

**The effects of scale differences on inferences in accounting research:
coefficient estimates, tests of incremental association,
and relative value relevance[†]**

Kin Lo^{*}

Sauder School of Business, The University of British Columbia, Vancouver, Canada, V6T 1Z2
and
MIT Sloan School of Management, Cambridge, MA, USA, 02142

June 4, 2004

Abstract

Firms' financial data vary considerably with the size of their operations. Such scale differences potentially confound several types of inferences, of which this paper analyzes three. This paper evaluates two potential solutions to these inference problems suggested by theory: (i) deflating the data by a proxy for scale; and (ii) including a scale proxy as an independent variable. First, simulations show that deflating the data more effectively mitigates coefficient bias than including that proxy as an independent variable. Reconciling this result with the opposing conclusion of Barth and Kallapur (1996, *Contemporary Accounting Research*) reveals that the prior results depend on assumptions that are economically and statistically unreasonable. Second, the deflation approach results in more accurate tests of incremental association in terms of mean squared error. Third, deflating by a scale proxy results in well-specified tests of relative association using Vuong's (1989) Z-statistic for non-nested models whereas including the scale proxy as an independent variable results in overstated significance. Given the additional advantages of deflation with respect to heteroscedasticity and the coefficient of determination (R^2) demonstrated in prior studies, researchers should generally deflate their models when scale differences exist in the data.

JEL Classification: C51, G10, G38, M41

Key Words: Econometric models, Capital markets, Financial reporting, Equity valuation

[†] This manuscript has benefited from discussion with S.P. Kothari and Thomas Lys. This project was supported by funding from the CA Education Foundation of BC, The Social Sciences and Humanities Research Council of Canada, and the KPMG Research Bureau at UBC.

^{*} Tel.: +1-847-491-2664; fax +1-847-467-1202

E-mail address: kin.lo@sauder.ubc.ca (Kin Lo).

1. Introduction

Scale is a pervasive notion in accounting research. Simply stated, scale refers to the size of an observation. It is a variable that affects all of the analysis variables (dependent and independent), but scale itself is seldom a variable of interest in the sense that the research question does not concern the marginal impact of scale on the experiment.¹ For example, In capital markets research, large firms have high values for equity market value, equity book value, income, losses, and so on. If one were to analyze the association of equity market value with its book value, the likely positive association would be partly attributable to the fact that large firms have both large market values and book values. In other words, the association is not entirely due to the relation between market values and book values *per se*.

The clearest example where scale confounds economic relationships is provided by the association between market values and losses. We expect a positive relation between earnings and value, so that firms with larger losses should be worth less, *ceteris paribus*. However, with sufficiently large variations in firm size (i.e., all else is *not* equal), we observe that firms with larger losses have larger market values of equity than firms with smaller losses. For example, Figure 1a illustrates the hypothetical positive relation in scale-free data, while Figure 1b shows that the relation turns negative when half of the data is scaled up by a factor of 5. This negative association obtains not because losses add value *per se*, but because larger firms have larger market values and they have sufficient capital to incur large losses. Thus, if research data is affected by scale, such as those illustrated in Figure 1b, then inferences are potentially confounded by scale.

¹ A case when scale is a variable of interest is, of course, analyses related to economies-of-scale. In the natural sciences, the mass of an object is of interest in the study of gravity, but in most other contexts, experiments equalize or otherwise control for mass.

An informal survey of the most current four issues of *The Accounting Review* demonstrates the pervasiveness this scale issue. As summarized in Table 1, at least seven of the forty articles published in the journal in the 12 months up to April 2004 use variables that are affected by scale, including analysts' forecast errors, market value of equity, stock price, earnings, stock option compensation cost, deferred taxes, brand value, and environmental capital expenditures. Also shown in Table 1 are the different ways in which the studies define their variables; some deflate the scale-affected variables and others do not.

The differing treatments of scale in these studies reflect the considerable disagreement that currently exists among researchers over what is the best, acceptable, or appropriate specification of variables for use in capital market research. Some advocate for the inclusion of a scale proxy as an independent variable (Barth and Kallapur, 1996), while others recommend that scale-affected variables should be deflated by a scale proxy (Christie 1987 and Brown, Lo, and Lys, 1999), and yet others argue that variables should be deflated by the dependent variable on the basis that it is the best measure of scale (Easton 1998, Easton and Summers 2003). This range of recommendations arises primarily because the studies focus on mitigating different inference problems such as coefficient bias, R^2 bias, and heteroscedasticity.²

The following analyses will show that, in most circumstances, coefficient bias is more effectively mitigated by deflating by a scale proxy than including that scale proxy as an independent variable. The contrast between this finding and Barth and Kallapur (1996) is due to this paper's use of assumptions that are economically and statistically more reasonable. In particular, Barth and Kallapur's assumptions result in up to 24% of observations attaining negative values of the scale proxy, which can be interpreted as negative prices and market

² Easton and Sommers (2003) consider the issue of influential observations, which is a combination of heteroscedasticity and nonlinearity (i.e., misspecification). The latter is an issue that is not directly related to scale since it is a potential issue in any research context, so the focus here is on the heteroscedasticity component.

values. In contrast, this paper's assumptions ensure that simulated prices and market values are positive.

Additional simulations using these revised assumptions show that tests of incremental association in multivariate settings and tests of relative value relevance using Vuong's (1989) statistic are more accurate using deflated models compared with undeflated models that include a scale proxy as a regressor. Combined with prior findings that R^2 bias and heteroscedasticity are mitigated by deflation, and further that any residual heteroscedasticity in deflated equations can be corrected by White's (1980) adjustment to standard errors, I conclude that deflation by a scale proxy is unambiguously the preferred solution to the three problems arising from cross-sectional scale differences.

The next section reviews the background related to scale issues and the context in which such issues arise. Section 3 replicates the key simulation results from Barth and Kallapur (1996) while Section 4 analyzes those simulations. Using modified assumptions, Section 5 shows that coefficient bias is more effectively mitigated using the deflated model compared with the undeflated model. Section 6 compares the efficacy of deflation in tests of incremental and relative association. Section 7 concludes the paper.

2. Background and development

Since the 1980's, there has been a large number of studies in capital markets research that in whole in part use a "levels approach" for their analyses.³ This approach uses variables measured at the firm level or as per share values so that the magnitudes of the dependent variables depend to a large extent (but not solely) on the scale of the observations. For studies that use firm level data, scale reflects the size of the firm; larger firms tend to have large values of many variables,

³ Recent papers in this long list include Begley and Feltham (2002), Bowen, Davis, and Rajgopal (2002), Bryant (2003), and Kallapur and Kwan (2004).

such. For studies that use per share values, scale reflects the size of the shares; some firms' shares are more valuable than others' shares because the former have fewer shares issued and outstanding, independent of other valuation characteristics such as the amount of anticipated future cash flows. In practically all instances, scale is a nuisance variable – one in which the researcher is not directly interested. However, the fact that scale has affected the observations leads to a number of econometric problems, including coefficient bias, R^2 bias, and heteroscedasticity. These issues, which are described more fully below, have been together described as “scale effects.”⁴

Following Barth and Kallapur (1996, BK hereafter) and Brown et al. (1999), consider the bivariate relation between $\mathbf{z} = (z_1, \dots, z_n)$ and $\mathbf{w} = (w_1, \dots, w_n)$. Assuming that the relation between the two variables is linear with a normally distributed disturbance, we can write:

$$z_i = \alpha + \beta w_i + \varepsilon_i. \quad (1)$$

Equation (1) models the relation “free of scale.” One might think of \mathbf{z} as market values at the end of a period that result from investing \$1 in each of the n assets. However, a researcher may not be able to observe the data free of scale but instead observes data affected by a scale factor $\mathbf{s} = (s_1, \dots, s_n)$ which reflects differential amounts of investment, resulting in:

$$s_i z_i = \alpha s_i + \beta s_i w_i + s_i \varepsilon_i. \quad (2)$$

The theoretically correct regression equation that satisfies the specification of (2) is

$$y_i = b + \alpha s_i + \beta x_i + \xi_i, \quad (3)$$

where $y_i = s_i z_i$, $x_i = s_i w_i$, $\xi_i = s_i \varepsilon_i$.

Equations (2) and (3) are identical, except for the addition of the intercept term b in (3) to

⁴ The generic term “scale effect” does not refer to any single problem and the use of this term has led to some confusion in the past. For example, Easton and Sommers (2003) state, “The overwhelming influence of large firms in these regressions is referred to as the ‘scale effect’,” but that definition is considerably different from that used in other studies. In this paper, I deliberately avoid this general term, and instead use the more specific names: coefficient bias, R^2 bias, and heteroscedasticity.

maintain the econometric consistency of the estimated coefficients (see Kennedy 1992, 111).

Suppose that, instead of the theoretically correct specifications of (1) or (3), researchers omit the unobservable variable s and estimate:

$$y_i = b_0 + b_1 x_i + \eta_i. \quad (4)$$

Brown, Lo, and Lys (1999) analyze the impact of scale on the coefficient of determination, or R^2 , resulting from equation (4). They analytically show that R^2 is a function of the coefficient of variation of the scale factor, and in the context of capital market research, that relation is usually positive. They conclude that cross-sample comparisons of R^2 are ill-advised without correcting for the differential variation in scale across samples.

The presence of scale in the data also leads to heteroscedastic errors in equation (4). Intuitively, larger observations have a higher likelihood of having errors with magnitude higher than a particular cutoff value. For example, there is a high likelihood of a valuation error of more than \$1 billion for a firm with market value of a \$100 billion, whereas that likelihood would be quite small for a firm with \$100 million market capitalization. Technically, from Equations (4) and (1), $\eta_i = s_i(\alpha + \varepsilon_i) + s_i(\beta - b_1)w_i - b_0$. Therefore, if the underlying errors ε_i have constant variance, then η_i will have variance proportional to the variance of s_i , which can be assumed to be increasing in s_i . Insofar as there are standard corrections for heteroscedasticity (e.g., standard errors computed according to White 1980), the researcher can accept the presence of heteroscedasticity. Alternatively, the researcher can divide the observed variables by a proxy for s and then estimate an approximate version of equation (1), thereby mitigating heteroscedasticity.⁵ This latter alternative is particularly useful when the requisite test statistics

⁵ Any remaining heteroscedasticity from a deflated model could of course be accommodated by the same adjustments as in the undeflated model. In addition, if homoscedasticity is achieved by deflation, then there is no need to incur the efficiency loss of using non-OLS methods such as White's (1980) correction.

do not (yet) have standard corrections for heteroscedasticity. For example, the Vuong (1989) statistic for non-nested models does not explicitly contemplate heteroscedastic errors, and Section 6 examines the degree to which inferences using the Vuong's Z-statistic is affected using a deflated model compared with an undeflated model that includes a scale proxy as a regressor.

The third problem arising from scale differences is coefficient bias. Estimation of equation (4) will result in coefficients that are biased due to the omission of s_i , which is clearly correlated with the included variable x_i , since $x_i = s_i w_i$. Brown et al. show that the estimated coefficients from Equation (4) will be:

$$\begin{aligned}\hat{\mathbf{b}} &= \begin{bmatrix} b \\ \beta \end{bmatrix} + \alpha(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{s} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e} \\ E(\hat{\mathbf{b}}) &= \begin{bmatrix} 0 \\ \beta \end{bmatrix} + \alpha\mathbf{c}\end{aligned}\tag{5}$$

$$\begin{aligned}\text{where } \hat{\mathbf{b}} &= (\hat{b}_0 \quad \hat{b}_1)' & \mathbf{X} &= (\mathbf{1} \quad \mathbf{x}) & \mathbf{e} &= \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \\ \mathbf{c} &= E(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{s} = (c_0 \quad c_1)' & & & E(\mathbf{X}'\mathbf{e}) &= 0.\end{aligned}$$

In other words, the amount of bias in the slope estimate ($E(\hat{b}_1) - \beta$) is increasing in α , the intercept from the original model, and c_1 , the degree of association between the scale factor and the independent variable \mathbf{x} . This latter association will increase with the variation in the scale factor.

The above equations point to two approaches to mitigate coefficient bias: deflating by a scale proxy or including a scale proxy as an independent variable. By deflating both dependent and independent variables by a scale proxy \mathbf{s}' , the first approach attempts to recover the original equation (1), although with error. In the second approach, adding an independent variable \mathbf{s}' results in an approximate version of equation (3). Barth and Kallapur (1996) find that the former approach is inferior to the latter approach for scale proxies that are as much as 95% correlated

with the true scale factor. The next section replicates the principal simulation results from BK, followed by a re-examination of this conclusion in Section 4.

3. Replication of results in Barth and Kallapur (1996)

The main conclusion arising from BK, that including a scale proxy as an independent variable is more effective at mitigating coefficient bias than deflating by the same scale proxy, is derived from regressions using simulated data. The simulations begin with 1990 Compustat firms with book value of equity and net income greater than 0.01 (\$ million), from which were selected the 500 firms with the largest total assets. A summary of BK's other assumptions are as follows (with relabelling to be consistent with the notation above), with σ^2 denoting the variance of the subscripted variable, and ρ the correlation between \mathbf{s} and \mathbf{s}' :⁶

A1. $z_i = \alpha + \beta w_i + \varepsilon_i$ (i.e., $\beta = 7$)

A2. $\alpha = \{1500, 150, 15\}$

A3. $w_i \sim N(200, 100^2)$

A4. $\varepsilon_i \sim N(0, 700^2)$

A5. s_i = book value of equity is the assumed scale factor

A6. s_i is independent of both z_i and w_i

A7. $s'_i = s_i + v_i$ is the simulated scale proxy

A8. $v_i \sim N(0, \sigma_{v_i}^2)$ where $\sigma_{v_i}^2$ satisfies $\text{plim} \frac{\sigma_{s_i}^2}{\sigma_{s_i}^2 + \sigma_{v_i}^2} = \rho^2$ and $\sigma_{v_i}^2 \propto s_i^2$

This condition restricting the ratio of the variances is derived from the fact that this ratio is the R^2 from a simple regression of A7, which equals the square of the correlation between s'_i and s_i

⁶ These assumptions are described on p. 538 and 540 of BK, although A8 will not be evident from a casual reading. For brevity, the analysis focuses on a subset of the most important simulation results reported in BK.

