The relative standard error as an easy index for checking the reliability of regression coefficients

Research · August 2015

DOI: 10.13140/RG.2.1.2123.6968

1 author:

Gudeta Weldesemayat Sileshi
None
274 PUBLICATIONS **4,511** CITATIONS

SEE PROFILE

The relative standard error as an easy index for checking the reliability of regression coefficients

Gudeta W. Sileshi

Lusaka, Zambia; Email: sileshigw@gmail.com

Abstract

Regression analysis is one of the powerful tools used for either explanatory or predictive modelling. Unfortunately, many things can go wrong and invalidate the analysis and the conclusions drawn from regression analysis. One of the common mistakes even among experienced practitioners is the temptation to include many predictors and trying to run complicated models with a limited sample size. As the number of predictors increases, correct estimation of coefficients becomes increasingly challenging. The data become collinear and the coefficient of determination (R^2) tends to 1 (and AIC decreases) giving the impression of a better model. However, some or all of the coefficients may be poorly estimated or unreasonable. This makes the interpretation of the coefficients and the ranking of predictors for a given outcome difficult. Thus the whole model becomes useless either for explanation or prediction. The aim of this paper is therefore to demonstrate this common, but often overlooked, problem and provide a rule of thumb for the relative standard error to quickly check and rectify errors in estimation.

Key words: attenuation, orthogonal, over-fitting, spurious regression

1. Introduction

Regression analysis is one of the powerful tools used in a variety of disciplines for either explanatory or predictive modelling. The goal of explanatory modelling is to determine changes in a response variable (Y) based on a set of predictors (X_i) and how much variation in *Y* is explained by the X_i. Here the emphasis is on understanding the individual relationships of X_i with Y. On the other hand, the goal

of predictive modelling is to select a single model that provides the most accurate prediction of Y at the relevant scale. In either case a bivariate or multivariate regression analysis may be applied and a null hypothesis is tested about the parameters (β_i) in models expressed as:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \epsilon$

This is usually done by comparing the *t*-statistics associated with the partial regression coefficients (β_i) and the 95% confidence limits (hereafter 95% CL). If we reject the null hypothesis (i.e. we accept $\beta_i \neq 0$), we can then move on to judging the marginal effects and the relative importance of the predictors[1]. However, this seemingly easy task can be daunting as all or some of the β_i could be non-significant (i.e. $\beta_i = 0$) while the omnibus test is still significant. This way a study could contradict itself, earlier findings or theoretical predictions. This brings the validity[2] of the published model into question. I have seen quite a number of models with questionable validity published in refereed journals.

This article is motivated by these observations and that crucial problems are overlooked when regression models are published. Here, I will focus on some of the problems that affect non-linear, simple linear and multiple regression in ordinary least square (OLS) and generalized linear modelling (GLMs) frameworks. I will not deal with other GLMs such as Poisson regression as I have dealt with some of the problems in earlier publications (e.g. Sileshi, 2006; 2008; Sileshi et al., 2009). In order to demonstrate the common problems, I will use examples from studies on insect-plant interaction, forest biomass estimation and technology adoption with which I am familiar with. I chose examples from published studies only because the studies reported the requisite data in a transparent and readily available manner. These examples should not be construed as examples of bad studies. For brevity, hereafter I will refer to all regression coefficients (i.e. β_0, β_1 ,… β_n), their standard errors and sample sizes as $\bar{\theta}_i$, $s_{\theta i}$ and N, respectively. I will refer to the partial coefficients (i.e. β_1 , β_2… β_n) as β_i. In the literature β_0 is called the constant and

[1] *The relative importance of predictors can be judged using beta weights (also called standardized coefficients). Beta weights are coefficients derived by standardizing the predictors to mean zero and standard deviation of 1 so that the coefficients can be directly compared and ranked. The greater the absolute value of the beta weight, the greater is the contribution of the predictor.*

[2] *Validity addresses questions of whether a causal effect is credibly estimated or not and that the statistical inferences are valid for the population being studied.*

some researchers do not care much about its value. However, β_0 is crucial in predictive models, thus it needs to be estimated correctly.

2. What can go wrong?

The reliability of regression technique is critically reliant on satisfying the underlying assumptions. For example, OLS regression assumes: (1) multivariate normality, (2) linear relationship between Y and X_i, (3) constant variance (homoscedasticity), (4) no autocorrelation in predictors, (5) independence of errors and predictors, (6) orthogonality (no multicollinearity among predictors), (7) reliability of measurements and (8) no outliers. In addition, it assumes that the model is correctly specified. Although OLS regression is robust to minor violations of one or a few of these assumptions, the consequences can be severe if there are multiple violations. I have summarized the consequences of violations of these assumptions and methods for detecting violations in Table 1.

Table 1. Common violations of regression assumptions and consequences

Violation	**Consequence**	**Detection method**
Non-normality of errors	*F*- and *t*-tests unreliable	-Normal Q-Q plots -Formal tests of normality
Non-linearity	$s_{\theta i}$ unreliable	-Partial residual plots -Scatter plots of Y on X_i
Heteroscedasticity	$s_{\theta i}$ unreliable	-Residual plots -White's test
Autocorrelation	$s_{\theta i}$ unreliable	-Residual plots -Durbin-Watson test
Multicolinearity	$s_{\theta i}$ unreliable *t*-test non-significant	-Variance inflation factor (VIF) -Tolerance
Measurement error s	Errors in X_i bias $\bar{\theta}_i$ towards 0 Errors in Y inflate $s_{\theta i}$	-Small *t*-values; $P > 0.05$ -Inflated $s_{\theta i}$ values
Outliers	$s_{\theta i}$ unreliable	-Studentized residuals -Scatter plots of Y on X_i

Some of the assumptions are violated by the temptation to include as many predictors as possible in regression models with small sample sizes (N). As the number of predictors (P) approaches N, data become increasingly collinear and R^2 tends to 1. As R^2 gets bigger, the $s_{\theta i}$ get larger and $\bar{\theta}_i$ fail to attain statistical significance. Even experienced researchers have been fooled into believing they

are getting good models when they see large R^2 or small AIC values. The increase in R^2 is a product of spurious regression, a problem known as the Freedman Paradox after Freedman (1983) and many statisticians advise against reliance on R^2 (Helland, 1987). Information criteria (e.g. AIC and BIC) may also be misleading if the true model is not in the set being compared.

In my experience working with datasets from observational studies and surveys, the biggest problem has been correct estimation of $\bar{\theta}_i$ and $s_{\theta i}$. In observational studies (e.g. forest inventories, household surveys, etc.) data are often collected using non-random (e.g. convenience) and multistage sampling designs. If not appropriately accounted for, this will results in incorrect estimates of $s_{\theta i}$ and 95% confidence intervals (CI). When such data are analysed using maximum likelihood estimation methods, convergence problems also arise. For example, in logistic regression, the negative of the Hessian may not be positive definite (i.e. not invertible). This happens because collinearity creates a situation called ill-conditioning of the X'X matrix. Thus the process for inverting the matrix and calculating $\bar{\theta}_i$ becomes unstable (Gill & King, 2004). Another common problem is quasi-complete or complete separation of observations (Albert and Anderson, 1984). This happens when the predictor (e.g. marital status) is split into many categories (e.g. unmarried, married, widowed, separated, divorced, etc.) with small N. Some software packages detect and warn about these problems. Others simply report $\bar{\theta}_i$ and $s_{\theta i}$ even if the algorithm does not converge. If one or more of the β_i are not correctly estimated, not only their interpretation but also the ranking of predictors for a given outcome becomes difficult.

