"PRICE-LEVELS" REGRESSIONS: "SCALE-EFFECT" OR "DISTRIBUTION EFFECT"?

REGRESIONES CON NIVELES DE PRECIOS: ¿"EFECTO-ESCALA" O "EFECTO DISTRIBUCIÓN"?

PASCUAL GARRIDO MIRALLES*, University of Alicante, Spain
PABLO J. VÁZQUEZ VEIRA*, University of Alicante, Spain

ABSTRACT

Easton and Sommers (ES) (2003) document the existence of an overwhelming influence of large firms in 'price-levels' regressions on US data (as do Akbar and Stark (2003) on UK data). They refer to this overwhelming influence as the 'scale effect'. ES argue that the scale effect is caused by non-linearlities in the relationship between market value and the accounting variables. But non-linearities are only one possibility. We posit that the scale effect documented by ES is a pure econometric phenomenon. Typical variables used in Market-based Accounting Research (i.e. market value, book value, total assets, positive earnings, losses....) follow distributions that are very strong skewed, that is, distributions with a single, long tail. We argue that the scale effect is related to the presence of this large tail. When we apply a logarithmic transformation (as recommended by the literature in the case of highly skewed distributions, tending to restore normality), the 'scale effect' disappears.

KEY WORDS: 'Scale-Effect', 'Price-levels' Regressions, Lognormal Distribution, Logarithmic Transformation.
JEL Classification: M41, C10.

(*) We gratefully acknowledge the helpful comments and suggestions from Peter Joos and Stuart McLeay, and the workshop participants at the 2005 European Accounting Association Congress. *Contact information:* Pablo J. Vázquez Veira; Universidad de Alicante; Departamento de Economía Financiera, Contabilidad y Marketing; Ctra. San Vicente del Raspeig s/n; 03690 San Vicente del Raspeig (Alicante); Spain.

Enviado/Submitted: 27/10/2010
Primera revisión/First revision: 11/7/2011
Aceptado/Accepted: 6/10/ 2011

RESUMEN

Easton y Sommers (ES) (2003) demuestran que las observaciones de mayor tamaño (empresas grandes) sesgan, de manera recurrente, los resultados de las regresiones precio-magnitudes contables, para una muestra de datos norteamericanos. Akbar y Stark (2003) hallan los mismos resultados en el mercado del Reino Unido. ES (2003) argumentan que este efecto se debe a una relación no lineal entre las variables del estudio. Sin embargo, esa es sólo una posibilidad. Nosotros defendemos que el efecto escala que encuentran ES (2003) es un efecto puramente econométrico. Las variables comúnmente utilizadas en la investigación contable orientada hacia el mercados de capitales (por ejemplo, precio, fondos propios, beneficios, etc.) siguen distribuciones fuertemente asimétricas, esto es, distribuciones con una única cola muy alargada. A nuestro juicio, el efecto escala está relacionado con esta característica econométrica. Una vez aplicada una transformación logarítmica (tal y como recomienda la literatura en el caso de distribuciones asimétricas para restablecer normalidad), el efecto escala desaparece.

PALABRAS CLAVE: Efecto-escala, Regresiones con niveles de precios, Distribución lognormal, Transformación logarítmica.

1 INTRODUCTION

Easton and Sommers (2003) state:

> *"The influence of the largest observations, the scale effect, is not just heterocedasticity. Increasing variance of the regression residual with increases in scale (that is, heteroscedasticity) would be manifested in higher absolute values of studentized residuals for larger firms. But, if the relation between the variables for larger firms differs from the relation for smaller firms (that is, there is non-linearity in the data related to scale), we would also see groups tending to 'pull' the regression plane more in one direction than the other".*

On the other hand, Abbar and Stark (2003), in their discussion of Easton and Sommers (2003), posit that:

> *"If this (the scale effect) is caused by non-linearities in the relationship between market value and the accounting variables, then this, in turn, raises an interesting question. Is it fundamentally scale which is driving non-linearities? If so, why? If not, what is scale capturing?"*

Easton and Sommers (ES) (2003) document the existence of an overwhelming influence of large firms in 'price-levels' regressions on US data (as do Akbar and Stark (2003) on UK data). They refer to this overwhelming influence as the 'scale effect'. ES argue that the scale effect is caused by non-linearities in the relationship between market value and the accounting variables. But non-linearities are only one possibility. We posit that the scale effect documented by ES is a pure econometric phenomenon. Typical variables used in Market-based Accounting Research (i.e. market value, book value, total assets, positive earnings, losses....) follow distributions that are very strongly skewed, that is, distributions with a single, long tail. We argue that the scale effect is related to the presence of this long tail. When we apply a logarithmic transformation (as recommended by the literature in the case of highly skewed distributions, tending to restore normality), the 'scale effect' disappears.[1]

As in Barth and Kallapur (1996) and Easton and Sommers (2003), this study also focuses on scale. But our interpretation of scale is quite different from both. Nor a correlated omitted variable nor the idea that market capitalization is scale -with the underlying assumption that relation between variables for larger firms differs from the relation for smaller firms. As posted, we simply consider the scale effect to be an econometric phenomenon.

(1) Previous research has well documented a different relationship between price and earnings depending on whether the latter is negative or positive (Hayn, 1995; Burgstahler & Dichev, 1997; Collins et al, 1997; Collins et al. 1999; Joos & Plesko, 2005; among others). So, if we are interested in testing the value relevance of earnings we must distinguish between loss and profit firms. Since we have to isolate loss firms, we can take absolute values of losses, run regressions applying the logarithmic transformation and then multiply the coefficient estimate by minus one to obtain the right coefficient.

As Easton and Sommers (2003) demonstrated, once they deflate by market capitalization via a weighted least squares regression, the scale effect is removed. Nevertheless, this methodology could lead to misleading results. As all observations are weighted with a weight equal to the inverse of the square of market capitalization, large firms will received very low weight in comparison to small firms. We are also losing information. Results will be driven by the set of small firms. Applying data transformation, we can still use the classical least squares specification. Also, under the classical framework, all observations are equally weighted so no information is wasted. Furthermore, as long as we are interested in applying robust estimation techniques (based in some kind of weighted least regression), the possibility of using weights remains available.

Regarding the interpretation of coefficients estimates when a logarithmic transformation is applied, we rely on the option-style valuation model developed by Burgstahler and Dichev (1997).