(i.e., ρ^2). It is this correlation that will be varied in the simulations. Note also that the variance of both scale s and the scale proxy s' are non-constant and depend on i . In A8, let k be the proportionality constant so $\sigma_{v_i}^2 = ks_i^2$. Then this assumption is equivalent to:

$$\text{A8. } v_i \sim N(0, \sigma_{v_i}^2) \text{ where } \sigma_{v_i}^2 = ks_i^2 \text{ where } k = \left(\frac{1}{\rho^2} - 1 \right) \frac{\sigma_s^2}{\mu_s^2 + \sigma_s^2}$$

For example, BK (p. 538) calculate that with a desired correlation of $\rho = 0.95$, and with in-sample estimates of $\mu_s = 3181$, $\sigma_s = 4625$, it is necessary that $\sigma_{v_i}^2 = \left(\frac{1521}{5616} \right)^2 s_i^2$. It should further be noted that the results replicated here also apply to the special but less descriptive case when s has constant variance.

BK examine the ability of a number of different models to mitigate coefficient bias, by comparing the distributions of the estimates for β to the true coefficient in equation (1). Of these various models, this paper examines the two models that are most important and relevant, compared with the benchmark Model 1, as follows:

$$\text{Model 1: } z_i = a_1 + b_1 w_i + e_{1i} \quad (6)$$

$$\text{Model 2: } y_i / s'_i = a_2 + b_2 x_i / s'_i + e_{2i} \quad (7)$$

$$\text{Model 3: } y_i = a_3 + b_3 x_i + c_3 s'_i + e_{3i} \quad (8)$$

Model 2 deflates the observed variables by a scale proxy, whereas Model 3 uses the observed variables, undeflated, but includes the scale proxy as a regressor.

Table 2 shows the descriptive statistics for the sample of 500 firms used to replicate BK's results and for all simulations in this study. The mean and standard deviation of s (i.e., book value) are slightly higher than those reported in BK, at 3393 and 4729, respectively. However, these minor differences should not affect the results materially.

Figures 2a and 2b shows the results of replicating BK's Figures 1A and 1C for $\alpha = 1500$

and $\alpha = 15$, respectively. (The intermediate value of $\alpha = 150$ has been omitted in this paper for brevity.) The results charted are, for each of Models 2 and 3, the 5th percentile, median, and 95th percentile of estimated slope coefficients less the true value of 7, for 250 randomizations using the above assumptions A1 to A8. A comparison with BK shows that their results have been successful replicated.

4. Analysis

Figures 2a and 2b show that Model 3 generally results in less coefficient bias compared with Model 2. Furthermore, Model 3 is more efficient since the distribution of coefficients from Model 3 has lower variance. Thus, just as BK conclude, including a scale proxy as a regressor (Model 3) is superior to deflating by the same scale proxy (Model 2). However, this conclusion are only warranted if the assumptions that generate the data are reasonable. While on the surface all the assumptions appear benign, detailed analysis of A7 and A8 show that the scale proxies generated have unusual characteristics.

Examine the 10 mini-charts in Figure 3. At the top is a representative plot of the underlying (scale-free) data for the 500 observations generated for one iteration of the simulation using A1 to A4. Visual inspection confirms that this data is generally well-behaved. The nine charts in the three rows below this graph plot relationships with $\rho = \{0.95, 0.75, 0.50\}$. First, focus on the left column. These three charts show that, when the correlation between \mathbf{s} and \mathbf{s}' is 0.95, the scale proxy has appropriate characteristics of (i) increasing one-to-one, on average, with the true scale factor; and (ii) positivity. However, when the correlation declines to 0.75 and 0.50, a substantial number of observations have negative values of the scale proxy. Deflating by negative values changes the sign of the observations in Model 2, and essentially moves data that generally reside in first Cartesian quadrant to the third quadrant, as illustrated in the three charts

in the central column. Such negative values do not make economic sense if the dependent variable reflects market value or stock price, for examples. The implication of the negative values of the scale proxy is that the estimated intercept is biased toward zero (when it should be 1500 in this case), while the slope coefficient is upwardly biased.⁷ In addition, these three charts show that the domain and range of the deflated data become increasingly large as ρ declines, due to the increasing likelihood of deflators that are close to zero when the true scale is not close to zero, which results in some extreme and influential observations which magnify the bias created by the observations in the third quadrant. These unusual values of the scale proxies explain why deflation according to Model 2 does poorly except when the correlation is extremely high. In contrast, the undeflated data in the three charts on the right column of Figure 3 show no substantial changes as ρ declines, except for some increased heteroscedasticity. Consequently, a regression of Model 3 does consistently well.

The plots in Figure 3 provide a visual impression of the data, but it is difficult to gauge more precisely the frequency or likelihood that the simulations generate the problematic observations. Consequently, Figure 4 shows the expected probability of observations that have negative scale proxies, small scale proxies (relative to the true scale factor), or both. For example, the probability of a negative scale proxy is computed as follows:

⁷ To see this, consider three observations with coordinates $(w, z) = \{(0, 1500), (500, 5000), (500, 5000)\}$. The estimated intercept equals 1500, and slope equals 7, with zero error. Suppose the $s/s' = \{1, 1, -1\}$. Therefore, the deflated data has coordinates $(x/s', y/x') = \{(0, 1500), (500, 5000), (-500, -5000)\}$. The estimated intercept is biased downward to 500 and the slope is biased upwards to 10.

$$\begin{aligned}
pr(s'_i < 0 | s_i) &= pr(s_i + v_i < 0 | s_i) \\
&= pr(v_i < -s_i | s_i) \\
&= pr\left(\frac{v_i}{\sigma_{v_i}} < \frac{-s_i}{\sigma_{v_i}} \middle| s_i\right) \\
&= pr\left(Z_i < \frac{-s_i}{\sqrt{ks_i^2}} \middle| s_i\right), \text{ where } Z_i \sim N(0,1) \\
&= pr\left(Z_i < -\left(\frac{\rho^2}{1-\rho^2} \frac{\mu_s^2 + \sigma_s^2}{\sigma_s^2}\right)^{\frac{1}{2}}\right)
\end{aligned} \tag{9}$$

For $\rho = 0.95$, $\mu_s = 3181$, $\sigma_s = 4625$, we obtain $pr(s'_i < 0 | s_i) = pr(Z_i < -3.69) = 0.01\%$, a probability that is so negligible that one does not expect to see even one instance of this occurrence in 500 observations. However, when $\rho = 0.50$, on average 121 of 500 observations are expected to have a negative scale proxy, since $pr(s'_i < 0 | s_i) = pr(Z_i < -0.700) = 24.2\%$. Even when the correlation is at a reasonably high level of 0.75, the probability is a non-trivial 8.4%. The heavy solid line in Figure 5 shows these probabilities at values of ρ from 0.50 to 0.99.

The probabilities for deflators (i.e., scale proxies) that are small relative to the true scale factor can be computed using calculations similar to Equation (9). “Small” here means that the ratio $|s'_i / s_i|$ is below a specified cutoff. For example, if $|s'_i / s_i| < 0.5$, then deflation will make the value of the observation *at least* twice as large as the true data. Again, Figure 5 shows that for all except the highest correlations, there is a substantial probability that the scale proxy will take values that are small. For a cutoff value of 0.5, the gap between the topmost and bottommost lines shows the probability to be about 20% and almost constant until ρ reaches 0.85. Likewise, for a cutoff value of 0.20, there is roughly a 9% chance that deflation will make

observations at least five times as large as the true data. Although not shown in the chart, there is in fact a 2% probability (10 out of 500 observations) that the deflator will make observations *more than 20 times* as large as the true data for correlations up to 0.85. Even more problematic is that about half of each of these probabilities (20%, 9%, and 2%) are for scale proxies that are both small *and* negative. For example, there is about a 1% chance that deflating by the scale proxy will change the sign of the observation and move it 20 times away from the origin (i.e., $s'_i / s_i \in (-0.05, 0) \Leftrightarrow x_i / s'_i < -20w_i$) for ρ up to 0.80. Therefore, the scale proxy can be very poor even while there is a reasonably high stated correlation between \mathbf{s} and \mathbf{s}' .

The source of the discrepancy between the stated correlations and the resultant quality of the scale proxy lies in BK's approach to generating the scale proxy \mathbf{s}' . Referring back to assumptions A7 and A8, noise (\mathbf{v}) is added to the true scale factor \mathbf{s} to obtain the proxy \mathbf{s}' , with the amount of noise determined by the desired correlation between \mathbf{s} and \mathbf{s}' . While simple, this is not the usual approach to generate correlated variables in monte carlo simulations. (Additional discussion of a better approach appears in the next section.) The biggest problem with this approach is that the scale proxy necessarily has higher variance than the true scale factor. When the desired correlation is high, say 0.95, $\sigma_{s'}^2 = (1/0.95)^2 \sigma_s^2$, there is not much difference in the variances and, if given a high enough mean (μ_s), there is a very low chance that the scale proxy will reach small or negative values. However, when the correlation is 0.50, then

$\sigma_{s'}^2 = (1/0.50)^2 \sigma_s^2$; the scale proxy will have *four times* the variance of the true scale factor.

Had BK extended the analysis to $\rho = 0.10$, the variance of the scale proxy would be 100 times as high. (Keep in mind that the mean is not changing, so the coefficient of variation increases along with the variance.)

While it is arguably an empirical issue whether such a relation is valid, a priori it does not

appear reasonable. The high variances result in substantial probabilities that the scale proxy becomes relatively small or negative, resulting in many observations that do not make economic sense. One may argue that it does not matter as long as all competing models use the same scale proxy. However, the approach used to construct the scale proxy handicaps Model 2 (deflation) but doesn't do so for the Model 3 because of the error assumption in A7 is additive (i.e., $s'_i = s_i + v_i$). Consequently, the distributions are reasonable in a linear sense: by definition, $s'_i - s_i$ is distributed normally because v_i is assumed Normal, so that extreme values of $s'_i - s_i$ are always limited to Normal probabilities. In contrast, the distribution of s_i / s'_i is highly non-Normal, as shown by the above probabilities of extreme values. In fact, the distribution of s'_i / s_i is $N(1, k)$, so the inverse s_i / s'_i has a Cauchy distribution, which has a non-existent mean and infinite variance. To justify the approach in BK, one may also argue these outcomes are appropriately part of the simulation. This is not reasonable because negative scale proxies are non-sense deflators that either do exist in the data, and when they do exist, are not used by researchers. For examples, a common deflator is market value of equity, which is always positive. When researchers use other deflators that could have negative values, such as total assets or book value of equity, they generally exclude observations with such negative deflators.

To obtain an idea of the impact of the small and negative values of the scale proxy on the estimates of coefficient bias shown in Figures 2a and 2b, one can simply eliminate observations that have such characteristics. Admittedly, this process is ad hoc, but this analysis is simply intended to gauge whether the unusual deflators *could* be driving the poor performance Model 2. A technically sound approach requires re-generating the data with other, more reasonable assumptions, which will be provided in the next section.

Figure 5a and 5b are analogous to Figures 2a and 2b, except that observations with $0.5 <$

$s'_i / s_i < 2$ (i.e., values of the scale proxy and the true scale factor are no more than 100% away from each other). These figures show that, once unusually large or small scale proxies are excluded, Model 2 (deflating) performs as well as or better than Model 3 (including a scale proxy as regressor). Specifically, there is little difference between the two methods on average, and the distribution of coefficient bias is much tighter for Model 2. As the next section will show, these findings are not just due to the ad hoc way in which poor values of the scale proxy have been discarded. Rather, the results from these reduced samples reflect more general outcomes when the distribution of the scale proxy is more reasonably specified.

5. Alternative assumptions to generate scale proxies

The previous section showed that assumptions A7 and A8 result in scale proxies that have unreasonable characteristics. To reiterate, those assumptions inflate the variance of the scale proxy as the specified correlation decreases, and this increased variance results in non-trivial numbers of observations with values of the scale proxy that are negative, or approach zero arbitrarily closely.⁸ An approach that avoids this problem is to use variations of a Cholesky decomposition that is used in many monte carlo simulations.

For technical reasons as shown in Appendix A, a standard Cholesky decomposition is not appropriate for the present circumstances. Instead, a modified approach that uses the general framework of the Cholesky method is required. Essentially, for a two variable case, the decomposition starts with two independent and identically distributed variables (call these \mathbf{z}_1 and \mathbf{z}_2). One of the variables of interest (\mathbf{s} in the present case) is set to equal \mathbf{z}_1 . The second variable of interest (\mathbf{s}') is defined as a weighted average of \mathbf{z}_1 and \mathbf{z}_2 , with the weights set precisely to attain the required degree of correlation (see Appendix A for derivation of these weights). In a

⁸ Of course, since the distributions are continuous, there is zero probability of the scale proxy taking the value of exactly zero, in which case the deflated values would not exist.

standard decomposition, both the mean and variance \mathbf{s}' would equal those of \mathbf{s} . However, the fact that \mathbf{s} is always positive results in simulated \mathbf{s}' that only preserve the mean or the variance, but not both.⁹

Two weighting schemes are thus employed to generate scale proxies, as follows, replacing A7 and A8:

$$\text{A9. Scheme A: } s'_i = \lambda s_i + (1 - \lambda)t_i$$

$$\text{A10. Scheme B: } s'_i = \rho s_i + \sqrt{1 - \rho^2} t_i$$

The random variable t_i is independent of and has the same distribution as s_i . The simplest way to generate t_i is to sample from the actual \mathbf{s} vector, randomly with replacement. The first weighting scheme is just a linear combination of two independent variables with the same distribution, so

$$E(s'_i) = E(s_i), \text{ but the variance will be } \text{Var}(s'_i) = (1 - 2\lambda + 2\lambda^2)\text{Var}(s_i) < \text{Var}(s_i) \text{ for } \lambda \in (0,1).^{10}$$

Scheme B uses the Cholesky weights based on the specified correlation ρ . This weighting scheme results in equal variances, with $\text{Var}(s'_i) = \rho^2\text{Var}(s_i) + (1 - \rho^2)\text{Var}(s_i) = \text{Var}(s_i)$, but unequal expected values, with $E(s'_i) > E(s_i)$ for $\lambda \in (0,1)$. There are no strong justifications to prefer equal means or equal variances, so the analysis below examines both.

