UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

New Insights into Equity Valuation Using Multiples

PhD Thesis submitted to the Faculty of Economics and Business

Information Management Institute

University of Neuchâtel

For the degree of PhD in Statistics

by

Jian KANG

Accepted by the dissertation committee:

Prof Catalin Starica, University of Neuchatel, thesis director

Prof Kilian Stoffel, University of Neuchatel

Prof Alain Schatt, University of Lausanne

Prof Lucie Courteau, University of Bozen-Bolzano, Italy

Prof Jacob Thomas, Yale University, USA

Defended on 28th, Jan, 2016

Faculté des Sciences Economiques
Avenue du 1er-Mars 26
CH-2000 Neuchâtel
www.unine.ch/seco

UNIVERSITÉ DE
NEUCHÂTEL

IMPRIMATUR POUR LA THÈSE

New insights into equity valuation using multiples

Jian KANG

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,
sur le rapport des membres du jury

Prof. Catalin Starica (directeur de thèse, Université de Neuchâtel)
Prof. Kilian Stoffel (président du jury, Université de Neuchâtel)
Prof. Alain Schatt (Université de Lausanne)
Prof. Lucie Courteau (University of Bozen-Bolzano)
Prof. Jacob Thomas (University of Yale)

Autorise l'impression de la présente thèse.

Neuchâtel, le 16 février 2016

La doyenne

Carolina Salva

New insights into equity valuation using multiples

JIAN KANG

Abstract

The thesis focuses on the equity valuation using multiples. Based on the notion of stochastic dominance, a relative valuation framework for comparing accuracy is developed and employed in the thesis.

In the first paper, we investigate the feasibility of a relative valuation framework for tracking developments of value relevance of earnings and book values across time.

The second paper focuses on the performance of EV/EBITDA, a multiple that gained in popularity with practitioners during the last decade, relative to that of the traditional multiple P/E.

The last paper explores the possibility of applying non-linear methods to improve the precision of multiple valuation.

Keywords: Equity valuation, multiples, accuracy, stochastic dominance, non-linear, P/E, P/B, EV/EBITDA

Acknowledgment

The experience of my PhD study at the University of Neuchatel is one of the largest assets in my life. First and foremost, I am grateful to my supervisor Catalin Starica for his guidance, supervision and support for last 66 months. His enlightening comments and advises guided me to discover and solve interesting questions, not only in academic, but also in life. It is my great honor and fortune to have met Catalin. The thesis can never be completed without his help and contributions, and I can never stand where I am today without his lead.

I would like to thank my co-author Alain Schatt, who inspired me how to compose pieces of evidence from a perspective of accounting. The paper we worked on together started my journey in the field of accounting. Thanks to Kilian Stoffel for having accepted to be the president of my dissertation committee, Lucie Courteau and Jacob Thomas for have accepted to review and comment this thesis.

I also thank the University of Neuchatel for providing me a terrific research environment. I would like to express my gratitude to my colleagues in the university.

Finally, I want to thank my wife, Li Gao, and my parents for their continuous support and selfless dedication.

Contents

1

Summary of papers

1.1 Paper I

The aim of this paper is to investigate the feasibility of a relative valuation framework for tracking developments of value relevance of earnings and book values across time. In such a framework, the measure of relevance of earnings or book values is the precision of the equity valuation using multiple P/E, respectively P/B. A loss of relevance of the accounting variable would be documented by a decrease through time of the corresponding multiple valuation precision.

To render such an approach operational, we optimize the implementation of the valuation by multiples with respect to precision of valuation (characteristics and number of peers, aggregation of peer multiples). Extending the existing literature that compares accuracy using particular statistics of the distribution of valuation errors, we develop a framework for comparing accuracy based on the notion of stochastic dominance. The relative valuation framework offers an alternative to the linear regression approach impaired by a number of econometric weaknesses when employed in the frame of value relevance studies.

Using this framework, we document a reduction of valuation accuracy of both P/E and P/B multiples that affect mostly the large firms after 1990. In the relative valuation framework for value relevance, we interpret this finding as evidence for a significant reduction of the value relevance of earnings and book values for large firms in the last twenty five years.

1.2 Paper II

This paper focuses on the equity valuation using multiples P/E and EV/EBITDA. We compare the accuracy of firm valuation based on EV/EBITDA, a multiple gained in popularity with practitioners during the last decade, with that of the traditional multiple P/E.

The paper is motivated by the evolution in analysts' preferences. Recent surveys denote that the preferences of choosing multiples of practitioners are changed in the United States (Block, 2010) and in Europe (Bancel & Mittoo, 2014), i.e. more analysts tend to use the multiple EV/EBITDA in equity valuation.