The remainder of the paper is organized as follows. Section 2 reviews the notion of scale effect (as described by Easton and Sommers (2003)), and provides an example with simulated data where the scale effect is caused by non-linearities in the relationship between variables. Section 3 matches the scale effect with the lognormal distribution and provides two examples where the scale effect also appears even if variables are independent. Section 4 defends the logarithmic transformation as an alternative for addressing the scale effect. Using US data, section 5 confirms the existence of scale effect when a standard 'price level' regression is run ≡market value against earnings and book value ≡and shows how this effect is removed when a logarithmic transformation is applied. Relying on the Burgstahler and Dichev (1997) option-style valuation model, section 6 discusses the economic interpretation of coefficients estimates obtained in section 5. Section 7 presents conclusions and implications for future research.

2 SCALE EFFECT UNDER NON-LINEARITIES

Easton and Sommers (ES) (2003) document the overwhelming influence of large firms in 'price-levels' regressions US data (as do Akbar and Stark (2003) on UK data[2]). They refer to this overwhelming influence as the scale effect. This prevents the researcher from obtaining unbiased coefficient estimates. Removal of large firms does not remove the coefficient bias caused by the scale effect. In the sample that remains, the group of firms with the largest scale (larger market capitalization) becomes the group of observations exerting significant influence. They also show that it takes the removal of approximately the top 60% of market value observations before this effect disappears.

(2) We decline to use the Barth and Clinch (2009) approach since they do not use real data but simulate. Regarding Wu and Xu (2008), and following Barth and Kallapur (1996), we are not interested in deflators since the use of them could potentially exacerbate scale bias.

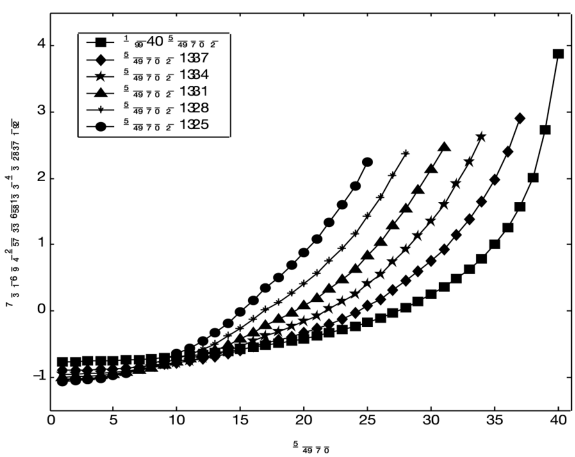
ES argue that the scale effect is caused by non-linearties in the relationship between market value and accounting variables. Figure 1 shows this possibility. Consider two variables (x, y) with a non-linear relationship: variable x is a (0,1) normal random variable and variable y equals x^2. We extract 5,000 observations from variable x and analyze the presence of the scale effect estimating regression (1):

$$y_i = \alpha_0 + \alpha_1 \times x_i + u_{1i} \tag{1}$$

Following ES, we demonstrate the influence of scale in regression (1) using studentized residuals, which are measures of the influence of individual observations. After studentized residuals are obtained from each observation, we form 40 groups based on y values (group 1 being the smallest group and group 40 the largest). For each group, a mean of the studentized residuals statistics is calculated. The graph labeled "all 40 groups" in Figure 1 is the plot of these means (on the y-axis) against the variable y group on the x-axis. In order to analyze the overwhelming influence of "large" (in y) observations in regression (1), we then delete the 3 groups of observations with higher values of y, re-estimate the regression, and re-calculate the mean of studentized residuals for each of the remaining groups. These data are labeled in Figure 1 according to the number of groups remaining.

FIGURE 1

Coefficient Bias in OLS Regression $y_i = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_i$ –variable x is a (0,1) normal random variable and variable y equals x^2– After Progressively Deleting Firms with Largest y
(The bias is measured using means of studentized residuals for groups formed on y)



Notes:
Regression (1) is conducted and the studentized residual is calculated for each observation. We then form 40 groups based on y (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression and divided by the standard error of the regression. The graph labeled "All 40 Groups" is the plot of these means (on the y-axis) against the y group on the x-axis. Then we delete the three groups of observations with the largest y, re-estimate the regression, and re-calculate the mean of the studentized residuals for each of the remaining groups. These data are labeled above according to the number of groups remaining.

Considering the full sample (all 40 groups), the mean of the studentized residuals is between -0.77 and 0.79 for each of the 34 groups with the smallest values of variable y. It then increases to 2.01 for the group with the third-to-largest y value, 2.73 for the group with the second-to-largest y value, and 3.87 for the group of observations with the largest y. This indicates that observations with large values of y exert undue influence on regression (1). When deleting the fortieth group we find that the mean of studentized residuals for the largest remaining group increases to 3.39, which reflects its significant influence on regression estimates. After 3 rounds of deletions, the largest remaining group now has a mean of studentized residuals of 2.91 (labeled "Groups 1-37" in Figure 1).

To achieve a mean of studentized residuals under 1.96 in the largest remaining group, 29 groups must eliminated. This is exactly the same phenomenon described by Easton and Sommers (2003) using real data.

3 SCALE EFFECT AND LOGNORMAL DISTRIBUTION

As we have pointed out, Easton and Sommers (2003) argue that the scale effect is caused by non-linearities in the relationship between variables. But non-linearities are only one possibility. We argue that the scale effect could also appear when at least one of the variables follows a lognormal distribution, even if variables are independent. The idea behind the scale effect is the overwhelming influence of outliers, that is, observations far from the majority of the data with a rather large influence on the regression plane. In Market Based Accounting Research, researchers often tackle extremes by deleting the top and bottom 1%, 2% or 3% of the distribution of all regression variables. But this solution will work in a short-tailed distribution. If the distribution is lognormal[3] (distributions with a single, long tail), the classical removal procedure will fail.