Figure 6a shows the results of 250 repetitions of applying Scheme A at each of 20 values of $\lambda = \{0.05, 0.10, \dots, 0.90, 0.95, 0.99\}$, for the scenario with potentially high coefficient bias because of a large intercept of $\alpha=1500$. As a reminder, each repetition involves regressions with 500 observations. The scatter plot shows the amount by which the estimated slope coefficients

⁹ Intuition would suggest that the mean can be stabilized by first de-meaning s_i and t_i , then applying weighting Scheme 2, and finally adding back the mean of s_i . However, this is inappropriate because, similar to BK's approach, the distribution of s'_i would include values that are negative and approach zero arbitrarily closely.

¹⁰ Readers will recognize this effect as the variance reduction from portfolio diversification. This lower variance results in correlations (in probability limit) that are higher than the weight λ under Scheme A; this correlation can be computed as $\text{plim corr}(s, s') = \lambda / \sqrt{1 - 2\lambda + 2\lambda^2} > \lambda$. In contrast, $\text{plim corr}(s, s') = \rho$ under weighting Scheme B.

from Model 2 or Model 3 deviate from the true value of 7 for each empirical correlation of s and s' on the horizontal axis. (The empirical correlations are used instead of the specified weights so that observations do not cluster at 20 locations on the x-axis and obscure the chart.) The plot shows that Model 2 (deflation) generally results in less coefficient bias, and the coefficients that are more tightly distributed compared with Model 3. Incorporating both bias and efficiency, the mean squared error criterion can be used to rank the two models, and this is shown in Figure 6b. This chart shows the squared errors (i.e. $(\text{estimated slope} - 7)^2$) for each of the $20 \times 250 = 5000$ repetitions, and a least squares polynomial of the fifth degree to ease interpretation. The results show that the squared errors are on average lower for Model 2 than for Model 3 except for a small range from about 0.8 to 0.97.¹¹

Figures 7a and 7b repeat this analysis for the scenario with a small intercept of $\alpha = 15$. Because of the smaller potential coefficient bias, Figure 6a shows that there is little discernible difference between the two models in terms of bias, but Model 2 clearly dominates Model 3 in terms of efficiency. As a result, Figure 6b shows that the difference in squared errors (Model 2 less Model 3) are predominantly negative throughout the entire range of correlations, and the mean squared error illustrated by the least squares polynomial is always negative in favor of Model 2.

The previous 4 figures are based on weighting Scheme A. Very similar results are obtained using Scheme B. For brevity, I show only the plots of the differences in squared errors and not the individual slope coefficients, for the two cases of $\alpha = \{1500, 15\}$. Figures 8a and 8b look in all material respects identical to Figures 5b and 6b. Therefore, the general dominance of Model 2 over Model 3 is not sensitive to the weighting scheme employed.

¹¹ Similar results obtain if mean squared errors are computed at each of the 20 values of λ instead of fitting a polynomial to the scatter plot.

Taken together, the above results show that deflating by a scale proxy in general results in less coefficient bias than including the same scale proxy as a regressor. This is distinctly opposite the conclusion drawn in BK. The different conclusions result from the way the scale proxies are generated. BK's approach can result in some non-sensible negative values of scale proxies, and proxy values that deviate enormously from the true scale values even when there are high stated correlations between the scale proxy and the true scale factor. The current approach aims to generate scale proxies that come from the same distribution as the true scale factor, resulting in values of the scale proxy that are more reasonable.

6. Further analyses: tests of incremental and relative association

The last section has shown that it is more often than not the case that deflating by a scale proxy will better mitigate coefficient bias in the case of a single independent variable of interest. However, researchers often desire to answer research questions that differ from this simple scenario. For example, the above scenarios could be interpreted as regressions of equity market values on net income (deflated or not). The research question is simply whether net income is associated with equity market values (i.e., whether net income is “value relevant”). A different research question is whether another income statement item (e.g., foreign exchange gain/loss, other comprehensive income, difference in US and non-US GAAP income) is incrementally associated with equity market values (i.e., whether these items are *incrementally* value relevant). Yet another question is which of two methods of computing income is relatively more highly associated with equity market values (i.e., whether one method has higher *relative* value relevance). The following analyses address the impact of scale on these two types of questions.

6.1. Tests of incremental association

The following three equations are analogous to those in presented in Section 3, with the addition

of variables w_2 , x_2 / s' , and x_2 for Models 1, 2, and 3, respectively. In incremental tests, what are of interest are the coefficients for these variables (i.e., b_{12} , b_{22} , or b_{32} in the following equations), and whether these values are significantly different from zero (or possibly some other value under the null hypothesis).

$$\text{Incremental Model 1: } z_i = a_1 + b_{11}w_{1i} + b_{12}w_{2i} + e_{1i} \quad (10)$$

$$\text{Incremental Model 2: } y_i / s'_i = a_2 + b_{21}x_{1i} / s'_i + b_{22}x_{2i} / s'_i + e_{2i} \quad (11)$$

$$\text{Incremental Model 3: } y_i = a_3 + b_{31}x_{1i} + b_{32}x_{2i} + c_3s'_i + e_{3i} \quad (12)$$

The presence of a scale factor in Models 2 and 3 can bias the estimated coefficient $b_{\bullet 2}$ in a manner similar to the effect on $b_{\bullet 1}$. In particular, b_{22} or b_{32} could be found to be significant even if \mathbf{z} is not associated with \mathbf{w}_2 , but because scale induces an association between $\mathbf{y} = \mathbf{s}'\mathbf{z}$ and $\mathbf{x}_2 = \mathbf{s}'\mathbf{w}_2$. To investigate which of Incremental Models 2 or 3 is less likely to suffer from this problem, I conduct a set of simulations similar to those in the previous section.

In addition to assumptions A1 to A6, A9, and varying the values of λ (which induces various level of correlation between \mathbf{s} and \mathbf{s}'), the multivariate regressions in Equations (11) and (12) requires the specification of \mathbf{w}_2 and how this variable is correlated with \mathbf{w}_1 . Thus, it is necessary to replace A1 and A3 with the following, respectively:

$$\text{A11. } z_i = \alpha + 7w_{1i} + 0w_{2i} + \varepsilon_i \quad (\text{i.e., } \beta_1 = 7, \beta_2 = 0)$$

$$\text{A12. } w_{1i} = 200 + 100v_{1i} \quad \text{where } v_{1i} \sim N(0,1)$$

$$w_{2i} = 100 + 100\pi v_{1i} + 100\sqrt{1-\pi^2} v_{2i} \quad \text{where } v_{2i} \sim N(0,1), v_{1i} \perp v_{2i}$$

Assumption A11 corresponds to the null hypothesis that the second independent variable is not incrementally associated with the dependent variable. A12 specifies the two independent variables with correlation π using the Cholesky decomposition as described in Appendix A. The

choice of $E(w_{2i}) = 100 = \frac{1}{2} E(w_{1i})$ is empirical descriptive in that the incremental variable will usually have a lower mean than then the first (non-incremental) variable. For examples, foreign exchange gain/loss, other comprehensive income, difference in US and non-US GAAP income will usually have smaller expected values than net income.¹²

Similar to Figures 6a and 7a, Figure 9a plots the estimates for coefficient b_{22} and b_{32} against the empirical correlations between the scale factor and the scale proxy, when π , the correlation between w_1 and w_2 , equals zero, and the intercept $\alpha = 1500$. This chart shows that Incremental Model 2 is superior. Figure 9b then extends this analysis over 19 values of $\pi = \{-0.9, -0.8, \dots -0.1, 0, 0.1, \dots 0.8, 0.9\}$ and 20 values of $\lambda = \{0.05, 0.10, \dots 0.90, 0.95, 0.99\}$ using weighting Scheme A, for a total of 380 combinations.¹³ This chart shows the differences in the mean squared errors of Incremental Model 2 less those of Incremental Model 3 for coefficient $b_{\bullet 2}$. Each mean squared error is computed using estimated coefficients $b_{\bullet 2}$ from 250 iterations (as in the previous section). Figure 9b shows that, in general, the differences in mean squared errors are negative. Incremental Model 2 (deflating by scale proxy) is less biased or more efficient, or both, compared with Model 3 (including scale proxy as regressor). Only for extreme values of π , positive or negative, and a small range of λ , are the mean squared errors lower for Model 3.

Figures 10a and 10b repeats this analysis for $\alpha = 15$. Strikingly, in all cases, the mean squared errors of the incremental coefficients are lower when the variables are deflated by a scale proxy, compared with when the same scale proxy is included as a regressor. This result holds regardless of the extent to which the scale proxy approximates the true scale factor, and

¹² The results below are qualitatively similar for $E(w_{2i}) = \{50, 100, 150\}$

¹³ Because the results of the previous section shows that both weighting scheme result in almost identical inferences, only Scheme A is used in this section.

regardless of the degree to which the independent variables are correlated. Taken together, these results show that it is in general more effective to deflate by a scale proxy if a researcher is interested in testing for incremental associations.

6.2. Tests of relative association

Tests of relative association arise frequently in accounting. For example, a researcher may be interested in whether one measure of income is superior to another. Such a question involves a comparison of two non-nested models, in which one model cannot be expressed as a constrained version of the other (as was the case in Section 6.2). Vuong (1989) provides a Z-statistic for testing such non-nested models.¹⁴

Vuong's model assumes that errors are homoscedastic. However, since the presence of a scale factor can result in heteroscedasticity, Vuong's Z is potentially mis-specified when scale affects the regression variables. The following analysis examines which of Models 2 or 3 is better specified: which model better matches the theoretical rejection rates under the null hypothesis of equal explanatory power for two alternative (sets of) independent variables.

To be precise, recall assumptions A1 to A8 and as adjusted in Section 5. The true data generating process is given by A1 through A4. Suppose a researcher is interested in comparing the explanatory power of the following two equations:

$$\text{Relative Model 1':} \quad z_i = a_1' + b_1'w_i' + e_{1i}' \quad (13)$$

$$\text{Relative Model 1'':} \quad z_i = a_1'' + b_1''w_i'' + e_{1i}'' \quad (14)$$

Neither w' nor w'' is the true independent variable w . Rather, under the null hypothesis, both variables are equally correlated with w . Again utilizing the Cholesky decomposition in

Appendix A, the variables are constructed as follows, replacing assumption A3:

¹⁴ Vuong (1989) provides statistics for both nested and non-nested models, although the latter have proven most useful because of the availability of other statistics for nested models.

$$\begin{aligned}
& w_i = 200 + 100v_{3i} \\
\text{A13. } & w'_i = 200 + 100\theta v_{3i} + 100\sqrt{1-\theta^2}v_{4i} \quad \text{where } \begin{bmatrix} v_{3i} \\ v_{4i} \\ v_{5i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \\
& w''_i = 200 + 100\theta v_{3i} + 100\sqrt{1-\theta^2}v_{5i}
\end{aligned}$$

Here, θ is the desired correlation between the true variable \mathbf{w} and either of the actual regressors \mathbf{w}' and \mathbf{w}'' . Under these assumptions, Vuong's Z-statistic for Model 1' vs. Model 1'' (denoted as VZ_1) has a standard normal distribution.

Of interest to this study are the distributions of the Z-statistics (denoted as VZ_2 and VZ_3) for the following two sets of equations:

$$\text{Relative Model 2':} \quad y_i / s'_i = a_2 + b_2 s_i w'_i / s'_i + e'_{2i} \quad (15)$$

$$\text{Relative Model 2'':} \quad y_i / s'_i = a_2 + b_2 s_i w''_i / s'_i + e''_{2i} \quad (16)$$

$$\text{Relative Model 3':} \quad y_i = a_3 + b_3 s_i w'_i + c_3 s'_i + e'_{3i} \quad (17)$$

$$\text{Relative Model 3'':} \quad y_i = a_3 + b_3 s_i w''_i + c_3 s'_i + e''_{3i} \quad (18)$$

The distributions of these VZ statistics are simulated using A1, A2, A4 to A6, A9, A13, with intercept $\alpha = \{1500, 15\}$, correlation between \mathbf{w} and \mathbf{w}' (or \mathbf{w}'') $\theta = \{0.50, 0.95\}$, and $\lambda = \{0.50, 0.95\}$. To obtain a sufficient degree of accuracy in the simulated probabilities for nominal p-values as low as 0.001, the statistics are computed over 10,000 iterations.

