University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
HIRSCHBERG, JG;LYE, JN

Title:
Interactions in regressions

Date:
2005

Citation:
HIRSCHBERG, J. G. & LYE, J. N. (2005). Interactions in regressions. Department of Economics, The University of Melbourne.

Persistent Link:
https://hdl.handle.net/11343/34153

ISSN 0819-2642
ISBN 0 7340 2609 9



THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 952

NOVEMBER 2005

INTERACTIONS IN REGRESSIONS

by

Joe Hirschberg
&
Jenny Lye

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

Interactions in Regressions

J. Hirschberg and J. Lye [1]

Department of Economics
University of Melbourne

November 2005

Abstract

Regression specifications frequently employ regressors that are defined as the product of two other regressors to form an interaction. Unfortunately, these models have a number of potential difficulties when it comes to interpretation. In this paper, we discuss two common aspects of these specifications that can lead to difficulties. The first is the change in parameter estimates and *t*-statistics when different definitions of dummy variables are used. The second involves an application of Fieller's theorem to draw appropriate inferences from a regression with interaction variables.

Key Words: Interaction effects; dummy variables; linear transformation, Fieller method
JEL classification: C12; C51

[1] Corresponding author, jnlye@unimelb.edu.au, Department of Economics, University of Melbourne, Melbourne, Victoria 3010, Australia. 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. We wish to thank David Moreton for helpful comments.

1 Introduction

Regression models with interaction effects are often proposed to allow for the marginal effect of one explanatory variable to depend on another. In this paper, we discuss two aspects of interactions in regression specifications. In Section 2, we show how dummy variables are prone to linear affine transformations and discuss the implications for interpreting an interaction variable in the regression equation when the interaction variable involves a dummy variable. An example is included to highlight the impact on parameter estimates and *t*-statistics. In Section 3, the interpretation of regressions with interactions is proposed by using an application of Fieller's theorem. Two empirical examples are presented. Conclusions are presented in Section 4.

2 Variable Transformations, Interactions and Dummy Variables

Griepentrog et al. (1982) (GRS) show that for models that include an interactive term an affine linear transformation of one variable in the interactive term affects the *t*-statistic for the lower order of the other variable. In this paper, we relate this result to the case in which the interactive term involves a dummy variable. Following GRS, consider a linear model of the form:

$$y = \mathbf{X}\beta + \varepsilon \tag{1}$$

where y is an $n \times 1$ matrix of observations on the dependent variable, ε is an $n \times 1$ matrix of disturbances, X is an $n \times k$ matrix of observations on all of the explanatory variables, that is, $\mathbf{X} = \begin{bmatrix} \mathbf{1} & \mathbf{x} & \mathbf{w} & \mathbf{xw} & \mathbf{v}_1 \cdots \mathbf{v}_{k-4} \end{bmatrix}$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \cdots \beta_k \end{bmatrix}'$ is a $k \times 1$ vector of coefficients.

Suppose that z_t, the t^{th} observation, is formed by an affine linear transformation of x_t that takes the form:

$$z_t = a + bx_t . \tag{2}$$

where *a* and *b* are constants. Consider an alternative model to (1) defined in terms of **Z** as:

$$y = \mathbf{Z}\gamma + \varepsilon \tag{3}$$

where $\mathbf{Z}$ is an $n \times k$ matrix of observations on all of the explanatory variables, that is,

$\mathbf{Z} = \begin{bmatrix} \mathbf{1} & \mathbf{z} & \mathbf{w} & \mathbf{zw} & \mathbf{v}_1 \cdots \mathbf{v}_{k-4} \end{bmatrix}$ and $\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 \cdots \gamma_k \end{bmatrix}'$ is a $k \times 1$ vector of coefficients. The relationship between (1) and (3) is given by:

$$y = \mathbf{XA}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{4}$$

where $\mathbf{A}$ is the non-singular matrix, $\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}'_{12} & \mathbf{A}_{22} \end{bmatrix}$, in which $\mathbf{A}_{11} = \begin{bmatrix} 1 & a & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & 1 & a \\ 0 & 0 & 0 & b \end{bmatrix}$,

$\mathbf{A}_{12} = \mathbf{0}_{4,k-4}$, $\mathbf{A}_{22} = \mathbf{I}_{k\text{-}4,k\text{-}4}$. From (4) it follows that $\hat{\boldsymbol{\gamma}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the ordinary least squares estimator of $\boldsymbol{\beta}$.