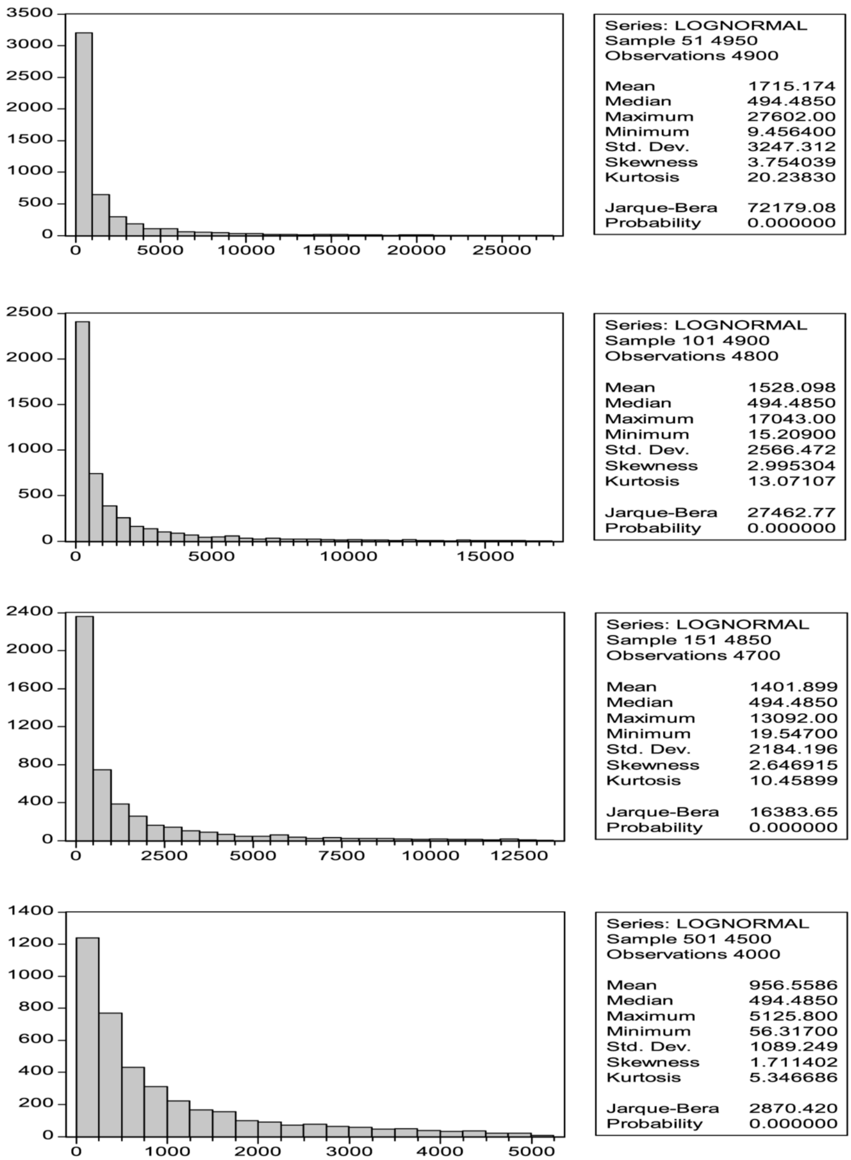
As an example, consider the distribution of a lognormal random with mean 6.25 and standard deviation 1.75[4]. From Figure 2 it is clear that classical removal procedure is not a feasible solution for the problem of outliers. The results of deleting the top 1% of the distribution are reported in Panel A. The distribution is still strongly skewed to the right (skewness 3.75, very far from zero, the normal case) –around 3% of the date rely out of interval [$\mu \pm 3 \cdot \sigma$], all of them on the right side. Deleting the top 2%, 3% or even 10% of the distribution (Panels B, C and D, respectively) does not solve the problem of outliers.

(3) In Economics, lognormal distribution is typically used for representing distributions of some economic variables (i.e. income).

(4) We have observed that a lognormal distribution with those parameters fits our data quite well. However, real values of those parameters are unobservable, and we are only interested here in pointing out the implications of working with data that follows any lognormal distribution.

FIGURE 2

Distribution and Descriptive Statistics of a (6.25, 1.75) lognormal random variable (n=5000) after deleting the top and bottom 1%, 2%, 3% and 10%; respectively



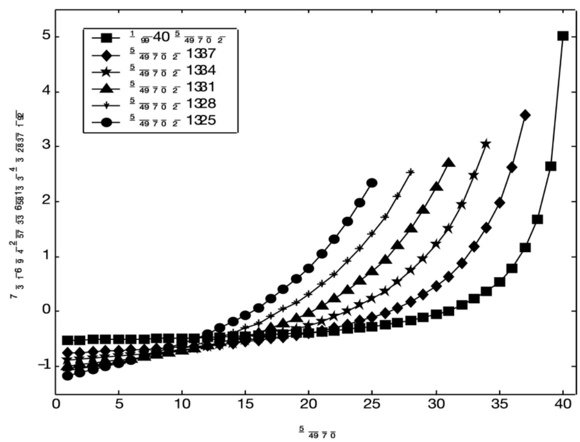
We argue that scale effect also appears if at least one of the regression variables follows a lognormal distribution. Relying on this argument, we are able to simulate a scale effect even if variables are independent. Figure 3 shows this possibility. Figure 3 is constructed in the same way as Figure 1, but in this case variables are independent. We run the following regression:

$$Logn_i = \alpha_0 + \alpha_1 \times x_i + u_{2i} \tag{2}$$

Variable x remains the same, that is, a (0,1) normal random variable, while *Logn* is a lognormal random variable with mean 6.25 and standard deviation equal to 1.75. As in previous simulations, we extract 5,000 observations from both variables[5].

FIGURE 3

Coefficient Bias in OLS Regression $Logn_i = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_i$ − x is a (0,1) normal random variable and *Logn* is a (6.25,1.75) lognormal random variable− After Progressively Deleting Firms with Largest *Logn*
(The bias is measured using means of studentized residuals for groups formed on *Logn*)



Notes:
Regression (1) is conducted and the studentized residual is calculated for each observation. We then form 40 groups based on y (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression and divided by the standard error of the regression. The graph labeled "All 40 Groups" is the plot of these means (on the y-axis) against the y group on the x-axis. Then we delete the three groups of observations with the largest y, re-estimate the regression, and re-calculate the mean of the studentized residuals for each of the remaining groups. These data are labeled above according to the number of groups remaining.

(5) We choose these parameters because they seem to approximate quite well market value annual distributions for US firms (period 1992-2002).

Considering the full sample[6] (all 40 groups), the mean of the studentized residuals is between -0.52 and 0.79 for each of the 36 groups with smallest values of y. It then increases to 1.16 for the groups with the fourth-to-largest values of y, 1.68 for the groups with the third-to-largest values of y, 2.64 for the groups with the second-to-largest values of y, and 5.02 for the groups of observations with the largest y. After 3 rounds of deletions, the largest remaining groups now have a mean of studentized residuals of 3.57 (labeled "groups 1-37" in Figure 3). To achieve a mean of studentized residuals under 1.96 in the largest remaining groups, 24 groups must be eliminated. It is clear that Figures 1 and 3 are describing the same phenomena.