Our analysis shows that valuations using P/E is more accurate than EV/EBITDA when aggregating over all observations. The result is robust to the implementation of the valuation method (characteristics and number of peers, aggregation of peer multiples), the type of multiple drivers (historical numbers or forecasts), the period (recent years or not), and the industry type (capital intensive industry or not). When different firm characteristics are taken into account, the relation becomes more intricate. For companies with low debt, EV/EBITDA is at least as precise as P/E. Moreover, EV/EBITDA is significantly more accurate than P/E when valuing firms that report largely negative special items and/or non-operating items. This finding is in line with the idea that not all items convey information to the investors. Overall, we conclude that EV/EBITDA leads to more accurate valuation in some specific cases.

1.3 Paper III

This paper revisits the price-to-earnings multiple valuation approach. Interpreting the median industry earning ratio as a quantity close to the expected return on equity and heeding the prescriptions of valuation models commonly used in the accounting literature, we motivate the need for a non-linear prediction of the multiple. We propose the use of multiple industry median ratios as predicting variables.

We show that non-linearly forecasting the multiple using three industry median multiple constructed according to different depth definition of industry (one, two or three SIC digits) sharply improves the accuracy of the prediction of the multiple. Since the relative pricing error for multiples that are ratios of prices to value driver is equal to the relative prediction error for the multiple, the non-linear prediction of the multiple greatly enhances pricing precision. We report a reduction of 47% of the median absolute pricing error, of 45% for the 75%-ile, and of 40% for the 90%-ile of the absolute error with respect to plain-vanilla multiple pricing.

Our findings reposition the multiple pricing with respect to the more comprehensive valuation approaches. Although not evaluated on the same sample, the performance of the non-linearly enhanced price-to-earnings multiple pricing compares favorable with that of more comprehensive valuation methods as reported in the literature.

We argue that, while price-to-earnings valuation might be an 'imperfect heuristic', when strengthened by the understanding of accounting valuation models, it becomes a precise pricing method.

2

Included Papers

2.1 Paper I: Have earnings and book values lost relevance? A relative valuation framework

Have earnings and book values lost relevance? A relative valuation framework

Jian Kang and Catalin Starica

School of Economics and Business, University of Neuchâtel

Abstract

We inquire into the suitability of a relative valuation framework for tracking developments of value relevance of earnings and book values across time. In such a framework the measure of relevance of earnings or book values is the precision of P/E, respectively P/B, valuation. A loss of relevance of the accounting variable would be documented by a decrease through time of the corresponding multiple valuation precision.

To render such an approach operational we optimize the implementation of the valuation by multiples with respect to precision of valuation. We argue that the concept of stochastic dominance gives the right statistical frame for measuring and comparing the accuracy of equity valuation.

We document a reduction of valuation accuracy of both P/E and P/B multiples that affect mostly the large firms after 1990. In the relative valuation framework for value relevance, we interpret this finding as evidence for a significant reduction of the value relevance of earnings and book values for large firms in the last twenty five years.

Keywords: Valuation, multiples, P/E, P/B, accuracy, stochastic dominance, value relevance.

1 Introduction

Temporal changes have been a recurrent issue in the extensive literature on value relevance of accounting information. In a thorough discussion on interpreting value relevance in academic research, Francis and Schipper (1999) identifies four main alternatives. In the frame of studying developments in value relevance, the most common among the four is that of association between financial information (i.e. accounting variables) and share prices or returns (Interpretation 4). This specific interpretation of value relevance allows for different operationalisations based on two main choices. First, the researcher needs to chose a way to formalize the relationship between value and financial information. Second, a measure of relevance needs to be defined. Developments in the value relevance are inferred from the time evolution of the proposed measure. By far, the most common choice on the first issue is that of a linear link between accounting variables and value based on Ohlson's accounting based valuation model. To study if one particular type of accounting information is significantly related to the market value, one linearly regresses prices on the accounting data and evaluates the significance level of individual regression coefficients and/or the explanatory power of the regression model.

The aim of this paper is to investigate the feasibility of a relative valuation framework as an alternative to the linear regression set-up. In such a framework, the relation between the accounting variable and value is established through the relative valuation process based on the corresponding multiple. The relevance of earnings or book values is reflected in the precision of the P/E, respectively P/B, valuation. A loss of relevance would be measured as a decrease over time of the multiple valuation precision.

To render such an approach operational one needs to first optimize the implementation of valuation by multiples. The optimal choices would result in a valuation that is most accurate and whose errors do not vary due to the details of the implementation of the method.