We can define the *t*- statistics for testing the null hypotheses $\mathrm{H}_0 : \beta_i = 0,\ i = 1, \cdots, k$ in (1) as $t_i = \dfrac{\hat{\beta}_i}{\sigma\sqrt{c_{ii}}}$, where c_{ij} is the *ij*th element of the matrix $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. The corresponding *t*–statistics for testing the null hypothesis $\mathrm{H}_0 : \gamma_i = 0$ in (3) are $t_i^* = \dfrac{\hat{\gamma}_i}{\sigma\sqrt{d_{ii}}}$, where d_{ii} is the diagonal element of the matrix $\mathbf{A}^{-1}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{A}^{-1}\right)'$. It can be shown that $t_i^* = sign(b)\ t_i$ when i = 2 and 4 and $t_i^* = t_i$ when $i = 5 \cdots k$. However, the *t*-statistics for i = 1 and 3 are defined as

$t_i^* = \dfrac{\hat{\beta}_i - \left(a/b\right)\hat{\beta}_{i+1}}{\sigma\sqrt{c_{ii} - 2\left(a/b\right)c_{i(i+1)} + \left(a/b\right)^2 c_{(i+1)(i+1)}}}$. In particular, the *t*-statistic for testing the hypothesis $\mathrm{H}_0 : \gamma_3 = 0$, which is a test of the significance of the coefficient for variable *w* in (3), is not equivalent to the *t*-statistic for testing the hypothesis $\mathrm{H}_0 : \beta_3 = 0$, that is, the test statistic for the same variable in (1). Thus, a transformation of one of the variables in the interactive term has implications for the *t*-statistic relating to the linear effect of the other variable in the interactive term.

This result is of particular interest when the interacted variable is a dummy variable since these are often subjected to linear affine transformations. In particular, if x_t = 1 when observations

represent the presence of a particular characteristic, we could just as easily define another variable z_t = 1 when the characteristic is not present, that is, $z_t = 1 - x_t$ or in this case, a = 1 and b = –1.

2.1 An Application of Models with Interactive Dummy Variables

This example is based on data from Berndt (1991, p. 193) in the form of observations on 550 individuals from the May 1985 Current Population Survey. The natural logarithm of average hourly earnings in dollars is the dependent variable, *ED* is years of schooling, *EX* is potential years of experience, *MR* is a dummy variable that takes the value 1 if the individual is married and *F* is a dummy variable that takes the value 1 if the individual is female. The following two models are estimated in which being female and being male are used as indicators of gender, respectively:

$$Log(Wage_t)_t = \beta_1 + \beta_2 F_t + \beta_3 MR_t + \beta_4 \left(F_t \times MR_t\right) + \lambda_1 ED + \lambda_2 EX + \lambda_3 EX^2 + \varepsilon_t \qquad (5)$$

$$Log(Wage_t) = \gamma_1 + \gamma_2 \left(1 - F_t\right) + \gamma_3 MR_t + \gamma_4 \left(\left(1 - F_t\right) \times MR_t\right) + \lambda_1 ED + \lambda_2 EX + \lambda_3 EX^2 + \varepsilon \qquad (6)$$

The results are reported in Table 1.

Table 1. A comparison of the results of the specifications (5) and (6)

Dependent Variable: LOG(WAGE)				
Sample: 1 534				
	Gender =F (5)		*Gender* = (1–F) (6)	
Variable	*Coeff*	*t-stat*	*Coeff*	*t-stat*
C	0.562	4.688	0.434	3.375
Gender	–0.128	–1.924	0.128	1.924
MR	0.143	2.471	–0.052	–0.852
Gender×*MR*	–0.195	–2.382	0.195	2.382
ED	0.089	11.153	0.089	11.153
EX	0.034	5.957	0.034	5.957
EX^2	–0.001	–4.202	–0.001	–4.202
R-squared	0.306	S.E. of regression		0.442
Adjusted *R*-squared	0.298	Sum squared resid		103.003

A comparison of the coefficient estimates and *t*-statistics for the Gender variable and the interaction between *Gender* and *MR* in Table 1 indicates that they are of the same magnitude but of opposite sign. However, in model (5), the coefficient estimate and the *t*-statistic associated with *MR* indicate that marriage has a positive and significant impact whereas in (6), the estimated parameter for the marriage dummy is negative and insignificant. Thus, while the coefficient on the redefined dummy variable is expected to change sign, the change in the effect of the unchanged variable (*MR*)

is not intuitive. From Section 2.1, although the parameter on *MR* is expected to change according to $\gamma_3 = \beta_3 - {}^{a}\!/_{b}\beta_4$ (in this case, –.052 = .143 – .195), the change in the *t*-statistic is inversely related to the covariance between $\hat{\beta}_3$ and $\hat{\beta}_4$.

