Spurious Correlation due to Scaling

February 2019

Abstract: Scaling is common in empirical accounting research. It is often done to mitigate heteroskedasticity or the influence of firm size on parameter estimates. However, using analytic results and Monte Carlo simulations we show that common forms of scaling induce substantial spurious correlation via biased parameter estimates. Researchers are typically better off dealing with both heteroskedasticity and the influence of larger firms using techniques other than scaling.

JEL classification: C15, C18, M41

Keywords: scaling, deflating

1 Introduction

When accounting researchers estimate models they frequently divide variables in a regression model by another variable associated with how large the company is. This is commonly referred to as "deflating" or "scaling." Scaling is typically done to address two concerns. First, researchers are worried that large firms may have larger error terms and small firms may have smaller error terms (i.e., errors are heteroskedastic). Second, researchers worry that the relation between the variables of interest is different for large firms compared to small firms (i.e., non-constant marginal effects).

The presence of scale effects, and the optimal methods to mitigate such effects, have been debated in the accounting literature since the early 1980's. Earlier studies worried that scaling may introduce spurious dependence (Lev and Sunder 1979). Later studies are more concerned with the choice of correct "deflators," arguing that deflating by some variables is more appropriate then by others (Christie 1987). Some argue that large firms have undue influence on estimated coefficients (Easton and Sommers 2003), while others argue that this is not necessarily the case (Barth and Clinch 2009).

We contribute to the scaling debate by showing that regardless of the chosen deflator, researchers may unintentionally induce spurious dependence via scaling in certain circumstances. Using analytical as well as Monte Carlo results, we demonstrate two facts that, we hope, will inform the scaling debate. First, we show that one common form of scaling, where all variables but the intercept are deflated, introduces spurious dependence for both levels models and returns models. According to our Monte Carlo simulations for returns models, if we generate four independent variables $X1$, $X2$, Y, and Z and estimate $\frac{Y}{Z} = \beta_0 + \beta_1 \frac{X1}{Z} + \beta_2 \frac{X2}{Z} + \epsilon$, the estimates of β_1 and β_2 will be significantly different from zero between 64% and 99% of the time rather than the expected 5% of the time under the traditional assumptions and critical values. This process is repeated for levels models and similar rejection rates are observed. Complementing our Monte Carlo simulations, we show analytically that for the true model $Y = \beta_0 + \epsilon$, the estimate of the slope of the regression $\frac{Y}{Z} = \beta_0 + \beta_1 \frac{X}{Z} + \epsilon$, $\hat{\beta}_1$ is biased and OLS standard errors are inappropriate.

Second, we show that if a researcher wants to scale or deflate variables, the appropriate form of scaling is to deflate the constant term along with the other regressors as follows:

$$\frac{Y}{Z} = \frac{\beta_0}{Z} + \beta_1 \frac{X1}{Z} + \beta_2 \frac{X2}{Z} + \epsilon,$$

and to use robust standard errors. According to our Monte Carlo results, if we generate four independent variables $X1$, $X2$, Y, and Z, estimate $\frac{Y}{Z} = \frac{\beta_0}{Z} + \beta_1 \frac{X1}{Z} + \beta_2 \frac{X2}{Z} + \epsilon$ and use robust standard errors, the estimates of β_1 and β_2 will be significantly different from zero about 10% of the time, which is much closer to the expected 5%. Also, we show analytically that for the true model $Y = \beta + \epsilon$, the expected value of the estimate of the slope in $\frac{Y}{Z} = \frac{\beta_0}{Z} + \beta_1 \frac{X}{Z} + \epsilon$, $\hat{\beta}_1$ is unbiased and equal to 0 and robust standard errors are appropriate.

Our results are alarming for two reasons. First, because scaling is prevalent in empirical accounting studies, inappropriate scaling may affect numerous studies. For example, Volume 93, Issues 1-5 of recently published articles in *The Accounting Review* contained 50 archival papers, 36 of which scaled, and 9 of these scaled in a manner that induces spurious correlation. Thus, 25% of modern papers which used scaling in an elite journal scaled in a manner which could invalidate hypothesis testing. Research designs in elite journals are often emulated by other researchers so it is likely that these scaled model specifications will be used again in subsequent publications. Second, we demonstrate that the magnitude of the researcher-induced bias is substantial and meaningfully influences hypothesis tests. In fact, our models indicate that nearly three quarters of the time researchers may erroneously conclude that independent random variables are correlated simply due to improper scaling. Thus, we recommend against scaling. If scaling is unavoidable, we recommend using Weighted Lease Squares with heteroskedasticity-robust standard errors for cross-sectional data or clustered for panel data.

2 Literature Review

The literature that investigates the statistical consequences of scaling has a long history. Almost forty years ago, Lev and Sunder (1979) noted that ratios (one variable scaled by another) are frequently used in accounting and finance. They suggest that the major reason for using scaling is to control for size: either firm size or the size of the error variance (heteroskedasticity). Lev and Sunder (1979) demonstrated that scaling is appropriate only under very strict conditions: when a theory implies strict proportionality, $y = \beta x$; there is no error term; no intercept; and no dependence of y on other variables. Only when all four conditions are satisfied is scaling appropriate. Lev and Sunder (1979) also show that scaling introduces heteroscedasticity when an additive homoscedastic error term is present in the model (i.e., $y = \beta x + \epsilon$). Further, scaling introduces bias if the model has an intercept, $y = \beta_0 + \beta_1 x$. Finally, Lev and Sunder (1979) observed that estimation results are difficult to interpret in the presence of other variables, (i.e., $y = \beta_0 + \beta_1 x + \beta_2 z$) or nonlinearities.

Lev and Sunder (1979) were also concerned that scaling may result in spurious inference. They showed that the correlation between scaled variables $\frac{x}{z}$ and $\frac{y}{z}$ may be greater or smaller than the correlation between the unscaled variables x and y. Lev and Sunder (1979) argue that two conditions are necessary for the two correlations to be close. First, the variance of z has to be small. Second, both x and y have to be linear and homogeneous functions of z. If the homogeneity assumption is not met, the correlation of scaled variables will be a biased estimate of the correlation of unscaled variables.

Lev and Sunder conclude:

> *"One should not carelessly deflate observations by some intuitively appealing variable without reference to the objective of investigation... It is a matter of concern that the use of financial ratios appears to be often motivated more by custom and tradition than by explicit reference to the specific hypothesis."*

An early paper that investigates the effect of various forms of scaling empirically is Lustgarten (1982). While studying the information content of ASR 190 (which requires firms to disclose replacement costs) he compares estimates of models scaled by four different variables: earnings, total assets, shares outstanding, and market value. After scaling all variables in the model but the constant (Ordinary Least Squares regression), he reports that unanticipated replacement costs are significant in three of the five models. Scaling all variables including the constant (Weighted Least Squares regression), he finds that unanticipated replacement costs are significant in four of the five models. Lustgarten suggests that the insignificant WLS estimate is due to spurious correlation caused by market value being part of the denominator of the dependent and independent variables. This explanation is something that Christie (1987) cites and takes issue with.

Christie (1987) investigates the effect of scaling in market-based accounting research. Christie observes that scaling is common both in return and levels studies and notes that the common stated objective of deflation is heteroskedasticity of the error term. He posits that market value is the correct deflator in return studies because market value is the denominator of returns in the theoretical model. By definition, returns are the sum of expected dividends and changes in stock price divided by current stock price. Christie (1987) argues against Lustgarten's (1982) spurious inference concern and states that nothing is wrong with the same scaling variable appearing in the denominators of both the dependent and independent variables. In fact, he concludes that the beginning period market value is correct deflator precisely because it does appear in the denominator of the mathematical calculation of returns. Christie suggests that use of any other scaling variable leads to biased and inconsistent estimators due to the omitted variable bias. To illustrate, if the true model is $u = \beta_0 + \beta_1 \frac{x}{S} + \epsilon$, then estimating $u = \beta_0 + \beta_1 \frac{x}{W} + \epsilon$ omits a variable equal to $\frac{x}{S} - \frac{x}{W}$. In level studies, Christie recommends scaling by a function of an independent variable. He argues that scaling by any other variable results in incorrect inference.