The multiple valuation approach estimates a firm's stock price as the product of a financial measure from either the income statement or the balance sheet, the so-called value driver, and the corresponding multiple based on the ratio of stock price to value driver for a group of comparable firms. The foundation of the multiple approach is the law of one price: perfect substitutes should earn the same price.

While intuitively appealing, the implementation of a multiple based valuation is not straight forward. It requires three choices with no clear-cut guidance. First is that of the quantity to price. This choice is made qualitatively among measures of revenues, earnings or book value. The multiple is then the market price of a single unit. The second choice regards the substitutes for the firm to be priced, i.e. firms comparable to the target company and it includes two aspects: the criterion for peer selection as well as the size of the set of comparable companies. The substitutes are commonly firms within the same industry or firms with the same risk, growth or cash flow profiles or intersections of these criteria. The decision on the size of the set of substitutes is made qualitatively. Finally, the method to calculate the target's multiple from peer multiples needs to be decided. Common choices are the arithmetic mean, the median or the harmonic mean.

The first goal of the paper is hence to optimize these choices with respect to the precision of valuation[1]. Performing the optimization hinges on a new overall measure of accuracy based on the notion of stochastic dominance. The methodological improvements are illustrated by a detailed analysis of the valuation accuracy of the price-to-earnings and price-to-book multiples, ratios that will play a key role in the construction of the relative valuation framework for value relevance.

When all choices are made optimally and a comprehensive measure of valuation accuracy is in place, relative valuation can be used to investigate the time evolution of the value relevance of the value driver. The second goal of the paper is to operationalize such a relative valuation approach for tracking developments in the value relevance of earnings and book values. The relative valuation framework we introduce offers an alternative to the linear regression approach known to be impaired by a number of econometric weaknesses when employed in the frame of value relevance studies (see Brown, Lo, and Lys (1999), Easton and Sommers (2003), Gu (2007)).

The methodological adjustments we propose with regard to employable multiple

[1] To clarify, valuation performance here does not refer to picking mis-priced stocks. We focus instead on how close valuations based on multiples are to traded prices.

valuation are motivated by the observation that the existing literature does not address sufficiently two important aspects of this type of valuation. First one concerns how to optimally implement the choices that makes multiple valuation operational. In particular, no comprehensive study exists on the effect on valuation accuracy of the number of companies in the set of comparable firms.

The second account concerns the evaluation of accuracy. Fundamental question as: what is a comprehensive measure of precision? and how to comprehensively test the relative performance of different valuations? do not receive satisfactory answers. Existing literature discusses accuracy using only partial information about the valuation errors: particular statistics of the distribution of errors, like the median, the interquartile range, the standard deviation or differences between two quantiles, or particular cumulative estimated probabilities are used to asses and to compare performance. The statistical tests are based on ad-hoc choices of measures of performance: median absolute prediction errors (Alford, 1992 and Cheng & McNamara, 2000) or interquartile range (Liu, Nissim, & Thomas, 2002 and Liu, Nissim, & Thomas, 2007).

Our contribution to the literature on multiple valuation is hence two fold. Firstly, we conduct a systematic investigation on the impact of all implementation choices on the precision of multiple valuation. We propose a method for an optimal choice of the number[2] of firms in the set of comparable companies. We systematically investigate the impact of the method of constructing the target multiple from the multiples of the comparable companies.

Secondly, we develop a framework for judging accuracy that allows for an overall evaluation of performance. We argue that the concept of stochastic dominance gives the right statistical frame for measuring and comparing the accuracy of equity valuation. This concept is used to, first, determine the optimal number of comparable companies to use for each peer selection approach and second, to compare the overall distributions of pricing relative errors. The stochastic dominance frame allows for rigorous testing of relative performance and is relevant to all comparisons of valuation approaches, independent of their relative or absolute nature.

[2]As we will see in the sequel, the optimal number of comparable firms depends on the peer selection approach but not on the multiple used.

We find that the valuation precision is affected by both the choice of multiple and that of the peer selection criteria. For a given multiple/peer selection criterion, a wise choice of peer selection criterion/multiple improves the performance. Somewhat surprising, the optimal number of comparable firms to be used in valuation is peer selection criterion specific and independent of the multiple. For a given criterion, the optimal choice of the size of the set of peers to be considered in valuation does not depend on the multiple. The use of the harmonic mean for constructing the target multiple improves the precision of valuation only marginally.

Unlike previous studies but in line with recent theoretical work and empirical evidence regarding the importance of book value, we find that the P/B multiple is at least as good as the popular P/E ratio[3]. In line with more recent multiple valuation literature, we find that industry membership is not the most effective criteria[4] for selecting comparable companies. Earnings growth or a combination of earnings growth with either size or industry are better peer selection criteria.

