

*I<sup>2</sup>SDS*  
*The Institute for Integrating Statistics in Decision Sciences*

*Technical Report TR-2008-8*  
**June 11, 2008**

**Ceteris Paribus and Goodness of Fit for the Coefficient  
in Simple Linear Regression**

David Zalkind  
*Department of Decision Sciences*  
*The George Washington University, USA*

## *Ceteris Paribus* and Goodness of Fit for the Coefficient in Simple Linear Regression

David Zalkind

Department of Decision Sciences, George Washington University, Washington D.C. 20052

DLZ@gwu.edu

### **Introduction**

When results are obtained from data based modeling or a study, there may be observable or unobservable factors that could affect the findings if data were to be obtained again and the modeling effort or the study were to be repeated. The notion of *ceteris paribus*, or "all things being the same", refers to such factors not changing enough to affect the results. In order that a model or study results be considered to apply beyond the data from which they were derived, the *ceteris paribus* assumption might be invoked to assert that the findings would remain undisturbed.

Some examples of the appearance of this phrase in recent Management Science papers are

- (i) "The positive effect of age suggests that, the longer the retail unit has existed, the higher was its aspiration level, *ceteris paribus*." (Mezias et al. 2002)
- (ii) "If cardiac surgeons at a particular hospital achieve significantly better outcomes with CABG than cardiologists do with PTCA, one would expect that facility to have a higher CABG rate *ceteris paribus*." (Huckman 2003)
- (iii) "The significant coefficient of 0.0114 on MAS% indicates that, *ceteris paribus*, a first-mover firm with an extra growth of 1% in MAS% from 1995 to 1999 experienced an incremental productivity growth of 0.97%." (Banker et al. 2005)

(iv) "LOANREV is an ongoing periodic revenue that positively affects future profits, and DEPEXP is an ongoing periodic cost that negatively affects future profits, *ceteris paribus*." (Nagar and Rajan 2005)

(v) "Suppose there is a shock to the expected return, *ceteris paribus*. In this example a change of 1% in  $r$  leads to approximately a 20% change in asset value." (Ferson et al. 2005)

In each of these examples, the invocation of *ceteris paribus* appears to be for the purpose of extending a researcher's claims from (apparent) internal validity of a model or study to external validity, by resorting to "hand-waving" rather than pursuing formal or even informal analysis as a foundation for the expanded claims. We believe that at best this type of use of *ceteris paribus* is tautological or vacuous. Furthermore, it can be misleading or wrong. In the context of obtaining data and constructing simple linear regression models, we derive counterintuitive results when a particular implementation of *ceteris paribus* is invoked.

Typically, goodness of fit of a simple linear regression model is measured with respect to the observed data points and relates primarily to the response variable. The coefficient of determination,  $R^2$ , is calculated to reveal how much of the variation of the response variable is "explained" by variation in the predictor variable. All things being the same for a given sample size and correctly specified model, one might suppose that the greater the value "explained" the better. Also, a standard error is calculated for the slope in the regression model,  $SE(b)$ . All things being the same for a given sample size and correctly specified model, one might suppose that the smaller the standard error the better, since the size of the standard error is related to the width of the confidence interval for the corresponding population parameter, i.e., the true slope.

Dallal (2008) makes the following observations regarding simple linear regression models:

Sometimes the same model is fitted to two different populations. For example, an [*sic*] researcher might wish to investigate whether weight predicts blood pressure

in smokers and nonsmokers and, if so, whether the regression model fits one group better than the other. The problem with questions like this is that the answer depends on what we mean by *better*.

It is common to hear investigators speak of the model with the larger coefficient of determination,  $R^2$ , as though it fits better because it accounts for more of the variability in the response. However, it is possible for the model with the smaller  $R^2$  to have the smaller standard error of the estimate and make more precise predictions.

He provides a small dataset that illustrates this behavior and discusses possible meanings of the model fitting one group better than the other. When a regression modeler wants to understand how the response variable (dependent variable) changes with changes in a predictor variable (independent variable), we believe that an appropriate measure of goodness of fit should relate to the coefficient of the predictor variable.

In the context of *ceteris paribus*, our results are stronger than those alluded to by Dallal. In the context of simple linear regression for models derived from two sample, using a plausible interpretation of *ceteris paribus* we show that not only does the model with the larger  $R^2$  not necessarily provide the better fit, but that the model with the smaller standard error of the coefficient provides a poorer fit of the coefficient in the model to the corresponding parameter in the population.

Standard formulas for analyses using simple linear regression may be found in texts such as Draper and Smith (1998) and Chatterjee and Price (1977). We rely on such formulas as the starting points for our analyses and do not specifically reference sources for them.

For convenience, we use the following notation conventions throughout the paper:

(a) Capital English letters or small Greek letters, such as  $Y$  or  $\beta$  represent entities related to the *population* such as a variable or a parameter.

(b) Small bolded English letters, such as  $\mathbf{x}$ , represent an ordered set of observations for a variable. Assume that the sample size is  $n$ . Small subscripted unbolded English letters, such as  $x_i$ , represent individual observations, e.g.,  $\mathbf{x} = (x_1, \dots, x_n)$

(c) We use suggestive, not necessarily standard, representations of statistical concepts, such as  $\text{Var}()$  for variance,  $\text{SE}()$  for standard error,  $\text{Cov}()$  for covariance,  $\text{Corr}()$  for correlation, and  $\text{sqrt}()$  for square root. Note:  $\text{Var}(X)$ , using capital  $X$ , is the variance of the variable  $X$  in the population.  $\text{Var}(\mathbf{x})$ , using small bolded  $\mathbf{x}$ , is the estimate of the population variance based on the sample  $\mathbf{x}$ . We use  $r^2_{xy}$  and  $R^2$  interchangeably.

