

Biases and Error Measures: How to Compare Valuation Methods¹

Ingolf Dittmann² Ernst Maug³

This Draft: August 25, 2008

¹We are grateful to Malcolm Baker, Christian Leuz, James Ohlson, Dennis Oswald, Jack Stecher and seminar participants at the American Accounting Association Congress (Chicago), the European Accounting Association Congress (Dublin), the British Accounting Association Congress (London), and Instituto de Empresa (Madrid) for clarifying discussions and suggestions on an earlier draft of this paper. We thank Christian Weiner for extensive research assistance. We gratefully acknowledge financial support from the Rudolf von Bennigsen-Foerder Foundation and from the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk.”

²Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands. Email: dittmann@few.eur.nl. Tel: +31 10 408 1283.

³Universität Mannheim, D-68131 Mannheim, Germany. Email: maug@corporate-finance-mannheim.de, Tel: +49 621 181 1952.

Biases and Error Measures: How to Compare Valuation Methods

Abstract

We investigate biases of valuation methods and document that these depend largely on the choice of error measure (percentage vs. logarithmic errors) used to compare valuation procedures. We analyze four multiple valuation methods (averaging with the arithmetic mean, harmonic mean, median, and the geometric mean) and three present value approaches (dividend discount model, discounted cash flow model, residual income model). Percentage errors generate a positive bias for most multiples, and they imply that setting company values equal to their book values dominates many established valuation methods. Logarithmic errors imply that the median and the geometric mean are unbiased while the arithmetic mean is biased upward as much as the harmonic mean is biased downward. The dividend discount model dominates the discounted cash flow model only for percentage errors, while the opposite is true for logarithmic errors. The residual income model is optimal for both error measures.

JEL Classification: G10, M41

Keywords: Valuation, Financial Ratios, Multiples, Dividend Discount Model, Discounted Cash Flow Model, Residual Income Model

1 Introduction

This paper analyzes a methodological question that turns out to be of primary importance for valuation research: what are the consequences of different error measures when comparing alternative valuation methods. Of the fourteen papers on horse races of multiples and present value methods we are aware of, nine measure valuation accuracy based on the percentage difference between estimated values and market values, whereas another five use log errors, defined as the logarithm of the ratio of the estimated value to the market value.¹ Only two of the articles that use percentage errors motivate their choice, and no paper explicitly recognizes the choice of error measure as a critical decision in the research design.² Also, no paper reports results for both error measures. In this paper we show that the researcher's choice of error measure is critical. This choice determines whether a valuation method produces a bias or not and therefore predisposes the conclusion in favor of certain types of valuation methods.

Percentage errors penalize overvaluations more than undervaluations. While undervaluations in excess of -100% are impossible by virtue of limited liability, overvaluations are not limited and often much more extreme than +100%. As a consequence, judging valuation methods on the basis of percentage errors favors methods that avoid large overvaluations. In contrast, logarithmic errors create more symmetric distributions of valuation errors because for each overvaluation there exists an undervaluation of equal absolute size. Statistically, logarithmic error distributions are closer to satisfying the normality assumptions often made for statistical inference.³

¹Alford (1992), Beatty, Riffe and Thompson (1999), Dechow, Hutton and Sloan (1999), Bhojraj and Lee (2002), Cheng and McNamara (2000), Francis, Olsson and Oswald (2000), Liu, Nissim and Thomas (2002a, b), and Penman and Sougiannis (1998) use percentage errors, while Gilson, Hotchkiss and Ruback (2000), Herrmann and Richter (2003), Kaplan and Ruback (1995), Kim and Ritter (1999), and Lie and Lie (2002) use log errors.

²Alford (1992) argues that absolute percentage errors put equal weight on positive and negative errors. Beatty, Riffe, and Thompson (1999) also provide an explicit justification.

³To the best of our knowledge, only Baker and Ruback (1999) have explicitly tested if the normality assumption applies to their sample and they could not reject it. However, their sample of 225 observations is rather small by the standards of the valuation literature. Kaplan and Ruback (1995), Lie and Lie (2002), and Herrmann and Richter (2003) explicitly motivate the use of log errors with the skewness of percentage errors or the distributions of the underlying fundamental variables.

Error measures are inherently subjective as they are determined by the loss function of the researcher or analyst who needs to choose a valuation procedure.⁴ Therefore, our analysis cannot establish which error measure *should* be used. Instead, our objective is to highlight the *effects* of the choice of error measure, so that researchers and analysts alike can draw their own conclusions about the error measure and, eventually, about the valuation methods they wish to use. To this end, we revisit two important questions in valuation analysis.

In our first application, we compare four methods for averaging multiples: the arithmetic mean, median, harmonic mean, and the geometric mean. The use of averaging procedures in academic research does not reveal a consensus: median, arithmetic mean, and harmonic mean are used by different researchers, and some papers use several averaging procedures simultaneously without providing the reader with explicit guidance as to which one is preferable. Several researchers have recently argued in favor of the harmonic mean as the best choice as it avoids the apparent upward bias of the arithmetic mean.⁵ Our analysis replicates the finding that the harmonic mean is less biased than the arithmetic mean, the geometric mean, or the median if percentage errors are used. For logarithmic errors, however, the harmonic mean is biased downward as much as the arithmetic mean is biased upward, whereas the geometric mean and the median are unbiased.

Our second analysis compares the dividend discount model, the residual income model, and the discounted cash flow model. We show that the ranking of these three models in terms of forecast accuracy depends on the error measure used for the comparison. While the residual income model turns out to be most accurate according to both measures (which is in line with Penman and Sougiannis, 1998, and Francis, Olsson and Oswald, 2000), the dividend discount model dominates the discounted cash flow model if percentage errors are used, whereas logarithmic errors arrive at the opposite conclusion. The reason is that value esti-

⁴Basu and Markov (2003) and Rodriguez (2005) infer the loss function of analysts regarding their earnings forecasts from their forecasts. Their results do not carry over to valuations, however, because earnings forecasts and valuations are conceptionally different quantities with different properties: Earnings can become negative while values cannot.

⁵Baker and Ruback (1999) argue that the harmonic mean is an ML-estimator in a model where valuation errors are normally distributed. Liu, Nissim and Thomas (2002a, 2002b) provide a derivation that supports the use of the harmonic mean as a viable and unbiased estimator. Beatty, Riffe, and Thompson (1999), Bhojraj and Lee (2002), and Herrmann and Richter (2003) also use the harmonic mean.

mates based on dividends strongly underestimate the value of the firm. This undervaluation is penalized more severely by logarithmic errors than by percentage errors.

We also include two *ad-hoc* methods in the two applications, in order to establish to what extent the biases of valuation procedures are a consequence of the error measure chosen rather than of the valuation procedure itself. We show that an *ad hoc* method that ignores all comparable and analyst information and sets the predicted market value equal to the firm's book value turns out to be just as good or even better than any of the four comparable procedures and the discounted cash flow method when percentage errors are used. Moreover, ignoring all information and arbitrarily setting the predicted firm value equal to \$1 leads to comparatively low percentage errors and – in some situations – turns out to be the best valuation method when judged by percentage errors.

Our explanation for these results is simple. Setting market values equal to book values severely undervalues companies on average as the market-to-book ratio is 1.9 for the typical company in our sample, but this *ad hoc* procedure avoids large overvaluations. The same is true for the more extreme approach of setting company values equal to \$1. Effectively, this sets all percentage errors equal to –100% by fiat. However, all averaging methods produce percentage errors in excess of +100% between one fifth and one third of the time, and errors exceeding 200% or more are not uncommon. The "method" of setting company values arbitrarily equal to \$1 conveniently avoids percentage errors of this magnitude and allows this procedure to perform relatively well compared to proper valuation methods. On the other hand, logarithmic errors arrive at exactly the opposite conclusions: With logarithmic errors, both *ad hoc* methods turn out to be clearly inferior to any valuation method that uses comparable companies or analyst forecast data. Altogether our results demonstrate that the error measure is a critical design feature in a horse race between different valuation methods.

The following Section 2 contains our comparison of four averaging methods for multiples. We first establish our empirical results with a sample of 52,112 U.S. firm-year observations from 1994 to 2003. We then use theory and simulations to generalize our empirical results and

to generate further insights. Section 3 contains a similar empirical comparison of dividend, free cash flow, and residual income valuation methods based on analyst forecasts. Section 4 concludes and offers some recommendations for the research design when valuation methods need to be compared. Technical material is deferred to the appendix.

2 Comparing four averaging methods for multiples

2.1 Dataset

Our analysis is based on annual data from Compustat between 1994 and 2003. We select all companies domiciled in the United States whose sales and total assets both exceed \$1 million. We also require that the market value of equity four months after the fiscal year end is available. The four months lag ensures that the company’s financial statements have been publicly available to investors and are therefore reflected in the market value. We exclude those companies where the SIC code is either not available or equals 9999 (not classifiable).⁶ We are left with a final dataset with 52,112 firm-year observations.

We focus on three widely used multipliers:⁷

- market-to-book ratio, defined as the market value of equity divided by the book value of equity.
- value-to-sales ratio, defined as the ratio of enterprise value to sales, where enterprise value is the market value of equity plus total debt.
- price-earnings ratio, defined as market value of equity divided by net income.

A multiple that is negative according to these definitions is set to a missing value. We also set the market-to-book ratio equal to a missing value if shareholders’ equity is smaller than \$1 million. We can compute the market-to-book ratio for 47,614 firm-year observations,

⁶We use the historical SIC code (SICH) when available. If the historical SIC code is not recorded on Compustat, we use the current SIC code (SIC).