Unaware of these pitfalls, many researchers look at the model fit statistics (R^2, AIC, etc.) and proceed to publishing the model without scrutinising the $\bar{\theta}_i$. This lack of attention to statistical detail has often resulted in publication of unreliable models even in journals with very high impact factors. I have also seen some models accompanied by inflated interpretation (*sensu* Ioannidis, 2008) such as over-stated generalization, equating association with causation, selective reporting and downplaying limitations of the models. In many cases the $s_{\theta i}$, t-statistics and other diagnostics are rarely reported. My advice to researchers is to (1) take extreme caution when making inferences about $\bar{\theta}_i$ and (2) report the $s_{\theta i}$

and *P*-values in a transparent manner. This will save the author from a lot of embarrassment (someone is likely to discover the errors) and the audience from being misled. In order to help authors to make informed decisions, in the following sections I will focus on uncertainty about regression coefficients and ways to detect common problems.

3. Uncertainty around regression coefficients ($\bar{\theta}_i$)

The primary interest in regression is correct inference about the true population parameter (θ) based on sample values ($\bar{\theta}_i$). Because of sampling variability, $\bar{\theta}_i$ estimates may be higher or lower than θ. Therefore, $\bar{\theta}_i$ always has some uncertainty about θ. The standard errors of $\bar{\theta}_i$ (i.e. $s_{\theta i}$) give us an indication of how much $\bar{\theta}_i$ are likely to vary from the corresponding θ value. For example, the standard error of the slopes (s_{bi}) gives us an indication of how much the sample estimates ($\bar{b}_i$) are likely to vary from the corresponding population value (β_i). In general, s_{bi} is calculated from the standard deviation of the regression error (S_ϵ), the coefficient of determination (R^2), the variance of X_i (S_{Xi}^2) and the sample size (N) as follows:

$$s_{bi} = \frac{S_\epsilon}{\sqrt{(1-R^2)\times S_{Xi}^2 \times (N-1)}}$$

or the following in bivariate regression

$$s_{bi} = \sqrt{\frac{1-R^2}{N-2}} \times \frac{SD_Y}{SD_X}$$

The standard error of the intercept (s_{b0}) is given as:

$$s_{b0} = S_e \sqrt{\frac{1}{N} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}}$$

There are many ways of computing $s_{\theta i}$ for survey data. For brevity, here I will use the standard definition $s_{\theta i} = \frac{\sigma}{\sqrt{N}}$ where σ is the standard deviation for survey data. The size of the uncertainty is usually judged using $s_{\theta i}$ and the 95% CI (≈ $1.96 s_{\theta i}$). The $s_{\theta i}$ also measures the extent to which the predicted mean ($\bar{Y}_i$) deviates from the true population value being predicted (i.e. μ). If N is large enough, according to the Central Limit Theorem, the distribution of $\bar{\theta}_i$ and $\bar{Y}_i$ will tend to be Gaussian and thus we can use $s_{\theta i}$ as a measure of uncertainty around $\bar{\theta}_i$ or $\bar{Y}_i$. However, there

are times when we cannot be sure whether $s_{\theta i}$ is small enough. Remember, we also have uncertainty around the $s_{\theta i}$ value itself.

The smaller $s_{\theta i}$ (the narrower the CI) the more confidence we can place on the model for explaining or predicting the phenomenon under study. If $s_{\theta i}$ is large relative to $\bar{\theta}_i$, the 95% CIs will be wide, and it will not provide us with accurate information about the location of θ. If the CI are correct, they represents the range in which θ is likely to be found. They also provide a useful way of assessing the quality of $\bar{Y}_i$. In predictive regression modelling CIs may be constructed for prediction of a single future value of $\bar{Y}_i$ corresponding to a chosen value of X or the regression line as a whole. When the entire line is of interest, a confidence region permits one to simultaneously make statements about estimates of Y for a number of values of the predictor variable X. That way we can judge the reliability of our predictions, i.e. the agreement between predicted and observed values. The reliability of an estimate is usually given as the fraction (f) of the mean (m).

The relative standard error (RSE) is one of such measures (Saydah et al., 2013; Sileshi, 2014). For model parameters ($\bar{\theta}_i$), Sileshi (2014) proposed the PRSE to be calculated as $100\left(\frac{s_{\theta i}}{|\bar{\theta}_i|}\right)$ with the absolute value in the denominator since $\bar{\theta}$ can take negative values. The PRSE gives an easier way for judging the reliability of $\bar{\theta}_i$ compared to either the $s_{\theta i}$ or 95% CI. Note that PRSE is related to the 95% CI and the *t*-statistic as follows:

$$PRSE = 100 \times \frac{\left(\frac{U-L}{3.92}\right)}{|\bar{\theta}_i|} = 100 \times \frac{1}{t}$$

where U and L are the lower and upper 95% confidence limits, respectively. We divide UCL-LCL by 3.92 because the two-tail $t_{\alpha=0.05}$ is 2x1.96 = 3.92. The standard error of the estimated coefficient (s_{bi}) may be thought of as the reciprocal of the signal-to-noise ratio for observing the effect of X on Y. The larger the s_{bi} is for any given regression coefficient, the *less precise* the coefficient is. This means that there is too much noise than signal in the data. In many practical applications, estimates are considered unreliable when PRSE exceeds 25%, i.e. 25% noise-to-signal ratio. However, the cut-off point for characterizing an estimate as reliable can be as low as 10% depending on the application. For example, in

integrated pest management (IPM)[3] PRSE of 10% has been recommended for parameter estimation and life-table studies (Southwood, 1978). PRSE <25% has been used for comparing the efficiency of sampling methods (Buntin, 1994). Estimation of pest population densities for assessment and control has been shown to be sufficiently accurate when PRSE<25% (Southwood, 1978). A smaller value of PRSE has also been suggested for forest biomass estimation. For example, Stellingwerf (1994) suggested that the 95% CI should not exceed ±20% of the mean. This translates to PRSE of 10% because $1.96s_{\theta i}$ = 95CI. In the National Health and Nutrition Examination Survey of USA, estimates become questionable when PRSE>20% (Saydah et al. 2013). In health statistics, estimates with PRSE >30% are deemed unreliable and not allowed to be published in reports (CDC, 2010). Picard et al. (2015) claimed that a PRSE of 25% is unreasonably low because it corresponds to a significance level of 10^{-5} to 10^{-4}. That argument is neither based on theory nor empirical data. In the following section I will further illustrate the validity of the 25-30% rules of thumb for PRSE of $\bar{\theta}_i$ based on two considerations. Firstly, we want to be sure that $\bar{\theta}_i$ is an accurate estimate of θ because it will eventually be used either in explanatory or predictive modelling. In that sense a PRSE of 30% would mean θ is within $0.15\bar{\theta}_i$, a reasonably narrow 95% CI. Secondly, a PRSE of <30% is likely to ensure significance of $\bar{\theta}_i$ at least at a=0.01 even with N ≤30 commonly used in practical applications. I will focus on demonstrating this second consideration using the *t*-distribution.