Suppose that there is a linear relationship in the population that may be represented as

$$Y = \beta X + D, \quad (1)$$

where  $Y$  is the response variable,  $X$  is the predictor variable,  $\beta$  is a parameter that is the average change in  $Y$  for a unit change in  $X$ , and  $D$  is a variable representing "displacements" that cause  $Y$  not to be merely a linear transformation of  $X$ . We assume that  $X$  has a finite mean and a non-zero finite variance. We make the same assumptions for  $D$ . The mean of the displacements variable is not restricted to be 0, so we do not lose generality by omitting an explicit constant term in equation (1).

In order that the model we derive be *correctly specified*, we assert that  $\beta$  fully captures the linear relationship between  $X$  and  $Y$ . This assumption implies that  $\text{Corr}(X,D) = 0$ , as shown in the following lemma.

**Lemma 1**

If equation (1),  $Y = \beta X + D$ , represents the true linear relationship in the population, then

$$\text{Corr}(X,D) = 0.$$

Proof: Express the full linear relationship (if any) between  $D$  and  $X$  as  $D = \beta' X + D'$ , where

$\text{Corr}(X,D') = 0$ . Substituting, in equation (1), we obtain

$$Y = \beta X + D = \beta X + (\beta' X + D') = (\beta + \beta') X + D'. \quad (1a)$$

Since  $\beta$  captures the true linear relationship between  $X$  and  $Y$ , we must have  $\beta + \beta' = \beta$ , implying  $\beta' = 0$ . It follows that  $D = D'$ . So  $\text{Corr}(X,D) = \text{Corr}(X,D') = 0$ .

QED

Assuming a sample size  $n$ , let  $\mathbf{x}$  and  $\mathbf{y}$  represent vectors of length  $n$  of observed values of  $X$  and  $Y$ , respectively. Let  $\mathbf{d}$  be a vector of length  $n$  representing corresponding values of  $D$ . These values of  $\mathbf{d}$  "occur", but are not observable by the typical researcher. However, when we do computer simulation studies, we will generate  $\mathbf{d}$  and  $\mathbf{x}$  and calculate  $\mathbf{y}$  for a chosen value of  $\beta$ . Thus, we can observe  $\mathbf{d}$ .

For any sample we have a "true" population relationships corresponding to equation (1):

$$\mathbf{y} = \beta\mathbf{x} + \mathbf{d} . \tag{1b}$$

A standard OLS approach produces a regression model of the form

$$\mathbf{y}_{\text{est}} = \mathbf{b}\mathbf{x} + \mathbf{a} , \tag{2}$$

where  $\mathbf{b}$  and  $\mathbf{a}$  are derived constants and  $\mathbf{y}_{\text{est}}$  represents the estimate for  $\mathbf{y}$  based on the regression model. There is a corresponding "full" simple linear regression model of the form

$$\mathbf{y} = \mathbf{b}\mathbf{x} + \mathbf{a} + \mathbf{e} , \tag{2a}$$

where  $\mathbf{e} = \mathbf{y} - \mathbf{y}_{\text{est}}$  is the vector of residuals.

Assume that two samples are taken from the population described by equation (1) and that models as described by equation (2) are constructed from each of these samples. Suppose that we invoke *ceteris paribus* to assure that the second model would be numerically "substantially" the same as the first. What factors do we mean to be equal from the first sample to the second? Do we merely have the uninformative condition that whatever factors need to be substantially the same *are* substantially the same so that the numerical results for the two models are substantially the same? If so, the concept of *ceteris paribus* is not helpful. It is vacuous or tautological.

There may be many identifiable conditions in this context that could plausibly be invoked by the *ceteris paribus* concept. For our explorations, the *ceteris paribus* concept is that  $\sqrt{\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x})}$  is the same for both samples, i.e., the ratio of the standard deviations of the

displacements to the carrier variables in each sample are the same. Theorem 1, regarding the distance between  $b$  and  $\beta$ , lays the ground work for exploring the implications of this assumption.

**Theorem 1:**

$$b - \beta = \text{Corr}(\mathbf{x}, \mathbf{d}) \sqrt{(\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x}))} = \text{Corr}(\mathbf{x}, \mathbf{d})w = cw, \quad (3)$$

where  $c = \text{Corr}(\mathbf{x}, \mathbf{d})$  and  $w = \sqrt{(\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x}))}$ .

NOTE: Fixing  $w = \sqrt{(\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x}))}$  is our *ceteris paribus* invariant.

Proof: Taking the covariance with respect to  $\mathbf{x}$  of both sides of equation (1b), we have

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = \text{Cov}(\mathbf{x}, \beta \mathbf{x} + \mathbf{d}) = \beta \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{d}). \quad (3a)$$

Similarly taking the covariance with respect to  $\mathbf{x}$  of both sides of equation (2a) we have

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = b\text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{a}) + \text{Cov}(\mathbf{x}, \mathbf{e}). \quad (3b)$$

The last term in equation (3b) is 0 by construction of the regression equation and the next to last term is also 0 since the covariance of a variable with a constant is 0. Thus we have the well known relationship (usually written with terms rearranged)

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = b \text{Var}(\mathbf{x}) \quad (3c)$$

Since the left hand sides of equations (3a) and (3c) are the same, we equate the right hand sides, obtaining

$$\beta \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x}, \mathbf{d}) = b \text{Var}(\mathbf{x}) \quad (3d)$$

Rearranging terms in equation (3d) we can obtain

$$\begin{aligned} b - \beta &= \text{Cov}(\mathbf{x}, \mathbf{d})/\text{Var}(\mathbf{x}) = (\text{Corr}(\mathbf{x}, \mathbf{d}) \sqrt{(\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x}))}) / \text{Var}(\mathbf{x}) \\ &= \text{Corr}(\mathbf{x}, \mathbf{d}) \sqrt{(\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x}))} = \text{Corr}(\mathbf{x}, \mathbf{d})w = cw \end{aligned} \quad (3e)$$

QED

**Corollary 1.1:**

$$\text{abs}(b - \beta) = \text{abs}(c)w \quad (3f)$$

Proof: This follows immediately from equation (3).

