval: the Delta method/nonlinear regression; the Fieller method/confidence interval for the first derivative function; the likelihood ratio test interval; the *percentile* and *studentized* bootstrap confidence intervals; and a Bayesian posterior distribution. These methods have been proposed in related literature on statistical calibration, slope-ratio assay, parallel-line assay, cost-effectiveness and surrogate endpoints.

From a simulation experiment based on typical data, we found that the methods that rely on a first order Taylor series expansion to approximate the variance of a ratio of regression parameters (namely the Delta method and the *studentized* bootstrap), have poor coverage of the confidence interval; particularly the *studentized* bootstrap. When the coefficient on the square of the regressor (β_2 in equation 1) is estimated with *p*-values of up to .001, coverage for the Delta method is similar to that of any other method. However, when this is not the case, we recommend that researchers avoid relying on the readily available Wald test for a ratio of parameters of a regression or a nonlinear regression in which the extremum is a parameter.

The alternative methods can be categorized by the level of computational and programming effort required to implement them and whether they provide finite intervals in every case. Both the Bayesian method proposed here and the *percentile* bootstrap method, yielded finite confidence intervals with good coverage in the simulation. The Fieller/first derivative function method and the likelihood ratio method also provided good coverage, and they may produce confidence intervals in which one or both of the limits are infinite.

Of the methods we applied, the *percentile* bootstrap relies on the fewest assumptions but requires a high level of computational effort. The Bayesian method assumes specific distributions and requires the use of a complex expression and numerical integration. Alternative Bayesian methods, such as, those employing Gibbs Sampling would involve more computations and specialized software. The likelihood ratio interval for a linear regression with a normal error can be obtained similarly to the way the Fieller interval is obtained by solving a quadratic equation. Any

likelihood function can be used to generate bounds by searching over the repeated evaluation of the likelihood ratio with alternative values of the extremum. A computational advantage of the Fieller method is that the equivalent first derivative function confidence bounds can be plotted by generating out-of-sample predictions and confidence bounds from the estimated regression.

A crucial feature of the Fieller-related and likelihood ratio methods is their ability to generate bounds that may not be finite. This feature may be more useful than the estimation of very large finite bounds. In Section 4.2, we analyzed a case in which the Fieller and the likelihood ratio methods implied the existence of a plateau, which is a value of the regressor beyond which it has no relationship to the dependent variable. Of the techniques that provide good coverage, the method based on the estimated first derivative function is the easiest to implement with existing computer software.

References

Berger, M. C. and Leigh, J. P. (1988), "The Effects of Alcohol use on Wages", *Applied Economics*, 20, 1343-51.

Briggs, A., Mooney, C. and Wonderling, D. (1999), "Constructing Confidence Intervals for Cost-Effectiveness Ratios: An Evaluation of Parametric and Non-Parametric Techniques using Monte Carlo Simulation", *Statistics in Medicine*, 18, 3245-62.

Buonaccorsi, J. P. (1979), "On Fieller's Theorem and the General Linear Model", *The American Statistician*, 33, 162.

Buonaccorsi, J. P. and Gatsonis, C. (1988), "Bayesian Inference for Ratios of Coefficients in a Linear Model", *Biometrics*, 44, 87-101.

Davison, A. C., Hinkley, D. V. and Schechtman, E. (1986), "Efficient Bootstrap Simulation", *Biometrika*, 73, 555-566.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife", *Annals of Statistics*, 7, 1-26.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.

Eviews 5 (2004), *User's Guide*, Quantitative Micro Software, LLC, Irvine CA, USA.

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.

Freedman, D. A. (1981), "Bootstrapping Regression Models", *Annals of Statistics*, 9, 1218-1228.

Freedman, L. (2001), "Confidence Intervals and Statistical Power of the 'Validation' Ratio for Surrogate or Intermediate Endpoints", *Journal of Statistical Planning and Inference*, 96, 143-153.

Ghosh, M., Yin, M. and Kim, Y. (2003), "Objective Bayesian Inference for Ratios of Regression Coefficients in Linear Models", *Statistica Sinica*, 13, 409-422.

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

Jones, G., Wortberg, M., Kreissig, S., Hammock, B. and Rocks, D. (1996), "Application of the Bootstrap to Calibration Experiments", *Analytical Chemistry*, 68, 763-770.

Koop, G. (2000), *Analysis of Economic Data*, Chichester, UK: Wiley.

Kuznets, S. (1955), "Economic Growth and Income Inequality", *American Economic Review*, 45, 1-28.

Lye, J. N. and Hirschberg, J. (2004), "Alcohol Consumption, Smoking and Wages", *Applied Economics*, 36, 1807-1817.

Mikulich, S., Zerbe, G., Jones, R. and Crowley, T. (2003), "Comparing Linear and Nonlinear Mixed Model Approaches to Cosinor Analysis", *Statistics in Medicine*, 22, 3195-3211.

Miller, R. G. (1981), *Simultaneous Statistical Inference*, 2nd edition, Springer-Verlag, New York

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

Morgan, B. J. T. (1992), *Analysis of Quantal Response Data*, New York: Chapman & Hall.

Polsky, D., Glick, H., Willke, R. and Schulman, K. (1997), "Confidence Intervals for Cost-Effectiveness Ratios: A Comparison of Four Methods", *Health Economics*, 6, 243-252.

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

Scheffé, H. (1970), "Multiple Testing versus Multiple Estimation. Improper Confidence Sets. Estimation of Directions and Ratios", *The Annals of Mathematical Statistics*, 41, 1-29.

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

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

Thompson Jr., H. G. and Wolf, L. L. (1993), "Regional Differences in Nuclear and Fossil-Fuel Generation of Electricity", *Land Economics*, 69, 234-248.

Williams, D. (1986), "Interval Estimation of the Median Lethal Dose", *Biometrics*, 42, 641-645.

Williams, E. (1959), *Regression Analysis*, New York: John Wiley and Sons Inc.

Zellner, A. (1978), "Estimation of Functions of Population Means and Regression Coefficients Including Structural Coefficients: A Minimum Expected Loss (MELO) Approach", *Journal of Econometrics*, 8, 127-158.

Zerbe, G. (1978), "On Fieller's Theorem and the General Linear Model", *The American Statistician*, 32, 103-105.

Zerbe, G. (1982), "On Multivariate Confidence Regions and Simultaneous Confidence Limits for Ratios", *Communications in Statistics Theory and Methods*, 11, 2401-2425.