The methodological improvements regarding employable multiple valuation open a fresh perspective on the issue of value relevance of earnings and book values.

As mentioned, most studies that investigate changes in value relevance across time use the linear regression to relate value to the financial information and the explanatory power of the regression model as a measure of value relevance (Collins, Maydew, & Weiss, 1997, Francis & Schipper, 1999, Ely and Waymire (1999), Lev & Zarowin, 1999, Dontoh, Radhakrishnan, & Ronen, 2004, Kim & Kross, 2005).

Recent literature has pointed, however, to several econometric challenges related to such a use of regression model's explanatory power in value relevance research. Brown et al. (1999) indicate that scale effects present in price regressions increase

[3]Cheng and McNamara (2000), the only study we are aware of that considers an exhaustive set of peer selection methods, and Liu et al. (2002) conclude that earnings perform better than book value. Cheng and McNamara (2000) use significantly fewer peers than the optimal number when selecting firms based on return on equity, the criterion that we find performs best for P/B. Liu et al. (2002) do not consider this criterion at all and focus only on choosing similar firms based on industry. For more details see Section 4.1.1.

[4]Our results confirm those of Cheng and McNamara (2000) who find that peer selection based on industry membership and earning growth is significantly better than all the other definition considered, industry included. Their analysis does not however consider the peer selection method that combines earning growth and risk as measured by Total Assets which we find superior when paired with the best performing ratios. Alford (1992) found a not significant superior accuracy of industry membership criteria for price-earnings valuations over other criteria. His results could be due, as argued in the sequel, to the small size of the sample on which the analysis was conducted.

the explanatory power. As a consequence, differences in R^2 between samples from different time periods may be due to differences in the scale factors and not to changes in value relevance. Gu (2007) argues that scale effects are not the only reason why explanatory power is not comparable across samples and should not be used to measure developments in the value relevance. He states that "the R^2s could be different even though the economic relation is entirely intact for each and every observation in two samples".

For questions related to developments of the relevance of financial statements across time, few value relevance measures alternative to price regression explanatory power have been proposed in the literature, all of them in the linear regression frame. Most prominently, Gu (2007) employs price regression's residual dispersion with proper control for scale as an alternative measure and robustly detects a decline of value relevance since the early 1970s.

The contribution of our paper the extensive literature on value relevance is two-fold and touches upon both choices faced by the researcher who studies the association amongst financial information and share prices (interpretation 4 in Francis and Schipper (1999)). First, we contribute by investigating the feasibility of a relative valuation framework as an alternative to price and return regression. In such a framework the relevance of earnings or book values is the gauged by the precision of the P/E, respectively P/B, valuation. A loss of relevance would be inferred from a decrease over time of the multiple valuation precision. We show that, in the relative valuation framework, the scale effects that plague the linear regression specifications are significantly reduced. The only reference we are aware of that uses a valuation framework is Chang (1998). Assessing value relevance based on discounted residual income valuation he concludes that the combined relevance of earnings and book values has decreased over time.

Second, we develop a more effective measure of value relevance based on the notion of stochastic dominance. Following Gu (2007) we argue that, after the systematic component, identified in our case by the relative valuation, is taken out, the distribution of the error term measures explicitly the unexplained noise level of the economic relation. As long as the economic relation does not change, the distribution of errors should remain constant. A weaker economic relationship would translate in larger errors. Gu (2007) proposes the residual standard deviation as a measure of relevance and infers on developments of value relevance based on the time evolution of the standard deviation of errors. We argue that a measure based on the notion of stochastic dominance generalizes the choice of standard deviation and gives a more comprehensive measure of precision that allows for accurate statistical testing of change.

More precisely, we analyze how the precision of the P/E and P/B valuations has evolved in the last forty five years. We bring evidence that the overall accuracy of the two valuations methods has declined, fact that, taken at face value, would be interpreted within the valuation framework as a decrease in value relevance of earnings and book values. Changes in the structure of the sample or scale effects could be responsible for the apparent loss of valuation accuracy. We show that the decrease in precision is not due to changes in the structure of the sample. The varying proportion of small/large firms or the changes in the proportion of firms that report one-time items do not contribute significantly to the reported decrease in valuation accuracy. We control for the scale effect by making inferences based on relative valuation errors, i.e. pricing errors are standardized with the level of the price to obtain comparable value relevance measures. We show that, in contrast to price regressions, the scale effect for the relative valuation framework is very weak, practically nonexistent. As a conclusion, the documented decrease in valuation precision seems real. We show that the reduction of accuracy affects mostly the large firms. While the decline seems to start in mid 70's, we show that it becomes more pronounced after 1990. In the relative valuation framework for value relevance, we interpret this finding as a significant reduction of the value relevance of earnings and book values for large firms in the last twenty five years.