QED

We will use  $\text{abs}(b - \beta)$  as the measure of goodness of fit of the model to the population, since we are interested in how well we estimate the change in the response variable with a unit change in the predictor variable. (Note: We could also use  $(b - \beta)^2$ .)

**Corollary 1.2:**

If  $w \neq 0$ , then  $b = \beta$  if and only if  $c = 0$ .

Proof: From equation (3), when  $w \neq 0$ , we see that  $b - \beta = 0$  if and only if  $c = 0$ . Thus,  $b = \beta$  if and only if  $c = 0$ .

QED

As a practical matter, the case where  $w=0$  would not occur, since that would require  $\text{Var}(\mathbf{d})=0$ .

**Corollary 1.3:**

The maximum possible value for  $\text{abs}(b - \beta)$  is  $w$ .

Proof: Since the maximum value of  $\text{abs}(c)$  is 1, Corollary 1.1 implies that the maximum possible value for  $\text{abs}(b - \beta)$  is  $w$ .

QED

We might intuitively use confidence intervals to get a sense of the distance between  $b$  and  $\beta$ . As a confidence level approaches 1.0, the width of the corresponding confidence interval approaches infinity, i.e., the possible distance between  $b$  and  $\beta$  is unbounded. However, using the concept of  $w$ , we can express a bound for the distance between  $b$  and  $\beta$ .

**Exploring SE(b)**

We now explore implications of the invocation of our implementation of *ceteris paribus* with regard to the relationship between  $\text{SE}(b)$  and  $\text{abs}(b - \beta)$ . The formula for the standard error of the regression coefficient  $b$  is typically given as

$$\text{SE}(b) = \sqrt{\text{SSE} / (n-2)} / \sqrt{\text{SS}_{xx}} \quad , \quad (4)$$

where  $SSE = \sum (y_i - \hat{y}_{est i})^2 = \sum e_i^2$  and  $SS_{xx} = \sum (x_i - \bar{x}_{avg})^2$ .

We can rewrite SSE as  $(n-1)\text{Var}(\mathbf{y})$  and  $SS_{xx}$  as  $(n-1)\text{Var}(\mathbf{x})$  and substitute into equation (4) to obtain

$$\begin{aligned} SE(b) &= \sqrt{\left( \frac{\text{Var}(\mathbf{e}) (n-1)}{(n-2)} \right) / \left( \text{Var}(\mathbf{x}) (n-1) \right)} \\ &= \sqrt{\left( \frac{1}{(n-2)} \right)} \sqrt{\text{Var}(\mathbf{e}) / \text{Var}(\mathbf{x})} \end{aligned} \quad (4a)$$

Thus, for a fixed sample size,  $n$ ,  $SE(b)$  is proportional to  $\sqrt{\text{Var}(\mathbf{e})/\text{Var}(\mathbf{x})}$ . We create the term

$$pSE(b) = \sqrt{\text{Var}(\mathbf{e}) / \text{Var}(\mathbf{x})} = [\sqrt{1/(n-2)}] SE(b) \quad (4b)$$

to mean "proportional to  $SE(b)$ " and use it in place of  $SE(b)$ . This suffices for the derivations below since we produce closed form results that are independent of sample size.

### **Lemma 2**

$$\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{x})(\beta^2 + 2\beta cw + w^2) \quad (4c)$$

Proof: From equation (1a) we have

$$\begin{aligned} \text{Var}(\mathbf{y}) &= \text{Var}(\beta \mathbf{x} + \mathbf{d}) = (\beta^2 \text{Var}(\mathbf{x}) + 2\beta \text{Cov}(\mathbf{x}, \mathbf{d}) + \text{Var}(\mathbf{d})) \\ &= (\beta^2 \text{Var}(\mathbf{x}) + 2\beta (b - \beta) \text{Var}(\mathbf{x}) + \text{Var}(\mathbf{d})) \\ &= \text{Var}(\mathbf{x})(\beta^2 + 2\beta (b - \beta) + \text{Var}(\mathbf{d})/\text{Var}(\mathbf{x})) \\ &= \text{Var}(\mathbf{x})(\beta^2 + 2\beta (b - \beta) + w^2) \end{aligned} \quad (4d)$$

Using Theorem 1, substitute for  $(b - \beta)$  in the right hand side of equation (4d) to get

$$\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{x})(\beta^2 + 2\beta cw + w^2)$$

QED

### **Theorem 2**

$$\text{Var}(\mathbf{e}) = (1 - c^2)\text{Var}(\mathbf{d}) \quad (5)$$

Proof: We begin with two standard expressions for  $R^2$ :

$$R^2 = r_{xy}^2 = \text{Corr}(\mathbf{x}, \mathbf{y})^2 = \text{Cov}(\mathbf{x}, \mathbf{y})^2 / (\text{Var}(\mathbf{x}) \text{Var}(\mathbf{y})) \quad (5a)$$

and

$$R^2 = 1 - \text{Var}(\mathbf{e})/\text{Var}(\mathbf{y}) \quad (5b)$$

Setting the right hand sides of equations (5a) and (5b) equal and rearranging terms, we can obtain

$$\text{Var}(\mathbf{e}) = \text{Var}(\mathbf{y}) - \text{Cov}(\mathbf{x},\mathbf{y})^2 / \text{Var}(\mathbf{x}) . \quad (5c)$$

Now, take the covariance of equation (1a) with respect to  $\mathbf{x}$  and then square both sides to obtain

$$\text{Cov}(\mathbf{x},\mathbf{y})^2 = (\beta \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x},\mathbf{d}))^2 . \quad (5d)$$