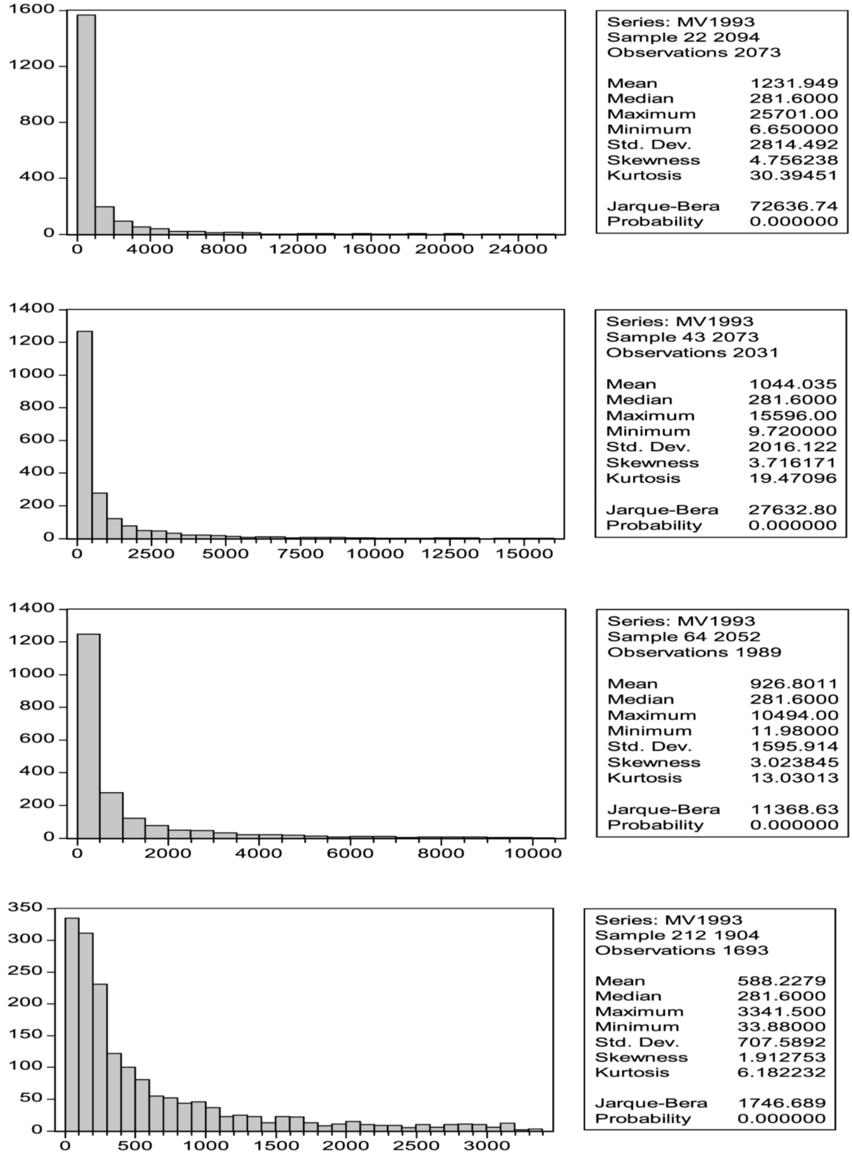
We have shown that the scale effect could also appear when even one of the regression variables follows a lognormal distribution, even if variables are independent. But the question is: can we match this finding with real data? Or stated differently, in "price level" regressions, does any variable follow a lognormal distribution?

Figure 4 examines the distribution of market capitalization for US firms (year 1993, n=2,115)[7]. As in Figure 2, it is clear that classical removal procedures (the top and bottom 1%, 2% or 3% of regression variables distribution) are not a solution to the problem of outliers. Even if we delete the top and bottom 10% of the distribution, the distribution is strongly right skewed. Analyzing the right tail, we find 54 observations with marker capitalization higher than $[\mu+3\cdot\sigma]$, that is, observations that could be considered as outliers in a normal distribution.

(6) After deleting the top and bottom 1% of the distribution of both variables.
(7) Empirical findings presented in subsequent sections are based on annual regressions. The same pattern is observed when considering the full sample or different years.

FIGURE 4

Distribution and Descriptive Statistics of Market Capitalization (year 1993; n=2115)
after deleting the top and bottom 1%, 2%, 3% and 10%; respectively



Now the second question is: could Easton and Sommers' (2003) results be driven by market capitalization distribution? Or, stated differently, is the scale effect a pure econometric phenomenon?

In order to address this possibility, we run the following regression:

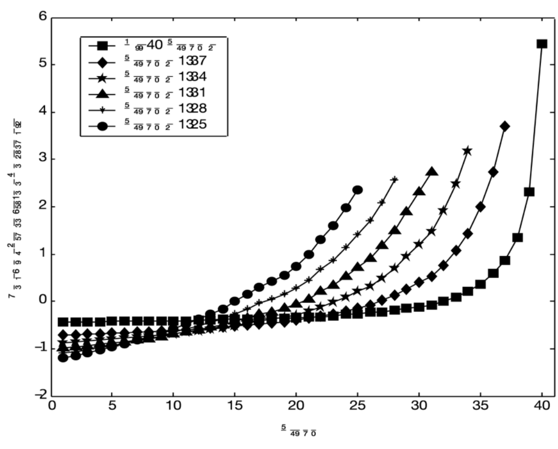
$$MC_{1993} = \alpha_0 + \alpha_1 \times x_i + u_{3i} \tag{3}$$

where MC_{1993} is the market capitalization for US firms (year 1993, n=2115); x_i is a (0, 1) normal random variable and ε the error term.

Figure 5 provides evidence for the existence of scale effect in regression (3). Figure 5 is constructed in the same way as Figures 1 and 3. Considering all observations[8] (labeled "all 40 groups" in Figure 5), the mean of the studentized residuals is between -0.43 and 0.58 for each of the 36 groups with the smallest market capitalization. It then increases to 0.86 for the groups with the fourth-to-largest market capitalization, 1.35 for the groups with the third-to-largest market capitalization, 2.31 for the groups with the second-to-largest market capitalization, and 5.44 for the group of observation with the largest market capitalization. Alter 3 rounds of deletions, the largest remaining group now has a mean of studentized residuals of 4.43. Again, to achieve a mean of studentized residuals under 1.96 in the largest group, 21 groups (more than 50% of the data) must be eliminated. Easton and Sommers (2003) would describe this phenomenon, as the scale effect.