Easton and Sommers (2003) also examine scaling, but their primary concern is not heteroskedasticity. Rather, Easton and Sommers (2003) argue that the relation between market value, book value and net income is different for large firms compared to small firms. To demonstrate that this is the case, Easton and Sommers estimate unscaled regressions, plot means of Studentized residuals of 40 groups stratified by market value. They report that means of the first 37 groups are between -0.1 and 0.1, as expected. On the other hand, the mean of the 39th group is 0.34 and that of 40th group is 2.6, much higher than one would expect from Studentized residuals. Based on these results, Easton and Sommers (2003) conclude that large firms have undue influence on estimated coefficients. To demonstrate that scaling addresses the problem, they estimate scaled regressions and report that the means of Studentized residuals of all 40 groups are bounded between -0.75 and 1.06. Easton and Sommers (2003) conclude that, first, scaling eliminates undue influence of large firms on estimated coefficients. Second, market value is the most appropriate deflator even in the models where market value is the dependent variable.

Barth and Clinch (2009) is the most recent and the most extensive study of the effect of scaling. They examine five potential scale effects that can cause inference problems: in addition to common concerns of scale-related heteroskedasticity and scale-varying coefficients, they also study additive scale effects, multiplicative scale effects, and survivor bias. Barth and Clinch (2009) use Monte Carlo simulations to generate artificial data with and without various forms of these five scale effects. First, they check if common diagnostic tools are able to correctly detect various scale effects and conclude that common diagnostics are typically ineffective. Second, they estimate the relation between market value, book value and net income comparing the performance of an unscaled model to five alternative specifications that are all intended to mitigate scaling effects: 1) scaling by number of shares, 2) scaling by book value, 3) scaling by lagged price, 4) using returns as the dependent variable rather than a "levels" specification which uses market value of equity, and 5) scaling by market capitalization.

Based on simulated scale-free data, Barth and Clinch (2009) report that firms size differences alone do not necessarily cause inference problems. They also report that an unscaled model performs the best for the scale free data. As for the data simulated to have some scale effects, Barth and Clinch (2009) report the following results: 1) the unscaled model and the model scaled by number of shares outstanding perform the best for the data generated with scale-varying parameters; 2) the model scaled by book value performs the best for the data generated to have heteroskedastic errors (with heteroskedasticity dependent on book value); 3) the model scaled by number of shares outstanding and the model scaled by lagged price perform the best for the data generated to have all five scale effects; 4) the unscaled model and the model scaled by number of shares outstanding perform the best for the data generated to have multiplicative scale effect, for the data generated to have additive scale effect, and for the data generated to have survivor bias.

3 Analytic Results

Following the literature we test the relation between y and x using OLS regression. The unscaled version of the regression is

$$y_i = \beta_0 + \beta_1 x_i + u_i. \tag{1}$$

The null hypothesis in this regression is $H_0 : \beta_1 = 0$. Our theoretical model of y implements this null hypothesis so that:

$$y_i = \beta_0 + \epsilon_i \tag{2}$$

where y has a non-zero mean and is independent of x. The theoretical model includes a non-zero mean because most of variables of interest in accounting models have means that are different from zero, thus, an appropriate null hypothesis of no relation between y and x should take into account the non-zero mean of y.

In this section we investigate analytically whether various forms of scaling may result in biased parameter estimates or inappropriate standard errors. Section 3.1 and 3.2 show consequences of estimating equation (2) with alternative scaling techniques. The scaled OLS regression in Section 3.1 is a version of Weighted Least Squares (WLS) because it scales all variables including the constant term. The scaled OLS regression in Section 3.2 deflates y and x by z, but does not deflate the constant term. Detailed results for each of these models follows.

3.1 Scaling all variables including the constant

First, we show that scaling all variables, including the constant term, in an OLS regression produces an unbiased estimate of the slope (i.e., $E[\hat{\beta}_1] = 0$). In other words, proper scaling does not introduce dependence where none exists. This result is unsurprising given that this is the Weighted Least Squares (WLS) estimator.

The scaled model is

$$\frac{y_i}{z_i} = \beta_0 \frac{1}{z_i} + \beta_1 \frac{x_i}{z_i} + u_i \tag{3}$$

By definition $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ where according to our regression equation (3)

$$X = \begin{bmatrix} \frac{1}{z_1} & \frac{x_1}{z_1} \\ \vdots & \vdots \\ \frac{1}{z_N} & \frac{x_N}{z_N} \end{bmatrix}, y = \begin{bmatrix} \frac{y_1}{z_1} \\ \vdots \\ \frac{y_N}{z_N} \end{bmatrix}$$

Using the above definitions of X and y we can re-write the general formula for coefficient estimates as

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} \frac{1}{z_1} & \cdots & \frac{1}{z_N} \\ \frac{x_1}{z_1} & \cdots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} \frac{1}{z_1} & \frac{x_1}{z_1} \\ \vdots & \vdots \\ \frac{1}{z_N} & \frac{x_N}{z_N} \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{1}{z_1} & \cdots & \frac{1}{z_N} \\ \frac{x_1}{z_1} & \cdots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} \frac{y_1}{z_1} \\ \vdots \\ \frac{y_N}{z_N} \end{bmatrix}$$

Multiplying the first and second matrices together and then multiplying the third and fourth matrices together yields

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} \sum \frac{1}{z_i^2} & \sum \frac{x_i}{z_i^2} \\ \sum \frac{x_i}{z_i^2} & \sum \frac{x_i^2}{z_i^2} \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum \frac{y_i}{z_i^2} \\ \sum \frac{x_i y_i}{z_i^2} \end{bmatrix}$$

where implementation of the inverse results in

$$\hat{\beta}_{OLS} = \frac{1}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2} \right)^2} \begin{bmatrix} \sum \frac{x_i^2}{z_i^2} & -\sum \frac{x_i}{z_i^2} \\ -\sum \frac{x_i}{z_i^2} & \sum \frac{1}{z_i^2} \end{bmatrix} \begin{bmatrix} \sum \frac{y_i}{z_i^2} \\ \sum \frac{x_i y_i}{z_i^2} \end{bmatrix}$$

It follows then that

$$\hat{\beta}_1 = \frac{\sum \frac{1}{z_i^2} \sum \frac{x_i y_i}{z_i^2} - \sum \frac{x_i}{z_i^2} \sum \frac{y_i}{z_i^2}}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2} \right)^2}$$

By substituting y_i above with our theoretical model, equation (1) we get

$$\hat{\beta}_1 = \frac{\sum \frac{1}{z_i^2} \sum \frac{x_i(\beta_0+\epsilon_i)}{z_i^2} - \sum \frac{x_i}{z_i^2} \sum \frac{(\beta_0+\epsilon_i)}{z_i^2}}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2} \tag{4}$$

Now, we can take the expectation of $\hat{\beta}_1$ taking into account that x_i and z_i are given and $E[\epsilon_i] = 0$ via the usual OLS assumptions.

$$E[\hat{\beta}_1] = \beta_0 \frac{\sum \frac{1}{z_i^2} \sum \frac{x_i}{z_i^2} - \sum \frac{1}{z_i^2} \sum \frac{x_i}{z_i^2}}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2} = 0$$

The expected value of $\hat{\beta}_1$ is equal to β_0 times a term whose numerator includes a difference of the same two products. β_0 is therefore being multiplied by zero, and this causes the expected value of $\hat{\beta}_1$ to also be zero.