We also use the valuation framework to asses the impact of known factors that affect the value-relevance of earnings and book values: investment in intangibles, non-recurring items, and firm size (Collins et al., 1997). Previous research suggests that, for smaller firms, book value takes on increased importance in valuation over earnings. Consistent with these facts, we find that the precision gap between small and large firms is wider for the P/E valuation compared to P/B valuation. We also document a positive trend in the P/B valuation precision gap between small and large firms. While in the beginning of the sample the difference in precision (in favor of big firms) was not statistically significant, in the end of the sample it became significant. The P/E precision gap has been significant all along and seemed to have widened towards the end of the sample.

Lev (1997) and Amir and Lev (1996) (among others) argued that accounting information is of limited value to market participants when valuing service and technology-based companies that invest in intangibles (e.g., research and development, human capital, and brand development). Our results are more nuanced. We find that while it is in the disadvantage of companies that invest in intangible, the difference in precision is, mostly, non-significant. The outstanding exception is that of the years of the dot com bubble when the valuation of companies that invest in intangibles is significantly less precise.

Consistent with previous research that hints at the fact that reporting one time items renders earnings more "transitory" and hence conveys less information to the investors, we bring evidence indicating that the P/E valuation of firms reporting non-zero special items is less precise than that of firms whose earnings do not contain a transitory component. The difference stops being significant after 2000. The P/B valuation does not show significant accuracy differences between the two classes of firms.

The rest of the paper is organized as follows. Section 2 describes the methodology, Section 3 discusses the sample while the empirical results are presented in Section 4. Section 5 concludes.

2 Methodology

2.1 Predicted price

For a multiple defined as the ratio between $P_{i,t}$, the actual stock price of firm i, year t and $Acc_{i,t}$, the value of the accounting variable of the firm in the denominator of the multiple, $\widehat{P}_{i,t}$, the predicted price, is given by:

$$\widehat{P}_{i,t} = Acc_{i,t} \times \widehat{\left(\frac{P}{Acc}\right)}_{i,t,C}, \tag{1}$$

where $\widehat{\left(\frac{P}{Acc}\right)}_{i,t,C}$ is the target firm(i)'s multiple at time t estimated on a set C of comparable firms (peers). In practice, the multiple is often estimated as the median of the multiples of peers. Since recent academic literature (see Baker & Ruback, 1999, Liu et al., 2002) documents accuracy gains from the use of the harmonic mean as an alternative to the median, we consider both definitions[5]:

$$\widehat{\left(\frac{P}{Acc}\right)}_{i,t,C} := \begin{cases} \underset{j\in C}{\text{median}} \left\{ \dfrac{P_{j,t}}{Acc_{j,t}} \right\} & \text{or} \\ \underset{j\in C}{\text{harmonic mean}} \left\{ \dfrac{P_{j,t}}{Acc_{j,t}} \right\}. \end{cases} \tag{2}$$

The value drivers we consider in this paper are the earnings and the book value.

2.2 Selection of comparable firms

Following common practice (Alford, 1992 and Cheng & McNamara, 2000), the set of peers of the target firm is constructed based on similarities in the line of business as measured by the industry codes (SIC) and on proximity in risk as measured by the Total Assets (TA) or earnings growth as measured by the Return on Equity (ROE). Besides the three criteria enumerated, we consider the following other three intersections: industry and size, industry and return on assets and size and return on equity.

Besides the selection criteria, the size of the set of peers is another parameter to be chosen in the implementation of any multiple valuation. The choice of the optimal number of peers is discussed in the end of the section.

[5]Section 4.1.4 contains a short discussion on the size of the accuracy gain when using the harmonic mean definition of the multiple estimate. While the quantitative results we present are based on pricing that applies the harmonic mean definition of the multiple (as being more precise), our qualitative findings are robust to the choice of the estimation method for the multiple. Moreover, the gain in precision due to the use of the harmonic mean estimator is small when compared to the overall lack of precision of the multiple valuation.

2.3 Accuracy evaluation

Since we are interested in measuring and comparing the accuracy in predicting the stock price of the target company, the natural measure of performance is the relative (percentage) prediction error.