⁷Our empirical results for these three ratios are exemplary. In Section 2.3, we show theoretically that the results carry over to any multiplier with a skewed distribution.

the value-to-sales ratio for 51,899 observations, and the price-earnings ratio for 33,753 observations. Finally, we winsorize the data separately for each multiple and each year at 1% and 99%.⁸ We report descriptive statistics for all three ratios and for their natural logarithms in Table 1.

[Insert Table 1 about here]

The table shows that the median market-to-book ratio in our sample is 1.87. The median value-to-sales ratio is 1.63, and the median price-earnings ratio is 17.1. Bhojraj and Lee (2002) report a similar median market-to-book ratio of 1.84. Their mean (2.26) is substantially lower than our mean, because they delete extreme values while we winsorize them; also their sample ranges from 1982 to 1998 and excludes the high valuation years 1999 and 2000 included in our sample.⁹ As usual, all distributions are highly skewed and means substantially exceed medians. Table 1 also reports considerable positive excess kurtosis for all three distributions, i.e. all distributions have fatter tails than the normal distribution.

2.2 Empirical analysis

We consider the financial ratio x_i of the market value of company i , denoted by MV_i , to some base B_i , where the base is the book value of assets, sales revenues, or net income, so that $x_i = MV_i/B_i$. Our aim is to estimate the market value of a target firm j , for which a set of comparable firms is available that does not include the target firm. The base B must be positive for the target and for all comparable firms. We compute an average financial ratio \bar{x}_j across all comparable firms and multiply it by firm j 's base B_j in order to obtain

⁸We repeated our analysis twice, once with a stronger winsorization at 2.5% and 97.5% and once without winsorization and obtained identical qualitative and similar quantitative results. We therefore only report the results for the 1% winsorization in our tables.

⁹Of the other comparable studies, Beatty, Riffe, and Thompson (1999) and Liu, Nissim, and Thomas (2002a) report all variables scaled by price, which is the inverse of our ratio and has different statistical properties. Alford (1992) and Cheng and McNamara (2000) work with much older samples. Lie and Lie (2002) work with different definitions (enterprise value to total assets instead of market-to-book) and only with 1998 data.

an estimate of firm j 's market value:

$$\widehat{MV}_j := \bar{x}_j \times B_j . \quad (1)$$

We consider four different averaging methods:

$$\text{Arithmetic Mean: } \bar{x}_j^A = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

$$\text{Harmonic Mean: } \bar{x}_j^H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} \quad (3)$$

$$\text{Geometric Mean: } \bar{x}_j^G = \prod_{i=1}^n x_i^{1/n} = \exp \left\{ \frac{1}{n} \sum_{i=1}^n \ln(x_i) \right\} \quad (4)$$

$$\text{Median: } \bar{x}_j^M = \text{Median}(x_i) \quad (5)$$

The second expression in (4) shows that the geometric mean can be interpreted as a retransformed arithmetic mean of the logs of the multiples x_i .

In order to compare the precision of the four averaging methods, we need an error measure that defines what we mean by ‘relative precision.’ Two error measures are commonly used in the literature on company valuation (see also footnote 1 above):

$$\text{Percentage Error: } e_{pct}(j) = \frac{\widehat{MV}_j - MV_j}{MV_j} = \frac{\bar{x}_j \times B_j - x_j \times B_j}{x_j \times B_j} = \frac{\bar{x}_j}{x_j} - 1 , \quad (6)$$

$$\text{Log Error: } e_{\log}(j) = \ln \left(\frac{\widehat{MV}_j}{MV_j} \right) = \ln(1 + e_{pct}(j)) = \ln \left(\frac{\bar{x}_j}{x_j} \right) . \quad (7)$$

Some of the literature following Alford (1992) has also looked at the mean or median absolute error and we will therefore also look at $E(|e_{pct}|)$.

The two error measures do not differ significantly for small errors as we have $e_{pct} = e_{\log} + O((e_{\log})^2)$ from a first order Taylor expansion. However, valuation errors with multiples are often large and then the two error functions generate some notable differences. In particular they differ in their relative treatment of over- and undervaluations. Percentage errors have the same absolute magnitude for overvaluations and undervaluations by the same dollar

amount, so for a firm with a value of 100 a valuation of 150 and a valuation of 50 generate the same absolute percentage error. By contrast, logarithmic errors have the same magnitude for overvaluations and undervaluations if the ratio is the same, so a twofold overvaluation (i.e., 200) generates the same log error as an undervaluation by a factor of two (i.e., 50).

Note that the function $\ln(1 + x)$ that maps percentage errors into log errors is monotonically increasing and sign preserving. Therefore, both measures provide the same ranking and the same signs of valuation errors. The two error measures only differ in the size of the errors relative to one another. Percentage errors state, for instance, that a forecast of 200 is twice as bad as a forecast of 150 for a firm with a true value of 100. By contrast, the logarithmic error suggests that a forecast of 200 is only about 70% worse than a forecast of 150. Note that both measures agree on the fact that 200 is worse than 150.

In addition to the four averaging methods (2) to (5), we also consider two benchmark valuation procedures. These procedures make no use of comparable information and we would expect that any valuation procedure that incorporates more information should also generate lower errors. In particular:

1. We set the value of the target company arbitrarily equal to its book value, which amounts to setting the market-to-book value of the target company equal to one. We therefore call this strategy in the tables below “MTB = 1.” This is clearly a very rough and imprecise valuation method that is based on only one piece of accounting information. Note that this method is biased downward as the median and mean market-to-book ratio are substantially larger than one in our sample.
2. We set the value of the company arbitrarily equal to a very small value close to zero. For this we choose \$1. We refer to this procedure as “Value = \$1” in the tables. This procedure is even worse than the first as it relies on no company information at all and any reasonable procedure should find it easy to beat this benchmark.

Both dummy-procedures lead on average to undervaluations and large biases. However, they generate very little dispersion and avoid overvaluations. The Value=\$1 procedure also

avoids errors of more than 100% in absolute value. We use these *ad hoc* procedures in order to establish more clearly which results should be attributed to characteristics of the valuation procedures, and which results should be attributed to the error measures chosen to characterize the valuation procedures.

For all firms in the dataset we select a set of at least five comparables from the same industry. Following Alford (1992), we start at the 4-digit SIC level. If we cannot find at least five comparable firms, we proceed to the 3-digit SIC level and, likewise, to the 2-digit SIC level. We delete all firms for which we cannot find at least five comparables at the 2-digit SIC level.¹⁰ We repeat this for every year from 1994 to 2003. For each firm-year and each averaging method, we then calculate the estimated firm value and the three error measures: the log error (7), the percentage error (6) and the absolute percentage error. Table 2 reports sample statistics for the three error measures for the market-to-book ratio.¹¹

[Insert Table 2 about here]

The table shows that the bias (i.e. the mean error) for percentage errors (Table 2, Panel B) is highest for the arithmetic mean (1.211), lowest for the harmonic mean (0.135), with the geometric mean and the median about halfway in between (0.538 and 0.530, respectively). We therefore reproduce the result stated in the previous literature that the harmonic mean dominates all other valuation methods in terms of percentage errors. In terms of log errors (Table 2, Panel A) however, the geometric mean is unbiased and the median exhibits a very small but statistically highly significant negative bias (-0.018). With a bias of 0.318, the arithmetic mean is biased upward by about as much as the harmonic mean is biased

¹⁰See Cheng and McNamara (2000) and Bhojraj and Lee (2002) for an analysis of more sophisticated methods to choose comparables.

¹¹In Tables 2, 3, and 4, we exclude observations with valuation errors larger than 1000 (i.e. 100,000%) under percentage errors before calculating the statistics shown in the table. Accordingly we exclude one observation for Table 2 (market-to-book), one observation for Table 3 (enterprise-value-to-sales), and 13 observations for Table 4 (price-earnings). These obvious outliers heavily influence standard deviations and RMSEs under percentage errors. On the other hand, they have only little effect on the numbers reported for log errors. As a robustness check, we repeated our analysis and excluded all observations with errors larger than 10 (i.e. 1,000%), which amounts to 31% of all observations for the market-to-book ratio (results not shown in the tables). Not surprisingly, average errors drop considerably, and median and geometric mean now show a negative bias for logarithmic errors. The rankings of the valuation methods according to bias or RMSE, however, do not change with the only exception that MTB=1 (and not anymore Value=1\$) is the best method according to the RMSE and percentage errors.

downward (-0.296). The MTB=1 method dominates the median, the geometric mean and the arithmetic mean in terms of percentage errors (Table 2, Panel B) and the bias for MTB=1 is only slightly higher in absolute value to that of the harmonic mean. However, MTB=1 is worse than all comparable methods for log errors (Table 2, Panel A). Interestingly, while the pattern of *mean* errors (i.e. bias) differs significantly between percentage errors and log errors, the *median* errors generate the same message for percentage errors and log errors: they are very small for the median and the geometric mean, positive for the arithmetic mean, and negative for the harmonic mean. This is due to the fact that the function that maps percentage errors into logarithmic errors is monotonic so that the median (and also other quantiles) produces by construction a similar ranking among the valuation methods.

Users of valuation methods are not only interested in the bias of valuation procedures but also in the dispersion of valuation errors. We measure these by the standard deviation and the root mean squared error (RMSE). The squared RMSE is equal to the variance plus the squared bias and can be rationalized from minimizing a quadratic loss function. While the RMSE combines bias and standard deviation in a convenient way, the bias may be more important than is reflected in RMSEs for many applications in practice. For example, for a successful acquisition strategy it may be more important to avoid consistently overpaying for acquisition targets. On the other hand, bidding for companies in an auction puts a higher emphasis on the dispersion of the valuations in order to avoid the winner's curse. We therefore also report biases and standard deviations throughout, as different applications may warrant different weights for the bias and the dispersion of valuation methods.