In sum, when x and y are independent, an OLS regression in which the constant and all other variables are scaled correctly estimates that a one-unit change in x does not result in any change in y. Scaling in this manner is equivalent to the Weighted Least Squares (WLS) estimator, and no bias is induced by the researcher via scaling.

Unbiased parameter estimates are a necessary for valid hypothesis tests, but calculating standard errors are equally important. This is because the test statistic is calculated as the ratio of the parameter estimate to its standard error. Using equation (4) we can find the variance of $\hat{\beta}_1$ and determine whether the usual OLS assumptions are appropriate:

$$Var(\hat{\beta}_1) = Var\left(\frac{\sum \frac{1}{z_i^2} \sum \frac{x_i(\beta_0+\epsilon_i)}{z_i^2} - \sum \frac{x_i}{z_i^2} \sum \frac{(\beta_0+\epsilon_i)}{z_i^2}}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2}\right)$$

Taking into account that x_i and z_i are given and that variance of a constant is zero:

$$Var(\hat{\beta}_1) = \frac{Var\left(\sum \frac{1}{z_i^2} \sum \frac{x_i}{z_i^2}\epsilon_i - \sum \frac{x_i}{z_i^2} \sum \frac{1}{z_i^2}\epsilon_i\right)}{\left(\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2\right)^2}$$

Opening up the parenthesis in the numerator:

$$Var(\hat{\beta}_1) = \frac{\left(\sum \frac{1}{z_i^2}\right)^2 \sum \frac{x_i^2}{z_i^4} Var(\epsilon_i) + \left(\sum \frac{x_i}{z_i^2}\right)^2 \sum \frac{1}{z_i^4} Var(\epsilon_i) - 2 \sum \frac{1}{z_i^2} \sum \frac{x_i}{z_i^2} \sum \frac{x_i}{z_i^4} Var(\epsilon_i)}{\left(\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2\right)^2}$$

Taking into account that $Var(\epsilon_i) = \sigma^2$ we get:

$$Var(\hat{\beta}_1) = \sigma^2 \frac{\left(\sum \frac{1}{z_i^2}\right)^2 \sum \frac{x_i^2}{z_i^4} + \left(\sum \frac{x_i}{z_i^2}\right)^2 \sum \frac{1}{z_i^4} - 2 \sum \frac{1}{z_i^2} \sum \frac{x_i}{z_i^2} \sum \frac{x_i}{z_i^4}}{\left(\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2\right)^2}$$

which is not equal to the second diagonal element of OLS variance-covariance matrix

$$\sigma^2 (X'X)^{-1} = \sigma^2 \frac{1}{\sum \frac{1}{z_i^2} \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i^2}\right)^2} \begin{bmatrix} \sum \frac{x_i^2}{z_i^2} & -\sum \frac{x_i}{z_i^2} \\ -\sum \frac{x_i}{z_i^2} & \sum \frac{1}{z_i^2} \end{bmatrix}$$

Thus, using OLS standard errors for the hypothesis test $\hat{\beta}_1 = 0$ will be incorrect. Instead one needs to use the robust standard error formula.

3.2 Scaling all variables excluding the constant

Now we show that scaling all variables, with the exception of the constant, results in a biased estimate of the slope (i.e., $E[\hat{\beta}_1] \neq 0$. In other words, improper scaling introduces dependence where none exists.

Our theoretical model remains equation (2) but now we estimate the following OLS regression

$$\frac{y_i}{z_i} = \beta_0 + \beta_1 \frac{x_i}{z_i} + u_i \tag{5}$$

Note that the only difference between equations (3) and (5) is that the constant is not scaled in (5).

Again, we use the definition $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ and our regression equation (5) where

$$X = \begin{bmatrix} 1 & \frac{x_1}{z_1} \\ \vdots & \vdots \\ 1 & \frac{x_N}{z_N} \end{bmatrix}, y = \begin{bmatrix} \frac{y_1}{z_1} \\ \vdots \\ \frac{y_N}{z_N} \end{bmatrix},$$

The X matrix above now contains a column of 1's rather than $\frac{1}{z_i}$. Using the above definitions of X and y we re-write the general formula for coefficient estimates as

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} 1 & \dots & 1 \\ \frac{x_1}{z_1} & \dots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} 1 & \frac{x_1}{z_1} \\ \vdots & \vdots \\ 1 & \frac{x_N}{z_N} \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \frac{x_1}{z_1} & \dots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} \frac{y_1}{z_1} \\ \vdots \\ \frac{y_N}{z_N} \end{bmatrix}$$

Multiplying the first and second matrices and then multiplying the third and fourth matrices yields

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} N & \sum \frac{x_i}{z_i} \\ \sum \frac{x_i}{z_i} & \sum \frac{x_i^2}{z_i^2} \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum \frac{y_i}{z_i} \\ \sum \frac{x_i y_i}{z_i^2} \end{bmatrix}$$

where implementation of the inverse results in

$$\hat{\beta}_{OLS} = \frac{1}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2} \begin{bmatrix} \sum \frac{x_i^2}{z_i^2} & -\sum \frac{x_i}{z_i} \\ -\sum \frac{x_i}{z_i} & N \end{bmatrix} \begin{bmatrix} \sum \frac{y_i}{z_i} \\ \sum \frac{x_i y_i}{z_i^2} \end{bmatrix}$$

It follows then that

$$\hat{\beta}_1 = \frac{N \sum \frac{x_i y_i}{z_i^2} - \sum \frac{x_i}{z_i} \sum \frac{y_i}{z_i}}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2}$$

By substituting y_i above with our theoretical model, equation (2) we get

$$\hat{\beta}_1 = \frac{N \sum \frac{x_i(\beta_0 + \epsilon_i)}{z_i^2} - \sum \frac{x_i}{z_i} \sum \frac{(\beta_0 + \epsilon_i)}{z_i}}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2} \tag{6}$$

Now we take expectation of $\hat{\beta}_1$ taking into account that x_i and z_i are given and $E[\epsilon_i] = 0$ (the usual OLS assumptions)

$$E[\hat{\beta}_1] = \beta_0 \frac{N \sum \frac{x_i}{z_i^2} - \sum \frac{x_i}{z_i} \sum \frac{1}{z_i}}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2} \neq 0$$

Unlike our result in the previous sub-section, the expected value of $\hat{\beta}_1$ is no longer equal to zero even though y and x are independent. The bias in the parameter estimate is a function of β_0, x, N, and z. This result is not conditional on any particular correlation structure between x and z, and assuming positive or negative correlations between x and z will not impact the analytical derivation of the bias.

In sum, when y and x are independent and y has a nonzero mean, an OLS regression of y on x will result in a biased estimate of $\hat{\beta}_1$ if the regression constant is not scaled. As shown in the previous section, no such bias is induced when the constant term is also scaled.