Use the expression for  $\text{Var}(\mathbf{y})$  from the first line of equation (4d) and substitute this and the right hand side of equation (5d) into equation (5c) to obtain

$$\begin{aligned} \text{Var}(\mathbf{e}) &= (\beta^2 \text{Var}(\mathbf{x}) + 2\beta \text{Cov}(\mathbf{x},\mathbf{d}) + \text{Var}(\mathbf{d})) \\ &\quad - (\beta \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x},\mathbf{d}))^2 / \text{Var}(\mathbf{x}) . \end{aligned} \quad (5e)$$

Expanding the square,  $(\beta \text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x},\mathbf{d}))^2$ , on the right hand side of equation (5e) and canceling terms we can get

$$\begin{aligned} \text{Var}(\mathbf{e}) &= (\beta^2 \text{Var}(\mathbf{x}) + 2\beta \text{Cov}(\mathbf{x},\mathbf{d}) + \text{Var}(\mathbf{d})) \\ &\quad - (\beta^2 \text{Var}(\mathbf{x})^2 + 2\beta \text{Cov}(\mathbf{x},\mathbf{d})\text{Var}(\mathbf{x}) + \text{Cov}(\mathbf{x},\mathbf{d})^2) / \text{Var}(\mathbf{x}) \\ &= (\beta^2 \text{Var}(\mathbf{x}) + 2\beta \text{Cov}(\mathbf{x},\mathbf{d}) + \text{Var}(\mathbf{d})) \\ &\quad - (\beta^2 \text{Var}(\mathbf{x}) + 2\beta \text{Cov}(\mathbf{x},\mathbf{d}) + \text{Cov}(\mathbf{x},\mathbf{d})^2 / \text{Var}(\mathbf{x})) \\ &= (\beta^2 \text{Var}(\mathbf{x}) - \beta^2 \text{Var}(\mathbf{x})) + (2\beta \text{Cov}(\mathbf{x},\mathbf{d}) - 2\beta \text{Cov}(\mathbf{x},\mathbf{d})) \\ &\quad + \text{Var}(\mathbf{d}) - \text{Cov}(\mathbf{x},\mathbf{d})^2 / \text{Var}(\mathbf{x}) \end{aligned} \quad (5f)$$

Converting  $\text{Cov}(\mathbf{x},\mathbf{d})$  to  $\text{Corr}(\mathbf{x},\mathbf{d})\sqrt{\text{Var}(\mathbf{d})\text{Var}(\mathbf{x})}$ , squaring and substituting into the right hand side of equation (5f), we obtain

$$\begin{aligned} \text{Var}(\mathbf{e}) &= \text{Var}(\mathbf{d}) - [\text{Corr}(\mathbf{x},\mathbf{d})^2 \text{Var}(\mathbf{x})\text{Var}(\mathbf{d})] / \text{Var}(\mathbf{x}) \\ &= \text{Var}(\mathbf{d}) (1 - c^2) = \text{Var}(\mathbf{d})(1 - \text{Corr}(\mathbf{x},\mathbf{d})^2) = \text{Var}(\mathbf{d}) (1 - c^2) \end{aligned} \quad (5g)$$

QED

One might intuit that minimizing  $\text{Var}(\mathbf{e})$  is a good thing. However, we can show

### **Corollary 2.1**

If  $c^2 = 1$ , then  $\text{Var}(\mathbf{e}) = 0$ .

Proof: This follows immediately Theorem 2.

QED.

Theorem 1 implies that for a given set of  $d_i$ 's and  $x_i$ 's in a sample, the worst fit that can be produced between  $b$  and  $\beta$  occurs when  $c = \text{Corr}(\mathbf{x}, \mathbf{d}) = 1$ . Corollary 2.1 shows us that for samples with the same  $\text{Var}(\mathbf{d})$ , which is a condition that is a bit different than our *ceteris paribus*,  $\text{Var}(\mathbf{e})$  is minimized when  $\text{abs}(b-\beta)$  is maximized. Of course, the maximized value of  $\text{abs}(b-\beta)$  also depends on  $\text{Var}(\mathbf{x})$ , as can be seen from Theorem 1.

**Corollary 2.2**

$$\text{pSE}(b) = \text{sqrt}((1-c^2)w) \quad (5h)$$

Proof: Divide both sides of equation (5) by  $\text{Var}(\mathbf{x})$  and take the square root of both sides.

QED

Recall that we can use  $\text{pSE}(b)$  in place of  $\text{SE}(b)$  until we need an explicit  $n$  (see equation (4b)).

One might anticipate that samples that produce simple linear regression models with smaller values of  $\text{SE}(b)$  would tend to yield better, i.e., smaller, values of  $\text{abs}(b-\beta)$ , since confidence intervals for  $\beta$  would be tighter (for given confidence levels). However, when our *ceteris paribus* condition holds, this is wrong.

**Theorem 3:**

Assume that equation (1) fully captures the linear relationship in the population and *ceteris paribus* ( $w$  is constant) holds for two samples of size  $n$ . Let  $b_1$  and  $b_2$  be the estimates for  $\beta$  from regression models from the two samples of size  $n$ . If  $\text{SE}(b_1) < \text{SE}(b_2)$  then  $\text{abs}(b_1 - \beta) > \text{abs}(b_2 - \beta)$ .

Proof: From Corollary 1.1 we have

$$\text{abs}(c) = \text{abs}(b-\beta)/w \quad (6)$$

Because the  $c$  is squared in equation (5h), we can substitute  $\text{abs}(c)$  for  $c$  in that equation, obtaining

$$pSE(b) = \sqrt{1 - \text{abs}(c)^2} w \tag{6a}$$

Substituting the right hand side of equation (6) for  $\text{abs}(c)$  into equation (6a) we have

$$pSE(b) = \sqrt{1 - [\text{abs}(b-\beta)/w]^2} w = \sqrt{w^2 - (b-\beta)^2} \tag{6b}$$

(from Lemma 2, we know that  $\text{abs}(b-\beta) \leq w$ , so the expression inside the radical of equation (6b) is never negative).