The *relative (percentage) error*[6] when predicting the stock price of the firm i in year t with a given multiple m and a given criterion c for selection of comparable firms, is defined as

$$E_{i,t}^{m,c} := \frac{\widehat{P}_{i,t}^{m,c} - P_{i,t}}{P_{i,t}}, \tag{3}$$

where $P_{i,t}$ is the actual stock price for the target firm i in year t while $\widehat{P}_{i,t}^{m,c}$ is the predicted stock price based on the multiple m and comparable firms selected by criterion c as defined by the equation (1), where $m \in \{P/E,\ P/B\}$ and $c \in \{I,\ TA,\ ROE,\ \ I + TA,\ \ I + ROE,\ TA + ROE\}$. For multiples based on the enterprise value, the debt is removed from the predicted firm value to obtain the predicted price.

In the sequel, the quantity defined by equation (3) will be referred as valuation (relative) error or as prediction (relative) error. We emphasize that the variable to predict in the frame of this study is the actual stock price and that this terminology does not mean that the target company is not appropriately valued by the market. We do not aim at picking mis-priced stocks. Our focus is on how close the multiple valuation is to market prices. Nevertheless, large relative errors in the sense of the definition in equation (3) could in fact signal a miss-pricing of the target company. This issue is the subject of further research and will be investigated elsewhere.

2.4 Dominance - an overall performance measure

Suppose that we consider two distributions X and Y, characterized respectively by their cumulative distribution functions F_X and F_Y. Then *distribution Y dominates*

[6] To avoid clutter in the displays, the graphs (and hence the tables) show always the relative errors while the discussion of results uses mostly the more intuitive percent error terminology. A relative error of -0.1 corresponds to a percent error of -10%, i.e. an under-pricing of 10%, a relative error of 0.5 corresponds to a percent error of 50%, i.e. an overpricing of 50%, etc.

distribution X stochastically at first order if, for any argument e,

$$F_X(e) \geq F_Y(e). \tag{4}$$

For our purposes, this definition is the wrong way round[7]. To keep with the intuition, we will say that valuation method X dominates valuation method Y (we write $X \geq Y$), if $F_X(e) \geq F_Y(e)$, for all errors e. F_X and F_Y denote here the cdfs of the absolute relative valuation errors $|E|$ of methods X and Y, respectively. In this case, we will prefer method X to method Y.

If e denotes an absolute relative error, say 20%, then the inequality in the definition, $F_X(e) \geq F_Y(e)$, means that the percentage of firms valued by method X within 20% of the actual price is greater than or equal to the percentage of such firms valued by method Y. In other words, there is at least as high a proportion of precisely valued firms by method X as by method Y (precision here means an absolute error smaller than e). If method X dominates method Y, then whatever error level we may choose, there is always more precision delivered by method X than by Y.

Definition (4) implies a clear relationship between the corresponding percentiles of the two distributions. Since for a $p \in [0,1]$, the $(p * 100)\%$ percentile of the distribution X is defined as $F_X^{-1}(p)$, if method X dominates method Y, i.e. the inequality (4) holds, then

$$F_X^{-1}(p) \leq F_Y^{-1}(p), \tag{5}$$

for all $p \in [0,1]$, i.e. all percentiles of X are smaller than the corresponding Y percentiles. In particular, the median absolute percentage error of method X is smaller than that of method Y.

Table 1 gives the interpretation of the relation of dominance that will be our main tool for comparing valuation accuracy.

We argue that the notion of dominance (as defined in Table 1) applied to the distribution of absolute errors $|E|$ yields the most exhaustive comparison of the accuracy of competing valuation methods. The accounting literature on the subject

[7]The definition fits the case where smaller probabilities of low values are desirable, like in the study of poverty. In our case small values mean higher precision and are hence desirable.

Relation	$dm(X,Y) :=$	Meaning
$X > Y$ or $F_X(e) > F_Y(e)$	$\sup_e(F_X(e) - F_Y(e)) > 0$	Method X is more precise than method Y
$X < Y$ or $F_X(e) < F_Y(e)$	$\inf_e(F_X(e) - F_Y(e)) < 0$	Method Y is more precise than method X
$X = Y$ or $F_X(e) = F_Y(e)$	0	Method X is as precise as method Y
Neither method dominates the other	not defined	The 2 methods cannot be compared

Table 1: **Dominance measure.** The table defines the dominance measure $dm(X,Y)$ between two valuation methods X and Y. F_X and F_Y denote here the CDF of the absolute valuation errors $|E|$ of methods X and Y, respectively.

uses particular statistics to capture the accuracy of different methods. A number of authors consider the distribution of absolute percentage errors $|E|$. Alford (1992) states that "the accuracy of the different methods [] is assumed to be captured by the median and 90th percentile of the distribution" of absolute errors $|E|$. Cheng and McNamara (2000) uses the same two quantiles for their comparisons. Kaplan and Ruback (1995), Kim and Ritter (1999), Gilson, Hotchkiss, and Ruback (2000), Lie and Lie (2002) focus on the "fraction of errors that are less than 15 percent", i.e. the 15th percentile of the distribution of $|E|$. The notion of dominance generalizes these approaches: if method X dominates method Y, all percentiles of X are smaller than those of Y. In particular, the median, the 15th or the 90th percentiles of the absolute percentage error of method X are smaller than those of method Y.