Similar to Section 3.1, we examine the impact that failing to scale the constant has on the regression standard errors. Using equation (6) we can find the variance of $\hat{\beta}_1$:

$$Var(\hat{\beta}_1) = Var\left(\frac{N\sum\frac{x_i(\beta_0+\epsilon_i)}{z_i^2} - \sum\frac{x_i}{z_i}\sum\frac{(\beta_0+\epsilon_i)}{z_i}}{N\sum\frac{x_i^2}{z_i^2} - \left(\sum\frac{x_i}{z_i}\right)^2}\right)$$

Taking into account that x_i and z_i are given and that variance of a constant is zero:

$$Var(\hat{\beta}_1) = \frac{Var\left(N\sum\frac{x_i}{z_i^2}\epsilon_i - \sum\frac{x_i}{z_i}\sum\frac{1}{z_i}\epsilon_i\right)}{\left(N\sum\frac{x_i^2}{z_i^2} - \left(\sum\frac{x_i}{z_i}\right)^2\right)^2}$$

Opening up the parenthesis in the numerator:

$$Var(\hat{\beta}_1) = \frac{N^2\sum\frac{x_i^2}{z_i^4}Var(\epsilon_i) + \left(\sum\frac{x_i}{z_i}\right)^2\sum\frac{1}{z_i^2}Var(\epsilon_i) - 2N\sum\frac{x_i}{z_i}\sum\frac{x_i}{z_i^3}Var(\epsilon_i)}{\left(N\sum\frac{x_i^2}{z_i^2} - \left(\sum\frac{x_i}{z_i}\right)^2\right)^2}$$

Taking into account that $Var(\epsilon_i) = \sigma^2$ we get:

$$Var(\hat{\beta}_1) = \sigma^2\frac{N^2\sum\frac{x_i^2}{z_i^4} + \left(\sum\frac{x_i}{z_i}\right)^2\sum\frac{1}{z_i^2} - 2N\sum\frac{x_i}{z_i}\sum\frac{x_i}{z_i^3}}{\left(N\sum\frac{x_i^2}{z_i^2} - \left(\sum\frac{x_i}{z_i}\right)^2\right)^2}$$

which is not equal to the second diagonal element of OLS variance-covariance matrix

$$\sigma^2(X'X)^{-1} = \sigma^2\frac{1}{\sum N\sum\frac{x_i^2}{z_i^2} - \left(\sum\frac{x_i}{z_i}\right)^2}\begin{bmatrix}\sum\frac{x_i^2}{z_i^2} & -\sum\frac{x_i}{z_i}\\ -\sum\frac{x_i}{z_i} & N\end{bmatrix}$$

Thus, as in previous subsection, using OLS standard errors for the hypothesis test $\hat{\beta}_1 = 0$ is incorrect. Again, one needs to use the robust standard error formula.

3.3 Scaling only independent variables

If a researcher scales only independent variables an OLS regression produces an unbiased estimate of the slope and standard error. In other words, with this type of scaling using the usual t-statics is appropriate.

As before, our theoretical model remains equation (2) but now we estimate the following OLS regression

$$y_i = \beta_0 + \beta_1 \frac{x_i}{z_i} + u_i \tag{7}$$

By definition $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ where according to our regression equation (7)

$$X = \begin{bmatrix} 1 & \frac{x_1}{z_1} \\ \vdots & \vdots \\ 1 & \frac{x_N}{z_N} \end{bmatrix}, y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Using the above definitions of X and y we can re-write the general formula for coefficient estimates as

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} 1 & \dots & 1 \\ \frac{x_1}{z_1} & \dots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} 1 & \frac{x_1}{z_1} \\ \vdots & \vdots \\ 1 & \frac{x_N}{z_N} \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \frac{x_1}{z_1} & \dots & \frac{x_N}{z_N} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

Multiplying the first and second matrices together and then multiplying the third and fourth matrices together yields

$$\hat{\beta}_{OLS} = \left(\begin{bmatrix} N & \sum \frac{x_i}{z_i} \\ \sum \frac{x_i}{z_i} & \sum \frac{x_i^2}{z_i^2} \end{bmatrix} \right)^{-1} \begin{bmatrix} \sum y_i \\ \sum \frac{x_i y_i}{z_i} \end{bmatrix}$$

where implementation of the inverse results in

$$\hat{\beta}_{OLS} = \frac{1}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2} \begin{bmatrix} \sum \frac{x_i^2}{z_i^2} & -\sum \frac{x_i}{z_i} \\ -\sum \frac{x_i}{z_i} & N \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum \frac{x_i y_i}{z_i} \end{bmatrix}$$

It follows then that

$$\hat{\beta}_1 = \frac{N \sum \frac{x_i y_i}{z_i} - \sum \frac{x_i}{z_i} \sum y_i}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i} \right)^2}$$

By substituting y_i above with our theoretical model, equation (2) we get

$$\hat{\beta}_1 = \frac{N \sum \frac{x_i(\beta_0+\epsilon_i)}{z_i} - \sum \frac{x_i}{z_i} \sum (\beta_0 + \epsilon_i)}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2} \tag{8}$$

Taking the expectation of $\hat{\beta}_1$

$$E\hat{\beta}_1 = \beta_0 \frac{N \sum \frac{x_i}{z_i} - N \sum \frac{x_i}{z_i}}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2} = 0$$

The expected value of $\hat{\beta}_1$ is equal to zero because the numerator of the product equals zero. Thus, scaling of only an independent variable results in an unbiased estimate of the slope.

Using equation (8) we can find the variance of $\hat{\beta}_1$:

$$Var(\hat{\beta}_1) = Var \left(\frac{N \sum \frac{x_i(\beta_0+\epsilon_i)}{z_i} - \sum \frac{x_i}{z_i} \sum (\beta_0 + \epsilon_i)}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2} \right)$$

Taking into account that x_i and z_i are given and that variance of a constant is zero:

$$Var(\hat{\beta}_1) = \frac{Var\left(N \sum \frac{x_i}{z_i} \epsilon_i - \sum \frac{x_i}{z_i} \sum \epsilon_i\right)}{\left(N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2\right)^2}$$

Opening up the parenthesis in the numerator:

$$Var(\hat{\beta}_1) = \frac{N^2 \sum \frac{x_i^2}{z_i^2} Var(\epsilon_i) + \left(\sum \frac{x_i}{z_i}\right)^2 \sum Var(\epsilon_i) - 2N \sum \frac{x_i}{z_i} \sum \frac{x_i}{z_i} Var(\epsilon_i)}{\left(N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2\right)^2}$$

Taking into account that $Var(\epsilon_i) = \sigma^2$ and simplifying we get:

$$Var(\hat{\beta}_1) = \sigma^2 \frac{N}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2}$$

which is equal to the second diagonal element of OLS variance-covariance matrix

$$\sigma^2 (X'X)^{-1} = \sigma^2 \frac{1}{N \sum \frac{x_i^2}{z_i^2} - \left(\sum \frac{x_i}{z_i}\right)^2} \begin{bmatrix} \sum \frac{x_i^2}{z_i^2} & -\sum \frac{x_i}{z_i} \\ -\sum \frac{x_i}{z_i} & N \end{bmatrix}$$

Thus, unlike in the previous two subsections, when we scale only the independent variable, OLS standard errors are appropriate to compute t-statistics.

3.4 Using the dependent variable as a deflator

Using the dependent variable as a scalar, as suggested by Easton and Sommers (2003), makes regression results impossible to interpret. Dividing the dependent variable by itself creates a vector of ones as the dependent variable. In other words, the dependent variable in the scaled regression is nonrandom, while a starting assumption of regression analysis is that dependent variable is random.

For instance, Casella and Berger(2001) state: "In particular, in simple linear regression we have a relationship of the form $Y_i = a + bx_i + e_i$, where Y_i is a random variable..." Later they continue: "The values $y_1, ..., y_n$ are observed values of uncorrelated random variables $Y_1, ..., Y_n$." Finally, they state: "The values $y_1, ..., y_n$ are observed values of random variables, $Y_1, ...Y_n$." Hogg and Craig (1995) state: "There is often interest in the relation between two variables, for example, a student's scholastic aptitude test sore in mathematics and this same student's grade in calculus. Frequently, one of these variables, say x, is known in advance of the other, and hence there is interest in predicting a future random variable Y." They continue: "To estimate $E(Y) = \mu(x)$, or equivalently the parameters α, β, and γ, we observe the random variable Y for each of n possibly different values of x, say $x1, ...xn$, which are not all equal. Once the n independent experiments have been performed, we have n pairs of known numbers $(x1, y1), ...(xn, yn)$. These pairs are then used to estimate the mean $E(Y)$. Problems like this are often classified under regression because $E(Y) = \mu_x$ is frequently called a regression curve." Finally they say: "... we assume that, $Y_1, Y_2, ...Y_n$ are independent normal variables ... Their joint p.d.f is therefore the product of the individual probability density functions; that is, the likelihood function..."