We can square both sides of equation (6b) and rearrange terms to get

$$(b-\beta)^2 = w^2 - pSE(b)^2 \tag{6c}$$

Equations (6b) and (6c) show that  $(b-\beta)^2$  is monotonically decreasing in  $pSE(b)^2$ . Equivalently,  $\text{abs}(b-\beta)$  is monotonically decreasing in  $pSE(b)$ . Since  $SE(b)$  is proportional to  $pSE(b)$ , we have the desired result. QED

We have shown that for a fixed sample size when the *ceteris paribus* condition is invoked, smaller values of  $SE(b)$  are *always* associated with larger values of  $\text{abs}(b-\beta)$ .

From Theorem 3 we infer that the relationship between  $\text{abs}(b-\beta)$  and  $pSE(b)$  does not depend on  $\beta$ . Figure 1 is an illustration of how  $\text{abs}(b-\beta)$  varies with  $pSE(b)$  for several values of  $[\beta/w]$ . We index the family of curves by using  $[\beta/w]$  because this is more general than fixing  $\beta$  and indexing by varying  $w$  or by fixing  $w$  and indexing by varying  $\beta$ .

-----

Figure 1 about here

-----

Rather than using  $pSE(b)$  on the horizontal axis of Figure 1, if we were to select a particular  $n$  and use  $SE(b)$  as the variable for the horizontal axis, the only change would be the scale for the axis, as can be inferred from equation (4b).

### Exploring $R^2$

One might believe that there is some connection in simple linear regression between goodness of fit of  $b$  to  $\beta$  and the size of  $R^2$ , since statistical significance for the model (using the F statistic for  $R^2$ ) matches statistical significance for  $b$  (using the t statistic). Such a belief is not correct. In addition to this, we find some interesting relationships between  $R^2$  and  $abs(b-\beta)$  by invoking *ceteris paribus* as we have done in the previous section.

$R^2$  can be computed for a simple linear regression model using the well known relationship

$$R^2 = r_{xy}^2 = [b \sqrt{\text{Var}(\mathbf{x})/\text{Var}(\mathbf{y})}]^2 \quad . \quad (7)$$

Corollary 1.1 provides the distance between  $b$  and  $\beta$  as a function of  $c$  and  $w$ . We can use this corollary to derive an expression for  $R^2$  as a function of  $c$  and  $[\beta/w]$ .

### Theorem 4:

$$R^2 = (c + [\beta/w])^2 / ([\beta/w]^2 + 2c[\beta/w] + 1) \quad (8)$$

Proof: Use  $\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{x})(\beta^2 + 2\beta cw + w^2)$  from Lemma 2 and  $b = cw + \beta$  based on

Theorem 1 to substitute into equation (7), obtaining

$$R^2 = [(cw + \beta) \sqrt{\text{Var}(\mathbf{x})} / (\text{Var}(\mathbf{x})(\beta^2 + 2\beta cw + w^2))]^2 \quad . \quad (8a)$$

Cancelling  $\text{Var}(\mathbf{x})$  from numerator and denominator, this simplifies to

$$R^2 = (cw + \beta)^2 / (\beta^2 + 2\beta cw + w^2) \quad (8b)$$

Dividing the numerator and denominator on the right hand side of equation (8b) by  $w^2$  we have the desired result.

QED

**Corollary 4.1:**

If  $\beta = 0$ , then  $R^2 = c^2$

Proof: Set  $\beta = 0$  in equation (8) and cancel terms.

QED

Corollary 4.1 together with Theorem 1 show that when  $\beta = 0$  and *ceteris paribus* is invoked, the larger that  $R^2$  is the worse the fit of  $b$  to  $\beta$  is. This matches our intuition, since larger  $R^2$  tend to correspond to larger values of  $\text{abs}(b)$ , which corresponds to  $b$  being farther away from 0. Our approach formalizes this, without having concern for  $\text{Var}(\mathbf{e})$ . We have certainty rather than tendency. (Note: We have *not* assumed that  $\text{Var}(\mathbf{e})$  is fixed.)

**Corollary 4.2:**

If  $b = \beta$ , then

$$R^2_{\text{opt}} = R^2 = \beta^2 / (\beta^2 + w^2) = [\beta/w]^2 / (1 + [\beta/w]^2) \quad (8c)$$

where  $R^2_{\text{opt}}$  is the value of  $R^2$  corresponding to the optimal value of  $b$ , i.e.,  $b = \beta$ .

Proof: A necessary condition for  $R^2_{\text{opt}}$  is to have  $b = \beta$ . When this is true, we can substitute 0 for  $b - \beta$  in the right hand side of equation (4d) to obtain

$$\text{Var}(\mathbf{y}) = \text{Var}(\mathbf{x}) (\beta^2 + w^2) \quad (8d)$$

Substituting the above expression for  $\text{Var}(\mathbf{y})$  and substituting  $\beta$  for  $b$  in equation (7) we have

$$\begin{aligned} R^2 &= [\beta \text{sqrt}(\text{Var}(\mathbf{x}) / (\text{Var}(\mathbf{x}) (\beta^2 + w^2)))]^2 \\ &= \beta^2 / (\beta^2 + w^2) \end{aligned} \quad (8e)$$

Divide numerator and denominator of the right hand side of equation (8e) by  $w^2$  to obtain the extreme right hand side of equation (8c).

QED

Corollary 4.2 shows that, *ceteris paribus*, the value of  $R^2$  that corresponds to the optimal  $b$ , i.e.,  $b = \beta$ , is not the largest possible value for  $R^2$ , which is 1.0, but is  $R^2 = \beta^2 / (\beta^2 + w^2)$ . This is a necessary condition for  $R^2$  for the model to produce the optimal  $b$ , but it is not sufficient. That is, it is possible for  $R^2$  to have this value, but  $b \neq \beta$ .