In statistical inferences regarding $\bar{\theta}_i$ we typically begin with the null hypothesis that $\bar{\theta}_i$ is equal to zero. However, if we draw repeated samples from a population with θ equal to zero we will obtain a range of sample values of $\bar{\theta}_i$ due to sampling error. Therefore, the significance of $\bar{\theta}_i$ must be evaluated using the standard *t*-test where *t* is expressed as:

$$t = \frac{\bar{\theta}_i - \theta}{s_{\theta i}}$$

This becomes $t = \frac{\bar{\theta}_i}{s_{\theta i}}$ since θ is assumed to be 0 for a regression of *Y* on *X*. As a rule of thumb if the *t*-statistic is greater than 2, we can reject H_0: $\bar{\theta}_i = 0$ (see Figure 1a). This means we can be fairly confident that $\theta \neq 0$ if $|\bar{\theta}_i|$ is at least $2s_{\theta i}$. Other things

[3]In the IPM literature PRSE is commonly referred to as relative variation (*RV*) statistic.

being equal, $|t|$ will be large whenever $|\bar{\theta}_i|$ is large. However, $|t|$ will not be large enough to be significant if $s_{\theta i}$ is large, hence it is possible to accept a false null hypothesis (β=0). Examination of PRSE in relation to the *t*-distribution indicates that PRSE must be less than 60% for $\bar{\theta}_i$to be statistically significant even with an infinitely large N (Figure 1). With N<50, $\bar{\theta}_i$ becomes significant at a=0.05 only when $|t|$ >1.68, i.e. PRSE<59.5 (Fig. 1b). For significance at a=0.01, $|t|$ should be >2.4 (or PRSE <41) with N<50 but for smaller N (e.g. <20), $|t|$ should be >2.5 (or PRSE <39%) for significance at a=0.05 level. For $\bar{\theta}_i$ to be significant at a=0.001 with N<50, $|t|$ should be >3.3 or PRSE <30%. For smaller N (e.g. <10), $|t|$ should be >4.3 or PRSE <23 (Fig. 1b). Therefore, the PRSE <25% is reasonable compromise for parameter estimation using the typically small sample sizes in many studies.

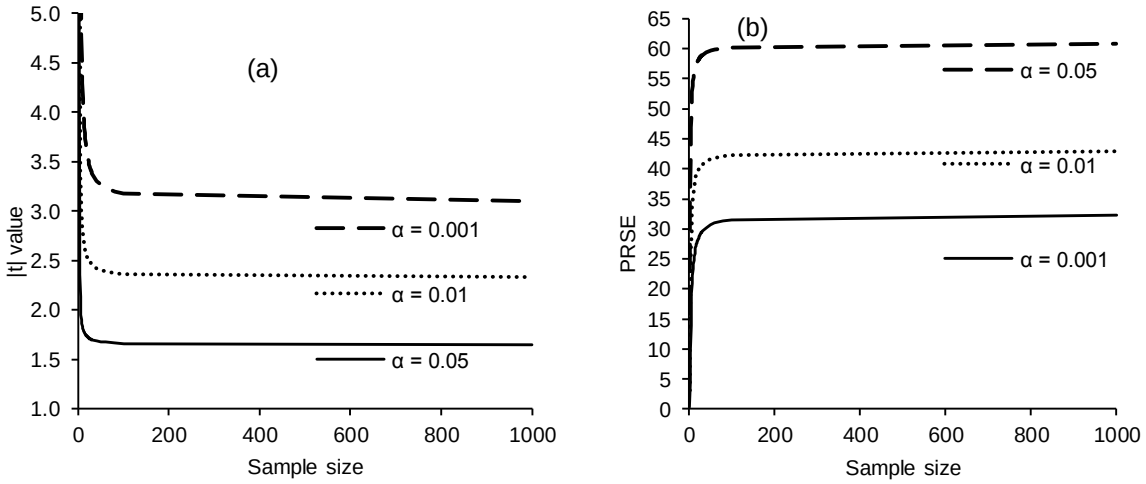

Figure 1. Variation in $|t|$ and PRSE with sample size and significance level. I drew

How do $\bar{\theta}_i$ values assume large PRSE and become unreliable? There are number of causes depending on whether one is applying bivariate or multivariate regression. Often, violations of regression assumptions (especially non-normality and heteroscedasticity), small N and measurement errors pose serious threats in regression. If the errors are non-normal or heteroscedastic, the $s_{\theta i}$ and the resultant 95%CIs are invalid. In the following sections I will presents examples from bivariate and multivariate regression.

3.1. Bivariate regression

Bivariate regression models can be expressed as:

$Y = \beta_0 + \beta_1 X + \epsilon$

Here the null hypothesis is that the slope is zero, i.e. H_0: β_1 = 0 and the alternative hypothesis H_a is $\beta_1 \neq 0$. We reject the null hypothesis when the omnibus test (e.g. F) and the t-test for the coefficients are non-significant at the prescribed level (usually P<0.05). However, there are situations where our sample estimates of a and β are too far from their respective population value. There are a number of problems in bivariate regression that can render this task difficult.

For example, small sample size (N), measurement errors, non-normality and heteroscedasticity are important problems. For the same value of θ, $s_{\theta i}$ is usually large for small N; but $s_{\theta i}$ declines as N increases. In order to demonstrate this, I will use the following four examples. In all examples, a power function $Y = \beta_0 X^{\beta_1}$ expressed in linear form as $\ln(Y) = \ln(\beta_0) + \beta_1(\ln D) + \epsilon$ taking the natural log of Y and X was used. This simple model is widely used in biology, physics, chemistry and many other fields where it is called power-law or allometric scaling rule. However, problems relating to the correct estimation of β_1 (and also β_0) have caused controversy among scientists in the last several decades. In the following section, I will use examples from forestry where this model is used for relating Y and X, and demonstrate how sample size affects uncertainty around β_1 and β_0.

Example 1

The first example comes from my own compilation of the variation of the allometric exponent (β_1) of the scaling of tree height (H) with tree diameter (D). I compiled (in Sileshi, 2014) the β values from a large number of studies that applied the following model: $\ln(H) = \ln(\beta_0) + \beta_1(\ln D) + \epsilon$. As described above, this is equivalent to drawing samples from a population with a known θ (in this case hypothesized to be 0.66). However, in each study a range of values of β_1 and corresponding values of s_β were obtained. I constructed the remaining three examples using values of β_1, their corresponding s_β and N extracted from Coomes & Allen (2009), Russo et al. (2007) and Rüger & Condit (2012). In each of these examples, researchers had independently analysed the scaling of tree diameter growth rate (GR) with tree

diameter (D) using the regression model $\ln(GR) = \ln(\beta_0) + \beta_1(\ln D) + \epsilon$. This way each worker obtained a range of β_1 values for different samples although the hypothesized population value of β is 0.33. Some of the β_1 values were significantly larger than 0.33 and others significantly smaller than 0.33. In Figure 2, I plotted s_β for each β_1 against the corresponding N value. Note that the value of s_β is less reliable and hence the uncertainty around β_1 is higher with small N compared to large N. In each example, s_β was too large when N was small but decreased rapidly as N increased (Figure 2).