Thus, given that the underlying assumption of regression analysis is violated when dependent variable is nonrandom, regression results can not be used for statistical inference when the dependent variable is a vector of 1's.

4 Data

For the Monte Carlo simulations we need to specify the distribution, mean, and standard deviation of each variable. We use the Compustat Fundamentals Annual file to obtain financial statement variables (e.g., MKVALT, CEQ, AT, SALE, AT, CSHO, EPSPX, and PRCC_F). All firms with DATADATEs between January 1980 and December 2017 make up the initial sample, and this results in 396,796 unique firm-year observations. A reconciliation of beginning and ending firm-year observations is included in Table 1. Values less than or equal to zero for MKVALT, CEQ, AT, and SALE are dropped. MKVALT is missing in approximately 119,000 observations and is therefore interpolated based on PRCC_F and CSHO in instances where MKVALT is missing in the data. MKVALT remains missing in 72,749 cases even after this adjustment. Additionally, missing values and the top and bottom 1% of EPSPX, MKVALT, CEQ, and NI are dropped. The above procedures yield 230,536 remaining firm-year observations.

Table 1: Reconciliation Sample of Financial Statement Variables

396,796	Firm-years with DATADATEs from January 1980 to December 2017
(50,012)	Less: Values of MKVALT, CEQ, AT, or SALE less than or equal to 0
(5,413)	Less: top and bottom 1% of EPSPX
(110,835)	Less: top and bottom 1% of MKVALT, CEQ, NI
230,536	Remaining firm-years

Equity returns, share prices, changes in share prices, and dividends are obtained from the CRSP Monthly Stock File. All firms with monthly returns between January 1980 and December 2016 result in 3,232,299 firm-month observations. Observations with negative dividends, share codes not equal to 10 or 11, and duplicates based on PERMNO and DATE are dropped. Table 2 contains a detailed reconciliation. Further, we drop the top 1% of dividends as the maximum value of dividends in the unadjusted data is $1,300. We do not drop the bottom 1% of dividends because values of zero for dividends are reasonable given that not all firms pay dividends. Share prices greater than \$500 are dropped, and observations with changes in share price between months in the top or bottom 1% are also dropped. The final sample includes 2,393,067 firm-month observations.

Table 2: Reconciliation Sample of CRSP Monthly Stock Variables

3,232,299	CRSP firm-months with DATEs between January 1980 and December 2016
(2)	Less: negative dividends
(781,260)	Less: share codes other than 10 and 11
(17,307)	Less: duplicate PERMNO and DATE observations
(3,027)	Less: top 1% of DIVAMT
(333)	Less: PRC greater than \$500
(37,303)	Less: top and bottom 1% of the difference in PRC between periods
2,393,067	Remaining firm-months

Table 3 reports the first and second moments of data as well as the functional form and parameters of the distributions of each variable. Based on histograms, PDIFF, EPSPX, and D.EPSPX appear to be normally distributed. All of the other variables appear to be log-normally distributed. The second and third columns in the table report the mean and standard deviation of each variable within the population of firms covered by Compustat or CRSP after retaining firm-years in the manner described above. The fourth column reports the functional form of the distribution for each variable. Finally, the last two columns contain the values of the parameters. For normally distributed variables we set μ equal to the mean and σ equal to the standard deviation in the data. For the log-normally distributed variables we solved the two equations $mean = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ and $s.d. = \sqrt{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)}$ for μ and σ.

5 Monte Carlo Simulations Results

Market-based accounting research estimates the relationship between market value and various financial statement measures. Researchers often use either a returns or levels

Table 3: Distributions

Variable	Data Summary Statistics		Functional form and parameters		
	mean	s.d.		μ	σ
MKVALT	2767	15038	Log-normal	6.22	1.85
CEQ	1068	6023	Log-normal	5.29	1.87
NI	117	1002	Log-normal	2.61	2.08
AT	6345	62693	Log-normal	6.46	2.14
SALE	1953	10247	Log-normal	5.9	1.83
PRCC_F	19.25	23.23	Log-normal	2.51	0.95
L.PRCC_F	19.11	23.05	Log-normal	2.5	0.95
DIVAMT	0.43	1.91	Log-normal	1.74	-2.36
PDIFF	0.11	2.23	Normal	0.11	2.23
EPSPX	0.73	1.68	Normal	0.73	1.68
CSHO	156.33	45998.55	Log-normal	-0.63	3.37
D.EPSPX	-0.01	1.05	Normal	-0.01	1.05

specification for the empirical models. The common estimated model for returns (e.g., equation 18 in Barth and Clinch (2009)) is

$$\frac{\triangle PRCC\_F_t + DIVAMT_t}{PRCC\_F_{t-1}} = \beta_0 + \beta_1 \frac{EPSPX_t}{PRCC\_F_{t-1}} + \beta_2 \frac{\triangle EPSPX_t}{PRCC\_F_{t-1}} + \epsilon_t,$$

where $PRCC\_F$ is share price, $DIVAMT$ is dividend, and $EPSPX$ is reported earnings per share. The variables in the model are deflated by $PRCC\_F_{t-1}$. The common estimated model for levels (e.g., equation 1 in Easton and Sommers (2003)) is

$$MKVALT_t = \beta_0 + \beta_1 CEQ_t + \beta_2 NI_t + \epsilon_t,$$

where $MKVALT$ is market value of equity, CEQ is book value of equity, and NI is net income. According to Barth and Kallapur (1996), common deflators for this model are total assets, sales, book value of equity, net income, number of shares outstanding, and share price, so that the estimated model becomes:

$$\frac{MKVALT_t}{D_t} = \beta_0 + \beta_1 \frac{CEQ_t}{D_t} + \beta_2 \frac{NI_t}{D_t} + \epsilon_t,$$

where D_t is a deflator. Finally, Easton and Sommers (2003) also recommend using market value of equity as a deflator in which case the estimated model becomes:

$$1 = \beta_0 \frac{1}{MKVALT_t} + \beta_1 \frac{CEQ_t}{MKVALT_t} + \beta_2 \frac{NI_t}{MKVALT_t} + \epsilon_t.$$

5.1 No scaling

To set the baseline, we begin our Monte Carlo simulations with models that have no scaling. As expected, when y and x are independent and we regress y on x, rejection rates for the regression slope are approximately 5% for OLS standard errors.

Our Monte Carlo simulation for the returns model proceeded as follows: In Step 1 we generated 10,000 observations of returns as a sum of changes in share price and dividends both divided by share price. We generated share price as a log-normal distribution with $\mu = 2.51$ and $\sigma = 0.95$; changes in share price as a normal distribution with $\mu = 0.11$ and $\sigma = 2.23$; and dividends as a log-normal distribution with $\mu = -2.36$ and $\sigma = 1.74$. We also generated 10,000 observations of earnings per share as a normal distribution with $\mu = 0.73$ and $\sigma = 1.68$. Finally, we generated 10,000 observations of the difference in earnings per share as a normal distribution with $\mu = -0.01$ and $\sigma = 1.05$. In Step 2 we regressed returns on earnings and difference in earnings:

$$\frac{\triangle PRCC\_F_t + DIVAMT_t}{P_{t-1}} = \beta_0 + \beta_1 EPSPX_t + \beta_2 \triangle EPSPX_t + \epsilon_t$$

In Step 3 we calculated t-statistics and compared them to the 5% critical value. In Step 4 we repeated steps 1, 2, and 3 10,000 times. In Step 5 we calculated rejection rates by dividing the number of times t-statistics were higher in absolute value than the critical value by 10,000. Our rejections rates, the first line in Table 4, were 5% for β_1 and 4.77% for β_2 using OLS standard errors, which is quite close to the theoretical 5%. Similar rejection rates are observed for robust standard errors.