Since our *ceteris paribus* assumption is that  $w = \sqrt{\text{Var}(\mathbf{d})/\text{Var}(\mathbf{x})}$  is invariant across samples, we can create a family of curves by varying  $\beta$  and leaving  $w$  fixed. Figure 2 displays the relationship between  $R^2$  and  $\text{abs}(b - \beta)$  for several values of  $[\beta / w]$ . The data points in the figure are generated by varying  $c$  from -1.0 to 1.0 in increments of 0.05 and using the formulas in Theorems 3 and 4 or their corollaries. To create these graphs we have used positive values for  $\beta$ . We would get similar results if negative values were used.

-----  
 FIGURE 2 ABOUT HERE  
 -----

Figure 2 illustrates that the curves relating  $\text{abs}(b - \beta)$  to  $R^2$  are not mathematical functions, since there are two values of the former relating to each value of the latter. For example, it can be seen that corresponding to  $R^2 = R^2_{\text{opt}}$  there is a value of  $\text{abs}(b - \beta)$  other than 0, illustrating that  $R^2 = \beta^2 / (\beta^2 + w^2)$  is a necessary but not sufficient condition for the optimal fit  $b = \beta$ . Figure 2 also illustrates that for  $R^2 > R^2_{\text{opt}}$ , increasing values of  $R^2$  yield worse fits of  $b$  to  $\beta$ , *ceteris paribus*. The slopes of each curve are positive in this range. For  $R^2 < R^2_{\text{opt}}$ , increasing values of  $R^2$  sometimes yield better fits of  $b$  to  $\beta$  and sometimes worse. The slopes of each curve can be both positive and negative in this range

So far, we have not needed to use the sample size,  $n$ , in our derivations or graphs, except to link  $SE(b)$  to  $pSE(b)$ . However, if we wish to consider *probable* values of  $c = \text{Corr}(\mathbf{x}, \mathbf{d})$ , not merely possible values, it is necessary to take  $n$  into account. For Figures 3a, 3b, and 3c, instead of enumerating values of  $c$  from -1 to +1 as we did for Figures 1 and 2, values of  $c$  are computer simulated. The results of the computer simulation depend on sample size. Figures 3a, 3b, and 3c display scatterplots for sample sizes  $n = 10, 30,$  and  $100,$  respectively. A family of curves of  $\text{abs}(b-\beta)$  vs.  $R^2$  are shown for each, based on generating 200 data points by computer simulation using Microsoft Excel. It can be seen by examining Theorems 1 and 4 and their corollaries that we only need to simulate values of  $c$  (based on generating  $\mathbf{x}$  and  $\mathbf{d}$ ) to formulate the scatterplots for the figures.

-----

Figures 3a 3b and 3c about here

-----

We see that the curves in Figures 3a, 3b, and 3c correspond to a portion of the equivalent curves in Figure 2. Larger sample size results in more compact distributions for corresponding curves. Unlike in Figure 2 where sample size is irrelevant, in Figure 3 we see that sample size can be sufficiently large so that the relationship between  $\text{abs}(b-\beta)$  and  $R^2$  is a mathematical function. As in Figure 2, for  $R^2 > R^2_{\text{opt}}$ , increasing values of  $R^2$  correspond to worse fits of  $b$  to  $\beta$ . However, when  $n$  is sufficiently large, for  $R^2 < R^2_{\text{opt}}$ , increasing values of  $R^2$  always correspond to better fits of  $b$  to  $\beta$ . Of course, in the field as opposed to in computer simulation experiments, we do not know what the value of  $R^2_{\text{opt}}$  is.

We now briefly explore the relationship between  $\text{abs}(b-\beta)$  and  $R^2$  without our *ceteris paribus* invocation. Figure 4 shows an example for which  $n = 10$ . As before, we have computer generated the data for  $\mathbf{x}$  and  $\mathbf{d}$  using normality assumptions for  $X$  and  $D$ . We have chosen values for  $\beta$ ,  $\text{Var}(D)$ , and  $\text{Var}(X)$  so that the square of the correlation between  $X$  and  $D$  *in the population* is 0.5. The concept of the square of the population correlation between  $X$  and  $Y$  corresponds to

$R^2_{\text{opt}}$  in the sample. We refer to this value as  $R^2_{\text{OPT}}$ . We have used 100 simulation trials. The fitted quadratic trendline shows that on average  $\text{abs}(b-\beta)$  is monotonically increasing in  $R^2$  for  $R^2 > R^2_{\text{OPT}}$  and  $\text{abs}(b-\beta)$  is monotonically decreasing in  $R^2$  for  $R^2 < R^2_{\text{OPT}}$ .

-----  
Figure 4 about here  
-----

Although in the field we do not know the value of  $R^2_{\text{OPT}}$ , this figure suggests that even without the *ceteris paribus* assumption that we have been invoking, it is not reasonable to assume that larger values of  $R^2$  correspond to better fits of  $b$  to  $\beta$ .

### **Discussion and Conclusion**

The use of *ceteris paribus* as a way of generalizing the applicability of a model or the result of a study (i.e., moving from internal validity to external validity) without the benefit of any additional evidence may be common. It is used in a fashion somewhat similar to "hand waving" in mathematics. If specifics are not provided with regard to what must be the same, the invocation of *ceteris paribus* is vacuous or tautological.

With malice aforethought, we have used simple linear regression modeling results to show that there can be potential issues even in basic modeling with regard to invoking *ceteris paribus* when specifics are not provided by the researcher.

The inclusion of the phrase *ceteris paribus* adds no insight to a discussion and may actually be misleading or false. It should be dropped unless it is explicitly supported by the research or modeling work.

## References

Banker, Rajiv D., Hsihui Chang; Ram Natarajan. Productivity Change, 2005. Technical Progress, and Relative Efficiency Change in the Public Accounting Industry.