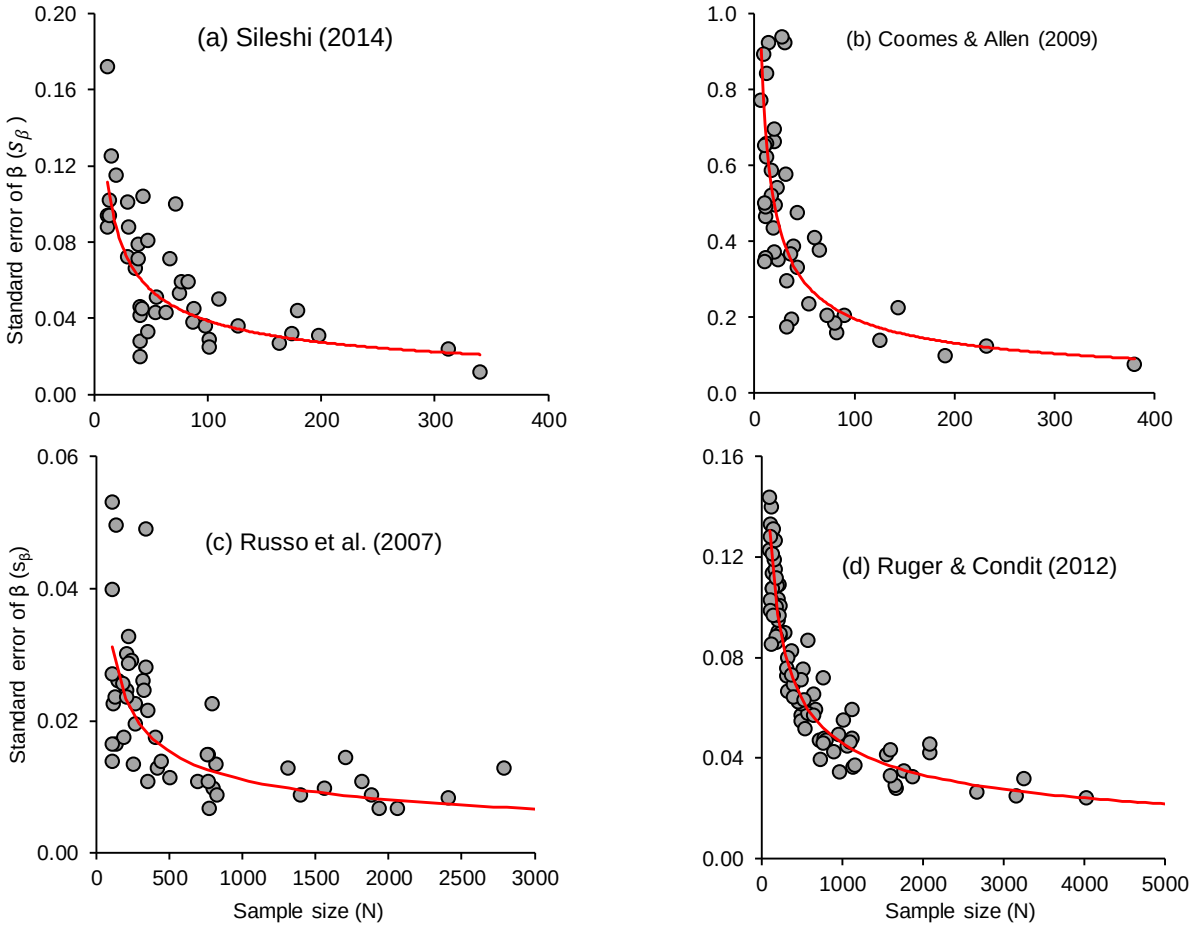

Figure 2. Variation in the standard error of the exponent (s_β) and sample size (N = number of trees) in a bivariate regression of the form $\ln(Y) = \ln(\beta_0) + \beta_1(\ln D) + \epsilon$. Two extreme values of N (>5000) were removed from Figure 2d for clarity.

Samples with small N have low statistical power, hence the estimated β_1 values are less reliable and significant differences may not be detected (i.e. $\beta_1 = 0$ may not be rejected). The β_1 values may also deviate from the hypothesized population value of 0.33 in underpowered samples. As a general rule of thumb N should be larger than 50 for bivariate regression to ensure statistical power

(VanVoorhis & Morgan 2007). The statistical power is the ability of a test to detect an effect, if the effect actually exists. But this concept[4] is not always clear and underappreciated (Jennionsa & Møller, 2003).

Example 2

Example 2 relates variations in β_1 with PRSE using H-D scaling. I collected the data in an earlier study on the fruit tree *Upaca kirkiana*. In Figure 3, I plotted β_1 against the PRSE and compared it with the theoretical value of 0.67 for H-D scaling. With increase in PRSE, β_1 approached zero (blue line) or assumed smaller values than predicted by theory (red dotted line in Figure 3a). Estimates with small values (β_1 = 0) and wider CL are symptomatic of regression dilution (attenuation). This is a statistical phenomenon where measurement errors in the predictor (X) cause flattening of the slope (Hutcheon et al., 2010). If there are measurement errors in the X, β_1 values tend to be underestimated. Errors in Y on the other hand tend to increase $s_{\theta i}$ (Hutcheon et al., 2010). In Figure 3b note that β_1 values significantly larger than the theoretical value (0.67) are associated with poorly estimated values of the constant (β_0). Note that reliable estimates of β_1 (with narrow CI) are found with PRSE of a<20%. Therefore, values of β_1 smaller or larger than theoretical values can be methodological and/or statistical artefacts.

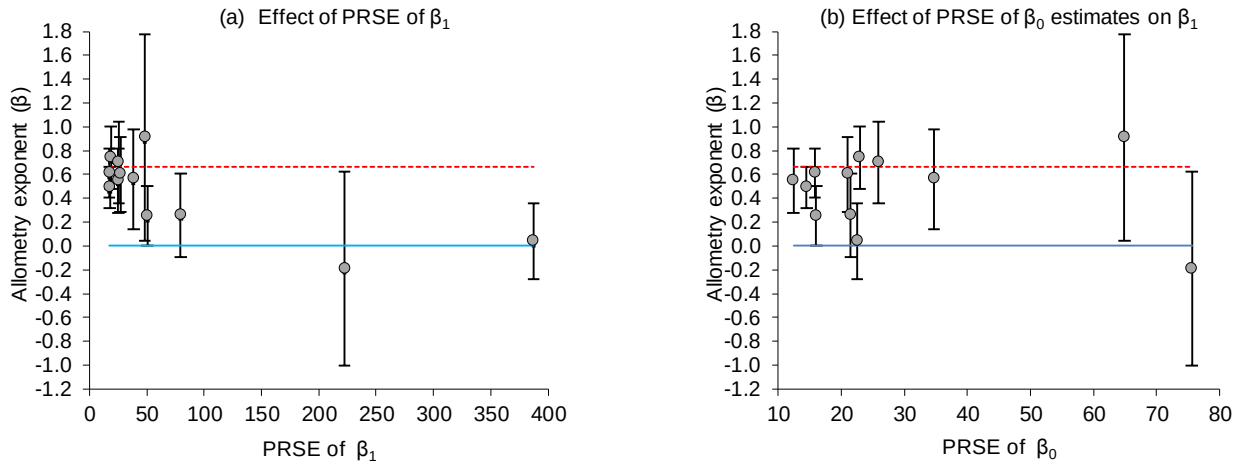

Figure 3. Decreases in the reliability of β_1 with increases in PRSE of β and a in a bivariate regression of tree height on diameter. Error bars represent the 95% CLs. Model expressed as $\ln(H) = \ln(\beta_0) + \beta_1(\ln D) + \epsilon$.

[4] The statistical power of a hypothesis test (1-β) is the probability of correctly rejecting a false null hypothesis. The test size or significance level (α) is the probability of rejecting the null hypothesis that is true.

3.2. Multivariate regression

In this section, I will discuss common problems with models with linear predictors, interaction terms and non-linear terms. Equations with only linear terms have the following form:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \epsilon$

where X_i are either raw or transformed predictors.

There also situations where inclusion of interaction and/or non-linear (e.g. quadratic) terms may be appropriate. A model with an interaction term may look like the following:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2) + \cdots \beta_n X_n + \epsilon$

where X_n may either be a main effect or an interaction term.

A model with non-linear (e.g. quadratic) terms may look like the following:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1)^2 + \beta_4 (X_2)^2 + \cdots \beta_n X_n + \epsilon$

where X_n may either be a main effect or a non-linear term.