Our Monte Carlo simulation for the levels model proceeded as follows: In Step 1 we generated 10,000 observations of market value of equity as a log-normal distribution with $\mu = 6.22$ and $\sigma = 1.85$; 10,000 observations of book value of equity as log-normal

Table 4: Returns Model Rejection Rates

	OLS standard errors		Robust standard errors	
Deflator	β_1	β_2	β_1	β_2
none	5.00	4.77	4.56	4.30
	A. Scaling excludes the constant			
$PRCC\_F_{t-1}$	98.93	63.94	85.27	5.35
	B. Scaling includes the constant			
$PRCC\_F_{t-1}$	57.71	57.78	5.37	5.44

Notes: $PRCC\_F_t$ is price of shares.

distribution with $\mu = 5.23$ and $\sigma = 1.87$; and 10,000 observation of net income as a log-normal distribution with $\mu = 2.61$ and $\sigma = 2.08$. In Step 2 we regressed market value of equity on book value and net income:

$$MKVALT_t = \beta_0 + \beta_1 CEQ_t + \beta_2 NI_t + \epsilon_t,$$

In Step 3 we calculated t-statistics and compared them to the 5% critical value. In Step 4 we repeated steps 1, 2, and 3 10,000 times. In Step 5 we calculated rejection rates by dividing the number of times the t-statistics were higher in absolute value than the critical value by 10,000. Our rejections rates, the first line in Table 5, were 3.03% for β_1 and 2.84% for β_2 using OLS standard errors, again, quite close to the theoretical 5%. However, the robust standard error rejection rates are quite a bit larger (i.e., 14.56% and 17.04%). These rejection rates are caused by the robust standard errors being smaller than the OLS standard errors which is theoretically possible but not typically expected by researchers. Robust standard errors may be smaller than OLS standard errors when 1) smaller residuals are observed for values of regressors that are far from the regressors' means and 2) larger residuals are observed for values of regressors near the regressors' means. This is the exact residual pattern detected in the randomly generated data used in our Monte Carlo simulations for the levels models. MKVALT, CEQ, and NI are all lognormally distributed, and this results in distributions with relatively long tails.

Standard regression post estimation procedures (e.g., plotting residuals versus predicted values and plotting residuals against regressors) reveals the presence of nonlinearity in the residuals of the unscaled levels model using robust standard errors. For example, the graph of the residuals versus predicted values closely resembles a graph of 1/x, where smaller residuals are associated with larger predicted values. Patterns such as this in the residuals are often interpreted by researchers as evidence of functional for misspecification. Accordingly, we repeated the Monte Carlo procedures above after estimating regression models using the logs of MKVALT, CEQ, and NI for the unscaled levels model with robust standard errors. The rejection rates observed after specifying the model using logged variables (untabulated) are 4.82% and 4.73% which round to the theoretically expected 5%.

To conclude, our baseline simulations indicate that if we generate independent y and x from distributions that are close to either the distributions of returns or the distribution of levels, our rejection rates are close to the theoretical 5% using OLS standard errors.

5.2 Scaling all variables including the constant

Our simulations suggest that when y, x, and z are independent and we regress $\frac{y}{z}$ on $\frac{1}{z}$ and $\frac{x}{z}$, rejection rates are far above expected 5% when OLS standard errors are used, while using robust standard errors substantially improves rejections rates.

For the returns model, we modified Step 2 of the unscaled simulation by regressing returns on a vector of ones divided by share price, earnings scaled by share price, and the difference in earnings scaled by share price:

$$\frac{\triangle PRCC\_F_t + DIVAMT_t}{PRCC\_F_{t-1}} = \beta_0 \frac{1}{PRCC\_F_{t-1}} + \beta_1 \frac{EPSPX_t}{PRCC\_F_{t-1}} + \beta_2 \frac{\triangle EPSPX_t}{PRCC\_F_{t-1}} + \epsilon_t.$$

Rejections rates, the third line in Table 4, were 57.71% for β_1 and 57.78% for β_2 using OLS standard errors. Using robust standard errors, rejection rates were 5.37% for β_1 and 5.44% for β_2, almost at the theoretical 5% rate.

For the levels models, we modified Step 2 of the unscaled simulation by scaling market

Table 5: Levels Model Rejection Rates

	OLS standard errors		Robust standard errors	
Deflator	β_1	β_2	β_1	β_2
none	3.03	2.84	14.56	17.04
	A. Scaling excludes the constant			
S_t	91.96	83.82	45.63	38.82
TA_t	92.85	86.22	41.61	35.81
CEQ_t	98.28	90.31	98.24	45.46
NI_t	95.68	95.91	46.88	95.38
$CSHO_t$	4.16	4.25	32.03	32.86
$PRCC\_F_t$	2.14	1.58	14.43	15.99
	B. Scaling includes the constant			
S_t	66.10	60.87	11.41	11.40
TA_t	73.50	69.37	13.28	12.83
CEQ_t	65.93	63.33	44.66	10.86
NI_t	72.99	65.59	12.57	48.91
$MKVALT_t$	65.49	64.03	30.55	33.21

Notes: S_t is sales, TA_t is total assets, CEQ_t is book value of equity, NI_t is net income, $CSHO_t$ is number of shares outstanding, $PRCC\_F_t$ is price of a share, $MKVALT_t$ is market value of equity.

value of equity, book value, net income, and a vector of ones by a deflator:

$$\frac{MKVALT_t}{D_t} = \beta_0 \frac{1}{D_t} + \beta_1 \frac{CEQ_t}{D_t} + \beta_2 \frac{NI_t}{D_t} + \epsilon_t.$$

As in the previous subsection, for deflators we used sales, total assets, book value, or net income. Part B of Table 5 summarizes our results. With sales as our deflator, our rejections rates were 66.1% for β_1 and 60.87% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 11.41% for β_1 and 11.4% for β_2. With total assets as a deflator, our rejections rates were 73.5% for β_1 and 69.37% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 13.28% for β_1

and 12.83% for β_2. With book value as a deflator, our rejections rates were 65.93% for β_1 and 63.33% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 44.66% for β_1 and 10.86% for β_2. Finally, with net income as a deflator, our rejections rates were 72.99% for β_1 and 65.59% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 12.57% for β_1 and 48.91% for β_2. Note that rejection rates with robust standard errors are substantially higher when the deflator is one of the independent variables, as is the case when either book value or net income are deflators. Scaling by one of the independent variables in the regression leads to an unscaled intercept term being included in equation above. It is this intercept term that is still rejected much more than expected even after using robust standard errors, and this indicates only that the regression line should not run through the origin. This is expected based on the means of the data used in the simulations.

In sum, when we generate independent y, x, and z and regress $\frac{y}{z}$ on $\frac{1}{z}$ and $\frac{x}{z}$, scaling the variables results in rejection rates between 61% and 73% when we use OLS standard errors. Using robust standard errors lowers rejection rates to between the 5% and 13% range, except for the cases when the deflator equals one of the independent variables. Our explanation for the slightly higher than expected rejection rates for the WLS model using robust standard errors is that the theoretical model has homoskedastic error terms. Estimating a WLS model on a homoskedastic process induces at least some heteroskedasticity, and it is known that robust standard error estimates may be biased downward in the presence of mild heteroskedasticity. Angrist and Pischke (2009) write, "Chesher and Jewitt (1987) show that as long as there is not "too much" heteroskedasticity, robust standard errors based on $\hat{\omega}$ are indeed biased downward." It is likely that some of the Monte Carlo repetitions resulted in mild heteroskedasticity after applying WLS to a homoskedastic process, and this resulted in the higher rejection rates. Thus, deflating all variables including the constant (WLS) together with using robust standard errors substantially improves rejection rates compared to not deflating the constant. However, using Weighted Least Squares along with robust standard errors still results in rejection rates nearly double the theoretically expected 5%.