Table 3 shows the results of the simulation. The reported results are for $\alpha = 1500$; those for $\alpha = 15$ are similar and have been omitted for brevity. The table shows the probabilities of obtaining a VZ statistic (frequency of occurrence \div 10,000) whose value is more extreme than the cutoff value for the corresponding p-value. First, observe that the benchmark VZ_1 indeed has a standard Normal distribution in accordance with theory. Second, we see that VZ_2 , computed from the deflated models, has an almost identical distribution. All of the empirical probabilities approximate those of the standard Normal p-values, the mean value of the statistics is close to

zero, and the standard deviation is close to unity. Interestingly, this is the case regardless of the value of θ and λ . In contrast, the VZ_3 statistics computed from the undeflated models have empirical probabilities that consistently exceed those of the Normal values. For instance, for $\theta = 0.50$, $\lambda = 0.50$, and a lower-tail p-value of 0.05, we observe VZ_3 exceeding in magnitude the theoretical cutoff value of -1.645 in 0.0951 fraction of the time (951 out of 10,000 iterations), almost twice the Normal p-value of 0.05. For the lower-tail p-value of 0.001, the empirical probability of 0.010 is 10 times as large. While VZ_3 is unbiased, with mean close to zero, the standard deviation is always significantly higher than unity in all four combinations of θ and λ . Taken together, these results show that deflating by a scale proxy results in VZ-statistics that are well-specified, whereas including the scale proxy as a regressor (and not deflating) results in VZ-statistics that overstate the degree of significance (i.e., understate the p-value).

7. Conclusions

The analysis shows that, in general, deflating by a scale proxy provides more accurate inferences than including the scale proxy as a regressor (and not deflating). This conclusion applies to simple regressions with one independent variable of interest, tests of incremental association, and test of relative association using Vuong's (1989) Z-statistic. This conclusion is based on results generated from a wide range of correlations between the scale proxy and the true scale factor, a wide range of correlations between independent variables (when there are two), and a wide range of correlations between proxies for the independent variables and the true variables. These results contrast with those found in Barth and Kallapur (1996) because the prior study uses assumptions that produce values of the simulated scale proxies that are economically unreasonable whereas no such unreasonable values arise from the modified assumptions in this study. While one may also disagree with the assumptions used in this paper, at the very least it

can be said that the conclusion in Barth and Kallapur (that deflation is an inferior approach) is not general.

Many settings in empirical accounting research involve the use of data affected by the differential sizes of the observations (firms). The market value of equity, net income, book value of equity, sales, accruals, and cash flows are all dependent on the scale of firms' operations. Thus, inferences on a diverse range of research questions involving equity valuation, alternative accounting standards, measurement of discretionary accruals, and the usefulness of accounting income versus cash flows are potentially affected by this issue.

Appendix A

A standard Cholesky decomposition generates J normal random variables with unit variance and correlation matrix Σ .¹⁵ This decomposition is achieved by factoring $\Sigma = \mathbf{A}'\mathbf{A}$, where \mathbf{A} is an upper triangular matrix. Letting \mathbf{z} be a $J \times 1$ vector of independent standard normal random variables, then $\mathbf{y} = \mathbf{A}'\mathbf{z}$ will be the desired vector of random variables with unit variance and correlation matrix Σ . For two variables, we have:

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & \rho \\ 0 & \sqrt{1-\rho^2} \end{bmatrix} \quad (\text{AP1})$$

Therefore, the two random variables are:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ \rho z_1 + \sqrt{1-\rho^2} z_2 \end{bmatrix}, \quad \text{and } \mathbf{y} \sim N(\mathbf{0}, \Sigma) \quad (\text{AP2})$$

Notice that this approach yields the desired correlations between two random variables while maintaining equal variances, in contrast to assumptions A7 and A8. This would be useful for the generation of scale proxies except that s_i and s'_i are not Normal variates; they do not even have zero expected values (because scale factors must be positive), which is the essential requirement of the decomposition. To see why, assume that Z_1 and Z_2 are two i.i.d. distributions with positive expected values. Then it follows from Equation (AP2) that $E(y_2) > E(y_1)$ for $\rho \in (0,1)$. Thus, applying a standard Cholesky decomposition would result in scale proxies that systematic over-estimate the true scale factor.

¹⁵ The notation in this appendix is for one observation at a time, as opposed to a vector of observations used in the main text of the paper.

References

- Aboody D. and B. Lev, 1998. The value relevance of intangibles: The case of software capitalization. *Journal of Accounting Research* 36 Supplement, 161-191.
- Barth, M, and S. Kallapur, 1996. The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13, 527-67.
- Begley, J., and G. Feltham, 2002. The relation between market values, earnings forecasts, and reported earnings. *Contemporary Accounting Research* 19, 1-48.
- Bowen, R., A. Davis, and S. Rajgopal, 2002. Determinants of revenue-reporting practices for internet firms. *Contemporary Accounting Research* 19, 523-562.
- Brown, S., K. Lo, and T. Lys, 1999. Use of R^2 in accounting research: measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics* 28, 83-115.
- Bryant, L., 2003. Relative value relevance of the successful effort and full cost accounting methods in the oil and gas industry. *Review of Accounting Studies* 8, 5-28.
- Christie, A., 1987. On cross-sectional analysis in accounting research. *Journal of Accounting and Economics* 9, 231-58.
- D'Souza J. and J. Jacob 2001. Electric utility stranded costs: Valuation and disclosure issues. *Journal of Accounting Research* 39, 495-512.
- Easton, P. , 1998. Discussion of revalued financial, tangible, and intangible assets: association with share prices and non-market-based value estimates. *Journal of Accounting Research* 36, 235-47.
- Easton, P., and G. Sommers, 2003. Scale and the scale effect in market-based accounting research. *Journal of Business, Finance and Accounting* 30, 25-55.
- Ely K., G. Waymire, 1999. Intangible assets and stock prices in the pre-SEC era. *Journal of Accounting Research* 37 Supplement, 17-44.
- Kallapur, S., and S.Y.S. Kwan, 2004. The value relevance and reliability of brand assets recognized by U.K. firms. *The Accounting Review* 79, 151-172.
- Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307-333.
- White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica* 48, 817-838.

Figure 1a: Hypothetical relation between negative earnings and market value of equity in scale-free data

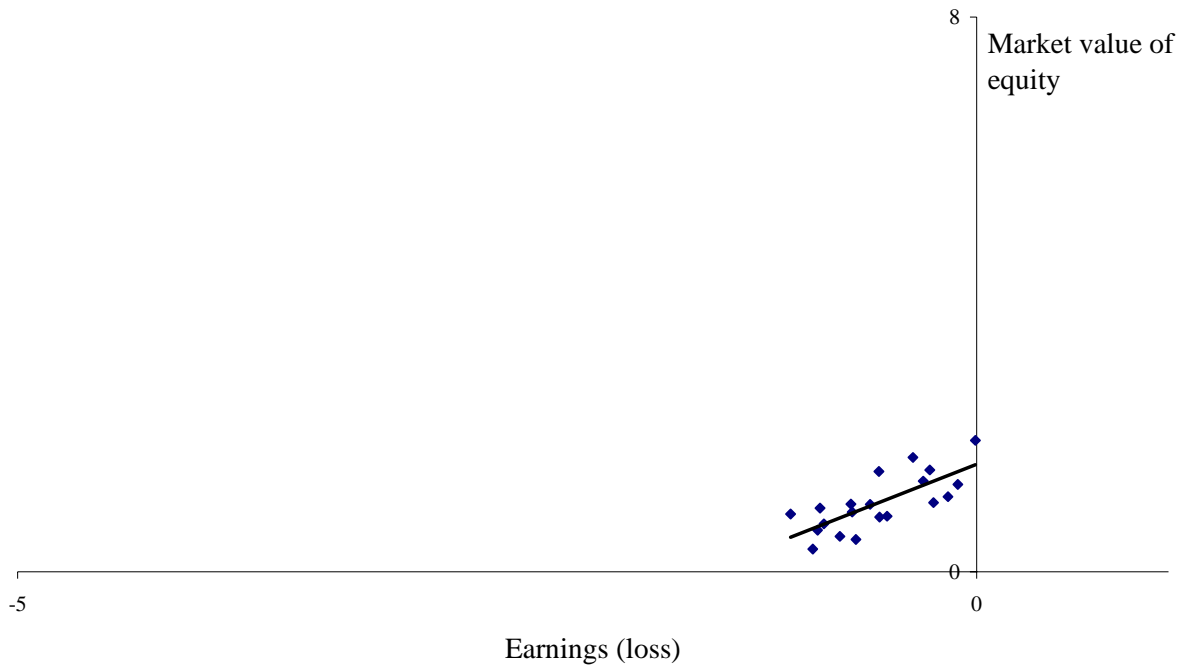
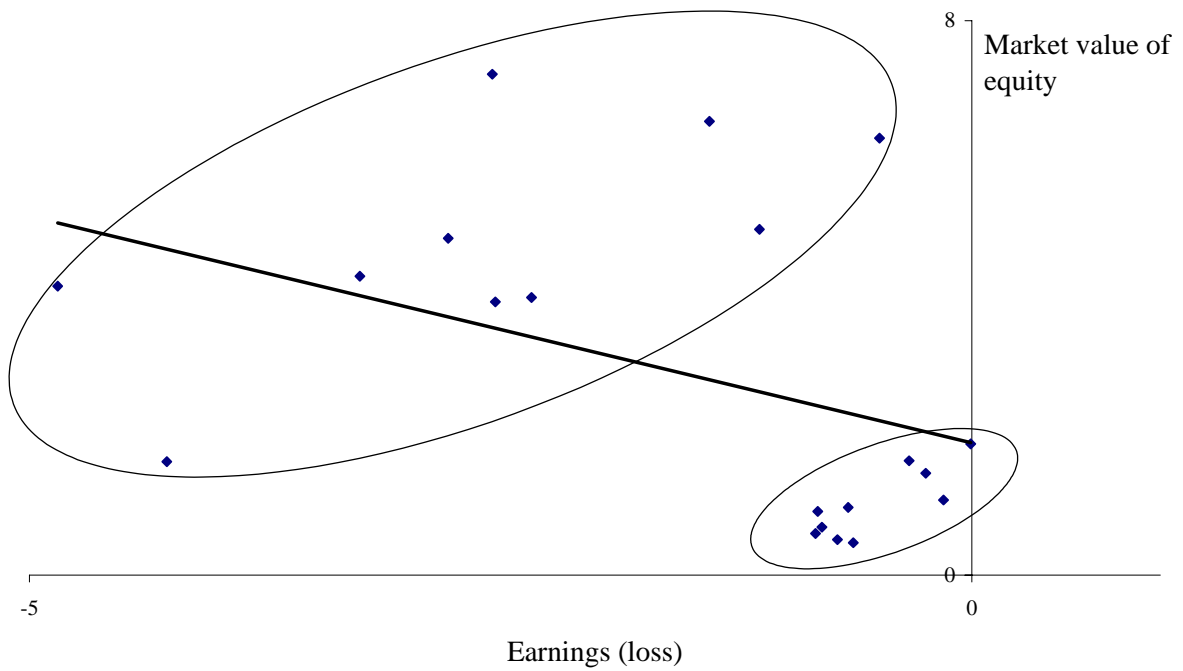


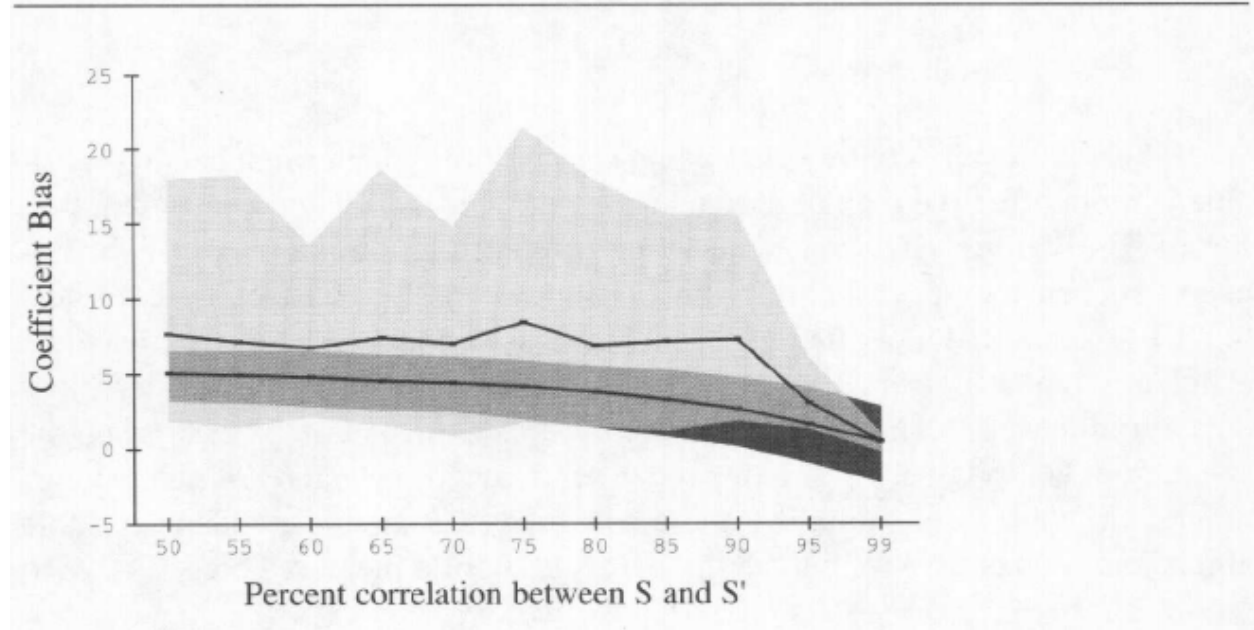
Figure 1b: Hypothetical relation between negative earnings and market value of equity when half of data are scaled up by a factor of 5



Extract from Barth and Kallapur (1996) - Figure 1 Panels A and C

The top (bottom) line and lightly (darkly) shaded area are, respectively, the mean bias and 95% confidence interval for Model 2 – deflated (Model 3 – scale proxy as regressor).