Other authors perform the comparisons based on the distribution of E, the percentage errors themselves. Liu et al. (2002) and Liu et al. (2007) focus on "the interquartile range as the primary measure of dispersion". It is not difficult to see that the interquartile range of the distribution of percentage errors can be written as the sum of two percentiles of the distribution of the absolute percentage errors[8]. Comparing hence the interquartile range of two distributions of errors of roughly the same shape is equivalent to comparing the sums of two percentiles of the corresponding absolute error distributions. The notion of dominance generalizes also this approach: if method X dominates method Y, all percentiles of X are smaller than those of Y. Hence, the sum of any two percentiles of X will be smaller than the sum of the corresponding percentiles of Y. In particular, the interquartile range of X will be smaller than that of Y.

In practice, before performing a comparison of the accuracy of two valuation methods X and Y, the cdfs of the corresponding absolute relative error $|E|$, F_X and F_Y, need to be estimated and the statistical error needs to be taken into account. The details of the methodology are presented in Section A in the Appendix.

2.5 Optimal number of peers

Next we turn to the choice of the optimal number of comparable companies. For a multiple valuation approach defined by a pair (m, c) of multiple and peer selection criterion, denote by $F_k^{m,c}$ the cdf of the pricing errors $|E^{m,c}|$ *when exactly k peers selected by the method c are used* to construct the multiple. The optimal number of comparable firms k_{opt}, to be used in estimating the multiple, is defined[9] by the condition:

$$F_{k_{opt}}^{m,c} \geq F_k^{m,c} \text{ for all } k \neq k_{opt}. \tag{6}$$

In words, the (m, p) valuation that uses k_{opt} peers, the optimal number of comparable companies, to estimate the multiple, is at least as accurate as any other valuation defined by the same pair (m, p) but with a different cardinal of the set of peers. Using the notion of dominance introduced in the previous section, this condition translates to:

$$dm(F_{k_{opt}}^{m,c}, F_k^{m,c}) \geq 0 \text{ for all } k \neq k_{opt}. \tag{7}$$

[8]If the distribution is symmetric, the interquartile range is twice the median absolute error.
[9]Note that the optimal k might not be unique.

When starting from the data, the cdfs of the error distribution need to be estimated and the expression in equation (7) needs to get an operational form. The details are presented in section B in the Appendix.

3 The sample

The accounting data was extracted from 96,598 financial annual reports covering the period between 1968 and 2012. The sample was obtained starting with all the annual reports available in the CRSP/Compustat merged database (initial sample size 372,647) by removing all the company/years for which

1. SIC code is from 6000 to 6799 (division Finance, Insurance and Real Estate),
2. at least one value of the eight variables needed to construct the multiples and to define the comparable companies was missing,
3. at least one of the six multiples was infinite due to division by 0,
4. at least one of the six multiples was negative,
5. at least one of the six multiples was among the largest and smallest 1% observations of the corresponding multiple,
6. the ones with the 1st digit SIC code that the number of peers in one year is less than 50.

More details about the sample are to be found in Section E in the Appendix.

4 Empirical results

4.1 Implementation and accuracy assessment of multiple valuation

This section applies the methodological improvements regarding the implementation of multiple valuation as well as the assessment of its accuracy presented in section 2 and prepare the tools for the relative valuation analysis of value relevance of earnings and book values.

4.1.1 Optimal choice of the number of comparable firms

The value of the multiple and hence the predicted price depends not only on the criterion of the choice of peers, like industry, risk, earnings growth, but also on the size of the set of comparable companies that one selects. To our knowledge there exists no systematic study in the literature that provides quantitative motivation for the choice of the size of the set of peers. Cheng and McNamara (2000) discuss the impact of restrictions on the number of firms for industry on valuation accuracy by looking at the mean and the standard deviation of the distribution of absolute valuation errors. However, they do not discuss the issue of the size of the set of peers for other peer selection approaches. This section improves on their analysis by conducting a systematic investigation for the six criteria of peer selection under study.

Figure 1 displays the results of the analysis of the effect on valuation's accuracy of the number of peers in the construction of the multiple. It shows, for every multiple and every peer selection approach, the median absolute error as a function of the size of the set of comparable companies used in valuation. The figure gives an overall idea about the size of the optimal choice of the cardinal of the set of peers. The full statistical motivation of the optimal choice is given in section B.2.