For percentage errors the RMSE generates a remarkable result: Estimating the target firm's market value by its book value (MTB=1) outperforms all other valuation methods. The more extreme *ad hoc* procedure of setting the target firm's market value to \$1 (Value=\$1) turns out to be even better. The reason is that percentage errors are bounded from below at -100% but they are not bounded from above. If errors on the unlimited upside are severe, methods that undervalue firms on average (or even set the error equal to the lower bound as Value=\$1 does) appear to be preferable. The second to last column of Table 2, Panel B

shows that, for more than a third of the firms, the percentage error of the arithmetic mean exceeds 100%, while the median and the geometric mean lead to percentage errors in excess of 100% for about one fifth of all firms. These high overvaluations are largely avoided by MTB=1 and completely eliminated by Value=\$1.

Log errors, on the other hand, have an unlimited downside and penalize large undervaluations as much as large overvaluations. Table 2, Panel A shows that the Value=\$1 procedure is heavily penalized for the extreme undervaluations it generates. The MTB=1 procedure has a higher RMSE than all comparables-based procedures because of its large downward bias, which is assessed at 69% in terms of log errors, but only 18% in terms of percentage errors. Absolute errors behave broadly similarly to percentage errors (see Panel C of Table 2). Note that MTB=1 also dominates the harmonic mean for absolute percentage errors.

The fact that over- and undervaluations are treated asymmetrically by percentage errors is also reflected in the skewness of valuation errors. Percentage errors are highly skewed while log errors are much closer to being symmetric. We suspect that the skewness of percentage errors caused by the limited downside and unlimited upside is the reason why many researchers who work with percentage errors or absolute percentage errors report medians (and sometimes other percentiles) rather than means of the error distribution. This approach ignores the large incidence of extreme overvaluations, however.

[Insert Figure 1 about here]

Figure 1 contains eight graphs that show the error distributions: each of the four rows corresponds to one of the four averaging methods (arithmetic mean, median, geometric mean, and harmonic mean). The left graphs give the distributions of percentage errors, whereas the right graphs show the distributions of log errors. Clearly, all distributions of percentage errors are highly skewed. Also, all distributions exhibit a significant proportion of percentage errors that exceed 100%. The graphs confirm our intuition that errors based on log transformations are much closer to the model of a normal distribution than percentage errors. Apparently, the log transformation is successful in generating a symmetric distribution centered around zero for the median and the geometric mean, whereas the distributions of log errors for the

arithmetic mean and the harmonic mean are not symmetric. The distribution of log errors for the arithmetic mean has an extremely fat right tail, and its mode is clearly positive. The harmonic mean, on the other hand, exhibits a fat left tail and a negative mode. Hence, the harmonic mean generates more undervaluations, whereas the arithmetic mean generates more overvaluations.

[Insert Tables 3 and 4 about here]

Tables 3 and 4 display the results for, respectively, the value-to-sales ratio and the price-earnings ratio. Instead of $MTB=1$ we use, respectively, $Value=Sales$ and $P/E=10$ as *ad hoc* valuation methods. We do not continue to use $MTB=1$, because setting the market value of the firm equal to the book value would not mean using *less* information than the averaging methods in the table, but would imply using *different* information. Tables 3 and 4 show that all results for the market-to-book ratio continue to hold for the other two ratios. Note that the statistics in Tables 2 to 4 are not comparable across tables, because they refer to slightly different sets of firms. For instance, firms with negative earnings are included in the samples analyzed in Tables 2 and 3 while they have been excluded for the calculations shown in Table 4.

2.3 Theoretical results

So far we have shown for a typical empirical application that the choice of error measure has an important influence on the ranking of valuation procedures, and that simple *ad-hoc* valuation procedures dominate more sophisticated multiple valuation methods for percentage errors, but not for logarithmic errors. In the remaining part of this section, we address the question whether these results are robust, i.e. whether they continue to hold for other multiples and for other, possibly more sophisticated comparable selection methods. We show theoretically that our qualitative results are robust to the choice of comparables and that they continue to hold for any multiple that has a similarly skewed distribution. In particular we show that the arithmetic mean has always a large positive bias according to both error

measures, that the geometric mean is unbiased for logarithmic errors and positively biased for percentage errors, and that the harmonic mean is downwards biased for logarithmic errors.

Our theoretical analysis also generates three important additional insights. First, it shows that it is the skewness of the multiple distributions that makes the geometric mean unbiased for log errors. Taking logs removes the skewness, and this is exactly what the geometric mean effectively does before averaging (see the second transformation in equation (4)). Second, our simulations show that, as the dispersion among comparable firms increases, all valuation methods - including MTB=1 - have a positive bias for percentage errors. So the finding that a valuation method has a positive bias according to percentage errors is not informative about the valuation method, as it might just have been caused by high dispersion among comparable firms. Third, our simulations reveal that the effect of the averaging method becomes quantitatively less important (for both error measures) if the variation among comparable firms becomes smaller. This result justifies the more sophisticated comparable selection methods used by analysts in practice.

We assume throughout this subsection that the target and its comparables have been drawn independently from an identical distribution.¹² We start this section by collecting some basic results about the relative size of the four averages (2) to (5). We give a short proof in the appendix, which follows directly from Jensen's inequality and the law of large numbers.

Result 1 (Means):

(i) The arithmetic mean always results in a higher market value estimate than the geometric mean or the harmonic mean, and the harmonic mean always results in a lower market value estimate than the geometric mean:

$$\bar{x}_j^H < \bar{x}_j^G < \bar{x}_j^A.$$

The relative position of the median, \bar{x}_j^M , depends on the distribution of the ratios.

¹²In addition, we assume that there are at least two comparables ($n \geq 2$) that differ from one another. We maintain the independence assumption only for expositional convenience. All our results can also be derived under weaker assumptions that allow for dependence between the x_j .

(ii) If the ratios have been drawn from a symmetric distribution, then the median is close to the arithmetic mean in sufficiently large samples. More formally: $(\bar{x}_j^M - \bar{x}_j^A) \rightarrow 0$ almost surely as $n \rightarrow \infty$.

(iii) If the ratios have been drawn from a lognormal distribution, then the median is close to the geometric mean in sufficiently large samples. More formally: $(\bar{x}_j^M - \bar{x}_j^G) \rightarrow 0$ almost surely as $n \rightarrow \infty$.

Hence, an analyst who uses the harmonic mean will *always* arrive at lower value estimates than her colleague who works with the geometric mean, given that they use the same set of comparable companies. Both will obtain lower estimates than a third analyst using the arithmetic mean. The differences between these three averages are larger if the variation in the sample is large. So for a set of comparables with little variation, the four methods arrive at similar results.

Result 2 summarizes our theoretical results about the biases of the different valuation methods with the two error measures. These results are based on the assumption that financial ratios are lognormally distributed. In Appendix A we compare the distributions of the standard financial ratios (market-to-book, value-to-sales, price-earnings), their inverses (i.e., book-to-market, etc.) and their log transformations (i.e. $\ln(\text{market-to-book})$, etc.). We show that the log-transformed ratios are much better described by a normal distribution than the original ratios or their inverses, even though formal tests reject the normality assumption for all three distributions. We therefore conclude that the lognormal distribution is a reasonable assumption to base our theory on.

Result 2 (Biases): Assume that x_j and all comparable ratios x_i are distributed lognormal with parameters μ and $\sigma^2 > 0$.

(i) For **percentage errors**, the geometric mean and the arithmetic mean are both biased upward. The bias of the arithmetic mean is stronger than that of the geometric mean.

$$0 < E(e_{pct}^G) < E(e_{pct}^A) .$$

The expected error of the harmonic mean is smaller than that of the geometric mean:

$E(e_{pct}^H) < E(e_{pct}^G)$, and the median is also biased upward in large samples.

(ii) For **log errors**, the geometric mean is unbiased, the arithmetic mean is biased upward, and the harmonic mean is biased downward.

$$E(e_{\log}^H) < E(e_{\log}^G) = 0 < E(e_{\log}^A) .$$

In absolute terms, the harmonic and the arithmetic mean are equally biased: $E(e_{\log}^H) = -E(e_{\log}^A)$. The median is unbiased in large samples.

The arithmetic mean has a positive bias for both definitions of valuation errors. The distribution of x_i is skewed and the arithmetic mean gives equal weight to all observations, including large positive outliers that necessarily occur with skewed distributions. The geometric mean is unbiased for log errors for reasons that are intuitive from looking at the second transformation in equation (4). The geometric mean is the retransformed arithmetic mean, applied to the logarithms of the financial ratios x_i , and $\ln(x_i)$ is distributed normal. Hence, this implicit logarithmic transformation removes the skewness of the original distribution and the transformed distribution is symmetric, which is exactly what is required for generating unbiased logarithmic errors. However, unbiased logarithmic errors are not equivalent to unbiased percentage errors. From (7) observe that $e_{pct} = \exp(e_{\log}) - 1$ and this convex transformation gives positive errors a larger weight than negative errors, creating an upward bias. By comparison the harmonic mean is biased downward for logarithmic errors. The skewness of the lognormal distribution implies that large outliers are balanced by a larger number of very small observations, which then create very large numbers $1/x_i$ from the definition of the harmonic mean. This inflates the denominator of \bar{x}_j^H and biases the valuation errors downward. Interestingly, the harmonic mean is biased downward just as much as the arithmetic mean is biased upward.