Panel A: Intercept, a , equals 1,500



Panel C: Intercept, a , equals 15

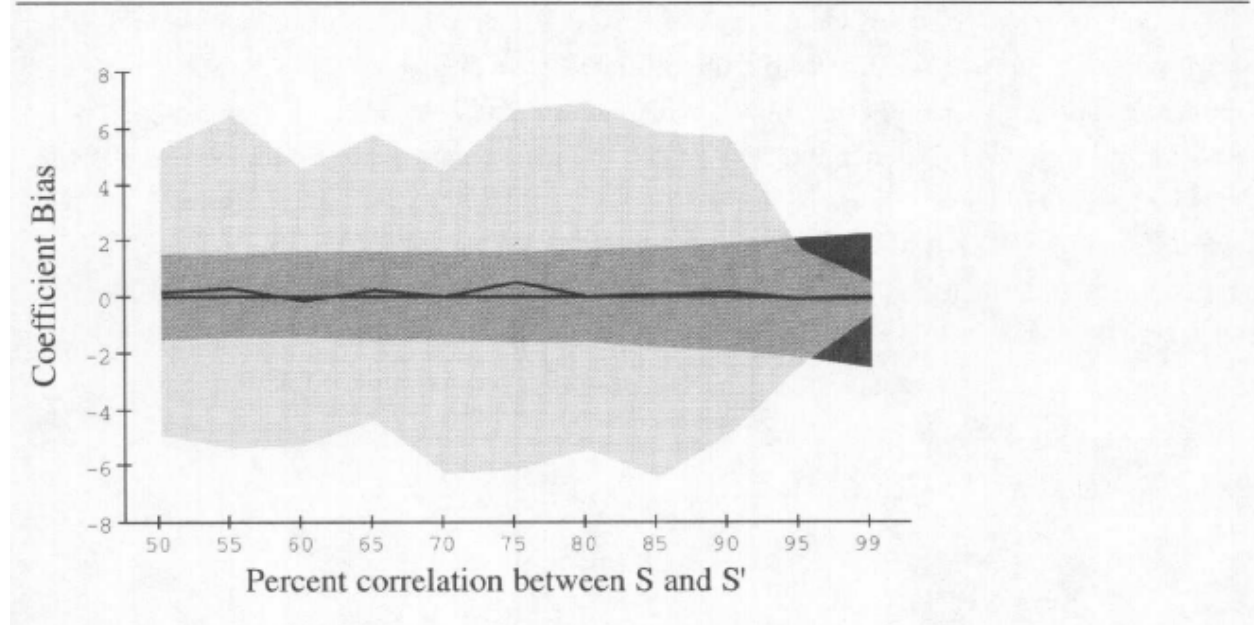


Figure 2a: Replication of BK Figure 1A
 Simulated distributions of coefficient bias when Intercept = 1500

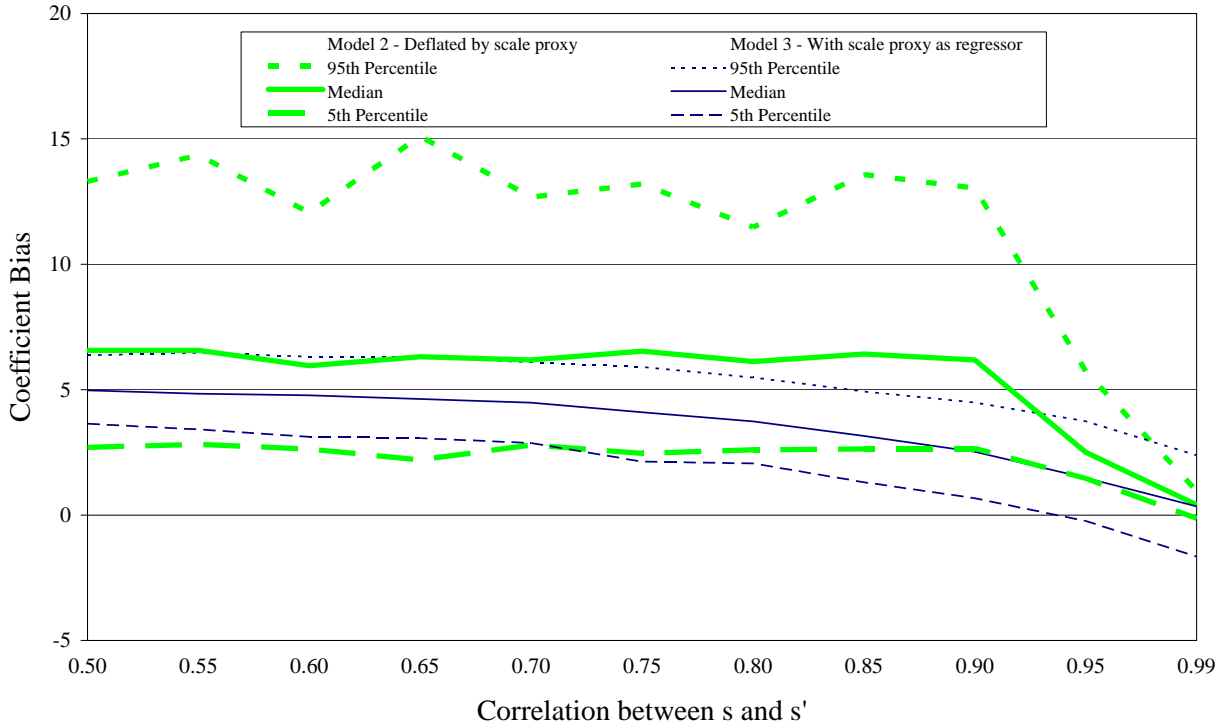


Figure 2b: Replication of BK Figure 1C
 Simulated distributions of coefficient bias when Intercept = 15

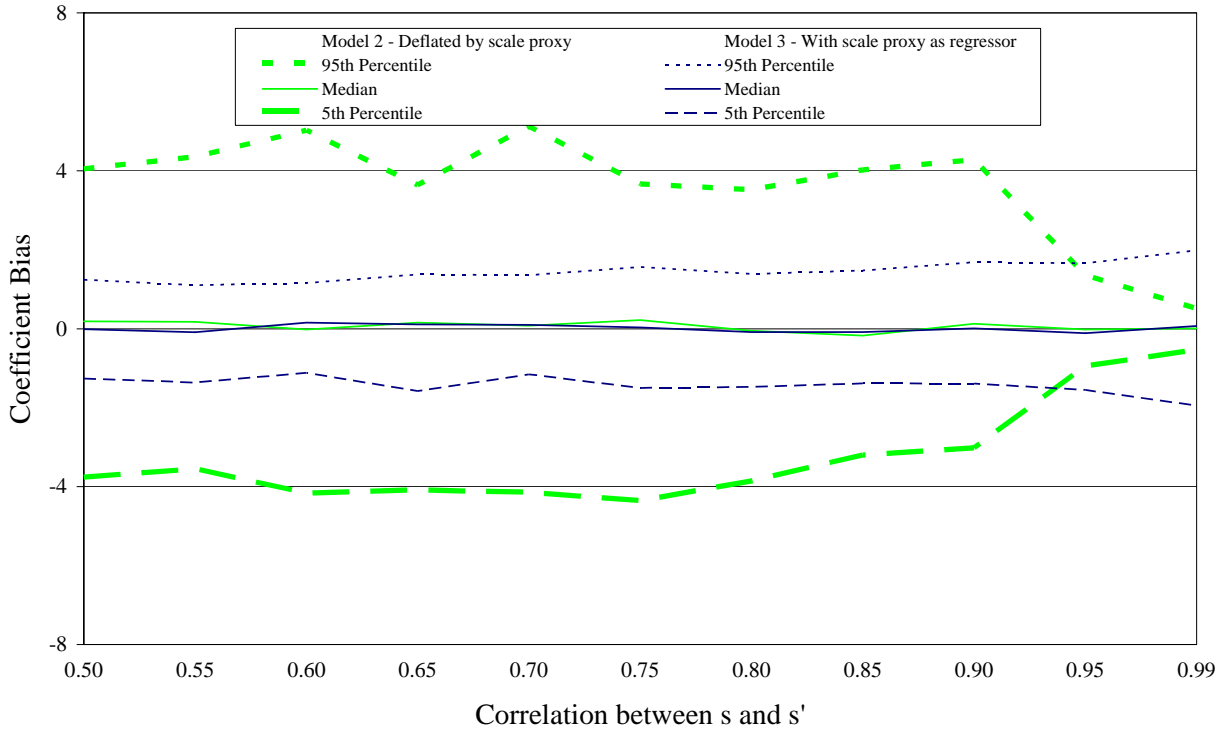


Figure 3 – Data plot of 500 simulated observations from one representative iteration at three levels of correlation between the scale proxy and the true scale factor

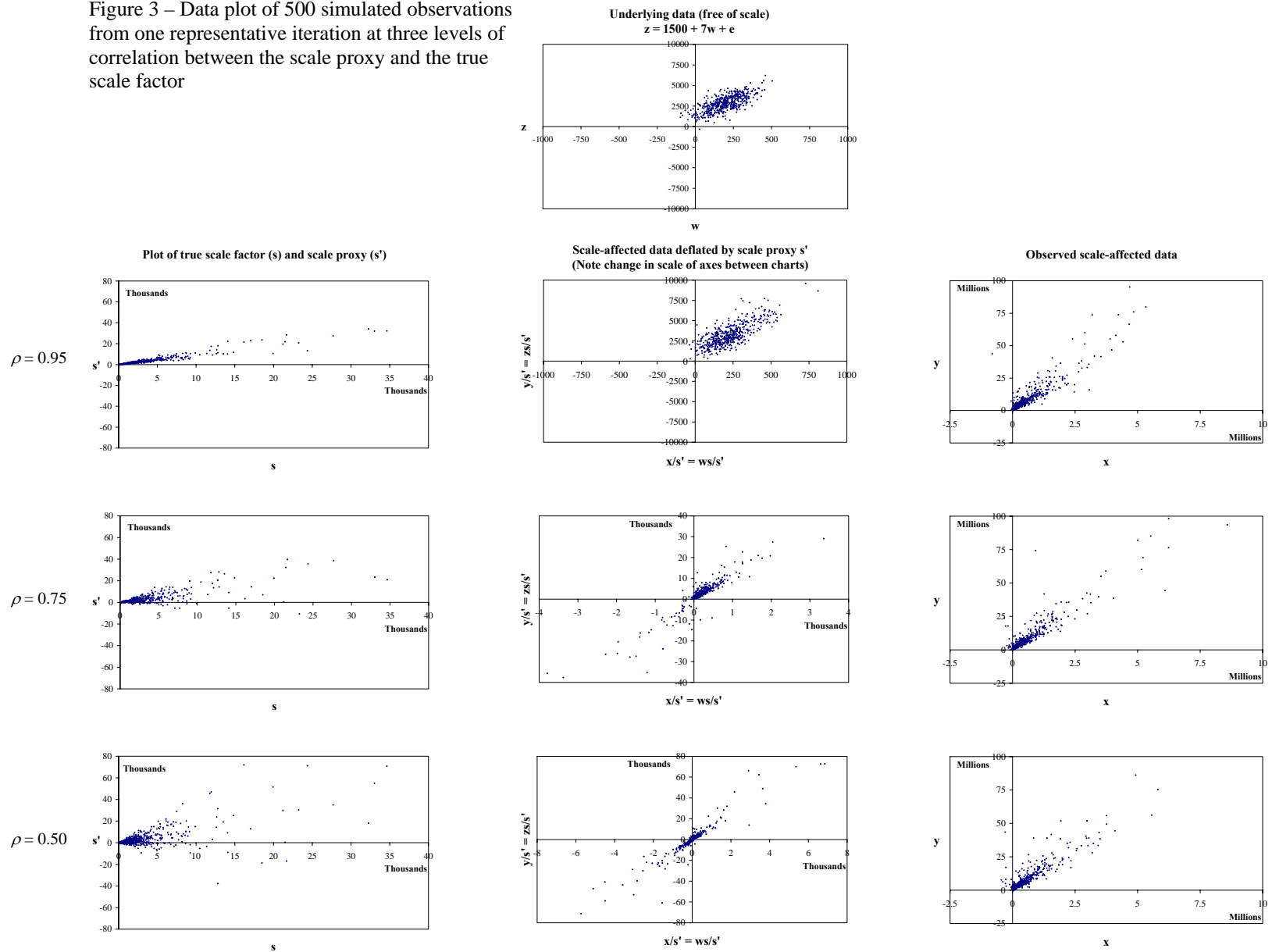


Figure 4: Probabilities of small and negative deflators at various correlations of the scale proxy to the true scale factor using BK's assumptions (A1 to A8)