3 The Partial Influence of Interacted Regressors

Consider the model:

$$y = \beta_1 + \beta_2 x + \beta_3 w + \beta_4 xw + \beta_5 v_1 + \cdots \beta_k v_{k-4} + \varepsilon \tag{7}$$

where *y* is the dependent variable, *x*, *w*, *xw*, $\sum_{j=5}^{k} v_{j-4}$ are explanatory variables and ε is the disturbance term. When *x* and *w* are both continuous variables, the marginal effect of a change in *x* is:[2]

$$\frac{\partial \mathrm{E}\left(y|x,w\right)}{\partial x} = \beta_2 + \beta_4 w \tag{8}$$

If *x* is a dummy variable and *w* is a continuous variable, we define the difference in the regression equation when *x* =1 and when *x* = 0 as:

$$\frac{\Delta \mathrm{E}\left(y|x,w\right)}{\Delta x} = \beta_2 + \beta_4 w \tag{9}$$

In the context of (8) and (9), the difficulty of choosing an appropriate value of the other regressor at which to evaluate these computations arises. One approach is to set the value of this other regressor to a particular value, such as the mean, and then make the computation. Another approach is to determine the value of the other regressor at which these definitions become zero or change sign. The value of *w* that results in $\frac{\partial \mathrm{E}(y|x,w)}{\partial x} = 0$ or $\frac{\Delta \mathrm{E}(y|x,w)}{\Delta x} = 0$ is $w^* = -\frac{\beta_2}{\beta_4}$ and is estimated by $\hat{w}^* = -\frac{\hat{\beta}_2}{\hat{\beta}_4}$, where $\hat{\beta}_i$ are the OLS estimates of β_i, i = 2, 4. A confidence bound for $\hat{w}^*$ can be constructed to determine if the impact of *x* on *y* is significant over the possible values of *w*.

[2] Note that a similar analysis can be performed on $\frac{\partial \mathrm{E}(y|x,w)}{\partial w}$

3.1 Confidence Intervals for the Partial Influence Function

For the regression model in (7), $\frac{\partial \mathrm{E}(y|x,w)}{\partial x}$ or $\frac{\Delta \mathrm{E}(y|x,w)}{\Delta x}$ is defined as $\beta_2 + \beta_4 w$. An estimate of this can be plotted with a $100(1-\alpha)$% confidence interval given by:

$$\mathrm{CI} = \left(\hat{\beta}_2 + \hat{\beta}_4 w\right) \pm t_{\alpha/2}\sqrt{\left(\hat{\sigma}_2^2 + 2w\hat{\sigma}_{24} + \mathrm{w}^2\hat{\sigma}_4^2\right)} \tag{10}$$

where $\hat{\sigma}_i^2$ is the estimated variance of $\hat{\beta}_i$, i = 2, 4, $\hat{\sigma}_{24}^2$ is the estimated covariance between $\hat{\beta}_2$ and $\hat{\beta}_4$. An estimate of the value of w ($\hat{w}^*$), where $\frac{\partial \mathrm{E}(y|x,w)}{\partial x} = 0$ or $\frac{\Delta \mathrm{E}(y|x,w)}{\Delta x} = 0$ is found by solving $\hat{\beta}_2 + \hat{\beta}_4 w = 0$. Similarly, the bounds defining the $100(1-\alpha)$% confidence interval on $\hat{w}^*$ are found by solving:

$$\left(\hat{\beta}_2 + \hat{\beta}_4 w\right) \pm t_{\alpha/2}\sqrt{\left(\hat{\sigma}_2^2 + 2w\hat{\sigma}_{24} + \mathrm{w}^2\hat{\sigma}_4^2\right)} = 0 \tag{11}$$

This is equivalent to solving the roots of the equation:

$$\left(\hat{\beta}_2 + \hat{\beta}_4 w\right)^2 - t_{\alpha/2}\left(\hat{\sigma}_2^2 + 2w\hat{\sigma}_{24} + \mathrm{w}^2\hat{\sigma}_4^2\right) = 0 \tag{12}$$

Equation (12) is equivalent to the quadratic equation, ($ax^2 + bx + c = 0$), where a, b and c are $a = \left(\hat{\beta}_4^2 - t_{\alpha/2}^2\hat{\sigma}_4^2\right)$, $b = 2\left(\hat{\beta}_2\hat{\beta}_4 - t_{\alpha/2}^2\hat{\sigma}_{24}\right)$ and $c = \hat{\beta}_2^2 - t_{\alpha/2}^2\hat{\sigma}_2^2$, $t_{\alpha/2}$ is the value from the *t*-distribution with an $\left(\alpha/2\right)$ level of significance and $T–k$ degrees of freedom.