FIGURE 5

Coefficient Bias in OLS Regression $MV_{1993} = \alpha_0 + \alpha_1 \cdot x_i + e_i$ –variable x is a (0,1) nomal random variable and MV_{1993} is Market Capitalization (year 1993, n=2115)–After Progressively Deleting Firms with Largest MV_{1993}
(The bias is measured using means of studentized residuals for groups formed on MV_{1993})



Notes:
Regression (2) is conducted and the studentized residual is calculated for each observation. We then form 40 groups based on MV_{1993} (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression and divided by the standard error of the regression. The graph labeled "All 40 Groups" is the plot of these means (on the y-axis) against the MV_{1993} group on the x-axis. Then we delete the three groups of observations with the largest y, re-estimate the regression, and re-calculate the mean of the studentized residuals for each of the remaining groups. These data are labeled above according to the number of groups remaining.

(8) Deleting the top and bottom 1% of the distribution of both variables.

4 ALTERNATIVES FOR ADDRESSING THE SCALE EFFECT: WEIGHTED LEAST SQUARES VS. LOGARITHMIC TRANSFORMATION

As Easton and Sommers (2003) pointed out, results of the regression of market capitalization on financial statement data are driven by a relatively small subset of the very largest firms in the sample. They refer to the overwhelming influence of the largest firms as the scale effect. Their argument that scale *is* market capitalization leads them to deflate the regression by market capitalization via a weighted least squares regression (WLS).

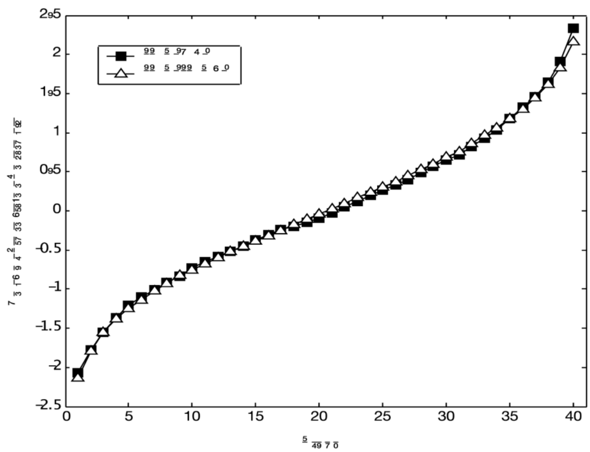
Nonetheless, if the scale effect is related to the lognormal distribution of one of the variables that enter in the regression, there is an alternative way to address the scale effect: *logarithmic transformation*. If we apply a logarithmic transformation to a variable that follows a lognormal distribution we will obtain a variable that is normally distributed. The presence of extreme values is drastically reduced. The long tail, typical of lognormal distributions, disappears. The transformed variable will not cause any scale effect.

Figure 6 confirms this intuition. We rerun regressions (2) and (3)[9] but now apply a logarithmic transformation to the dependent variable (a lognormal random variable and market capitalization, respectively). As in the previous example, we calculate the studentized residuals for each observation and group observations into forty groups based on the value of the dependent variable. Then we calculate the mean of studentized residuals for each size partition and plot results. The scale effect, as described by Easton and Sommers (2003), disappears. In both cases, the mean of studentized residuals increases monotonically, in a range that goes from -2.5 to 2.5. That is, small and large observations (identifying the dependent variable with size) exert the same influence on regression.

(9) Deleting the top and bottom 1% of the distribution of both variables.

FIGURE 6

Coefficient Bias in OLS Regressions (2) and (3) After Applying a
Logarithmic Transformation to the Dependent Variable
(The bias is measured using means of studentized residuals for groups formed on the value of the dependent variable)



Notes:
Regression (2): $Log(Logn_i) = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_i$; variable x is a (0,1) normal random variable and *Logn* is a (6.25,1.75) lognormal random variable (n=5000).
Regression (3): $Log(MV_{1993}) = \alpha_0 + \alpha_1 \cdot x_i + \varepsilon_i$; variable x is a (0,1) normal random variable and MV_{1993} is Market Capitalization for US firms (year 1993, n=2115).
In both cases, regression is conducted and the studentized residual is calculated for each observation. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression divided by the standard error of the regression. We then form 40 groups based on dependent variable (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. Graphs labeled "LOG(LOGN)" and "LOG(MV)" are the plot of these means (on the y-axis) against the dependent variable group on the x-axis for regressions (2) and (3), respectively.

We defend the superiority of the logarithmic transformation against the WLS procedure. A weighted regression involves giving a weight to each observation. ES (2003) propose the inverse of market capitalization as the appropriate weight. By doing so, they force large firms to enter the regression with less weight. The scale effect disappears but small firms will drive results; to some extent we are losing information. Applying a logarithmic transformation all observations are equally weighted[10], so no information is wasted. Furthermore, as long as we are interested in applying robust estimation techniques (based in some kind of weighted least regression), the possibility of using weights remains available.

(10) As an illustration, consider the following example. Assume two observations with market capitalization equal to 10 and 100. Easton and Sommers (2003) weighted regression will impound weights equal to $1/10^2$ and $1/100^2$, respectively. Thus the first observation receives a weight that is 100 times the weight impounded to the second one. If we apply logarithmic transformation, both observations will be equally weighted.

5 EMPIRICAL RESULTS

In this section, we first replicate Easton and Sommers' (2003) results concerning the existence of the scale effect when annual "price level" regressions are run using US data[11]. Secondly, we show how this effect is removed when applying a logarithmic transformation.

5.1 Scale effects in "price-levels" regressions (US data)

As in Easton and Sommers (2003), we run the following annual cross-sectional regressions:

$$MC_{it} = \alpha_0 + \alpha_1 \times BV_{it} + \alpha_2 \times NI_{it} + u_{4it} \tag{4}$$

where:

MC_{it} = is the market capitalization (price per share times number of shares outstanding) for firm *i* at the end of period *t*,

BV_{it} = is the book value of common equity reported in the balance sheet of firm *i* at the end of period *t*,

NI_{it} = is net income reported in the income statement firm *i* for the period *t-1* to *t*,

These regressions are conducted for each of the years 1993 to 2003. The sample consists of all US firms available on Compustat (Global Vantage) for the selected period. Banks and financial services firms and firms with negative book value were deleted. We also exclude firms with negative earnings because logarithmic transformation does not apply to negative numbers.[12] The results of these regressions are reported in Table 1.