More complicated models with two or more linear effects, their quadratic and interaction effects are best tackled with response surface regression, and will be beyond the scope of this short article. Examples of their application are available in Sileshi et al. (2010).

In all these we test the null hypothesis (H_0): $\beta_1 = \beta_2 = \beta_3 = \ldots \beta_n = 0$; and H_a: $\beta_1 \neq \beta_2 \neq \beta_3 \neq \ldots \beta_n \neq 0$. The H_0 is rejected when the *t*-statistic is large and the associated *P* is small for each coefficient. If $\beta_i = 0$, we can conclude that Y and X_i are not associated (but association does not mean causation). However, this task is more complicated because of the added dimensions of collinearity among the X_i, which is not a problem in bivariate regression. Outliers and leverage points also pose a common problem because each X_i comes with its own excess baggage. One of the cardinal assumptions of this type of regression is that X_i are orthogonal (not collinear). However, a variety of data come from observational studies or surveys where multicollinearity is difficult to avoid.

One of the indicators of collinearity is the variance inflation factor (VIF) or its inverse called Tolerance. A common rule of thumb for deciding when a given predictor causes collinearity is if VIF > 5. Nevertheless, values as low as 2 can have significant impacts on $\bar{\theta}_i$ (Graham, 2003). The value of $s_{\theta i}$ for a given coefficient is

doubled when VIF is 4.0 and is larger by 3.16 times when VIF = 10. Therefore, VIF=10 is the upper limit for indicating a definite problem for individual predictors. Together with small N, multicollinearity can inflate $s_{\theta i}$ and change the magnitude and signs of $\bar{\theta}_i$. This hinders interpretation of the coefficients in a meaningful way. Collinearity can also render the interpretation of beta weights difficult. When the predictors are correlated, standardizing does not disentangle the effects of X on Y from the standard deviations of X; in fact, it confounds them in the service of placing all weights on a z-score metric (Nimon and Oswald, 2013). There are different methods to overcome multicollinearity. For example, partial orthogonization of interaction and non-linear terms with respect to their row predictors is one method. Another method involves centred score regressions, which is constructed by subtracting the mean from the raw data for each variable. However, these methods cannot correct the problem if the model has not been correctly specified. I will demonstrate this problem with examples in the following sections.

Example 1

The first example comes from my own study (Sileshi et al., 2003) investigating the effect of nutritional and anti-nutritional factors in the leaves of *Sesbania* species on the weight of adult beetles (*Mesoplatys*) at emergence. The larvae of the beetle were cultured on genotypes of *Sesbania* with different foliar nitrogen (N), polyphenol (P) and lignin (L) contents affecting larval fitness. Adult fitness indexed by weight at emergence was positively correlated with larval weight gain. Therefore, in this study I set out to determine whether or not these factors can affect adult weight at emergence. Exploratory analyses had indicated that adult weight at emergence had a non-linear relationship with foliar N and P content. Therefore, I included quadratic and interaction terms in different models. In order to overcome multicollinearity, I used either orthogonal[5] or centred scores of the quadratic and interaction terms. In some models I only included raw predictors and their interaction or quadratic terms. In others I entered the linear and

[5] *This is partial Gram-Schmidt orthogonalization where the residuals replace the corresponding interaction or non-linear effects.*

orthogonal terms (see Table 2). I then compared these models in terms of the magnitude and signs of the partial regression coefficients ($\bar{\theta}_i$), VIF and AIC. The orthogonal and centred regression produced more or less the same outcome. Therefore, I will present only the orthogonal regression here (Table 2) and keep the centred regression in Appendix Table 1. In all cases, the omnibus F-test was highly significant (P<0.0001) indicating that the model as a whole is significant, i.e. the predictors, taken together, predict adult weight at emergence adequately. However, the regression coefficients in model #2 contradict this, i.e. none are significant. Although the R^2 values and AIC do not differ among model #1, #2, #3 and #4, close examination of the $\bar{\theta}_i$, $s_{\theta i}$ and VIF reveal fundamental differences. In the model with only raw predictors (#1) the VIFs were close to unity. When interaction terms were entered (#2), VIF dramatically increased for all $\bar{\theta}_i$, $s_{\theta i}$ were inflated and $\bar{\theta}_i$ did not differ from zero. In the model where interaction terms were orthogonal (#3), collinearity was reduced and two coefficients significantly differed from zero. However, note the change in the sign of the estimates for the different specifications. In the model with raw quadratic terms (#4), the VIF are large indicating multicolinearity. When the quadratic terms were made orthogonal (#5), four coefficients were significantly different from zero. Note also that the sign of the $\bar{\theta}_i$ changed from negative in #4 to positive in #5. This indicates that collinearity can make $\bar{\theta}_i$ unstable (unpredictable). The R^2 and AIC tell us that #4 and #5 are better than #1-3 thus painting a misleading picture.

According to the principle of parsimony, if two models explain the data about equally well, we choose the simpler one. A simplified model (#6a) without lignin and its non-linear terms appears to be better than all the others in terms of internal validity[6]. The best model (#6b) picked using all subset selection methods (R^2 and AIC) had unreliable coefficients (large VIF) for β_2 and β_3. From the forgoing discussion, it is apparent that correctly specifying the models and including orthogonal or centred predictors improve validity. Please do not entirely depend on R^2 and AIC to choose your models.

[6] *Although model validity is a complex concept, a model may be called internally valid if it (a) is biologically sensible and amenable to meaningful interpretation, (b) does not violating regression assumptions, (c) has coefficients (β_i) that are all significant, (d) has significant F-statistic and R^2, and (e) makes definite predictions.*

Table 2. Estimates of regression coefficients ($\bar{\theta}_i$), their standard errors ($s_{\theta i}$), significance levels and variance inflation factors (VIF) in various models relating adult weight at emergence with foliar nitrogen (N), polyphenol and lignin content. N was 33 observations.

Model	Predictors	Variables	$\bar{\theta}_i$	Estimate	$s_{\theta i}$	t-value	P>|t|	VIF	R^2_{adj}	AICc
#1	Linear terms (LT)	Intercept	β_0	0.109	0.043	2.5	0.018		0.554	-230.1
		N	β_1	0.040	0.012	3.3	0.003	1.0		
		P	β_2	-0.028	0.005	-5.6	<0.001	1.0		
		L	β_3	-0.001	0.001	-1.0	0.424	1.1		
#2	LT + Interactions (raw predictors)	Intercept	β_0	-0.452	0.325	-1.4	0.177		0.569	-238.5
		N	β_1	0.187	0.097	1.9	0.065	65.2		
		P	β_2	0.070	0.061	1.1	0.261	187.5	P	
		L	β_3	0.047	0.028	1.7	0.102	538.1		
		N*P	β_4	-0.022	0.017	-1.3	0.201	163.1		
		N*L	β_5	-0.011	0.008	-1.4	0.201	433.7		
		P*L	β_6	-0.004	0.003	-1.3	0.109	61.9		
#3	LT + Interactions (orthogonal)	Intercept	β_0	0.081	0.046	1.8	0.090		0.563	-238.9
		N	β_1	0.048	0.013	3.7	<0.001	1.2		
		P	β_2	-0.027	0.005	-5.4	<0.001	1.1		
		L	β_3	-0.001	0.001	-1.0	0.527	1.1		
		N*P	β_4	-0.019	0.016	-1.2	0.255	1.7		
		N*L	β_5	-0.010	0.008	-1.3	0.236	1.5		
		P*L	β_6	-0.004	0.003	-1.3	0.125	1.3		
#4	LT + Nonlinear (non-orthogonal)	Intercept	β_0	0.396	0.201	2.0	0.060		0.675	-241.0
		N	β_1	-0.063	0.135	-0.5	0.643	167.3		
		P	β_2	-0.121	0.030	-4.0	<0.001	59.9	P	
		L	β_3	-0.001	0.001	-1.0	0.521	1.1		
		N^2	β_4	0.017	0.021	0.8	0.432	165.0		
		P^2	β_5	0.014	0.004	3.5	0.003	58.2		
#5	LT + Nonlinear (orthogonal)	Intercept	β_0	0.091	0.037	2.5	0.022	0	0.675	-241.0
		N	β_1	0.043	0.011	3.9	<0.001	1.1		
		P	β_2	-0.026	0.004	-6.5	<0.001	1.2	P	
		L	β_3	-0.001	0.001	-1.0	0.519	1.1		
		N^2	β_4	0.017	0.021	0.8	0.420	1.3		
		P^2	β_5	0.014	0.004	3.5	0.003	1.1		
#6a	LT + Nonlinear (orthogonal)	Intercept	β_0	0.083	0.033	2.5	0.017		0.691	-242.8
		N	β_1	0.045	0.010	4.5	<0.001	1.0		
		P	β_2	-0.027	0.004	-7.1	<0.001	1.0		
		P^2	β_3	0.015	0.004	3.6	<0.001	1.0		
#6b	LT + Nonlinear (Raw predictors)	Intercept	β_0	0.235	0.051	4.6	<0.001		0.687	-245.2
		N	β_1	0.045	0.010	4.4	<0.001	1.0		
		P	β_2	-0.129	0.028	-4.6	<0.001	56.9		
		P^2	β_3	0.016	0.004	3.7	<0.001	56.8		