*Management Science*; 51, 2 291- 304. p. 300

Chatterjee. Samprit, Bertram Price. 1977. Regression Analysis by Example. Wiley.

Dallal, Gerald, The Little Handbook of Statistical Practice,

<http://www.tufts.edu/~gdallal/LHSP.HTM>. accessed 3-13-08

Draper, Norman, Harry Smith. 1998. Applied Regression Analysis. Wiley.

Ferson, Wayne E., Andrea Heuson, 2005. Tie Su, Weak-Form and Semi-Strong-Form Stock Return Predictability Revisited. *Management Science*, 51, 10. 1582- 1592.p. 1583

Huckman, Robert S. 2003. The utilization of competing technologies within the firm: Evidence from cardiac procedures. *Management Science*. 49, 5 599 - 617. p. 603

Mezias, Stephen J., Ya-Ru Chen; Patrice R Murphy. 2002. Aspiration-level adaptation in an American financial services organization: A Field Study. *Management Science*; 48, 10 1285-1300. p. 1295.

Nagar, Venky and Medhav V. Rajan, 2005. Measuring Customer Relationships: The Case of the Retail Banking Industry. *Management Science*, 51, 6 904-919. p. 910

## Figures

Figure 1

This is a family of curves illustrating the monotonically *decreasing* relationship between  $pSE(b)$ , which is proportional to  $SE(b)$ , and  $abs(b-\beta)$ . Data points are generated by iterating values of  $c = Corr(\mathbf{x}, \mathbf{d})$  from -1 to +1 in steps of 0.05. Formulas from Theorems 1 and 2 and their corollaries are used to calculate  $abs(b-\beta)$  and  $pSE(b)$ .

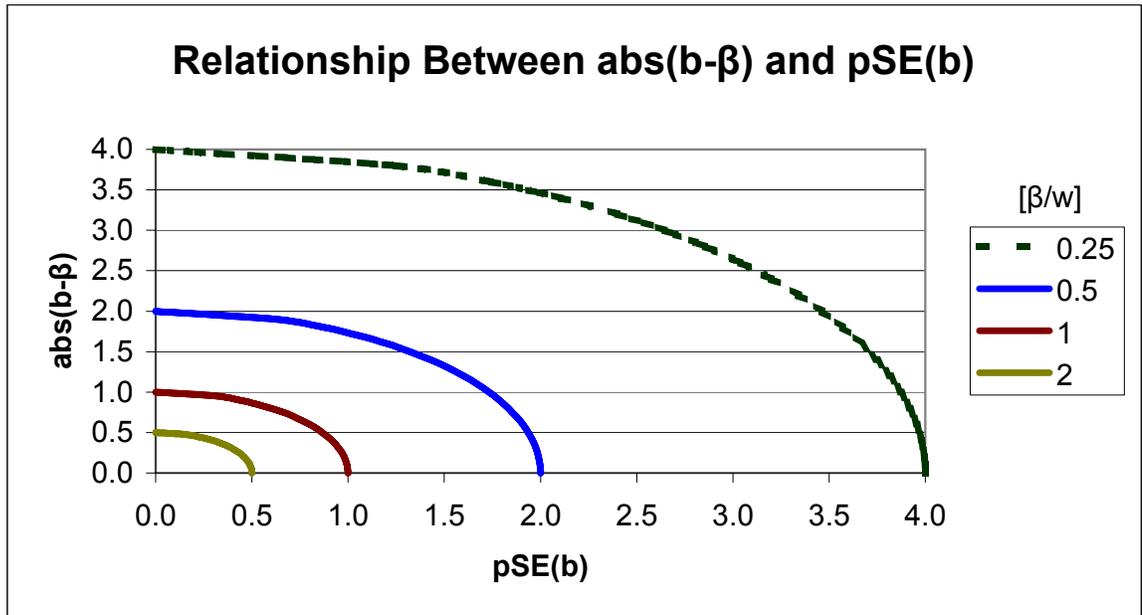
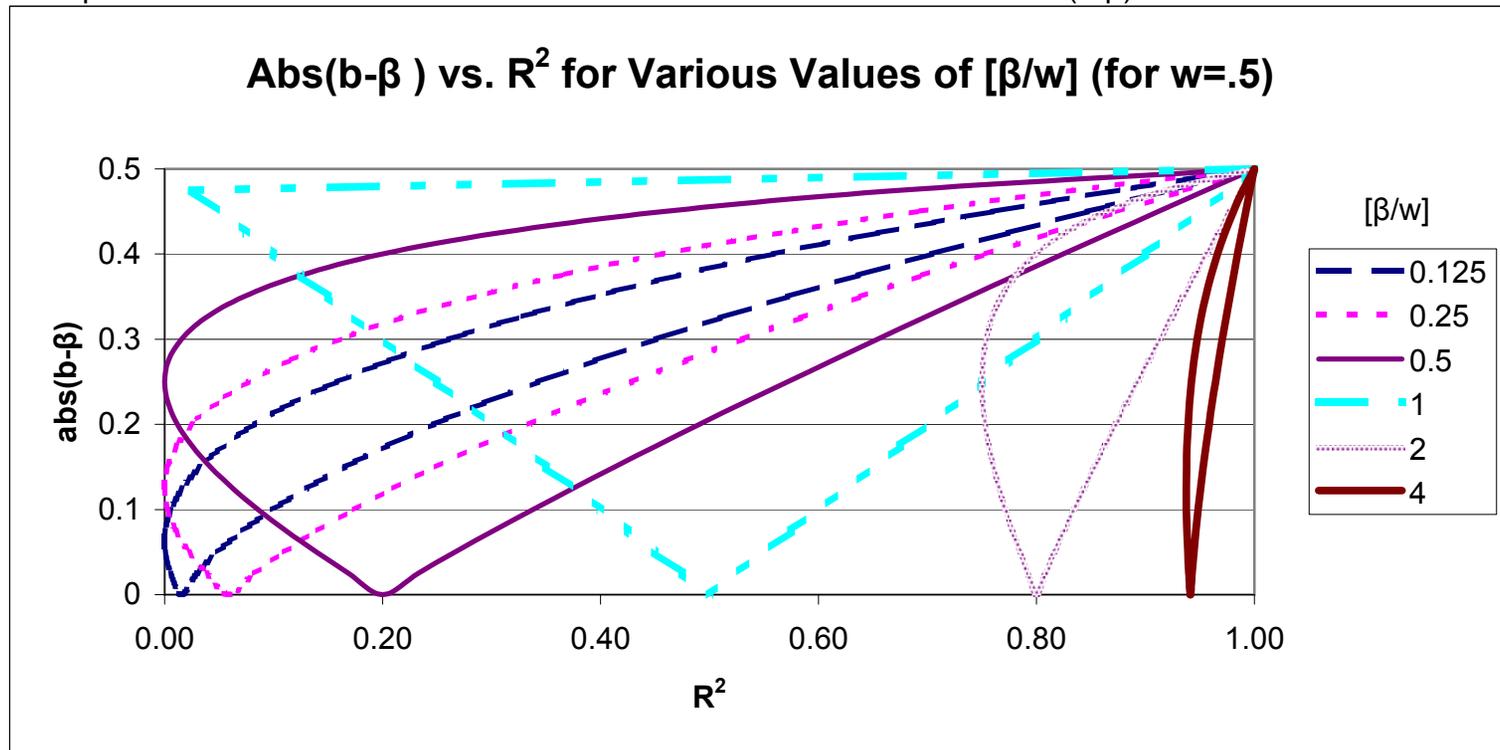