Comparison of the results presented in Table 1 (undeflated data) with those reported by Easton and Sommers (2003) shows three main differences. Looking at the period 1993-1999[13], we found a lower coefficient on book value for each year. In contrast, we found a higher coefficient on earnings for each year. The two differences are well explained by the absence of loss firms in our sample. Following the "Abandonment hypothesis" (Hayn, 1995; Berger et al., 1996; Bursgtahler and Dichev, 1997), firms with negative earnings are more likely to liquidate cease operations. In these types of firms, book value (acting as a proxy for liquidation value) will be the only relevant variable.

(11) It is usual in Market-based Accounting Research to delete banks, financial services and negative book values (Burgsthaler, D. and I. Dichev, 1997; Fama and French, 2002; among others). However, as an anonymous reviewer points out, Easton and Sommers (2003) do not make this deletion explicit, so we cannot compare results. Nonetheless, given the small percentage of this type of firms in relation to the whole sample, we argue that our findings are still robust.

(12) Nevertheless, if we are interested in loss firms, one possibility arises: isolating the firms and taking logs of absolute values.

(13) 1993-1999 is the period where the two works match.

TABLA 1.- ANALYSES OF ANNUAL OLS REGRESSION OF MARKET CAPITALIZATION ON BOOK VALUE AND NET INCOME

$$\text{Regression: } MC_{it} = \alpha_0 + \alpha_1 \cdot BV_{it} + \alpha_2 \cdot NI_{it} + \varepsilon_{it} \tag{4}$$

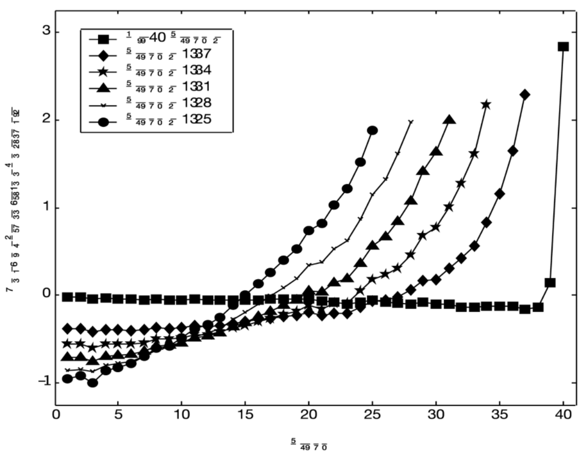
Year	*Coefficients (t-statistics)*			*Adjusted* R^2*%*	*N*
	α_0	α_1	α_2		
1993	17.46	0.88	12.52	0.83	2,037
	(0.66)	(18.68)	(35.74)		
1994	74.10	0.27	12.85	0.86	2,213
	(3.27)	(6.80)	(50.87)		
1995	43.91	0.67	12.20	0.80	2,212
	(1.32)	(11.78)	(36.16)		
1996	138.17	0.41	14.13	0.81	2,293
	(3.80)	(7.62)	(44.88)		
1997	-22.96	1.12	14.85	0.76	2,175
	(-0.41)	(17.22)	(37.48)		
1998	-142.07	1.02	18.01	0.67	2,028
	(-1.68)	(10.88)	(30.98)		
1999	-253.05	0.98	20.10	0.62	2,009
	(-2.01)	(8.41)	(28.75)		
2000	-71.21	0.69	18.38	0.68	1,837
	(-0.59)	(6.64)	(31.09)		
2001	-125.26	0.75	19.09	0.73	1,486
	(-1.03)	(8.76)	(34.60)		
2002	-45.31	0.19	17.94	0.83	1,510
	(-0.59)	(3.26)	(50.73)		
2003	32.34	0.63	16.70	0.79	1,540
	(0.29)	(9.39)	(38.20)		
Mean	-32.17	0.69	16.07		

Notes:
MC_{it} is the market capitalization (price per share times number of shares outstanding) for firm i at the end of period t,
BV_{it} is the book value of common equity reported in the balance sheet of firm i at the end of period t,
NI_{it} is net income reported in the income statement firm i for the period t-1 to t,
N is the number of observations.

In order to demonstrate the existence of a scale effect, we follow Easton and Sommers (2003) procedure. We form 40 groups within each set of annual observations based on market capitalization. The mean of the studentized residuals are then calculated for each group within each year. Means across the 11 years of data are then computed for each of the 40 groups. Finally, we repeatedly estimate equation (4) each time deleting the largest remaining group from the 40 groups we started with. Figure 7 presents the results after every third round of deletion. Considering the full sample (plot labeled "all 40 [14]groups") the mean of annual means of studentized residuals is between -0.15 and 0.14 for all groups except group 40. It then increases to 2.83 for the groups with the largest market capitalization. After 3 rounds of deletions, the largest remaining group must not have a mean of annual means of studentized residuals greater than 1.96, so 13 groups must be eliminated (this removes 32.5% of the data). As Easton and Sommers (2003) state: "clearly, simply deleting the largest firms is not a feasible solution to the problem of the scale effect".

FIGURE 7

Coefficient Bias in OLS Regression of Market Capitalization on
Book Value and Net Income After Progressively Deleting Firms
with Largest Market Capitalization
(The bias is measured using means of studentized residuals for groups formed on Market Capitalization)



Notes:
Regression (4) is conducted and the studentized residual is calculated for each observation. We then form 40 groups based on market capitalization (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression and divided by the standard error of the regression. The graph labeled "All 40 Groups" is the plot of these means (on the y-axis) against the market capitalization group on the x-axis. Then we delete the three groups of observations with the largest y, re-estimate the regression, and re-calculate the mean of the studentized residuals for each of the remaining groups. These data are labeled above according to the number of groups remaining.

(14) We do not report book value and earnings descriptive statistics, because they are qualitatively the same as those reported for market capitalization (Figure 4)

5.2. Removal of the scale effect when a logarithmic transformation is applied

Following the arguments presented in section 4, we hypothesize the scale effect will disappear if we apply a logarithmic transformation to all variables in equation (4). Those variables ≡market capitalization, earnings and book value ≡seem to follow distributions that approximate quite well to a lognormal one. Taking logs, distributions will reverse closely to a normal distribution[15] avoiding econometric problems related to strongly skewed distributions (i.e. the scale effect) and making the classical removal procedure a feasible solution to the problem of outliers ≡potential influential observations[16].