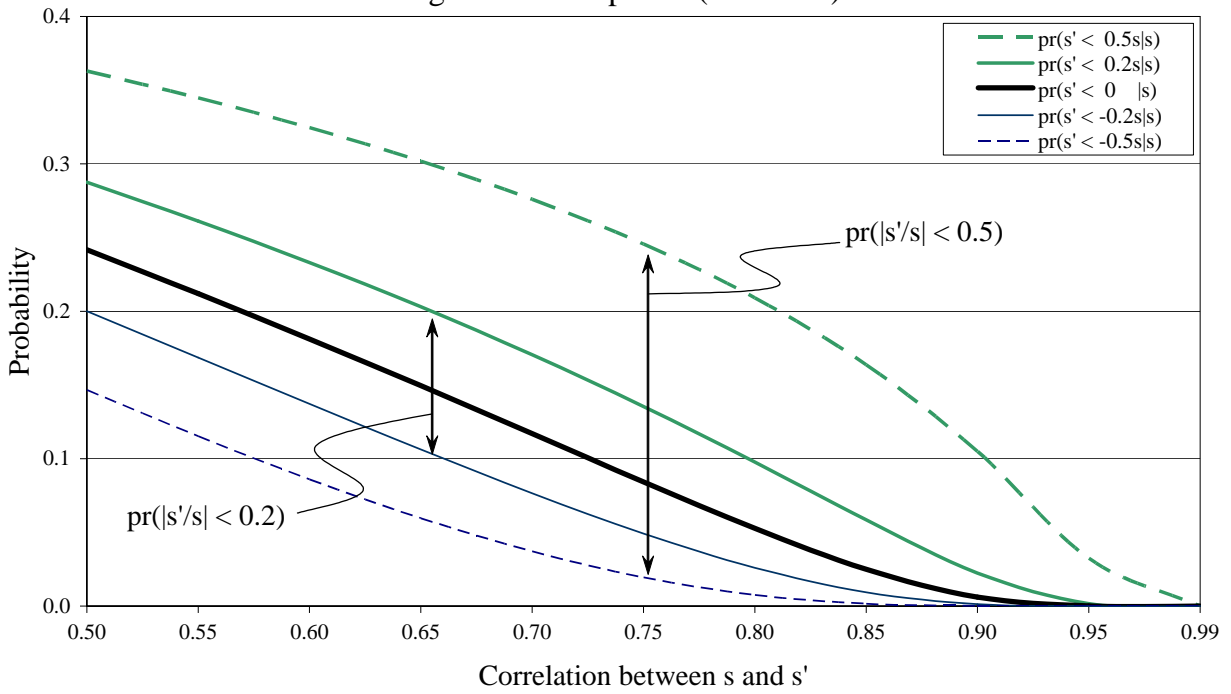


Figure 5a: Simulated distribution of coefficient bias when intercept = 1500 and including only observations with $0.5s < s' < 2s$

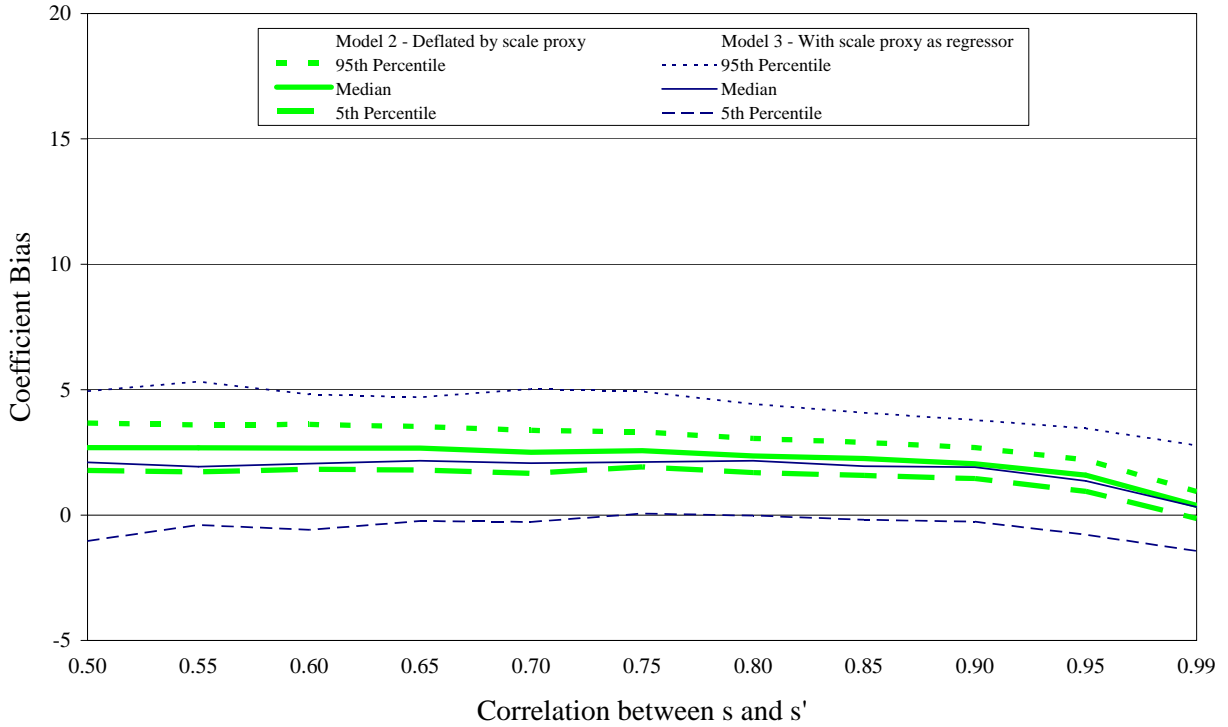


Figure 5b: Simulated distribution of coefficient bias when intercept = 15 and including only observations with $0.5s < s' < 2s$

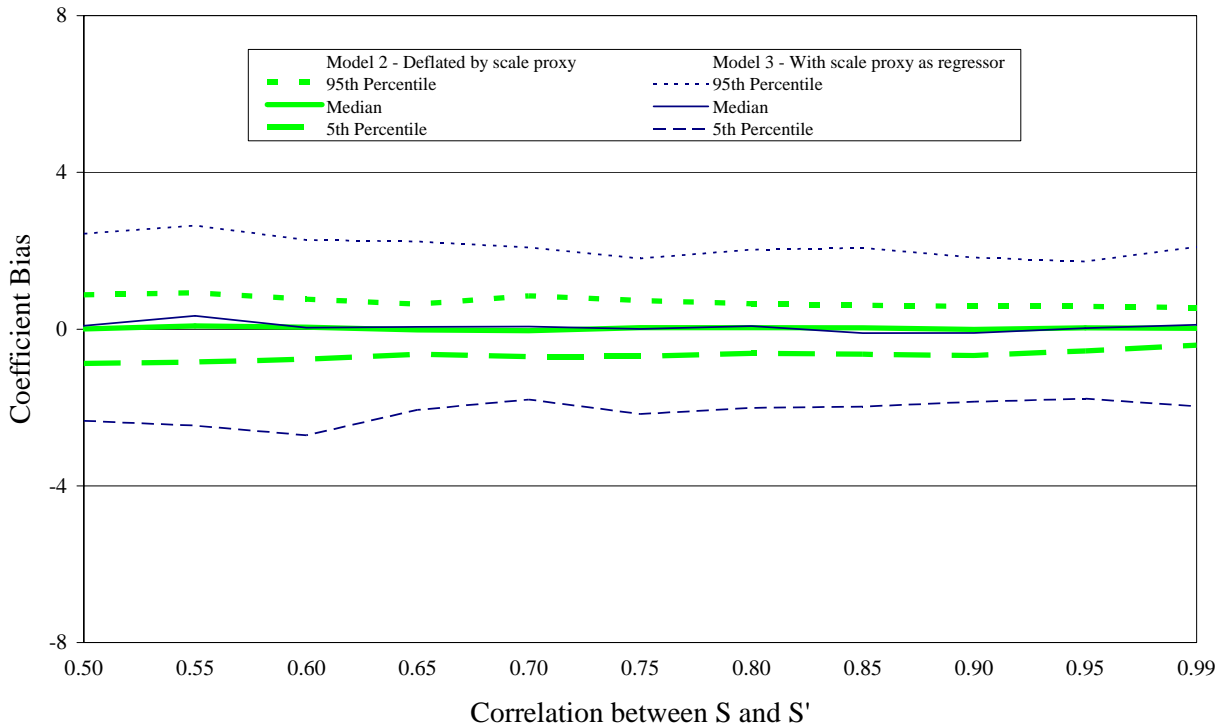


Figure 6a: Scatter plot of slope coefficients less true value of 7 from 250 repetitions at each of 20 weightings using Scheme A when intercept = 1500

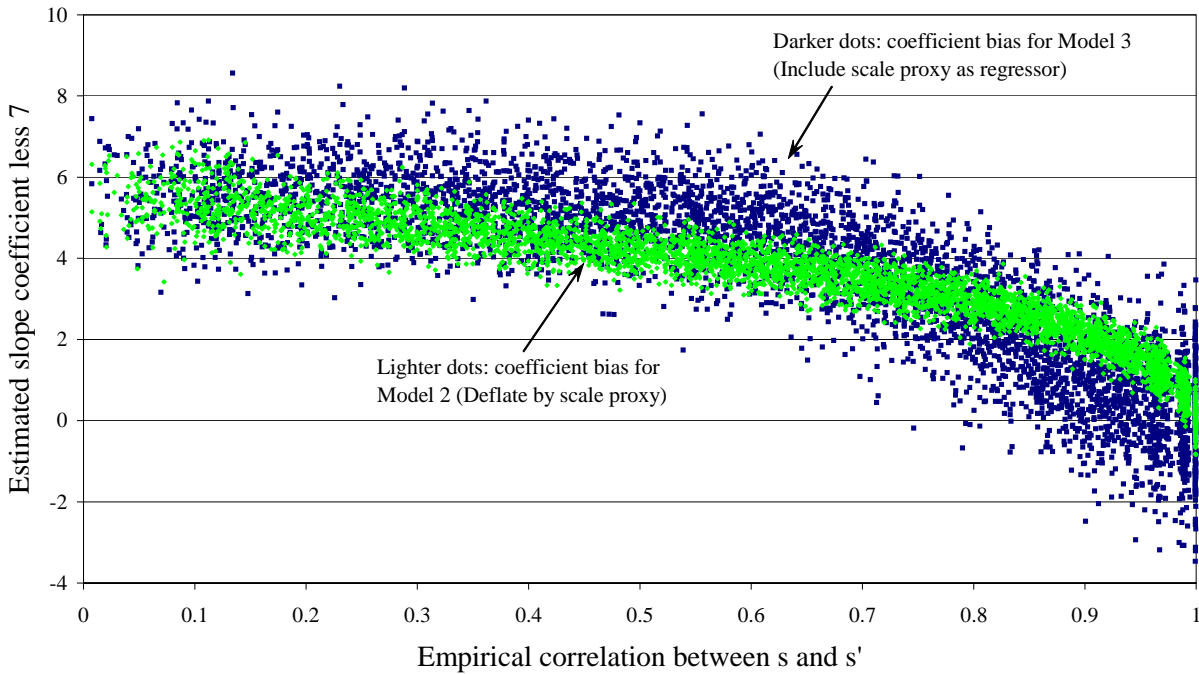


Figure 6b: Model 2 - Model 3 differences in squared errors of slope coefficients with s' computed using weighting Scheme A when intercept = 1500

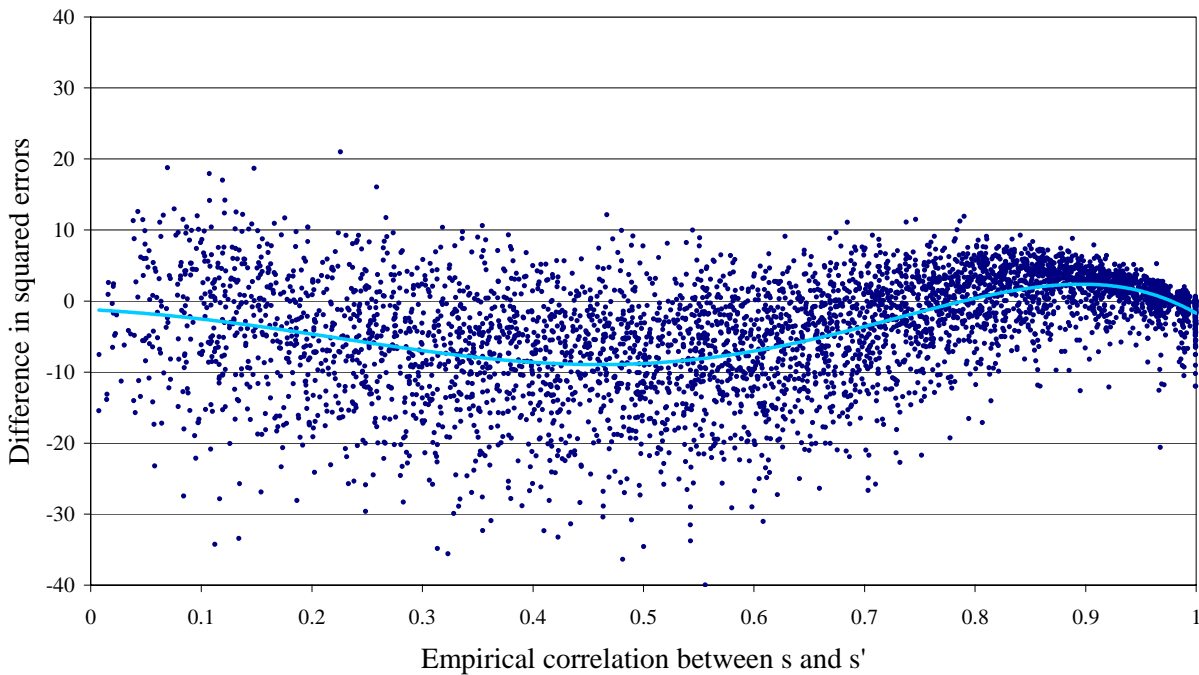


Figure 7a: Scatter plot of slope coefficients less true value of 7 from 250 repetitions at each of 20 weightings using Scheme A when intercept = 15