Note that all theoretical predictions from Result 2 are borne out by the empirical results shown in Tables 2 to 4. This implicitly supports our assumption that multiples are lognormally distributed and implies that the same qualitative results will obtain for any type of multiple that has a similar skewness as the three multiples analyzed above.

2.4 The impact of dispersion among comparable firms

An important determinant for the size of the bias of a particular averaging method is the dispersion of financial ratios among comparable firms. A good valuation procedure should be able to cope with industries that exhibit large dispersions of financial ratios. We investigate the effect of an increase in dispersion with a simulation study.

For each of the 100,000 runs of our base scenario, we draw 40 market-to-book ratios from a lognormal distribution with $\mu = 0.593$ and $\sigma = 0.838$. These parameters are the average industry mean and the average within-industry standard deviation of the log market-to-book ratio across the 608 industry-years in our sample. The median industry comprises 38 firms. We therefore consider industries with 40 firms in our simulations.¹³ Likewise, we draw 40 book-values from a lognormal distribution with $\mu = 18.329$ and $\sigma = 1.838$. Again these parameters have been estimated from our dataset. We analyze this simulated dataset in the same way as we analyzed the empirical data in Tables 2 to 4.

[Insert Table 5 about here]

The left part of Table 5 shows the results for this base scenario. The numbers are similar to those in Table 2, and all qualitative results are identical, which confirms that the lognormal distribution is an adequate model. The main difference is the smaller standard deviation of errors for the simulated data, evidently a consequence of the fact that the empirical distribution function has fatter tails than the lognormal distribution (see Figure 2 in Appendix A). The right part of Table 5 displays the results we obtain when we repeat this simulation with a standard deviation for the market-to-book ratio that is raised by 33% to $\sigma = 1.115$ (all other parameters stay the same). We choose a rather extreme increase in the standard deviation (only 7.2% of the 608 industry-years in our sample have a standard deviation that exceeds 1.115) in order to better demonstrate the effects of increased dispersion.

¹³The standard deviation is lower than the overall standard deviation in our sample (0.919, see Table 1), so choosing firms from the same industry reduces the dispersion as it removes between-industry variation. The mean is also different because we average first within each industry and then across industries, which gives firms in smaller industries a larger weight.

In many ways, an increase in dispersion magnifies the effects we have discussed previously. The bias, the standard deviation and the proportion of errors exceeding 100% increase dramatically for all valuation methods and all definitions of valuation errors. A notable exception is the geometric mean and the median with log errors, which remain unbiased. Interestingly, even the MTB=1 procedure now overvalues the target firm with percentage errors (on average by 2.9%), confirming our result above that overvaluation is a feature of percentage errors and not informative about the valuation procedures themselves. More importantly, MTB=1 now dominates all averaging methods including the harmonic mean, which has a bias of 5.8%. Note that MTB=1 still heavily underestimates company values according to log errors.

Investment bankers and practitioners who use multiples to value acquisition targets or IPOs seem to have an intuitive grasp of the necessity to reduce the variation among comparables. They typically inspect the distribution and eliminate what appear to be outliers that are not representative of the industry and can be attributed to circumstances inapplicable to the target firm. In the light of the evidence given in Table 5, this approach appears reasonable when the arithmetic or the harmonic mean are used. If practitioners would use the median or the geometric mean, then such an ex post pruning of the sample is not necessary and (because of the loss of information) not helpful.

We have also performed simulations where industries are smaller or larger than 40 firms (results not reported in the tables). It turns out that reducing industry size has a similar effect as increasing dispersion as shown in Table 5. The only exception is that the median has a positive bias in industries with less than 10 firms if the number of comparable firms (excluding the target firm) is even. The reason is that the median is equal to the arithmetic mean of the two most central observations if the number of comparables is even, so that the bias of the arithmetic mean carries over to the median if the number of comparables is small and even.

3 Comparing dividend, free cash flow, and residual income valuation methods

In this section, we show that the choice of error measures is also relevant when comparing more sophisticated valuation methods that are based on discounting future expected income flows. We compare the dividend discount model, the residual income model, and the discounted cash flow model. The analysis in this chapter is in the spirit of Francis, Olsson and Oswald (2000) who provide a comprehensive comparison of these three methods using percentage errors. We cannot replicate their results, as they work with a hand collected dataset from a different data source. We also keep this analysis deliberately simple, because our main aim is to demonstrate the impact of the error measure on the comparison of valuation methods.

We use the following specification of the dividend discount model (DDM):

$$\widehat{MV}^{DDM} = \sum_{t=1}^5 \frac{div_t}{(1+r_E)^t} + \frac{div_5(1+g)}{(1+r_E)^5(r_E-g)}. \quad (8)$$

Here, div_t is the forecasted annual dividend (IBES mean) for year t . We require that div_1 is available for all firms in our dataset. If div_2, \dots, div_5 are not available, we calculate them from the last available dividend forecast and from the IBES long-term earnings growth rate. r_E is the cost of equity capital, which we set equal to the risk-free rate (five year government bond rate) plus a risk-premium of 6%. The terminal growth rate g is set equal to 3%.

The residual income model (RIM) is given by

$$\widehat{MV}^{RIM} = bv_0 + \sum_{t=1}^5 \frac{ri_t}{(1+r_E)^t} + \frac{ri_5(1+g)}{(1+r_E)^5(r_E-g)}, \quad (9)$$

where residual income is defined by $ri_t = i_t - r_E bv_{t-1}$ with net income i_t and book value bv_t .¹⁴ The terminal growth rate g is equal to 3% as in the other two models. We require that bv_0 is available from Compustat (item 60) and that at least i_1 and the forecasted long-

¹⁴See Ohlson (1995) for a derivation and detailed discussion of this model.

term earnings growth rate are available on IBES for all firms in our sample. If i_2, \dots, i_5 are not available, we calculate them from the last available earnings forecast and the earnings growth rate. The book value is updated over time by assuming that 50% of the earnings i_t are reinvested, i.e., $bv_t = bv_{t-1} + 0.5i_t$. See Claus and Thomas (2001) for a justification of this approach.

Finally, we specify the discounted cash flow (DCF) valuation model

$$\widehat{MV}^{DCF} = \sum_{t=1}^5 \frac{lcf_t}{(1+r_E)^t} + \frac{lcf_5(1+g)}{(1+r_E)^5(r_E-g)}, \quad (10)$$

where lcf_t is the levered cash flow to equity, i.e. the unlevered free cash flow to the firm minus interest expenses plus the interest tax shield. The discount factor is consequently the cost of equity r_E . We calculate lcf_t as the IBES operating cash flow forecasts (which is before investment) minus a mechanical capital expenditure forecast. For year $t = 1$, capital expenditures are assumed to be equal to actual capital expenditures in year $t = 0$, which we obtain from Compustat (item 128).¹⁵ For the following years, we assume that the ratio of capital expenditures to operating cash flows remains constant, i.e. capital expenditures grow (or shrink) at the same rate as operating cash flows. We require that at least next year's cash flow forecast is available on IBES. If the forecasts for years 2 to 5 are not available, we calculate them from the last available cash flow forecast and the long-term earnings growth rate. Finally, we obtain the actual market value of equity MV from IBES.

There are 864 observations that fulfil our data requirements. We lose many observations, because cash flow and especially dividend forecasts are only infrequently available from IBES. In addition, we exclude 210 firm-year observations for which any of the three valuation methods yields a non-positive value.¹⁶ Our final dataset contains 654 observations.

¹⁵If capital expenditures are not directly available from Compustat, we calculate them as $PPE_t - PPE_{t-1} + DEP$, where PPE is property, plant and equipment (item 8) and DEP is depreciation and amortization (item 14). This gives us 80 additional observations that would otherwise be lost. Results do not change if these 80 observations are excluded from the sample.

¹⁶Francis, Olsson and Oswald (2000) set negative value forecasts equal to zero, which is not an option for our analysis, because the logarithm of zero is not defined. We argue that excluding non-positive estimates is reasonable because the typical user of these valuation methods is unlikely to infer from a negative value estimate that the value of the firm is zero. She would rather conclude that the valuation method is not applicable under the given circumstances.

[Insert Table 6 about here]

Table 6 presents our results. The residual income model (RIM) turns out to be the least biased valuation method according to all three error measures. It leads to an average overvaluation of 8% with log errors and 23% with percentage errors. This is in line with the findings of Penman and Sougiannis (1998) and Francis, Olsson and Oswald (2000). By contrast, the ranking of discounted cash flow model (DCF) and dividend discount model (DDM) depends on the error measure: With log errors (Table 6, Panel A), DCF turns out to be superior to DDM according to all criteria (mean, median, RMSE, standard deviation), whereas the opposite is true if percentage errors are used (Panel B).¹⁷ The reason for this reversal of the ranking is again the asymmetric treatment of positive and negative errors by the percentage error measure. The dividend discount model avoids large positive errors, whereas 52% of all DCF estimates have an error that exceeds +100%. Percentage errors punish positive errors more than negative errors, so that DCF appears much worse than DDM. Note that DDM is even superior to RIM if judged by the RMSE criteria for percentage errors.

We also include the two benchmark valuation methods that do not make use of any forecasts: "MTB=1" and "Value=1\$". If log errors are used, these two benchmark methods turn out to be much worse than any of the methods that are based on analyst forecasts. With percentage or absolute percentage errors, however, both methods are superior to DCF according to all criteria.