The confidence limits obtained in this way are identical to those found applying a version of the Fieller method (Fieller 1932, 1954) to a ratio of linear combinations of regression parameters (Zerbe 1978). In order to have real roots, that is, to have a finite interval, we require $a > 0$, which implies rejection of the null hypothesis that $\beta_4 = 0$ based on the *t*-statistic (Buonaccorsi 1979). Besides the finite interval, the resulting confidence interval may be the complement of a finite interval, ($b^2 - 4ac > 0$, $a < 0$), or the whole real line, ($b^2 - 4ac < 0$, $a < 0$). These conditions are discussed in Scheffé (1970) and Zerbe (1982).

If $\frac{\partial \mathrm{E}(y|x,w)}{\partial x}$ or $\frac{\Delta \mathrm{E}(y|x,w)}{\Delta x}$ is nonlinear and defined as $g(w)$, then its estimate, say $\hat{g}(w)$, can be plotted along with a $100(1-\alpha)$% confidence interval given by:

$$\text{CI} = \hat{g}(w) \pm z_{\alpha/2}\sqrt{\hat{\text{var}}\{\hat{g}(w)\}} \tag{13}$$

where $\hat{\text{var}}\{\hat{g}(w)\}$ is the estimated variance of $\hat{g}(w)$ and $z_{\alpha/2}$ is the value from the standardized normal distribution with an $(\alpha/2)$ level of significance. An estimate (or estimates) of the value of *w*, where $\frac{\partial \text{E}(y|x,w)}{\partial x}$ or $\frac{\Delta \text{E}(y|x,w)}{\Delta x}$ is found by solving $\hat{g}(w) = 0$; similarly, the bounds defining the $100(1-\alpha)$% confidence interval at this point(s) are found by solving (13) equated to zero. Because $\hat{g}(w)$ is nonlinear, there may be more than one solution and, hence, more than one set of confidence bounds. In many instances, it is possible to rule out a number of the solutions by examining the range of the data.

3.2 Example 1: The Demand for Economic Journals

Stock and Watson (2003, p. 227) analyse the relationship between the number of subscriptions to a journal at US libraries (*Y*) and its library subscriptions price by using data from 2000 on 180 economics journals. *Price* is measured in prices per citation, which represents an approximation of dollars per idea. Other explanatory variables include the age of the journal (*Age*) and the number of characters per year in the journal (*Char*). The regression equation is:

$$\begin{aligned}\text{Log}(Y) = \beta_0 &+ \beta_1 \text{Log}(Price) + \beta_2 \text{Log}(Age) + \beta_3 \left(\text{Log}(Price) \times \text{Log}(Age)\right) \\ &+ \beta_4 \text{Log}(Char) + u\end{aligned} \tag{14}$$

The regression results are reported in Table 2, which shows that the coefficient on the interactive term is highly significant. The price elasticity is:

$$\eta = \beta_1 + \beta_3 \left(\text{Log}(Age)\right) \tag{15}$$

which is estimated as:

$$\hat{\eta} = -0.899 + 0.141 \times \text{Log}(Age) \tag{16}$$

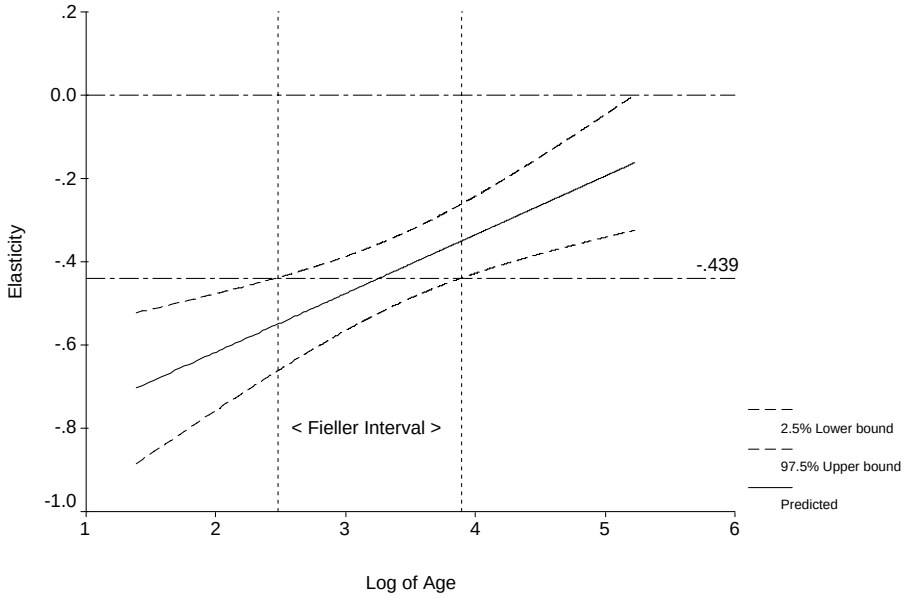
The ratio $\left(0.899/0.141\right)$ defines the value of Log(*Age*) at which $\hat{\eta} = 0$.

Table 2: Regression Results, Demand For Economic Journals

Dependent Variable: Log(Y)		
Sample: 1 180		
Variable	*Coefficient*	*t-Statistic*[3]
C	3.433	9.344
Log(Price)	–0.899	–6.214
Log(Age)	0.374	3.178
*Log(Price)*Log(Age)*	0.141	3.507
Log(Char)	0.229	2.378
R-squared	0.633 S.E. of regression	0.688
Adjusted R-squared	0.626 Sum squared resid	82.738

Figure 1 plots the estimated price elasticity and its 95% corresponding confidence interval. Since the actual values of *Age* in the data have a minimum value of 4 (Log(*Age*)=0.602) and a maximum value of 156 (Log(*Age*)=2.193), we conclude from Figure 1 that the price elasticity is significantly different from 0 for all values of *Age* in this data set.

Figure 1: Partial Influence Function, Demand For Economics Journals Example



However, it is common practice to report a summary price elasticity such as the value of the mean of all the price elasticities calculated by using (16), which in this example equals –0.439. Figure 1 plots a reference line at this value from which we can determine the values of *Age* for which the elasticity is significantly different from the mean of all the price elasticities. In this

[3] The *t*-statistics are based on White Heteroskedastic-Consistent Standard errors

example, this is the case for young and old journals, that is, those for which *Age* is less than 11.01 (Log(*Age*) = 2.4) or greater than 49.4 (Log(*Age*) = 3.9), respectively.

3.3 Example 2: California Test Data Scores

In this example, data from Stock and Watson (2003, p. 230) is used on test scores from 420 California school districts in 1998. The test score measure (*TESTSCR*) is the district-wide average of reading and math scores for fifth graders. Additional variables are the district-wide student teacher ratio (*STR*); the percentage of students qualifying for a subsidized lunch (*Meal_pct*); and the average annual per capita income in the school district measured in thousands of 1998 dollars (*Avginc*). Further, *HIEL* is a dummy variable that equals 1 if the percentage of students still learning English in the district is greater than 10%. The regression equation examines whether the effect of the student-teacher ratio depends not only on the value of the student-teacher ratio but also on the percentage of English learners. Interactions between *HIEL* and *STR*, STR^2 and STR^3 are included to allow the regression functions relating test scores and *STR* to be different for low and high percentages of English learners. The regression equation is:

$$\begin{aligned} TESTSCR = {} & \beta_0 + \beta_1 STR + \beta_2 STR^2 + \beta_3 STR^3 + \beta_4 HIEL + \beta_5 (STR \times HIEL) \\ & + \beta_6 (STR^2 \times HIEL) + \beta_7 (STR^3 \times HIEL) + \beta_8 Meal_Pct + \beta_9 \log(Avginc) + \varepsilon \end{aligned} \tag{17}$$

Table 3 reports the regression results. Individually, β_4, β_5, β_6, β_7 are significant at the 5% level and a joint F-test of $H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ results in a *p-value* of less than 0.01.