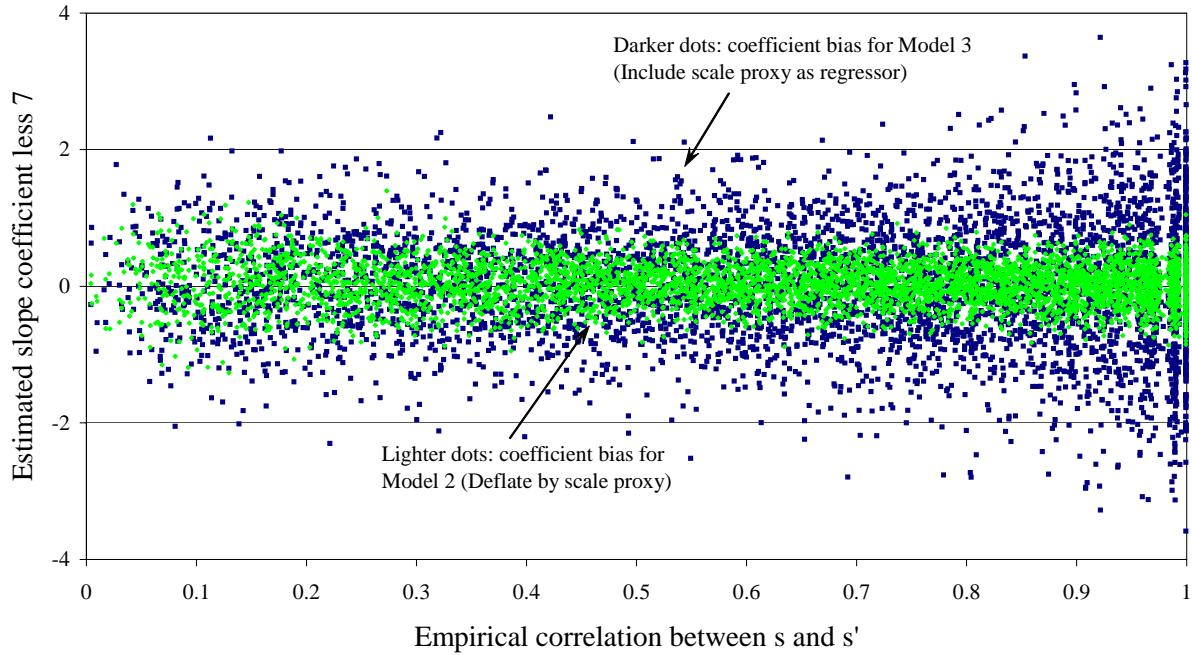


Figure 7b: Model 2 - Model 3 differences in squared errors of slope coefficients with s' computed using weighting Scheme A when intercept = 15

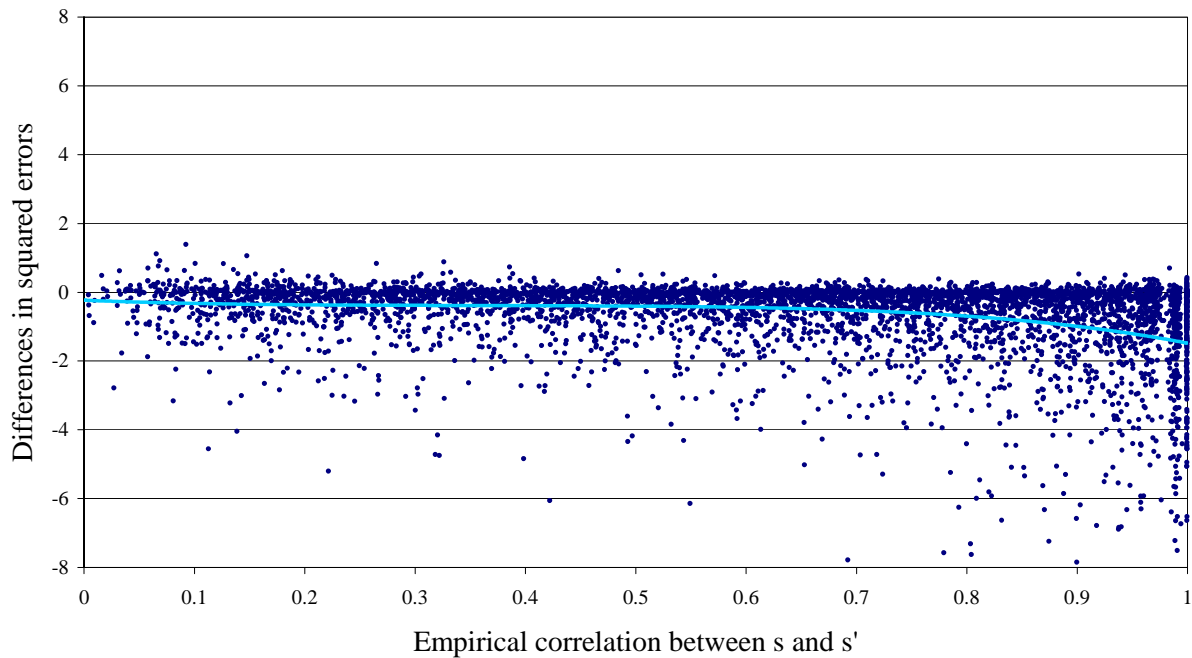


Figure 8a: Model 2 - Model 3 differences in squared errors of slope coefficients with s' computed using weighting Scheme B when intercept = 1500

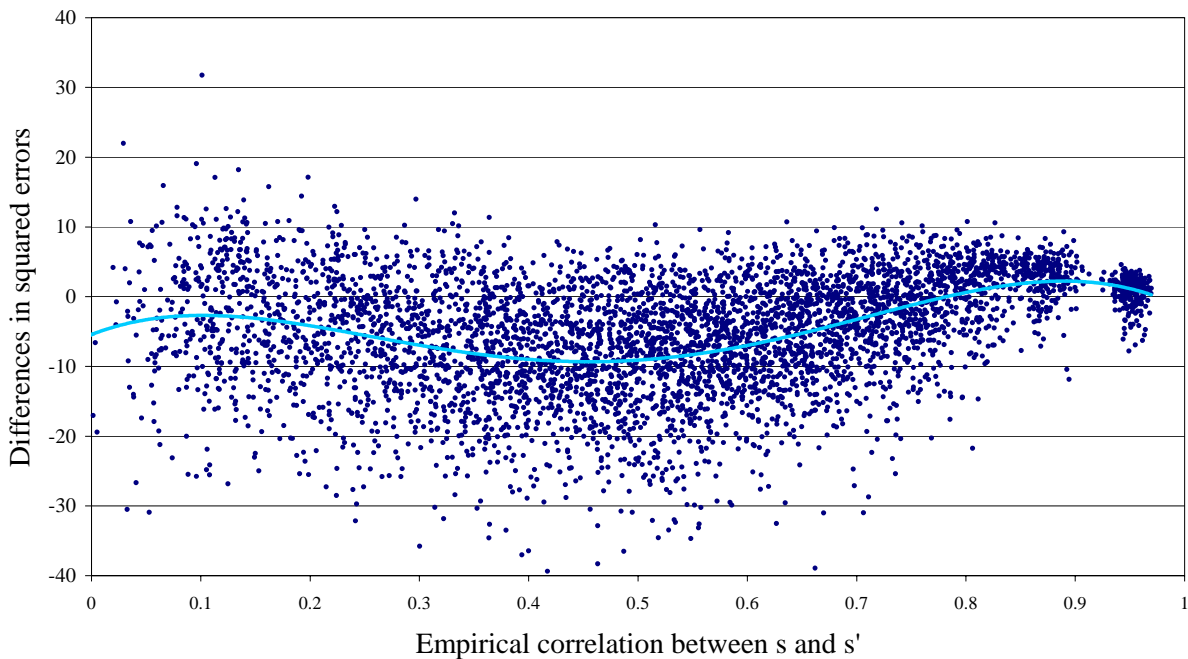


Figure 8b: Model 2 - Model 3 differences in squared errors of slope coefficients with s' computed using weighting Scheme B when intercept = 15

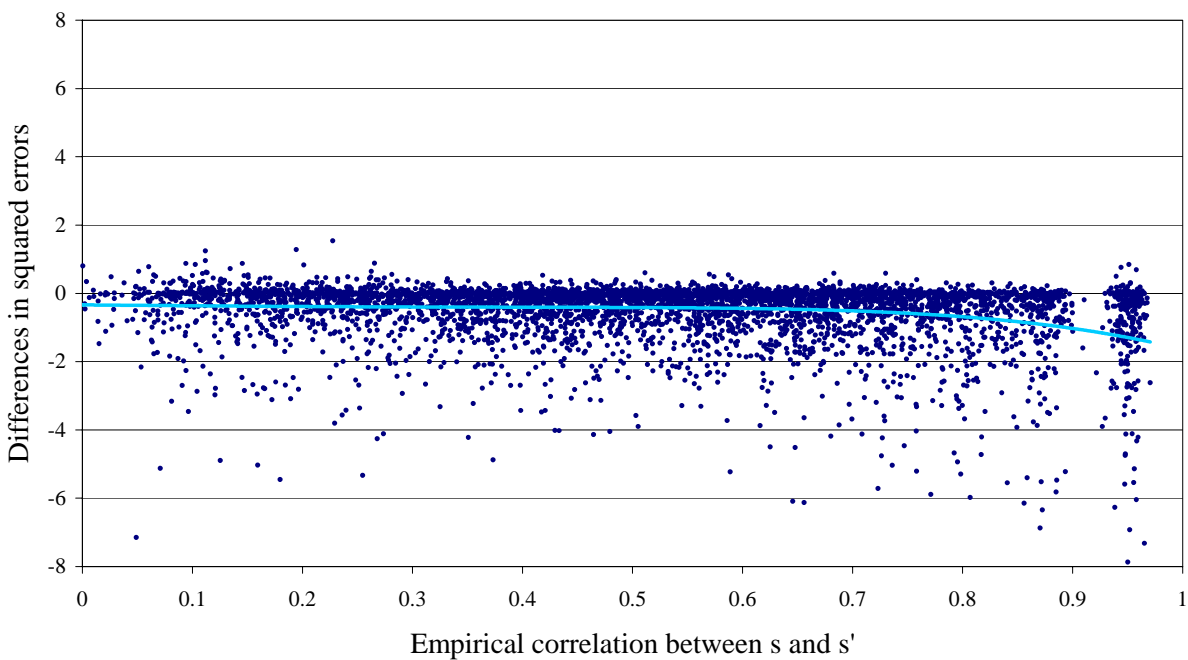


Figure 9a: Estimated coefficients for incremental variable from 250 iterations at each of 20 weightings using Scheme A when $\text{corr}(w_1, w_2) = 0$ and intercept = 1500

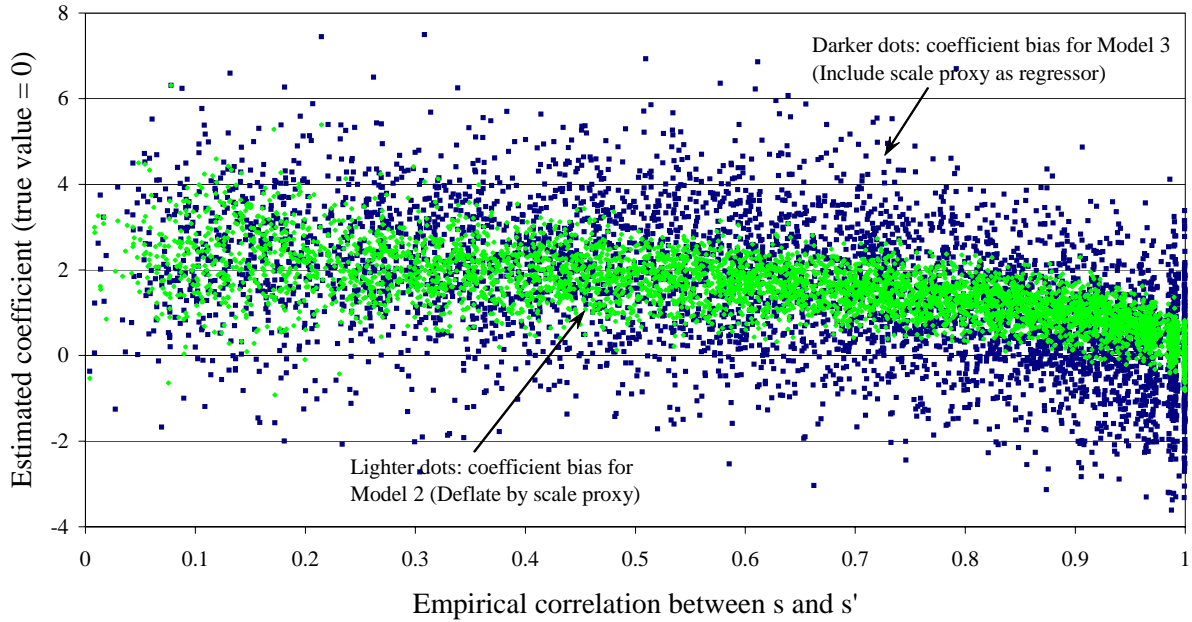


Figure 9b: Model 2 - Model 3 differences in mean squared error for 250 iterations at each of 20 values of lambda and 19 values of pi; intercept = 1500