5.3 Scaling all variables excluding the constant

Our Monte Carlo simulations suggest that when y, x, and z are independent and we regress $\frac{y}{z}$ on $\frac{x}{z}$, rejection rates are far above the expected 5%. This is expected given that our analytical results in section 3.2 show that scaling in this manner results in biased parameter estimates.

Our Monte Carlo simulation for the returns model was similar to the unscaled model, except we modified Step 2 by regressing returns on earnings scaled by share price and difference in earnings scaled by share price:

$$\frac{\triangle PRCC\_F_t + DIVAMT_t}{PRCC\_F_{t-1}} = \beta_0 + \beta_1 \frac{EPSPX_t}{PRCC\_F_{t-1}} + \beta_2 \frac{\triangle EPSPX_t}{PRCC\_F_{t-1}} + \epsilon_t.$$

Our rejections rates, the second line in Table 4, were 98.93% for β_1 and 63.94% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 85.27% for β_1 and 5.35% for β_2. The rejection rate for β_1 is alarming while the rejection rate for β_2 is surprisingly close to the theoretically expected 5%. The bias in the parameter estimates was previously shown to be a function of β_0, x, N, and z. The values of β_0, N, and z are the same for β_1 and β_2, but the values of x are not. The mean of EPSPX in the generated data is .73 while the mean of D.EPSPX is -.01. The mean of D.EPSPX being so close to zero likely results in a smaller bias for hypothesis tests related to β_2.

Our Monte Carlo simulation for the levels model was similar to the unscaled model, except we modified Step 2 by scaling all variables, market value of equity, book value, and net income, by a deflator:

$$\frac{MKVALT_t}{D_t} = \beta_0 + \beta_1 \frac{CEQ_t}{D_t} + \beta_2 \frac{NI_t}{D_t} + \epsilon_t.$$

For the deflators we used either sales (generated as log-normal with $\mu = 5.9$ and $\sigma = 1.83$), total assets (generated as log-normal with $\mu = 6.46$ and $\sigma = 2.14$), book value (generated as log-normal with $\mu = 5.23$ and $\sigma = 1.87$), or net income (generated as log-normal with $\mu = 2.61$ and $\sigma = 2.08$). Part A of Table 5 summarizes our results. With sales as our deflator, our rejections rates were 91.96% for β_1 and 83.82% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 45.63% for β_1 and

38.82% for β_2. With total assets as a deflator, our rejections rates were 92.85% for β_1 and 86.22% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 41.61% for β_1 and 35.81% for β_2 using OLS standard errors. With book value as a deflator, our rejections rates were 98.28% for β_1 and 90.31% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 98.24% for β_1 and 45.46% for β_2. Finally, with net income as a deflator, our rejections rates were 95.68% for β_1 and 95.91% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 46.88% for β_1 and 95.38% for β_2.

In sum, when we generate independent y, x, and z, regress $\frac{y}{z}$ on an unscaled constant and $\frac{x}{z}$ and use OLS standard errors, scaling the variables results in substantially higher rejection rates, at times reaching over 90%, than the expected 5% even though y and x are independent. Using robust standard errors brings the Type 1 rejection rates down substantially for either β_1 or β_2, but the null hypotheses are still rejected at an alarming rate. Christie (1987) spends much time discussing the use of levels vs. return studies, and our results in Tables 4 and 5 indicate that the returns model is less subject to Type 1 rejections using improper scaling than are the levels models (e.g., β_2 in the returns model is rejected at the 5% rate). These results hold regardless of the particular deflator used in the levels models.

5.4 Scaling only independent variables

For the model in levels, because by definition market value of equity is a product of number shares outstanding and share price,$DIVAMT_t = PRCC\_F_t * CSHO_t$, using either shares outstanding or price as a deflator presents a different case from other deflators. For example, when number of shares outstanding is a deflator the estimated equation becomes:

$$PRCC\_F_t = \beta_0 + \beta_1 \frac{CEQ_t}{CSHO_t} + \beta_2 \frac{NI_t}{CSHO_t} + \epsilon_t.$$

Note that the deflator does not enter dependent variables as was the case in previous subsections. The absence of a common deflator on both sides of the regression equation removes inferential problems in these models even though the constant is not scaled.

Our simulations confirm that when y, x, and z are independent and we regress y on $\frac{x}{z}$ rejections rates are close to the theoretical 5% rate using OLS standard errors.

In particular, we modified Step 2 of the unscaled simulation by scaling market value of equity, book value, and net income, by either number of shares outstanding:

$$PRCC\_F_t = \beta_0 + \beta_1 \frac{CEQ_t}{CSHO_t} + \beta_2 \frac{NI_t}{CSHO_t} + \epsilon_t,$$

or share price:

$$CSHO_t = \beta_0 + \beta_1 \frac{CEQ_t}{PRCC\_F_t} + \beta_2 \frac{NI_t}{PRCC\_F_t} + \epsilon_t.$$

We generated number of shares outstanding as a log-normal distribution with $\mu = -6.3$ and $\sigma = 3.37$. The last two lines of Part A of Table 5 summarize our results. With number of shares outstanding as a deflator, our rejections rates were 4.16% for β_1 and 4.25% for β_2 using OLS standard errors. With share price as a deflator, our rejections rates were 2.14% for β_1 and 1.58% for β_2 using OLS standard errors. The rejection rates using robust standard errors are between 14% and 33%. The larger rejections are again due to the robust standard errors being smaller than the OLS standard errors which is possible when values of the independent variables far from the their means are associated with smaller residuals.

5.5 Scaling by Market Value of Equity

Another deflator that represents a special case is scaling by market value of equity. As advocated by Easton and Sommers (2003), the estimated model becomes:

$$1 = \beta_0 \frac{1}{MKVALT_t} + \beta_1 \frac{CEQ_t}{MKVALT_t} + \beta_2 \frac{NI_t}{MKVALT_t} + \epsilon_t.$$

Unlike all of the previous subsections, using market value of equity as a deflator makes the dependent variable nonstochastic.

Our simulations indicate that when y, x, and z are independent and we regress 1 on $\frac{1}{y}$ and $\frac{x}{y}$ rejection rates are substantially higher than the theoretical 5% rate.

For this simulation, we modified Step 2 of the unscaled simulation by scaling market value of equity, book value, net income, and a vector of ones by market value of equity:

$$1 = \beta_0 \frac{1}{MKVALT_t} + \beta_1 \frac{CEQ_t}{MKVALT_t} + \beta_2 \frac{NI_t}{MKVALT_t} + \epsilon_t.$$

Our rejections rates, the last line of Part B of Table 5, were 65.49% for β_1 and 64.03% for β_2 using OLS standard errors. Rejection rates with robust standard errors were 30.55% for β_1 and 33.21% for β_2.

We conclude that deflating all variables by y substantially increases rejection rates even though y and x are independent. This conclusion holds even when robust standard errors are used.

6 Conclusions and Discussions

We can summarize our results as follows: First, when y, x, and z are independent, the slope of the regression of y/z on x/z and an unscaled constant is biased and OLS standard errors are inappropriate. Both the bias and the inappropriate standard errors increase rejections of the null hypothesis of no relation between y and x from the expected 5% to between 63% and 98%. Using heteroskedasticity-robust standard errors lowers rejection rates to between 5% and 98%, where the 5% is limited to one of the regressors in the returns model. The lower bound of more typical heteroskedasticity-robust standard error rejection rates are around 35%. In other words, scaling the independent variables but failing to also scale the constant produces strong evidence of dependence between y and x even though y and x are independent. Using robust standard errors lowers this range but does not address the problem adequately.