AICc is the small sample correction for AIC.

Example 2

Example 2 consists in entering linear and non-linear terms of tree diameter (D) and wood density (ρ) for prediction of above-ground biomass (AGB) of trees in a model expressed as: $ln(AGB) = ln(a) + b(lnD) + c(lnD)^2 + d(lnD)^3 + e(\ln \rho)$. This

equation was first proposed by Chave et al. (2005) and now applied for predicting forest biomass and carbon stocks (Sileshi, 2012). In order to demonstrate problem with this equation, I fitted it to two datasets. I chose these datasets because of sample size; the first dataset is medium sized (N = 72) and came from Kuyah et al. (2012) while the second is very large (N = 4004) dataset from Chave et al. (2014). In both datasets the explained variance (adjusted R^2) is very high. The omnibus *F*-statistic, which tests the joint significance of $\beta_i = 0$, is also large (P<0.0001). According to the *t*-test and the 95% CI, however, all $\bar{\theta}_i$ except for ρ are non-significant in Dataset 1. On face value we could accept H_0: $\beta_1 = \beta_2 = \beta_3 = 0$. In Dataset 2 β_2 and β_3 are non-significant; so we will reject H_0: $\beta_2 = 0$. When we compare the magnitude and signs of β_1, β_2 and β_3 in Dataset 1 and 2 they change erratically (see yellow and blue highlights in Table 3). Compare these also with the model by Feldspauch et al (2012); i.e. $\ln(AGB) = -1.822 + 2.337(lnD) + 0.163(lnD)^2 - 0.025(lnD)^3 + 0.979(\ln \rho)$. The *P*-values seemingly indicate that wood density is a better predictor of AGB than D in both datasets. This contradicts empirical data.

Table 3. Parameters of the model and their significance indicated by the P-value and 95% confidence intervals (CI). L95% and U95% represent the lower and upper CI

Dataset	*Predictors*	$\bar{\theta}_i$	*Estimate*	s_θ	*t-value*	*P-value*	*L95%*	*U95%*	R^2
1 (Kuyah)	$\ln(\beta_0)$	β_0	-0.44	0.90	-0.49	0.6267	-2.23	1.35	0.990
(N = 72)	ln(D)	β_1	1.36	0.97	1.39	0.1677	-0.58	3.30	
	$\ln(D)^2$	β_2	0.27	0.34	0.82	0.4170	-0.40	0.95	
	$\ln(D)^3$	β_3	-0.02	0.04	-0.64	0.5267	-0.10	0.05	
	ln(ρ)	β_4	0.74	0.13	5.59	<0.0001	0.48	1.01	
2 (Chave)	$\ln(\beta_0)$	β_0	-3.30	0.35	-9.35	<0.0001	-4.00	-2.61	0.941
(N = 4004)	ln(D)	β_1	3.50	0.37	9.53	<0.0001	2.78	4.22	
	$\ln(D)^2$	β_2	-0.24	0.12	-1.98	0.0480	-0.48	0.00	
	$\ln(D)^3$	β_3	0.02	0.01	1.55	0.1220	-0.01	0.04	
	ln(ρ)	β_4	0.66	0.03	23.35	<0.0001	0.60	0.71	

From Table 4 blow we see that the lack of significance (and instability) of coefficients is linked to the collinearity between the predictors. In the column marked PRSE, note that β_1, β_2 and β_3 have very large PRSE values (hence are unreliable). The columns marked VIF and DIs_θ shows that s_θ is inflated by several orders. For example, the variance of β_1 is 31 and 36 times larger than it would be if the predictors were orthogonal in Dataset 1 and 2, respectively. These problems

make such models useless for either explanation or prediction. However, in some publications authors have claimed their model is superior over others because it achieved high R^2 and small AIC. In reality, this model is not only statistically unstable but also biologically implausible (see Sileshi, 2014 for details). As such it has no predictive power and practically useless.

Table 4. Indicators of multicollinearity and unreliability of $\bar{\theta}_i$ in the model described in Table 3. Large percent relative standard error (PRSE) values indicate unreliability of $\bar{\theta}_i$ and the variance inflation factors (VIF) indicates multicollinearity. The degree of inflation (DIs_θ) indicates by how many times s_θ is larger than it normally should be.

Dataset	Predictors	$\bar{\theta}_i$	PRSE	VIF	DIs_θ
1	ln(β_0)	β_0	204.6		
	ln(D)	β_1	71.7	975.7	31.2
	ln(D)^2	β_2	122.5	4474.4	66.9
	ln(D)^3	β_3	157.1	1340.9	36.6
	ln(ρ)	β_4	17.9	1.6	1.3
2	ln(β_0)	β_0	10.7		
	ln(D)	β_1	10.5	1271.3	35.7
	ln(D)^2	β_2	50.6	5255.1	72.5
	ln(D)^3	β_3	64.7	1438.3	37.9
	ln(ρ)	β_4	4.3	1.0	1.0

*Unreliable estimates are indicated in red.

Example 3

The third example consists of fitting a linear model that includes D, tree height (H), ρ and crown area (CA) as predictors of AGB, i.e. $ln(AGB) = ln(a) + b(lnD) + c(lnH) + d(\ln \rho)$. Models like these are motivated by the argument that more accurate models of forest biomass are those that incorporate as many predictors as possible (e.g. Goodman *et al*., 2014). However, including many predictors has also its own problems. In order to demonstrate this problem, I fitted this model to the dataset from Kuyah et al. (2012) described above. The outcomes (Table 5) give the wrong impression that H is not a significant predictor of AGB contradicting theory and empirical data. This is partly due to the collinearity between the predictors (see large VIF for H and CR). The coefficients of ln(D) and ln(CR) are unreliable (large PRSE) because their variances have been inflated by up to 3 times than they would be if the predictors were orthogonal (se DI in Table 5). If the goal of this modelling exercise was to develop an explanatory model relating

variation in AGB using these predictors, the exercise has failed. It is well known that when too many predictors are entered in a model, some predictors can mask the truly significant ones. If two or more variables are highly correlated, they are essentially containing the same information and are measuring the same thing. For example, D, H and CR contain much of the same information about AGB. Therefore, the claim by some workers (e.g. Goodman *et al*., 2014; Picard et al., 2015) that including these variables as separate predictors improves biomass prediction is misleading. In some studies, H estimated from D has been included together with D in biomass estimation models. This practice constitutes "double accounting" besides the collinearity it presents.