4 Discussion and conclusion

This paper demonstrates that the choice of error measure can critically affect the ranking of valuation methods obtained in a horse race. Percentage errors have a limited downside and

¹⁷Francis, Olsson and Oswald (2000) find that the DCF model is superior to the DDM model with percentage errors. At least a part of this effect is due to the fact that they retain zero value estimates and set negative value estimates to zero. In their sample 19% of the DDM estimates are zero; if these were excluded, DDM would clearly improve. For DCF, on the other hand, the bias would increase if such zeroes were dropped. Please also see footnote 16.

an unlimited upside and therefore tend to favor methods that produce low valuations. We demonstrate this in two prominent applications. First, we compare four multiple valuation methods and replicate the finding that the harmonic mean turns out to be least biased when percentage errors are used, whereas the arithmetic mean is most biased. When this comparison is done with logarithmic errors, however, then the harmonic mean is biased downward as much as the arithmetic mean is biased upward while the geometric mean and the median are unbiased. In our second application, we compare three present value methods and establish that the dividend discount model is more precise than the discounted cash flow model when percentage errors are used, whereas logarithmic errors arrive at the opposite conclusion. According to both error measures, the residual income model is the most precise present value approach.

Error measures are determined by the researcher's or user's loss function which is part of their preferences. Therefore, we cannot make any general recommendation as to which error measure should be used. Nevertheless even if the choice of error measure has already been made, our analysis yields a number of insights that help to interpret the results of existing studies and to improve the design of future studies:

- The tendency of percentage errors to favor methods that undervalue becomes stronger as the dispersion among comparables increases. Our results imply that in industries or years with large dispersion the use of the book value as forecast of the market value might even turn out to be the most precise valuation method. Studies that use percentage errors should therefore make sure that they employ an elaborate comparable selection procedure that reduces this dispersion. For logarithmic errors, on the other hand, this is not necessary. We show that the results obtained with log errors are little affected by the dispersion among comparables, because log errors are more robust to outliers. While the removal of outliers is desirable in all cases, it might not be possible or economical; then results are more reliable when log errors are used.
- We demonstrate that the dispersion among comparables strongly affects the biases obtained with percentage errors. The higher the dispersion is (i.e. the less represen-

tative the comparable firms are) the higher is the bias for *all* valuation methods, even for the harmonic mean. In contrast, results obtained with logarithmic errors are not affected by changes in dispersion among comparable firms. Therefore, the result that a particular valuation method is unbiased according to logarithmic errors is much more informative than if it is unbiased according to percentage errors. We suggest that in studies that use percentage errors, the focus should be on the bias relative to other methods rather than on the level of the bias itself.

- In this paper we followed the convention to define the bias as the *mean* error. However, it is also possible to rank valuation methods according to the *median* error. For logarithmic errors, mean and median error always arrive at the same conclusions. For percentage errors, however, this is not true: the median percentage error comes to similar conclusions as the (mean or median) logarithmic error, whereas the mean percentage error yields very different results (as discussed above). As a consequence, studies with percentage errors should also report the median errors as these are indicative of the results that would be obtained with logarithmic errors. Several existing studies already do this.
- Percentage errors and logarithmic errors will produce different rankings among valuation methods whenever these valuation methods differ in the skewness of the error distribution they produce. If the methods compared in a horse race only differ in their precision (i.e. error variance) but not their skewness, the choice of error measure will not influence the results. For instance, if the forward looking P/E ratio is compared with the historical P/E ratio (all else being equal), then both error measures will arrive at the same conclusion, because the choice of the base alone does not affect the skewness (whereas the way of averaging comparable information does as Figure 1 demonstrates). Hence, the choice of error measure is only critical if the candidate methods differ in the skewness of their error distributions.

For researchers or users who are agnostic about the choice of error measure, our results

suggest that logarithmic errors should be preferred to percentage errors. The reason is that ranking valuation methods based on percentage errors leads to a number of counterintuitive consequences that few practitioners would endorse. In particular:

- Ignoring all valuation information except the book value and setting the target firm value equal to its book value results in more precise and less biased forecasts than using the arithmetic mean or the median and becomes optimal relative to any multiple for sufficiently dispersed samples. It also dominates the discounted cash flow model.
- Ignoring all information altogether and setting the target firm value equal to \$1 turns out to be a reasonable valuation method and even becomes optimal if the dispersion of the sample is sufficiently large.
- As the variation among comparable firms increases, the bias of *all* averaging methods increases. Eventually, even the harmonic mean, which tends to produce low valuations, features a large positive bias.

Logarithmic errors avoid these pitfalls. With logarithmic errors, both *ad hoc* methods turn out to be clearly inferior to any valuation method that uses comparable or analyst forecast data, and the rankings obtained in horse-races are independent of the variation among comparable firms. Any reader who subscribes to this view should avoid the arithmetic and the harmonic mean in multiple valuation and apply the geometric mean or the median instead.

A frequently voiced argument against logarithmic errors is that investors are interested in percentage errors. Consider an investor who uses a valuation method in order to form an equally weighted portfolio of undervalued firms. If market prices subsequently move towards the predicted prices, the portfolio return is equal to the initial average percentage error of portfolio firms. Note, however, that the investor is never interested in the bias, i.e. the percentage error of *all* (under- *and* overvalued) firms. Instead, the investor will pick a given number of the most undervalued firms or possibly all undervalued firms. As both error measures generate the same ordering among firms, the choice which firms enter

the investor's portfolio is independent of the error measure used. Therefore, the investor is effectively indifferent between using percentage or logarithmic errors.

Ultimately, the error measure chosen must depend on the application in question. The objective function for a bidder in an auction for a company may be different from that of a security analyst who values a market traded company for investment purposes. Depending on risk aversion, degree of diversification, asymmetric information, and other considerations, practitioners will give different weights to small valuation errors versus large errors. Also, they wish to equate either equal relative mispricings (the case for percentage errors) or equal multiplicative mispricings (the case for log errors), or treat undervaluations and overvaluations differently altogether. Rigorous answers to these questions can only be obtained based on an axiomatic approach that relates decision rules to preferences and to the salient features of the application. We are not aware that such an approach has ever been pursued and believe that this will be a fertile area for future research.

A Distribution of multiples

Figure 2 graphs the standardized distributions of the three multiples and their log transformations. We standardize each observation by deducting the industry-year mean and dividing by the industry-year standard deviation. In the left column of Figure 2, we always compare the empirical distribution of the untransformed data (solid line) with the best fits obtained for the normal distribution (dotted line). In the right column we compare the distribution of the logarithmic transformations of the original data with the normal distribution (dotted line).

[Insert Figure 2 about here]

Clearly, the lognormal distribution is a better model than the normal distribution for all three financial ratios. From a visual inspection of the graph, the lognormal distribution appears to be a reasonable model for the market-to-book ratio and the value-to-sales ratio, but a less convincing model for the price-earnings ratio. For all ratios, the empirical distributions are more skewed and exhibit fatter tails than the lognormal distribution.

We test the fit of the distributions to the data more formally by applying three standard tests for normality to the ratios and their log transformations.¹⁸ These results are reported in Table 7. We also analyze the inverse of the three ratios, because it is well known that the distribution of the earnings-price ratio is closer to a normal distribution than the distribution of the price-earnings ratio.

[Insert Table 7 about here]

All three tests reject the normal distribution as the correct model for the distributions of all three ratios as well as for their logarithmic transformations and their inverses. Hence, we also reject the lognormal distribution as the correct distributional model. This is unsurprising

¹⁸All three tests compare the empirical distribution function with the normal distribution function where the mean and the variance are estimated from the sample. The Kolmogorov-Smirnov test is based on the maximum absolute distance between the two distributions. In contrast, the Cramer-von Mises test and the Anderson-Darling test are based on the expected sum of squared distances under the normal distribution function. While the Cramer-von Mises test gives equal weight to all observations, the Anderson-Darling test gives higher weight to the tails of the distribution. See D’Agnostino and Stephens (1986), p.100.

given that we have a very large dataset, so that the tests have high power. However, a closer look at the test statistics supports the same conclusion already suggested by Figure 2: the lognormal distribution is a much better model for all three financial ratios than the normal distribution, given that the data are highly skewed. The Kolmogorov-Smirnov test statistic falls by between 70% (price-earnings) and 90% (value-to-sales), the Cramér-von Mises and Anderson Darling test statistics even fall by up to 99% for the value-to-sales ratio. By taking the inverses of the ratios, the test statistics can also be reduced, but not by nearly as much as by applying the logarithmic transformation. In all cases the relative improvement of the log transformation is largest for the value-to-sales ratio and smallest for the price-earnings ratio, which is consistent with our visual inspection of these distributions in Figure 2. We therefore conclude that the lognormal distribution works best for the value-to-sales ratio and worst for the price-earnings ratio.

There is also a theoretical reason why some financial ratios are better approximated by the lognormal distribution than others. Variables like market value, book value, sales, total assets, or the number of employees are measures of firm size. They can only be positive and their distribution is highly skewed. If these variables are lognormally distributed (which is an established assumption in empirical applications), then the ratio of two of these variables is also lognormally distributed. This argument does not hold for performance measures, like net income or EBIT.

B Proof of Theoretical Results

B.1 Proof of Result 1

(i) From Jensen's inequality we have:

$$\ln(\bar{x}_j^A) = \ln\left(\frac{1}{n}\sum_{i=1}^n x_i\right) > \frac{1}{n}\sum_{i=1}^n \ln(x_i) = \ln(\bar{x}_j^G). \quad (11)$$

Similarly, we can apply Jensen's inequality to the rewritten harmonic mean:

$$\ln(\bar{x}_j^H) = -\ln\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}\right) < -\frac{1}{n}\sum_{i=1}^n \ln(1/x_i) = \frac{1}{n}\sum_{i=1}^n \ln(x_i) = \ln(\bar{x}_j^G).$$

As the logarithm is a monotonic transformation, we obtain $\bar{x}_j^H < \bar{x}_j^G < \bar{x}_j^A$.