Applying a logarithmic transformation to Equation (4) yields the following regression:

$$Log\left(MC_{it}\right)=\alpha_0+\alpha_1\times Log\left(BV_{it}\right)+\alpha_2\times Log\left(NI_{it}\right)+u_{5it} \qquad (5)$$

Results from estimating Equation (5) are presented in Table 2. Before comparing results with those from Table 1 (untransformed data), we first analyze whether the scale effect has been mitigated through data transformation.

As in previous subsection, we first run annual regressions (5) and calculate studentized residuals for each observation. We form 40 groups within each set of annual observation based on market capitalization. The mean of the studentized residuals and also the mean of the absolute values of the studentized residuals (to make our results comparable with those from Easton and Sommers (2003)), are then calculated for each group within each year. Mean across the 11 years of data are then computed for each of the 40 groups.

(15) We do not posit that the selected variables are lognormal, however they are quite close to it. So if we take logs, we will obtain distributions quite similar to the normal one (results, not shown, confirm this aspect).
(16) Previous research has well documented a different relationship between price and earnings depending on whether the latter is negative or positive (Hayn, 1995; Burgstahler & Dichev, 1997; Collins et al, 1997; Collins et al. 1999; Joos & Plesko, 2005; among others). So, if we are interested in testing the value relevance of earnings we must distinguish between loss and profit firms. Since we have to isolate loss firms, we can take absolute values of losses, run regressions applying the logarithmic transformation and then multiply by minus one the coefficient estimate in order to obtain the right coefficient.

TABLA 2.- ANALYSIS OF ANNUAL OLS REGRESSION OF MARKET CAPITALIZATION ON BOOK VALUE AND NET INCOME AFTER APPLYING A LOGARITHMIC TRANSFORMATION

Regression: $Log(MC_{it}) = \alpha_0 + \alpha_1 \cdot Log(BV_{it}) + \alpha_2 \cdot Log(NI_{it}) + \varepsilon_{it}$ (5

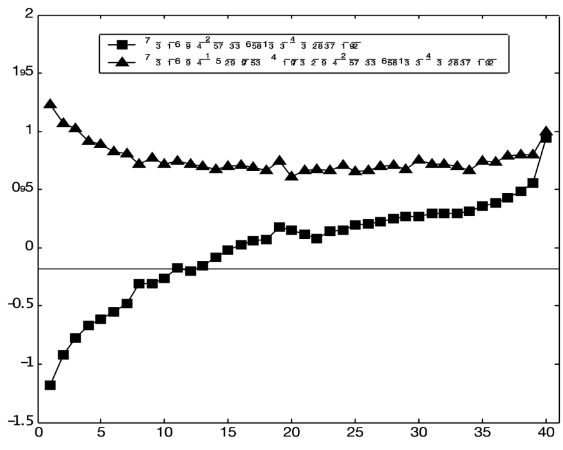
Year	*Coefficients (t-statistics)*			Adjusted R^2%	*N*
	α_0	α_1	α_2		
1993	1.79	0.56	0.43	0.88	2,037
	(31.49)	(31.20)	(26.96)		
1994	1.86	0.49	0.49	0.88	2,213
	(34.21)	(27.37)	(30.58)		
1995	1.97	0.52	0.44	0.85	2,212
	(33.17)	(27.24)	(26.33)		
1996	1.80	0.58	0.40	0.84	2,293
	(29.98)	(31.73)	(24.29)		
1997	1.80	0.61	0.38	0.85	2,175
	(29.87)	(34.15)	(24.48)		
1998	1.63	0.60	0.42	0.80	2,028
	(21.71)	(27.58)	(22.19)		
1999	1.80	0.60	0.37	0.71	2,009
	(19.29)	(23.46)	(16.65)		
2000	1.36	0.64	0.41	0.77	1,837
	(14.58)	(25.22)	(18.97)		
2001	1.48	0.65	0.39	0.82	1,486
	(16.38)	(28.58)	(20.46)		
2002	1.52	0.55	0.49	0.86	1,510
	(19.39)	(27.45)	(28.43)		
2003	2.07	0.51	0.47	0.86	1,540
	(26.77)	(24.99)	(26.58)		
Mean	1.73	0.58	0.43		

Notes:
$Log(MC_{it})$ is the logarithmic of market capitalization (price per share times number of shares outstanding) for firm *i* at the end of period *t*,
$Log(BV_{it})$ is the logarithmic of book value of common equity reported in the balance sheet of firm *i* at the end of period *t*,
$Log(NI_{it})$ is the logarithmic of net income reported in the income statement firm *i* for the period *t-1* to *t*,
N is the number of observations.

Figure 8 presents the results. We find that the means of annual means of studentized residuals are no longer significant at conventional levels, ranging from -1.28 for the group with the smallest market capitalization to 0.94 for the group with the largest market capitalization. Additionally, the means of annual means of absolutes values of studentized residuals, indicating heteroscedasticity, also are no longer significant, ranging from 0.61 for the twentieth group of firms to 1.22 for the group of the smallest firms. From Figure 8, it is clear that the scale effect has been successfully mitigated via data transformation and there are no longer signs of heteroscedasticity.

FIGURE 8

Coefficient Bias in OLS Regression of Market Capitalization on
Book Value and Net Income After Applying a Logarithmic Transformation
(The bias (heteroscedasticity) is measured using means of studentized residuals
(absolute values of studentized residuals) for groups formed on Market Capitalization)



Notes:
Regression (5) is conducted and the studentized residual is calculated for each observation. The studentized residual for an observation is the residual obtained using the coefficient estimated with this observation omitted from the regression and divided by the standard error of the regression. We then form 40 groups based on market capitalization (group 1 being the smallest group and group 40 the largest). Then for each group, a mean of the studentized statistics is calculated. The same analysis is performed using absolute values of studentized residuals.

6 ECONOMIC INTERPRETATION OF LOGARITHMIC TRANSFORMATION

Having documented the econometric justification of a logarithmic transformation when estimating a price level regression, in this section we provide an economic justification.