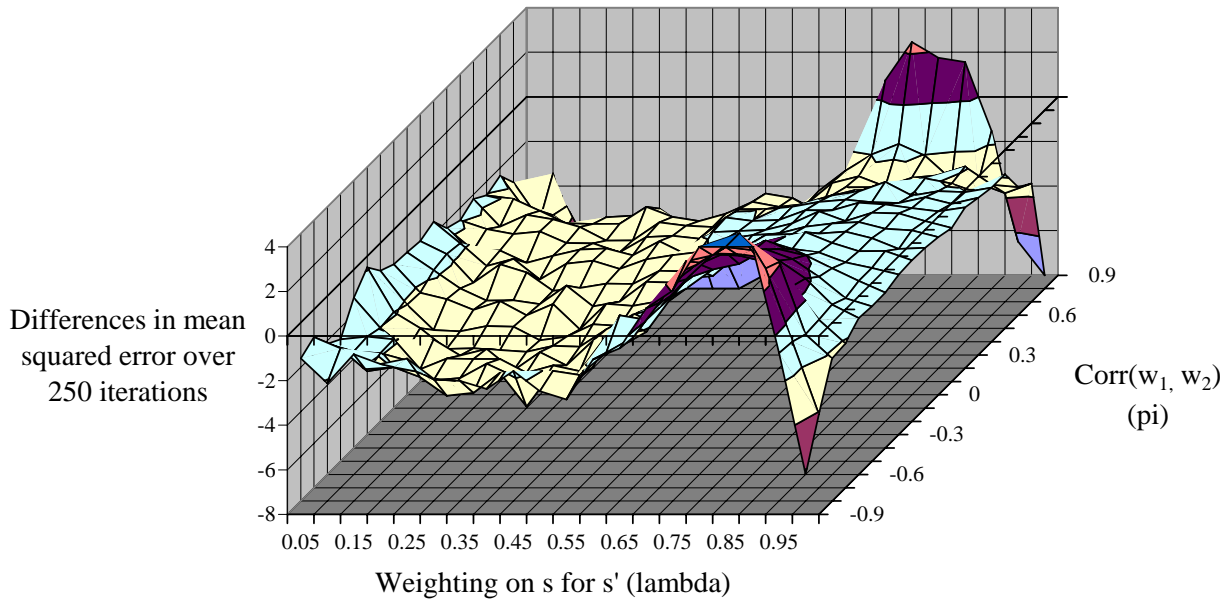


Figure 10a: Estimated coefficients for incremental variable from 250 iterations at each of 20 weightings using Scheme A when $\text{corr}(w_1, w_2) = 0$ and intercept = 15

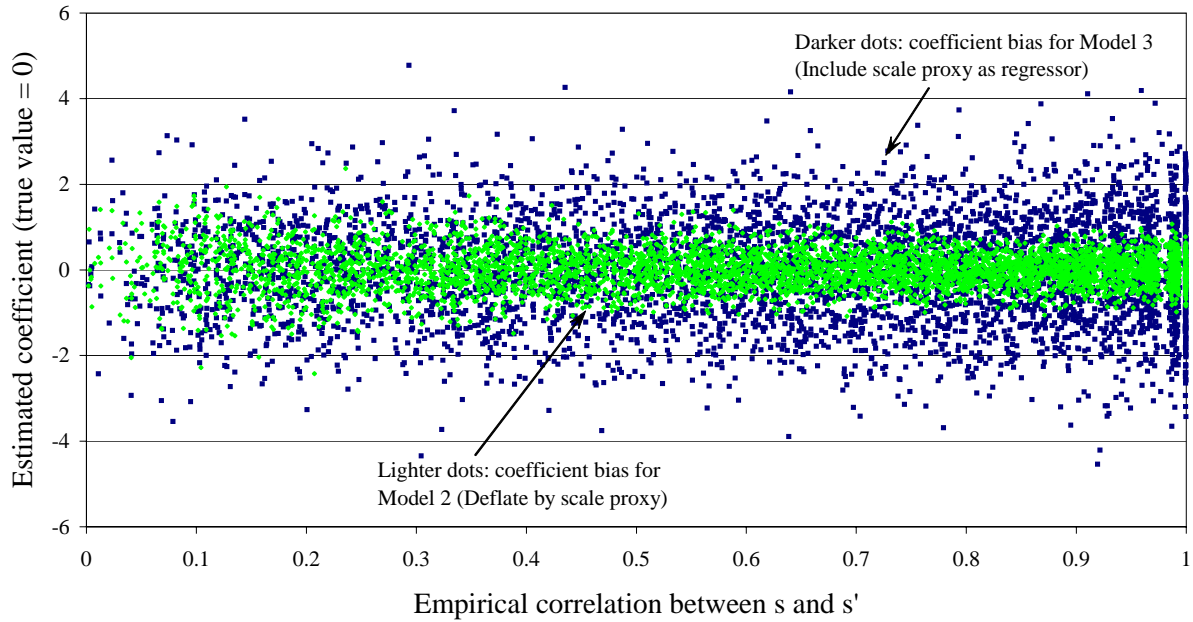


Figure 10b: Model 2 - Model 3 differences in mean squared error for 250 iterations at each of 20 values of lambda and 19 values of pi; intercept = 15

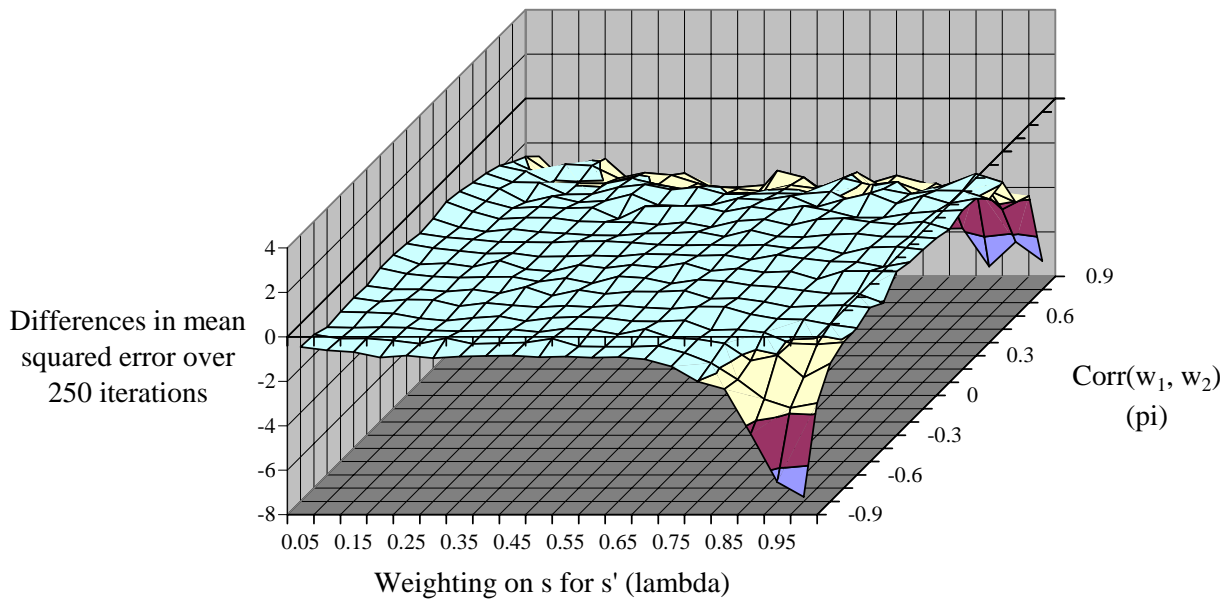


Table 1

Papers from *The Accounting Review*, 78(3) – 79(2), July 2003 – April 2004 for which variables are affected by scale.

Vol. (Issue)	Authors	Title, example regression specification, and approach of dealing with scale
78(3)	Eames and Glover	<p>“Earnings predictability and the direction of analysts’ earnings forecast errors”</p> <ul style="list-style-type: none"> ▪ Forecast Error = f(Earnings Unpredictability, Earnings, Firm Size, Value Line Timeliness Rank). Dependent variable is per share forecast error deflated by price at beginning of period.
	Swanson, Rees, and Juarez-Valdes	<p>“The contribution of fundamental analysis after a currency devaluation”</p> <ul style="list-style-type: none"> ▪ ΔEarnings = f(Pre-tax earnings, fundamental signals). Dependent variable is change in earnings per share deflated by price at beginning of period.
78(4)	Louis	<p>“The value relevance of the foreign translation adjustment”</p> <ul style="list-style-type: none"> ▪ Foreign Earnings_{t+1} = f(Foreign Earnings_t, Translation Adjustment_t). Variables are deflated by beginning market value.
79(1)	Gordon and Joos	<p>“Unrecognized deferred taxes: evidence from the U.K.”</p> <ul style="list-style-type: none"> ▪ Unrecognized Deferred Taxes = f(Operational Determinants, Opportunistic Determinants). Dependent variable is deflated by market value.
	Kallapur and Kwan	<p>“The value relevance and reliability of brand assets recognized by U.K. firms”</p> <ul style="list-style-type: none"> ▪ Market Value of Equity = f(Book Value of Equity, Earnings, Brand Assets, year and firm indicators). Variable are measured at firm level (not deflated).
79(2)	Aboody, Barth, and Kasznik	<p>“SFAS No. 123 stock-based compensation and equity market values”</p> <ul style="list-style-type: none"> ▪ Stock Price = f(Book Value of Equity, Earnings, Long-term Growth, Option Compensation Cost, industry indicators). Variables are per share values.
	Clarkson, Li, and Richardson	<p>“The market valuation of environmental capital expenditures by pulp and paper companies”</p> <ul style="list-style-type: none"> ▪ Market Value of Equity = f(Book Value of Equity, Abnormal Earnings, Environmental Capital Expenditures, High-pollution Indicator). Variables are measured at firm level and not deflated.

Table 2
Descriptive statistics

	Mean	Std. dev.	5 th %ile	Median	95 th %ile
Book value of equity (\$millions) (assumed scale factor)	3,393	4,729	234	1,978	11,570
Total assets (\$millions)	18,221	32,111	3,118	8,227	63,775
Net income before extraord. items (\$millions)	444	642	23	241	1,496
Number of shares (millions)	161	265	12	86	478
Stock price (\$)	50	298	9	30	87

Sample consists of 500 firms extracted from Compustat for the 1990 fiscal year. The top 500 firms ranked by total assets (annual item #6) with positive book value of equity (item #60) and net income before extraordinary items (item #18). Number of shares is item #25 and stock price is item #24.

Table 3

Distribution of Vuong's Z-statistic (VZ) for non-nested models

		$\theta = 0.50$					
		$\lambda = 0.50$			$\lambda = 0.95$		
One-tail p-value	Theoretical Cutoff (VZ*)	Model 1' vs. 1''	Model 2' vs. 2''	Model 3' vs. 3''	Model 1' vs. 1''	Model 2' vs. 2''	Model 3' vs. 3''
0.001	-3.090	0.0005	0.0011	0.0100	0.0010	0.0013	0.0027
0.010	-2.326	0.0098	0.0111	0.0319	0.0104	0.0111	0.0172
0.025	-1.960	0.0248	0.0254	0.0598	0.0268	0.0267	0.0408
0.050	-1.645	0.0505	0.0528	0.0951	0.0528	0.0529	0.0696
0.100	-1.282	0.1008	0.1006	0.1529	0.1081	0.1093	0.1352
0.500	0.000	0.4911	0.4920	0.5020	0.5019	0.5015	0.5061
0.100	1.282	0.1041	0.1052	0.1521	0.1004	0.1036	0.1368
0.050	1.645	0.0537	0.0557	0.0938	0.0493	0.0516	0.0753
0.025	1.960	0.0267	0.0280	0.0586	0.0256	0.0250	0.0428
0.010	2.326	0.0098	0.0124	0.0340	0.0099	0.0107	0.0188
0.001	3.090	0.0013	0.0013	0.0084	0.0017	0.0009	0.0027
Mean(VZ)		0.0097	0.0197	-0.0016	-0.0088	-0.0084	-0.0105
Standard deviation(VZ)		1.0067	1.0178	1.2722	1.0149	1.0193	1.1427
		$\theta = 0.95$					
		$\lambda = 0.50$			$\lambda = 0.95$		
One-tail p-value	Theoretical Cutoff (VZ*)	Model 1' vs. 1''	Model 2' vs. 2''	Model 3' vs. 3''	Model 1' vs. 1''	Model 2' vs. 2''	Model 3' vs. 3''
0.001	-3.090	0.0009	0.0008	0.0129	0.0017	0.0014	0.0049
0.010	-2.326	0.0106	0.0111	0.0370	0.0107	0.0131	0.0297
0.025	-1.960	0.0264	0.0259	0.0619	0.0266	0.0271	0.0608
0.050	-1.645	0.0522	0.0511	0.0994	0.0508	0.0518	0.0973
0.100	-1.282	0.1014	0.1030	0.1610	0.0971	0.1004	0.1541
0.500	0.000	0.4932	0.4958	0.5091	0.5002	0.5011	0.4986
0.100	1.282	0.1007	0.0979	0.1484	0.1018	0.1017	0.1553
0.050	1.645	0.0490	0.0473	0.0931	0.0490	0.0502	0.0918
0.025	1.960	0.0259	0.0243	0.0607	0.0244	0.0265	0.0553
0.010	2.326	0.0107	0.0100	0.0352	0.0092	0.0114	0.0304
0.001	3.090	0.0009	0.0004	0.0108	0.0012	0.0014	0.0055
Mean(VZ)		0.0047	-0.0058	-0.0258	0.0044	0.0019	0.0046
Standard deviation(VZ)		1.0044	0.9985	1.3021	1.0032	1.0135	1.2487

This table shows the empirical probabilities of obtaining a Z-statistic larger in magnitude than the theoretical cutoff for common p-values under the null hypothesis of equal explanatory power (Model 1' vs Model 1'', Model 2' vs Model 2'', Model 3' vs Model 3''). Distributions are

tabulated from 10,000 iterations at each of the four combinations of θ and λ , for a total of 40,000 iterations. Each iteration consists of three Vuong's Z-statistics, where each Z is computed from the residuals from two regressions (Model 1' and Model 1'', etc). Each regression has 500 observations. θ is the degree of correlation between \mathbf{w} and \mathbf{w}' and between \mathbf{w} and \mathbf{w}'' . λ is the weighting on \mathbf{s} in the construction of \mathbf{s}' . For assumption A2, $\alpha = 1500$.