Table 5. Example of collinearity in a biomass estimation model. Model parameters were estimated using sample size (N) of 72 trees. The raw data came from Kuyah et al. (2012).

Predictors	$\bar{\theta}_i$	*Estimate*	$s_{\theta i}$	*t-value*	*P-value*	PRSE	VIF	DI
ln(β_0)	β_0	-1.39	0.20	-6.85	<0.001	14.6		
ln(D)	β_1	2.08	0.11	19.8	<0.001	5.0	11.9	3.4
ln(H)	β_2	0.13	0.07	1.82	0.073	54.9	3.1	1.8
ln(CR)	β_3	0.12	0.05	2.44	0.018	41.1	7.2	2.7
ln(WD)	β_4	0.76	0.13	5.79	<0.001	17.3	1.7	1.3

The omnibus F-test is highly significant (P = <0.001).

This problem is not restricted to OLS regression but also GLMs. For example, the use of many predictors and dummy variables in logistic regression results in coefficients with inflated $s_{\theta i}$ and small χ^2 statistic. I will demonstrate this using the following two examples.

Example 4

Example 4 is based on a study by Gregory and Sewando (2013) who analysed the adoption of quality protein maize technology. The authors entered 13 predictor variables in a logistic regression model and inferred that β_3, β_8, β_9, β_{10}, β_{12} and β_{13} have influence on the rate of adoption. They deemed the rest of the variables non-significant. However, close examination of Table 6 reveals that most of the coefficients have inflated $s_{\theta i}$ and hence are unreliable. This is indicated by large PRSE values (marked red). Therefore, we cannot confidently say that the predictors with P>0.05 are truly non-significant. The lack of significance is the manifestation of multicollinearity among the predictors.

Table 6. Parameters estimates of a logit model for factors influencing the adoption of QPM technology in Tanzania (source Gregory and Sewando, 2013)

Predictors	$\bar{\theta}_i$	*Estimate*	$s_{\theta i}$	*P*-value	PRSE
Age of the household head	β_1	0.03	0.41	0.48	1366.7
Sex of the household head	β_2	0.88	1.28	0.49	145.5
Education level of household head	β_3	0.45	0.22	0.06	48.9
Number of people in household	β_4	0.07	0.18	0.69	257.1
Number of people working on the farm	β_5	-0.76	0.39	0.85	51.3
Farm size	β_6	0.12	0.16	0.46	133.3
Attendance of training	β_7	-1.47	1.25	0.24	85.0
Attendance of farmers field day	β_8	2.17	1.11	0.04	51.2
Livestock ownership	β_9	3.26	1.85	0.08	56.7
Participation in demonstration trials	β_{10}	4.75	1.52	0.00	32.0
Frequency of extension contact	β_{11}	-0.03	0.35	0.93	116.7
QPM marketability	β_{12}	-1.13	0.34	0.00	30.1
Access to credit	β_{13}	-3.82	1.37	0.03	35.9
Constant	β_0	-8.79	4.19	0.04	47.7

Values in red indicate unreliable estimates

Example 5

This example is based on a study by Raut et al. (2011) who analysed factors that influence agricultural intensification in central Nepal. They collected data from 310 households and entered 14 predictors in a logistic regression model. They then inferred that 5 out of the 14 predictors had a significant influence on adoption (Table 7 below). However, close examination of the $\bar{\theta}_i$ revealed that 12 out of the 14 estimates have inflated $s_{\theta i}$ and were unreliable. In this study N was large enough to correctly estimate the coefficients. Green (1991) suggests $N > 50 + 8p$ for testing individual predictors in multivariate regression situations. Based on this, N = 154 would have been adequate for this study. Regardless of the large N used, the variance inflation caused multicollinearity among the predictors has derailed the analysis. The authors chose the 14 predictors after correlation analysis of 15 predictors to determine the collinearity. They excluded predictors with correlation coefficient (r >0.5) with other predictors. Note however, using bivariate correlations is inadequate to detect multicollinearity. Therefore, one cannot confidently say that the non-significant variables are indeed not associated with adoption.

Table 5. Estimated parameters of factors that influence households' participation in Bt maize cultivation (source: Raut et al., 2011)

Predictors	$\bar{\theta}_i$	*Estimate*	$s_{\theta i}$	P value	PRSE
Caste	β_1	-0.07	0.35	0.836	483.6
Gender	β_2	0.31	0.36	0.384	114.8
Education	β_3	-0.22	0.24	0.351	107.1
Occupation	β_4	-0.12	0.82	0.880	664.5
Agricultural labor input	β_5	0.00	0.00	0.294	100.0
Cattle herd size (LSU)	β_6	0.02	0.14	0.879	652.4
Landholding size	β_7	3.65	1.77	0.039	48.6
Distance to farm plots	β_8	0.16	0.17	0.362	110.2
Crop yield	β_9	0.00	0.00	0.020	50.0
Net income	β_{10}	0.00	0.00	0.506	1.0
Irrigation frequency	β_{11}	0.99	0.12	0.000	12.0
Distance to fertilizer store	β_{12}	-0.23	0.12	0.042	49.4
Access to credit	β_{13}	1.01	0.52	0.050	51.1
Frequency of extension assistance	β_{14}	-0.20	0.39	0.620	201.0
Constant	β_0	-6.96	1.89	0.000	27.2

In analyses like Example 4 and 5, the algorithm may also fail to converge and return unreliable $\bar{\theta}_i$ values because of lack of invertible Hessians for some combinations of predictors. The most common advice for this situation is to rethink the model, simplify it, and rerun the analysis or get more data (Gill & King, 2004).

4. Helpful rules of thumb

In section 2 and 3 we have seen that a number of things can go wrong and make $\bar{\theta}_i$ unreliable. Many of the problems affect the $s_{\theta i}$. We have also seen that the PRSE can be used as an easy index for checking the reliability of $\bar{\theta}_i$. Estimates could become unreliable due to various factors. Therefore, it is important to consider diagnostic tests as an integral part of regression model-building. A few of the diagnostic criteria include the leverage statistic (*Hat value), COVRATIO, DFBETAS* and *Cook's D*. The Hat value is a measure of how far an observation is from the others in terms of the levels of the independent variables. *COVRATIO* measures the impact of each observation on the $s_{\theta i}$ of $\bar{\theta}_i$ and their co-variances. *DFBETAS* measure both leverage and the effect of a large residual, hence indicate how much an observation has affected the estimate of each $\bar{\theta}_i$. *Cook's D* is a measure of the aggregate impact of each observation on each $\bar{\theta}_i$ as well as the group of

fitted values. In the literature, various rules of thumb are available (Belsley et al., 1980). Note that rules of thumb and cut-off points are only guidelines. Therefore, it is a mistake to state them as a single value because the effects on the whole regression or each $\bar{\theta}_i$ depends on the number of predictors (P) and the sample size. Therefore, for example stating Hat value less than <0.5 are not a problem is inadequate. I have summarized the common rules of thumb proposed in the literature in Table 6 below.