Our results illustrate the following cautionary words from Ohlson (2015), "... the use of a common denominator across variables tends to potentially impart misleading correlations. If X and Y correlate, then X/Z and Y/Z may correlate more even if Z happens to be random." In fact, our results take this one step further by showing that even if X and Y are independent, a common form of scaling by random Z results in spurious inferences.

Second, when y, x, and z are independent, the slope of the regression of y/z on x/z and $1/z$ (Weighted Least Squares) is not biased, but OLS standard errors are still inappropriate. The absence of bias in WLS estimates decrease rejection rates to between 63% and 73% when using traditional OLS standard errors. Using robust standard errors lowers rejections rates further to between 5% and 48%. Thus, WLS with traditional standard errors still results in substantial evidence of dependence between y and x, while using robust standard errors ameliorates the spurious dependence in some cases almost completely.

Our results strongly suggest that choosing to scale both the independent and dependent variables in regressions with the same deflator comes with a high risk of inducing spurious dependence between the variables of interest. Thus, we would recommend practitioners not to scale variables. Scaling motivated by heteroskedasticity concerns is better treated without scaling and using robust standard errors. Scaling motivated by concerns of larger firms influencing regression results is better treated with alternative model specifications (e.g., including indicator variables for size and interactions with variables of interest or estimating a model in logs rather than levels). If scaling cannot be avoided, then we recommend using regression model with a scaled constant (WLS) together with heteroskedasticity-robust standard errors for cross-sectional data or clustered standard errors for panel data. The caution here is that even after doing this rejection rates are still twice as frequent as theoretically expected.

Our recommendations are very much in line with recent recommendations to economists by Solon, Haider, and Wooldridge (2015) who begin by observing that "In published research, top-notch empirical scholars make conflicting choices about whether and how to weight and often provide little or no rationale for their choices. And in private discussions, we have found that accomplished researchers sometimes own up to confusion or declare demonstrably faulty reasons for their weighting choices." In particular, as we do, they recommend either not to scale. "In our detailed discussions of each case, we have noted instances in which weighting is not as good an idea as empirical researchers sometimes think. ... In situations in which you might be inclined to weight, it often is useful to report both weighted and unweighted estimates and to discuss what the contrast implies to the interpretation of the results." Also, as we do, they recommend to use robust standard errors "And in many of the situations we have discussed, it is advisable to use robust standard errors."

Finally, we would like to make few comments on the most recent papers in the accounting literature that we have introduced in the literature review section. We would like to start by noting that scaling may have become a popular option to address heteroskedasticity concerns because heteroskedasticity-robust standard errors were not readily available in statistical packages in the 1980's. Most modern statistical software packages now include routines to automatically calculate robust or clustered standard errors. Using these alternative standard errors allows researchers to deal with heteroskedasticity without needing to scale variables.

Christie (1987) recommends using regression with a scaled constant (WLS) in levels studies (equation 3.4 on page 241) and regression with an unscaled constant in returns studies (equation 3.1 on page 238). For levels studies, our results suggest that WLS is better compared to regression with an unscaled constant. For returns studies, Christie's approach induces bias. Christie suggests that X_i should be scaled by S_i because U_i is already scaled by S_i. The model in equation 3.1, however, is equivalent to regression with an unscaled constant since there is no $\frac{1}{S_i}$ in the model. Additionally, for both level and return models we would recommend using robust standard errors.

Easton and Sommers (2003) advocate deflate all variables by the dependent variable. Our results, however, suggest that on one hand it creates a model in which the dependent variable is a vector on ones, which is not random. On the other hand, doing this increases rejection rates from 5% to between 30 and 60% depending on how standard errors are calculated.

Easton and Sommers (2003) also argue that large firms bias OLS estimates by calculating standardized residuals, grouping them into 40 groups by the size of the dependent variable and demonstrating that the average of the 40th group is far above zero. This is presented as evidence that size differences bias OLS estimates. We would like to make several observations regarding this result. First, this result should not be surprising. By construction, the highest values of the dependent variable would have the largest residuals. Barth and Clinch (2009) also observe that "... size differences across firms in and of themselves need not reflect scale effects that lead to incorrect inferences." Second, OLS regression by construction gives average marginal effects. If one desires to address the influence of large firms, including interaction effects into OLS is a better way to do that. Other standard econometric solutions include estimating models using the natural logarithms of variables rather than levels or including quadratic terms. Solon, Haider, and Wooldridge (2015) gives similar advice. Finally, scaling to address the influence of large firms on the estimate of the slope is ineffective. Again, Solon, Haider, and Wooldridge (2015) comment "... we urge practitioners not to fall pray to the fallacy that, in the presence of unmodeled heterogenous effects, weighting to reflect population shares generally identifies the population average partial effect."

Lastly, Barth and Clinch (2009) consider scale effects due to heteroskedasticity (including multiplicative and additive omitted scale factors), survivor bias, and scale varying coefficients. We have already discussed above how heteroskedasticity and scale varying coefficients should be addressed. As for survivor bias, we believe that this is a sample selection problem and cannot be solved by scaling. To address it one should not restrict the sample to firms that survive in all periods.

References

Akbar, S., and W. Stark. 2003. Discussion of scale and the scale effect in market-based accounting research. *Journal of Business Finance & Accounting* 30 (1): 57- 72.

Angrist, P., and J.Pischke. 2009. Mostly harmless econometrics. Princeton University Press.

Barth, M., and S. Kallapur. 1996. The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13 (2): 527-567.

Barth, M., and G. Clinch. 2009. Scale effects in capital markets-based accounting research. *Journal of Business Finance & Accounting* 36 (3) and (4): 253- 288.

Chesher, A., and I. Jewitt. 1987. The bias of heteroskedasticity consistent covariance estimation. *Econometrica* 55: 1217- 1222.

Christie, A. 1987. On cross-sectional analysis in accounting research *Journal of Accounting and Economics* 9: 231-258.

Easton, P., and G. Sommers. 2003. Scale and the scale effect in market-based accounting research. *Journal of Business Finance & Accounting* 30 (1): and (2) 25-55.

Gonedes, N. 1973. Evidence on the information content of accounting numbers: accounting-based and market-based estimates of systematic risk. *Journal of Financial and Quantitative Analysis* 407-443.

Gu, Z. 2005. Scale factor, R squared, and the choice of levels versus returns models. *Journal of Accounting, Auditing & Finance* 20 (1): 71-91.

Lev, B., and S. Sunder. 1979. Methodological issues in the use of financial ratios. *Journal of Accounting and Economics* 1: 187- 210.

Lys, T. and K. Lo. 2000. The Ohlson model: contribution to valuation theory, limitations, and empirical applications. *Journal of Accounting, Auditing & Finance* 15: 337-367.

Lustgarten, S. 1982. The impact of replacement cost disclosure on security prices. *Journal of Accounting and Economics* 4: 121- 141.

Miralles, P., and P. Veira. 2011. Price-levels regressions: Scale-effect or distribution effect? *Spanish Accounting Review* 14 (2): 35-57.

Ohlson, J. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research* 11 (2): 661- 687.

Ohlson, J. 2015. Accounting research and common sense. *Abacus* 51 (4): 525- 534.

Solon, G., S. Haider, and J. Wooldridge. 2015. What are we weighting for? *The Journal of Human Resources* 50 (2): 301- 316.

Wooldridge, J. 2003. Cluster-sample methods in applied econometrics. *The American Economic Review* 93 (2): 133- 138.