(ii) Let M denote the median and μ the mean of the distribution of the ratios x_i . As the distribution is symmetric, we have $M = \mu$. The law of large numbers implies that sample moments converge to population moments, so $\bar{x}_j^M - \bar{x}_j^A \xrightarrow{a.s.} M - \mu = 0$ as $n \rightarrow \infty$. The convergence is almost surely (a.s.) or with probability one.

(iii) Denote the parameters of the lognormal distribution from which the x_i have been drawn by μ and σ^2 . Then, the median of the distribution is $M = \exp\{\mu\}$ and $E(\ln(x_i)) = \mu$, so that $\bar{x}_j^M - \bar{x}_j^G = \bar{x}_j^M - \exp\left\{\frac{1}{n}\sum_{i=1}^n \ln(x_i)\right\} \xrightarrow{a.s.} M - \exp(\mu) = 0$ as $n \rightarrow \infty$ by the same argument as in (ii).

B.2 Proof of Result 2

We only need to show that $E(e_{\log}^G) = 0$, $E(e_{pct}^G) > 0$, and $E(e_{\log}^H) = -E(e_{\log}^A)$. The remaining statements of the proposition then follow immediately from Result 1.

From (4) we have:

$$\ln(\bar{x}^G) = \frac{1}{n}\sum_{i=1}^n \ln(x_i),$$

so $\ln(\bar{x}^G)$ is distributed normal with mean μ and variance σ^2/n . Hence, $e_{\log}^G = \ln(\bar{x}^G) -$

$\ln(x_j)$ is distributed normal with mean zero and variance $\frac{1+n}{n}\sigma^2$, so the geometric mean leads to unbiased estimates in terms of logarithmic errors. As a consequence, $1 + e_{pct}^G = \exp(e_{\log}^G)$ is distributed lognormal with parameters 0 and $\frac{1+n}{n}\sigma^2$, so we obtain:

$$E(e_{pct}^G) = \exp\left(\frac{1+n}{2n}\sigma^2\right) - 1 > 0,$$

as long as $\sigma^2 > 0$. This shows that the geometric mean leads to biased estimates in terms of percentage errors.

Showing that $E(e_{\log}^H) = -E(e_{\log}^A)$ requires a little more work: First note that

$$E(\ln(\bar{x}^H)) = -E\left(\ln\left(\frac{1}{n}\sum_{i=1}^n \exp(-\ln x_i)\right)\right) = -E\left(\ln\left(\frac{1}{n}\sum_{i=1}^n \exp(-u_i)\right)\right), \quad (12)$$

where $u_i = \ln x_i$, which is distributed normal with expectation μ and variance σ^2 . We expand this expression and perform the substitution $v_i = -u_i + 2\mu$ for all $i = 1..n$.

$$\begin{aligned} E(\ln(\bar{x}^H)) &= \\ &= -\left[\sigma\sqrt{2\pi}\right]^{-n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \ln\left(\frac{1}{n}\sum_{i=1}^n \exp(-u_i)\right) \prod_{i=1}^n \exp\left\{-\frac{1}{2}\left(\frac{u_i - \mu}{\sigma}\right)^2\right\} du_1 \dots du_n \\ &= -\left[\sigma\sqrt{2\pi}\right]^{-n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \ln\left(\frac{1}{n}\sum_{i=1}^n \exp(v_i - 2\mu)\right) \prod_{i=1}^n \exp\left\{-\frac{1}{2}\left(\frac{-v_i + \mu}{\sigma}\right)^2\right\} dv_1 \dots dv_n \\ &= 2\mu - \left[\sigma\sqrt{2\pi}\right]^{-n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \ln\left(\frac{1}{n}\sum_{i=1}^n \exp(v_i)\right) \prod_{i=1}^n \exp\left\{-\frac{1}{2}\left(\frac{v_i - \mu}{\sigma}\right)^2\right\} dv_1 \dots dv_n \\ &= 2\mu - E\left(\ln\left(\frac{1}{n}\sum_{i=1}^n \exp u_i\right)\right) = 2\mu - E(\ln(\bar{x}^A)) \end{aligned}$$

The second line rewrites (12) more explicitly and the third line applies the transformation $u_i = -v_i + 2\mu$. Here, we use the fact that $\frac{dv_i}{du_i} = -1$, which cancels with the factor -1 caused by the necessary transformation of the integration limits. The fourth line follows upon rearranging, and the fifth line rewrites the same expression using the expectations operator. Observe that v_i is also distributed normal with mean μ and variance σ^2 . Hence,

$E(\ln(\bar{x}^H)) - \mu = -[E(\ln(\bar{x}^A)) - \mu]$. As $E(\ln(x_i)) = \mu$, this implies $E(e_{\log}^H) = -E(e_{\log}^A)$ from the definition of log errors (7).

References

- [1] Alford, Andrew W., 1992, *The Effect of the Set of Comparable Firms on the Accuracy of the Price-Earnings Valuation Method*, Journal of Accounting Research 30, no. 1 , pp. 94-109
- [2] Baker, Malcolm, and Richard S. Ruback, 1999, *Estimating Industry Multiples*, Mimeo, Harvard University, (June)
- [3] Basu, Sudipta, and Stanimir Markov, 2003, *Loss Function Assumptions in Rational Expectations Tests on Financial Analysts' Earnings Forecasts*, Mimeo, Emory University
- [4] Beatty, Randolph P.; Susan M. Riffe, and Rex Thompson, 1999, *The Method of Comparables and Tax Court Valuations of Private Firms: An Empirical Investigation*, Accounting Horizons 13, no. 3 , pp. 177-199
- [5] Bhojraj, Sanjeev, and Charles M. C. Lee, 2002, *Who Is My Peer? A Valuation-Based Approach to the Selection of Comparable Firms*, Journal of Accounting Research 40, no. 2 (September), pp. 407-439
- [6] Cheng, C. S. Agnes, and Ray McNamara, 2000, *The Valuation Accuracy of the Price-Earnings and Price-Book Benchmark Valuation Methods*, Review of Quantitative Finance and Accounting 15, pp. 349-370
- [7] Claus, James, and Jacob Thomas, 2001, *Equity Premia as Low as Three Percent? Evidence from Analysts' Earnings Forecasts for Domestic and International Stock Markets*, Journal of Finance 56, pp. 1629-1666
- [8] D'Agostino, Ralph B.; Michael A. Stephens, 1986, *Goodness-of-fit techniques*, New York (Marcel Dekker)
- [9] Dechow, Patricia M.; Amy P. Hutton, and Richard G. Sloan, 1999, *An empirical assessment of the residual income valuation model*, Journal of Accounting and Economics 26, 1-34
- [10] Francis, Jennifer; Per Olsson, and Dennis R. Oswald, 2000, *Comparing the Accuracy and Explainability of Dividend, Free Cash Flow and Abnormal Earnings Equity Value Estimates*, Journal of Accounting Research 38, pp. 45-70
- [11] Gilson, Stuart C.; Edith S. Hotchkiss, and Richard S. Ruback, 2000, *Valuation of Bankrupt Firms*, Review of Financial Studies 13, no. 1 , pp. 43-74
- [12] Herrmann, Volker, and Frank Richter, 2003, *Pricing With Performance-Controlled Multiples*, Schmalenbach Business Review 55, (July), pp. 194-219
- [13] Kaplan, Steven N., and Richard S. Ruback, 1995, *The Valuation of Cash Flow Forecasts: An Empirical Analysis*, Journal of Finance 50, no. 4 (September), pp. 1059-1093
- [14] Kim, Moonchul, and Jay R. Ritter, 1999, *Valuing IPOs*, Journal of Financial Economics 53, pp. 409-437

- [15] Lie, Erik, and Heidi J. Lie, 2002, *Multiples Used to Estimate Corporate Value*, Financial Analysts Journal, pp. 44-54
- [16] Liu, Jing; Doron Nissim, and Jacob Thomas, 2002a, *Equity Valuation Using Multiples*, Journal of Accounting Research 40, no. 1 (March), pp. 135-172
- [17] Liu, Jing; Doron Nissim, and Jacob Thomas, 2002b, *International Equity Valuation Using Multiples*, Mimeo, University of California at Los Angeles
- [18] Ohlson, James A., 1995, *Earnings, Book Values, and Dividends in Equity Valuation*, Contemporary Accounting Research 11, pp. 661-687
- [19] Penman, Stephen H., and Theodore Sougiannis, 1998, *A Comparison of Dividend, Cash Flow, and Earnings Approaches to Equity Valuation*, Contemporary Accounting Research 15, no. 3 (Fall), pp. 343-383
- [20] Rodriguez, Marius del Giudice, 2005, *Financial Analysts' Incentives and Forecast Biases*, Mimeo, University of California at San Diego.

Table 1:
Descriptive Statistics for Multiples

This table displays descriptive statistics (number of observations, mean, median, standard deviation, skewness, excess kurtosis, and the 10% and 90% quantile) for the distributions of the market-to-book ratio, the value-to-sales ratio, and the price-earnings ratio for the pooled sample from 1994 to 2003. The lower part of the table shows the statistics for the natural logarithms of these ratios. For each multiple and each year, the multiples have been winzorized at the 1% and 99% quantile.