We rely on the Burgstahler and Dichev (1997) option-style valuation model. In this framework, the market value reflects an option-style combination of recursion value (capitalized expected earnings when the firm recursively applies its current business technology to its resources) and adaptation value (the value of the firm's resources adapted to an alternative use). The relative weights on the two components of value reflect the possibility that the firm exercises the option to adapt resources to an alternative use. Specifically, the firm will opt out of recursion value in favor of adaptation value when recursion value is low relative to the adaptation value. The market value of equity reflects the option to choose either the recursion value or the adaptation value (AV) whichever is larger, at some point in the future. That is,

$$MC\left(NI,AV\right)=E\left[max\left(c\times NI,AV\right)\right]=\int_{NI}\int_{AV}max\left(c\times NI,AV\right)\times f\left(NI,AV\right)\times dAV\times dNI \quad (6)$$

Burgstahler and Dichev (1997) derive an empirical version of equation (6), expressing the market value of equity (MC) as a function of both book value (BV) and net income (NI), where BV is a proxy for adaptation value and NI is a proxy for recursion value:

$$MC_{it}=\alpha_1\times BV_{it}+\alpha_2\times NI_{it}+\varepsilon_{it} \quad (7)$$

The empirical version of Equation (6) approximates the market value of equity for a given firm at time t as a linear combination of book value and expected earnings at time t[17]. As earnings become extremely low relative to book value, book value becomes the sole determinant of market value and α_2 approaches zero while α_1 approaches unity. In contrast, as earnings become extremely high relative to book value, earnings become the sole determinant of market value and α_1 approaches zero while α_2 approaches the earnings capitalization factor (c).

Theoretical Equation (6) can be redefined using transformed variables:

$$Log\left(MC\right)=E\left[max\left(Log\left(c\times NI\right),Log\left(AV\right)\right)\right]=\int_{NI}\int_{AV}max\left(Log\left(c\times NI\right),Log\left(AV\right)\right)\times f\left(NI,AV\right)\times dAV\times dNI \quad (8)$$

An empirical version of Equation (8) could then be:

$$Log\left(MC_{it}\right)=\alpha_1\times Log\left(BV_{it}\right)+\alpha_2\times\left[Log\left(c\right)+Log\left(NI_{it}\right)\right]+\varepsilon'_{it} \quad (9)$$

(17) Following Easton (1999), when we are valuing a firm we can consider two types of assets: on the one hand, those assets that, since they are expected to continue, are valued based on their future expected earnings (c·NI); and on the other, those assets that are valued at liquidation prices (BV), since they are related to discontinued operations. If all assets are expected to continue operations, then the value will be MC=c·NI. In contrast, if the firm discontinues all operations, then the value will be MC=BV. We can apply logs in each scenario so we will obtain log(MC)=log(c)+log(NI) in the former case and log(MC)=log(BV) in the latter. The final whole value of the firm can then be expressed as a linear combination of both, with coefficient estimates moving in a [0,1] range. Our results support this theoretical background.

or,

$$Log\left(MC_{it}\right) = \alpha_1 \times Log\left(BV_{it}\right) + \alpha_2 \times \left[Log\left(c\right) + Log\left(NI_{it}\right)\right] + \varepsilon'_{it} \tag{10}$$

where $\alpha_0 = \alpha_2 \cdot Log(c)$.

In this alternative setting, as earnings become extremely low relative to book value, a_2 approaches zero while α_1 approaches unity. On the other hand, as earnings become extremely high relative to book value, α_1 approaches zero while α_2 approaches unity.

We are now able to discuss coefficient estimates from regression (5) ≡equivalent to Equation (10). Results reported in Table 2 reveal that both book value and earnings coefficient estimates are consistent with predicted values. Coefficients are always located in a (0-1) range. Moreover, the sum of both coefficients averages just 1.01 across all individual year regressions, ranging from 0.97 to 1.05. This confirms the complementary valuation roles of book value and earnings as predicted by Burgstahler and Dichev (1997).[18]

In short, logarithmic transformation not only removes the scale effect; it also provides meaningful economic estimates.

7 CONCLUSIONS

We consider the scale effect as a pure econometric phenomenon. We relate this effect with regression variables distribution. Typical variables used in Market Based Accounting Research (i.e. market capitalization, book value, total assets, positive earnings, losses) seem to follow a distribution that approximates quite well to a lognormal one. The mean feature of this type of distribution is the presence of a single, long tail revealing the existence of a large amount of extreme observations. If a distribution is lognormal, classical removal procedures ≡the top and bottom 1%, 2% or 3% of the distribution ≡will fail. Applying a logarithmic transformation to a variable that follows a lognormal distribution we obtain a variable that is normally distributed. The presence of extreme values is drastically reduced and the transformed variable will not cause any scale effect.

Matching Easton and Sommers (2003) results, we document that untransformed price-levels regressions suffer from the problem of the scale effect. Once we apply a logarithmic transformation, the resulting regression specification no longer suffers from the coefficient bias and heteroscedasticity present in the undeflated regression. Easton and Sommers

(18) Logarithmic transformation could be useful in assessing relative value-relevance studies, since coefficient estimates on earnings and book value reflect relative weights.

(2003) also mitigate the scale effect deflating by market capitalization via a weighted least squares regression. However, this methodology may result in misleading empirical findings. When we give weights we are impounding subjective criteria in the data. Easton and Sommers (2003) rank the data by size (market capitalization), so results are driven by this selection: large firms will receive very low weight in comparison to small firms. Applying logarithmic transformation, all observations are equally weighted and no information is wasted.

Regarding the economic interpretation of coefficient estimates, we rely on the real-option valuation framework proposed by Burgstahler and Dichev (1997). Coefficient estimates are in accordance with theoretical predictions.

REFERENCES

Akbar, S. and A.W. Stark (2003), "Discussion of Scale and the Scale Effect in Market-Based Accounting Research", *Journal of Business, Finance & Accounting*, Vol. 30, pp. 57-72.

Barth, M. and S. Kallapur (1996), "The Effects of Cross-Sectional Scale Differences on Regression Results in Empirical Accounting Research", Contemporary Accounting Research, Vol. 13, pp. 199-233.

Berger, P., E. Ofek and I. Swary (1996), "Investor Valuation of the Abandonment Option", *Journal of Financial Economics*, Vol. 24, pp. 257-87

Burgsthaler, D. and I. Dichev (1997), "Earnings, Adaptation and Equity Value", *The Accounting Review*, Vol. 72, pp. 187-215.

Easton, P. D. (1999), "Security returns and the value relevance of accounting data", *Accounting Horizons* (December), pp. 399-412.

Easton, P and Sommers, G. (2003), "Scale and the Scale Effect in Market-Based Accounting Research", *Journal of Business, Finance & Accounting*, Vol. 30, pp. 25-56.

Hayn, C. (1995), "The Information Content of Losses", *Journal of Accounting and Economics*, Vol. 20, pp. 125-53.