Table 6. Summary of useful rules of thumb for correct regression diagnostics

Problem	Criteria	Rule of thumb
Autocorrelation	Durbin-Watson stat (d)	1.5 < d < 2.5
Collinearity	VIF	VIF > 5
	Conditioning index	>15
Outliers	Studentized residuals	>|2|
Leverage points	Hat diagnostic (H_i)	$H_i > \frac{2(P+1)}{n}$
Influence	Cook's distance (D_i)	$D_i > \frac{4}{(n-P-1)}$
	DFBETAS (Df)	$\|Df\| > \frac{2}{\sqrt{n}}$
	COVRATIO	$1 \pm \frac{3(P+1)}{n}$
Sample size	*N*	$N > 50 + 8P$

References

Albert, A. & Anderson, J.A. (1984) On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 71, 1-10.

Belsley, Kuh & Welch (1980) *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*.

Bruton, A, Conway, J H & Holgate, S T (2000). Reliability: What is it and how is it measured? *Physiotherapy* 86, 94-99.

Buntin, G.D. (1994) Developing a primary sampling program. *In*: Pedigo, L.P. and Buntin, G.D. (Eds.), *CRC Handbook of Sampling Methods for Arthropods in Agriculture*. CRC Press Inc., Boca Raton, Florida, pp. 99–115.

CDC/National Center for Health Statistics, 2010. Reliability of Estimates. http://www.cdc.gov/nchs/ahcd/ahcd_estimation_reliability.htm (downloaded on 15/06/2014).

Chave J, Réjou-Méchain ME, Búrquez A *et al* (2014) Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20, 3177–3190.

Coomes, D.A. & Allen, R.B. (2009) Testing the Metabolic Scaling Theory of tree growth. *Journal of Ecology* 97, 1369–1373.

Feldpausch, T.R., Lloyd, J., Lewis, S.L, et al. (2012) Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* 9, 3381–3403.

Freedman, D.A. (1983) A note on screening regression equations. *American Statistician* 37, 152–155.

Gill, J. & King, G. (2004) Numerical Issues Involved in Inverting Hessian Matrices. In: Altman, M., Gill, J. & McDonald, M.P. (ed.) *Numerical Issues in Statistical Computing for the Social Scientist*. John Wiley & Sons, Inc. pp 143-176.

Goodman RC, Phillips OL & Baker TR (2014) The importance of crown dimensions to improve tropical tree biomass estimates. *Ecological Applications* 24, 680–698.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology* 84, 2809–2815.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research 26,* 499-510.

Gregory, T. & Sewando, P. (2013) Determinants of the probability of adopting quality protein maize (QPM) technology in Tanzania: A logistic regression analysis. *International Journal of Development and Sustainability* 2(2): 729-746.

Helland, I.S. (1987). On the interpretation and use of R^2 in regression analysis. *Biometrics* 43, 61-69.

Higgins, J.P.T. & Green, S. (eds) (2011) *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, 672 pp.

Hutcheon, J.A, Chiolero, A. & Hanley, J.A. (2010) Random measurement error and regression dilution bias. *BMJ* 340:c2289.

Jennionsa M.D. & Møller, A.P. (2003) A survey of the statistical power of research in behavioral ecology and animal behaviour. Behavioral Ecology Vol. 14 No. 3: 438–445.

Kuyah S, Dietz J, Muthuria C *et al* (2012) Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. *Agriculture Ecosystems and Environment* 158, 216–24.

Larjavaara, M. & Muller-Landau, H.C., 2013. Measuring tree height: a quantitative comparison of two common field methods in a moist tropical forest. *Methods in Ecology and Evolution* 4, 793–801.

Liao, D. & Valliant, R. 2012. Variance inflation factors in the analysis of complex survey data. *Survey Methodology* 38: 53-62.

Nimon, K.F. & Oswald, F.L. 2013. Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods* 16: 650-674.

Palmer, P.B. & O'Connell, D.G. 2009. Regression analysis for prediction: understanding the process. *Cardiopulmonary Physical Therapy Journal* 20, 23-26.

Picard, N., Rutishauser, E., Ploton, P., Ngomanda, A. & Henry, M. 2015. Should tree biomass allometry be restricted to power models? *Forest Ecology and Management* 353, 156-163.

Raut, N., Sitaula, B.K., Vatn, A. & Paudel, G.S. (2011) Determinants of adoption and extent of agricultural intensification in the central mid-hills of Nepal. *Journal of Sustainable Development* 4, 47-60.

Rüger, N. & Condit, R. (2012) Testing metabolic theory with models of tree growth that include light competition. *Funct. Ecol.*, 26, 759–765.

Russo, S.E., Wiser, S.K. & Coomes, D.A. (2008) A re-analysis of growth–size scaling relationships of woody plant species. *Ecol. Lett.*, 11, 311–312.

Sileshi G (2006) Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological Research* 96: 479-488.

Sileshi G (2008) The excess-zero problem in soil animal count data and choice of models for statistical inference. *Pedobiologia* 52: 1-17.

Sileshi GW (2012) A critique of current trends in the statistical analysis of seed germination and viability data. *Seed Science Research* 22: 145–159.

Sileshi, G.W. (2014) A critical review of forest biomass estimation models, common mistakes and corrective measures. *Forest Ecology and Management* 329, 237–254.

Sileshi, G., Sithanantham, .S, Mafongoya, P.L., Ogol, C.K.P.O. & Rao, M.R. (2003) Biology of *Mesoplatys ochroptera* Stål (Chrysomelidae: Coleoptera), a pest of *Sebania* species in southern central Africa. *African Entomology* 11: 49-58.

Sileshi, G., Hailu, G., Nyadzi, G.I. (2009) Traditional occupancy–abundance models are inadequate for zero–inflated ecological count data. *Ecological Modelling* 220, 1764-1775.

Sileshi, G., Akinnifesi, F.K., Debusho, L.K., Beedy, T., Ajayi, O.C. & Mong'omba , S. (2010) Variation in maize yield gaps with plant nutrient inputs, soil type and climate across sub-Saharan Africa. *Field Crops Research* 116: 1-13.

Southwood, T.R.E. (1978) *Ecological methods*. London, Chapman Hall. 524p.

Stellingwerf, D. A. (1994). *Forest Inventory and Remote Sensing*. Enschede, Overijssel: International Training Centre for Aerial Survey (ITC).

VanVoorhis, C.R. W. & Morgan, B.L. (2007) Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology* 3: 43-50.

Appendix

Table 1. Estimates of centred and standardized regression coefficients ($\bar{\theta}_i$), their standard errors ($s_{\theta i}$), significance levels and variance inflation factors (VIF) of models of adult weight at emergence

Predictors	Variables	$\bar{\theta}_i$	Estimate	$s_{\theta i}$	t-value	P>|t|	PRSE	R^2_{adj}	AICc
LT + Interactions	N	β_1	0.040	0.014	2.9	0.010	35.0	0.569	-225.1
(centred)	P	β_2	-0.028	0.005	-5.6	<0.001	17.9		
	L	β_3	-0.001	0.002	-0.5	0.575	200.0		
	N*P	β_4	-0.022	0.017	-1.3	0.201	77.3		
	N*L	β_5	-0.011	0.008	-1.4	0.201	72.7		
	P*L	β_6	-0.004	0.003	-1.3	0.109	75.0		

Note that the standardized regression gives poorer results than the centred and orthogonal regression. In this procedure, β_0 and its corresponding s_θ were supposed to be zero. However, $s_{\theta i}$ is large. I do not know why. If the $\bar{\theta}_i$ estimates were correct, their magnitude will give an idea of their respective contribution to the variation in Y.