Multiple	# obs.	Mean	Std. dev.	Skewness	Kurtosis	P10	Median	P90
market-to-book	47,614	3.222	4.582	5.173	40.187	0.682	1.867	6.601
value-to-sales	51,899	4.061	9.690	9.493	129.104	0.392	1.634	8.027
price-earnings	33,753	34.755	71.490	5.980	40.997	7.520	17.073	57.317
log(market-to-book)	47,614	0.693	0.919	0.327	0.600	-0.382	0.624	1.887
log(value-to-sales)	51,899	0.553	1.216	0.325	0.165	-0.937	0.491	2.083
log(price-earnings)	33,753	2.947	0.935	0.698	2.471	2.018	2.837	4.049

Table 2:
Empirical error distributions for valuations
based on the market-to-book ratio

This table displays descriptive statistics of the valuation errors from six valuation methods based on the market-to-book ratio. It is calculated from 47,614 firm-year observations from 1994 to 2003, and shows the mean, median, root mean squared error, the standard deviation, skewness, the proportion of observations larger than or equal to +100%, and the t-statistic of the two sided t-test that the mean equals zero. For the methods 'arithmetic mean,' 'median,' 'geometric mean,' and 'harmonic mean,' the industry peer group market-to-book ratios are averaged with the respective method and the result is multiplied by the target firm's book value to arrive at a forecast of the target firm's market value. The method 'MTB=1' sets the target firm's market value of equity equal to its book value, and the method 'Value=\$1' sets the target firm's market value of equity equal to \$1. Panel A shows the results for log errors, Panel B for percentage errors, and Panel C for absolute percentage errors.

Panel A: Log errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	0.318	0.277	0.961	0.907	-0.002	0.202	76.48
Median	-0.018	0.002	0.891	0.891	-0.223	0.106	-4.47
Geometric mean	-0.001	0.018	0.882	0.882	-0.227	0.109	-0.34
Harmonic mean	-0.296	-0.228	0.944	0.896	-0.415	0.056	-71.95
MTB=1	-0.693	-0.624	1.182	0.958	-0.362	0.030	-157.67
Value=\$1	-18.788	-18.676	18.919	2.222	-0.247	0.000	-1841.93

Panel B: Percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	1.211	0.319	6.964	6.858	72.757	0.305	38.48
Median	0.530	0.002	5.341	5.315	94.589	0.180	21.71
Geometric mean	0.538	0.018	5.127	5.099	91.479	0.183	22.98
Harmonic mean	0.135	-0.204	3.879	3.877	99.864	0.106	7.61
MTB=1	-0.175	-0.464	3.873	3.869	107.339	0.055	-9.87
Value=\$1	-1.000	-1.000	1.000	0.000	26.598	0.000	-9.7E+08

Panel C: Absolute percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	1.467	0.554	6.964	6.808	74.267	0.305	46.95
Median	0.941	0.461	5.341	5.258	97.478	0.180	38.99
Geometric mean	0.938	0.463	5.127	5.040	94.456	0.183	40.56
Harmonic mean	0.712	0.461	3.879	3.813	104.492	0.106	40.70
MTB=1	0.692	0.555	3.873	3.811	111.660	0.055	39.56
Value=\$1	1.000	1.000	1.000	0.000	-26.598	0.000	9.7E+08

Table 3:
Empirical error distributions for valuations
based on the value-to-sales ratio

This table displays descriptive statistics of the valuation errors from six valuation methods based on the value-to-sales ratio. It is calculated from 51,899 firm-year observations from 1994 to 2003, and shows the mean, median, root mean squared error, the standard deviation, skewness, the proportion of observations larger than or equal to +100%, and the t-statistic of the two sided t-test that the mean equals zero. For the methods 'arithmetic mean,' 'median,' 'geometric mean,' and 'harmonic mean,' the industry peer group value-to-sales ratios are averaged with the respective method and the result is multiplied by the target firm's sales to arrive at a forecast of the target firm's enterprise value. The method 'Value=Sales' sets the target firm's enterprise value equal to its sales, and the method 'Value=\$1' sets the target firm's enterprise value equal to \$1. Panel A shows the results for log errors, Panel B for percentage errors, and Panel C for absolute percentage errors.

Panel A: Log errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	0.435	0.357	1.140	1.053	0.123	0.258	93.89
Median	-0.025	0.004	0.996	0.996	-0.395	0.122	-5.65
Geometric mean	0.004	0.024	0.988	0.988	-0.373	0.130	0.83
Harmonic mean	-0.367	-0.269	1.080	1.015	-0.718	0.062	-82.31
Value=Sales	-0.555	-0.491	1.370	1.253	-0.379	0.089	-100.71
Value=\$1	-19.009	-18.892	19.142	2.257	-0.268	0.000	-1915.35

Panel B: Percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	1.984	0.428	8.133	7.888	23.422	0.360	57.20
Median	0.638	0.004	3.962	3.911	39.864	0.198	37.09
Geometric mean	0.671	0.024	3.879	3.821	36.722	0.206	39.94
Harmonic mean	0.121	-0.236	2.409	2.406	42.597	0.110	11.48
Value=Sales	0.211	-0.388	3.500	3.494	43.786	0.150	13.77
Value=\$1	-1.000	-1.000	1.000	0.000	119.448	0.000	-7.3E+08

Panel C: Absolute percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	2.235	0.645	8.133	7.820	23.942	0.360	64.99
Median	1.067	0.492	3.962	3.816	42.582	0.198	63.61
Geometric mean	1.087	0.497	3.879	3.724	39.342	0.206	66.38
Harmonic mean	0.743	0.488	2.409	2.292	48.495	0.110	73.76
Value=Sales	0.999	0.683	3.500	3.355	48.758	0.150	67.70
Value=\$1	1.000	1.000	1.000	0.000	-119.448	0.000	7.3E+08

Table 4:
Empirical error distributions for valuations
based on the price-earnings ratio

This table displays descriptive statistics of the valuation errors from six valuation methods based on the price-earnings ratio. It is calculated from 33,753 firm-year observations from 1994 to 2003, and shows the mean, median, root mean squared error, the standard deviation, skewness, the proportion of observations larger than or equal to +100%, and the t-statistic of the two sided t-test that the mean equals zero. For the methods 'arithmetic mean,' 'median,' 'geometric mean,' and 'harmonic mean,' the industry peer group price-earnings ratios are averaged with the respective method and the result is multiplied by the target firm's net income to arrive at a forecast of the target firm's market value. The method 'P/E=10' sets the target firm's market value of equity equal to ten times its net income, and the method 'Value=\$1' sets the target firm's market value of equity equal to \$1. Panel A shows the results for log errors, Panel B for percentage errors, and Panel C for absolute percentage errors.

Panel A: Log errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	0.396	0.359	1.107	1.033	-0.173	0.221	70.10
Median	-0.054	0.004	0.968	0.967	-0.505	0.081	-10.14
Geometric mean	-0.001	0.056	0.970	0.970	-0.518	0.090	-0.17
Harmonic mean	-0.318	-0.188	1.049	1.000	-0.772	0.048	-58.28
P/E=10	-0.648	-0.535	1.198	1.007	-0.752	0.029	-117.70
Value=\$1	-19.194	-19.144	19.321	2.204	-0.140	0.000	-1594.55

Panel B: Percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	2.232	0.433	18.700	18.566	29.987	0.331	22.01
Median	0.878	0.004	11.641	11.608	39.050	0.137	13.85
Geometric mean	0.949	0.057	11.346	11.306	36.159	0.158	15.37
Harmonic mean	0.381	-0.171	8.009	8.000	41.555	0.083	8.72
P/E=10	0.036	-0.415	6.601	6.601	41.423	0.046	0.99
Value=\$1	-1.000	-1.000	1.000	0.000	31.443	0.000	-6.7E+08

Panel C: Absolute percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
Arithmetic mean	2.454	0.612	18.700	18.538	30.088	0.331	24.23
Median	1.265	0.382	11.641	11.572	39.313	0.137	20.02
Geometric mean	1.311	0.408	11.346	11.270	36.414	0.158	21.30
Harmonic mean	0.937	0.404	8.009	7.955	42.070	0.083	21.57
P/E=10	0.855	0.488	6.601	6.545	42.110	0.046	23.91
Value=\$1	1.000	1.000	1.000	0.000	-31.443	0.000	6.7E+08

Table 5:**The influence of an increase of dispersion on the error distributions**

This table displays descriptive statistics of the simulated valuation errors from six valuation methods based on the market-to-book ratio. It shows the mean, median, standard deviation, and the proportion of observations larger than or equal to +100%. In each of the 100,000 runs, we simulate an industry with 40 companies and value each of these companies using comparable information from the remaining 39 firms. For the methods 'arithmetic mean', 'median', 'geometric mean', and 'harmonic mean', the 39 market-to-book ratios are averaged with the respective method and the result is multiplied by the target firm's book value to arrive at a forecast of the target firm's market value. The method 'MTB=1' sets the target firm's market value equal to its book value, and the method 'Value=\$1' sets the target firm's market value equal to \$1. For the "Base scenario" we assume that market-to-book ratios are lognormally distributed with mean 0.593 and standard deviation 0.838, and that the book value is lognormally distributed with mean 18.329 and standard deviation 1.838. In the simulations with "Increased dispersion" we increase the standard deviation of the log market-to-book ratio to 1.115 (+33%). Panel A shows the results for log errors, and Panel B for percentage errors.