Figure 2

This is a family of curves illustrating the relationship between  $R^2$  and  $\text{abs}(b-\beta)$ .

Data points are generated by iterating values of  $c = \text{Corr}(x,d)$  from -1 to +1

in steps of 0.05. Formulas from Theorems 1 and 4 were used to calculate  $\text{abs}(b-\beta)$  and  $R^2$ .



Figures 3a, 3b, 3c

This is a family of scatter plots illustrating relationships between  $\text{abs}(b-\beta)$  and  $R^2$ . The scatterplot curves are indexed by values of  $[\beta/w]$ . As a practical matter, we have fixed the value of  $w$  and varied the value of  $\beta$ . Also, as a practical matter we have used normal distributions for  $X$  and  $D$  as the basis for generating values of  $x$  and  $d$  by computer simulation. The curves in the three figures vary according to the sample sizes used in the simulation:  $n = 10, 30, 100$ .

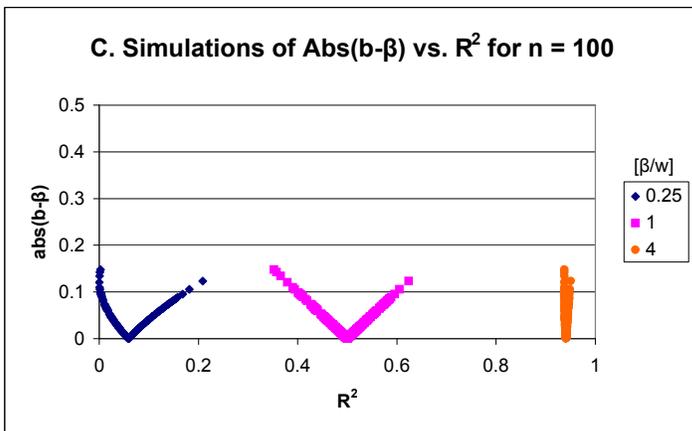
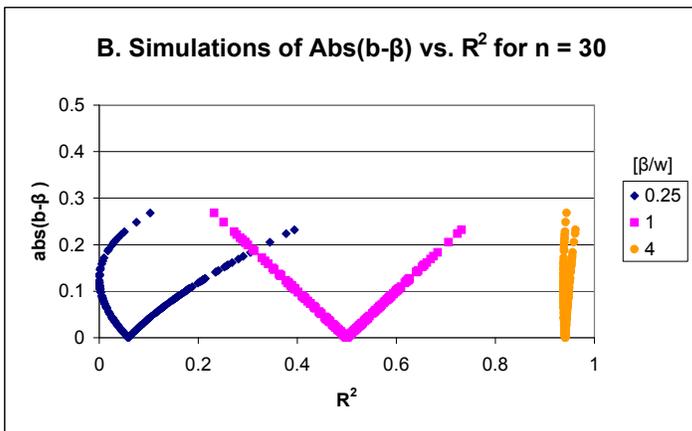
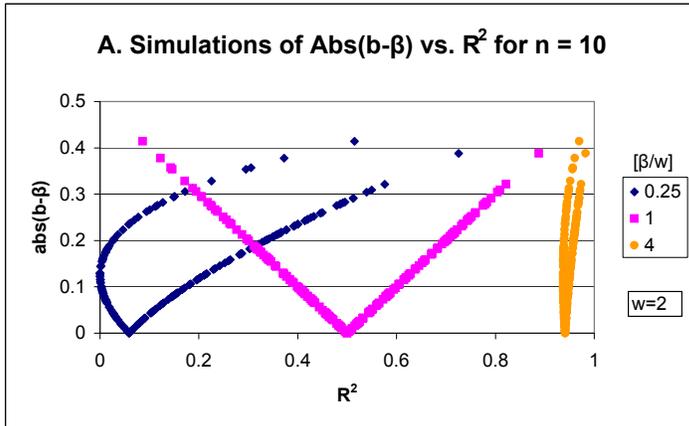


Figure 4

This is an illustration of the relationship between  $\text{abs}(b-\beta)$  and  $R^2$ , based on 100 computer generated samples from a virtual population. The quadratic trendline demonstrates the non-monotonic relationship.