The difference in the test scores when *HIEL* takes the value 1 and when it takes the value 0 is defined as:

$$\frac{\Delta E(Testscr|x)}{\Delta HIEL} = \beta_4 + \beta_5 (STR) + \beta_6 (STR^2) + \beta_7 (STR^3) \tag{18}$$

which is estimated as:

$$\widehat{\frac{\Delta E(Testscr|x)}{\Delta HIEL}} = 816.07 - 123.28 \times STR + 6.12 \times STR^2 - 0.10 \times STR^3 \tag{19}$$

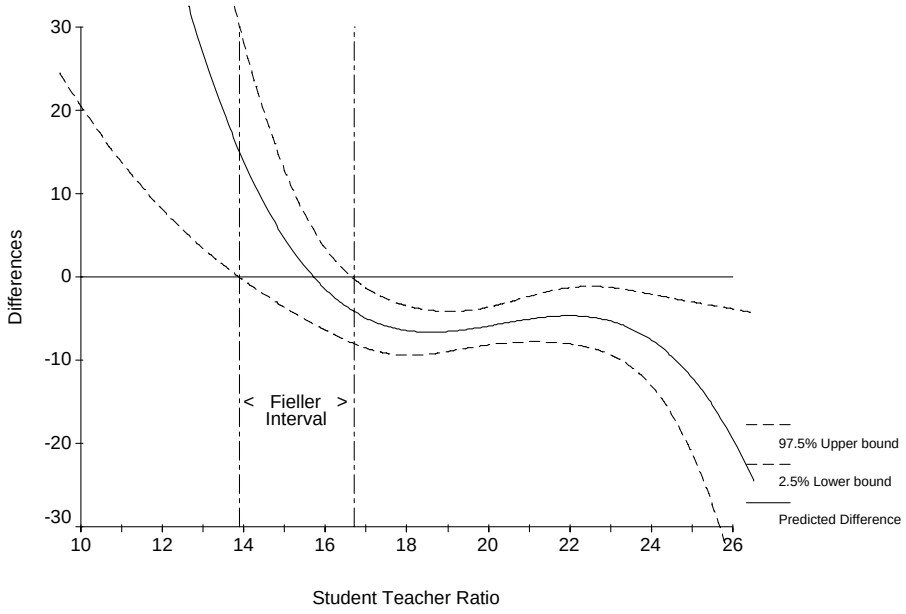
Equation (19) only has a real root when *STR* = 15.7.

Table 3: Regression Results, California Test Data Scores

Dependent Variable: *TESTSCR*			
Sample: 1 420			
Variable	*Coefficient*	*t-Statistic*[4]	
C	122.354	0.659	
STR	83.702	2.937	
STR^2	–4.381	–3.040	
STR^3	0.075	3.121	
HIEL	816.075	2.491	
*STR*HIEL*	–123.282	–2.455	
STR^2*HIEL*	6.121	2.408	
STR^3*HIEL*	–0.101	–2.366	
Meal_pct	–0.418	–14.556	
LOG(*Avginc*)	11.800	6..637	
R-squared	0.803	S.E. of regression	8.547
Adjusted R-squared	0.798	Sum squared resid	29954.410

Figure (2) plots (19) and the corresponding 95% confidence interval. We find that $\frac{\widehat{\Delta E(Testscr|x)}}{\Delta HIEL}$ is significantly different from zero for values of *STR* that are greater than 16.7 and less than 13.9. However, only the former range (values from 16.7) overlaps with the actual range of *STR* values (14.0 to 25.8). Thus, we can conclude that the effect of student-teacher ratios (*STR*s) in schools with a high proportion of students learning English is only negative for those schools that have *STR*s of over 16.7.

Figure 2: Partial Influence Function, California Test Data Scores Example



[4] The *t*-statistics are based on White Heteroskedastic-Consistent Standard errors.

4 Conclusions

Interactive terms are often included in regression equations to determine how the effect on the dependent variable of one independent variable depends on another independent variable. However, an affine linear transformation of one of the variables in the interactive term has the counterintuitive effect of changing the *t*-statistic on the linear effect of the other variable. We have shown the implications of this when interaction terms are based on dummy variables, which are prone to affine linear transformations. In addition, we have also shown that inferences can be drawn from a regression with interaction variables by examining the confidence bounds of the partial influence function. In the context of linear regression, this is equivalent to the application of Fieller's Method for the construction of confidence bounds for ratios of random variables.

References

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

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

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.

Griepentrog, G. L., J. M. Ryan and D. Smith (1982), "Linear Transformations of Polynomial Regression Models", *The American Statistician*, 36, 171–174.

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

Stock, J. H. and M. W. Watson, (2003), *Introduction To Econometrics*, Addison Wesley, Boston, USA.

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.