Panel A: Log errors

Method	Base scenario				Increased dispersion			
	Mean	Median	Std. dev.	Prop \geq 100%	Mean	Median	Std. dev.	Prop \geq 100%
Arithm. mean	0.338	0.338	0.852	0.219	0.593	0.593	1.140	0.360
Median	0.000	0.000	0.854	0.121	0.000	0.000	1.137	0.189
Geom. mean	0.000	0.000	0.848	0.119	0.000	0.000	1.129	0.188
Harm. mean	-0.338	-0.338	0.852	0.058	-0.592	-0.591	1.140	0.081
MTB=1	-0.593	-0.593	0.838	0.029	-0.593	-0.593	1.115	0.076
Value=\$1	-18.921	-18.921	2.035	0.000	-18.923	-18.922	2.164	0.000

Panel B: Percentage errors

Method	Base scenario				Increased dispersion			
	Mean	Median	Std. dev.	Prop \geq 100%	Mean	Median	Std. dev.	Prop \geq 100%
Arithm. mean	1.017	0.402	2.087	0.338	2.465	0.809	5.647	0.465
Median	0.441	0.000	1.493	0.208	0.908	0.000	3.087	0.271
Geom. mean	0.433	0.000	1.472	0.207	0.890	0.000	3.021	0.270
Harm. mean	0.026	-0.287	1.060	0.113	0.058	-0.446	1.715	0.130
MTB=1	-0.215	-0.447	0.793	0.062	0.029	-0.447	1.609	0.124
Value=\$1	-1.000	-1.000	0.000	0.000	-1.000	-1.000	0.000	0.000

Table 6:
Comparison between present value methods based on
dividends, residual income, or cash flows

This table displays descriptive statistics of the valuation errors from three present value methods: The dividend discount model (DDM), the residual income model (RIM), and the discounted cash flow model (DCF). In addition, results for two benchmark methods are presented: The method MTB=1 sets the target firm's market value of equity equal to its book value, and the method 'Value=\$1' sets the target firm's market value of equity equal to \$1. The table is calculated from 654 firm-year observations from 1994 to 2005, and shows the mean, median, root mean squared error, the standard deviation, skewness, the proportion of observations larger than or equal to +100%, and the t-statistic of the two sided t-test that the mean equals zero. Panel A shows the results for log errors, Panel B for percentage errors, and Panel C for absolute percentage errors.

Panel A: Log errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
DDM	-0.912	-0.743	1.322	0.958	-1.357	0.005	-24.35
RIM	0.078	0.089	0.524	0.518	-0.847	0.034	3.86
DCF	0.735	0.730	1.001	0.680	-1.105	0.329	27.66
MTB=1	-7.467	-5.077	10.268	7.054	-1.487	0.003	-27.07
Value=\$1	-22.594	-22.615	22.640	1.446	0.035	0.000	-399.62

Panel B: Percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
DDM	-0.435	-0.525	0.620	0.443	2.095	0.009	-25.13
RIM	0.231	0.093	0.823	0.791	6.453	0.070	7.46
DCF	1.605	1.076	2.844	2.350	7.142	0.523	17.46
MTB=1	-0.913	-0.994	0.956	0.283	8.053	0.003	-8.2E+01
Value=\$1	-1.000	-1.000	1.000	0.000	6.067	0.000	-2.8E+10

Panel C: Absolute percentage errors

Method	Mean	Median	RMSE	Std. dev.	Skewness	Prop \geq 100%	T-test
DDM	0.539	0.552	0.620	0.308	1.094	0.009	44.73
RIM	0.432	0.254	0.823	0.701	8.609	0.070	15.78
DCF	1.661	1.076	2.844	2.310	7.479	0.523	18.39
MTB=1	0.939	0.994	0.956	0.180	1.126	0.003	1.3E+02
Value=\$1	1.000	1.000	1.000	0.000	-6.067	0.000	2.8E+10

Table 7:
Tests for Normality

This tables shows the test statistics of three tests for normality (Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling) applied to the pooled sample (1994-2003) of three different financial ratios, their logarithmic transformations, and their inverses. All observations have been standardized by deducting the industry-year mean and then dividing by the industry-year standard deviation, where the industry is given by the two-digit SIC code. For the Kolmogorov-Smirnov test, the table displays the usual test statistic multiplied by the square root of the number of observations in order to make comparisons across samples meaningful. For the thus transformed Kolmogorov-Smirnov test statistic, the 1% critical value is 1.035. For the Cramér-von Mises and the Anderson-Darling test the 1% critical value are 0.179 and 1.035, respectively. Critical values have been obtained from D'Agostino and Stephens (1986), p.123.

Multiple	Test statistics		
	Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
market-to-book	40.81	683.06	3,734.04
value-to-sales	48.07	910.45	4,800.33
price-earnings	50.21	957.28	4,862.20
log(market-to-book)	4.61	7.16	44.87
log(value-to-sales)	5.04	8.76	58.02
log(price-earnings)	14.56	74.61	426.01
(market-to-book) ⁻¹	31.31	390.97	2,201.48
(value-to-sales) ⁻¹	37.45	579.93	3,209.81
(price-earnings) ⁻¹	31.11	371.94	2,048.24

Figure 1:
Empirical error distributions

The figure shows the empirical error distributions based on the market-to-book ratio, calculated with 100 histogram intervals from 47,614 firm-year observations from 1994-2003. The left column shows the percentage errors, whereas the right column shows log errors. The rows correspond to one averaging method each. The data have been truncated to generate meaningful plots.

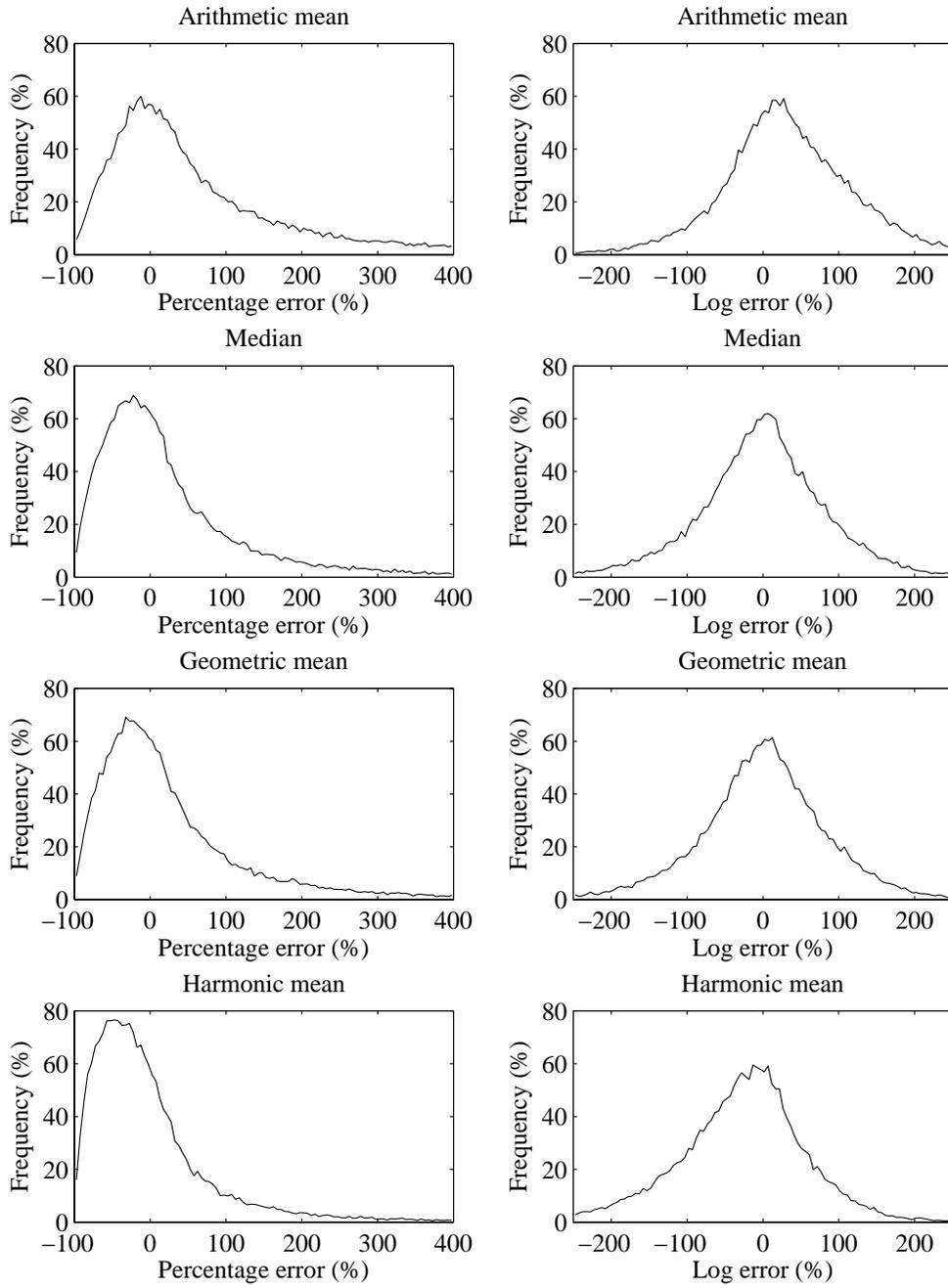


Figure 2:
Empirical density functions of financial ratios

The figure shows the empirical distributions of the financial ratios (left column) and their logarithmic transformations (right column), calculated with 100 histogram intervals. In each plot, the solid line shows the actual distribution of the data and the dotted line shows the density function of a normal distribution with mean and variance fitted to the data. All observations have been standardized by deducting the industry-year mean and then dividing by the industry-year standard deviation, where the industry is given by the two-digit SIC code. The data have been truncated to generate meaningful plots.