The graphs in Figure 1 show that an appropriate choice of the size of the set of comparable companies is essential for a fair comparison of the performance of different approaches.

Each graph in Figure 1 reports two median absolute error curves corresponding to valuation using one of the six peer selection approaches. Note that the shape of the function does not depend on the multiple. As a result the optimal number of comparable firms will depend only on the peer selection approach and not on the multiple used. The vertical line in each graph indicates the optimal number of peers for each method of peer selection as explained in section B.2 of Appendix.

The graphs show a consistent ordering of the multiples (according to the median absolute error of the valuation) for I, TA and I+TA criteria with P/B displaying higher median absolute error. For the other criteria the two curves are very close to each other.

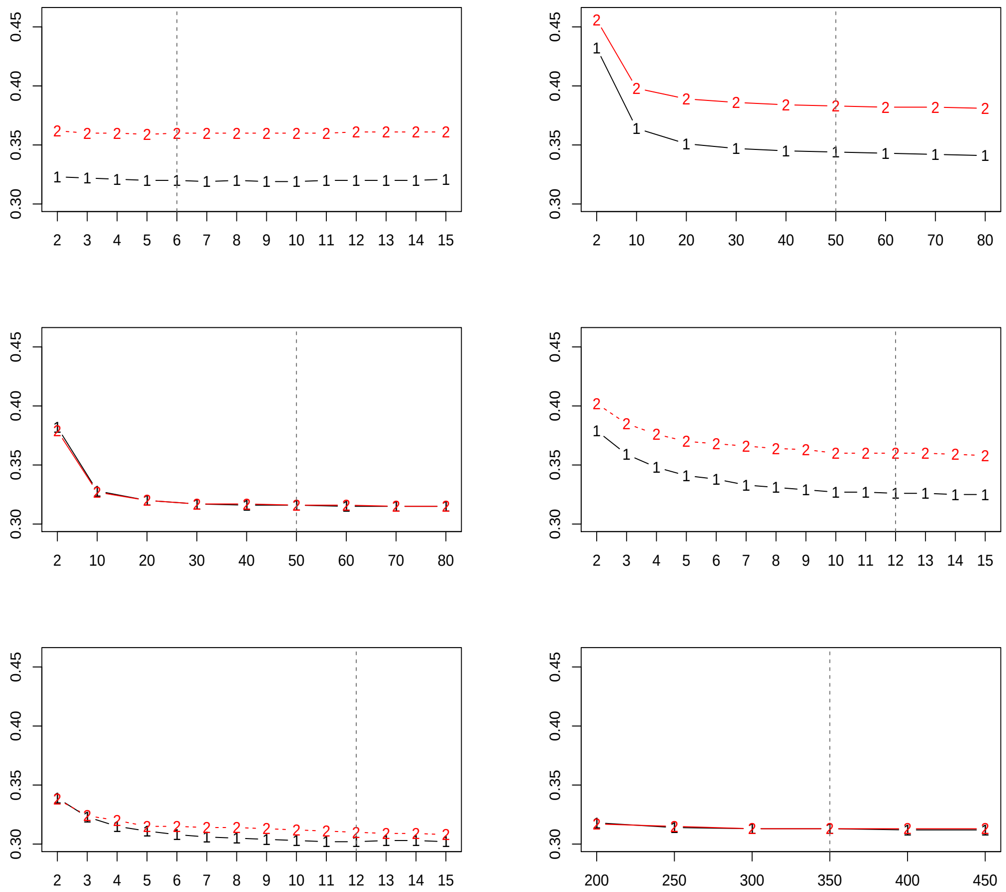


Figure 1: Median absolute relative pricing error $|E|$ (see definition (3)) as a function of the number of peers used in the construction of the multiple. Each graph reports two median absolute error curves corresponding to valuation using each of the six peer selection approaches: *First row:* I (*left*), TA (*right*); *Second row:* ROE (*left*), I+TA (*right*); *Third row:* ROE+I (*left*), ROE+TA (*right*). The curves are numbered as follows : (1) - P/E, (2) - P/B. The graphs show the same patterns for all the multiples (and a given peer selection approach). The vertical lines indicate the optimal number of peers, k_{opt}.

Next we summarize the results of the statistical implementation of the optimal choice of the number of peers described in detail in section B.2 of the Appendix. For the Industry criterion of peer selection, a choice of the number of peers k' between 3 and 9 yield equally good accuracy while a number of peers greater than 10 is sub-optimal. Alford (1992) uses $k = 6$ while Cheng and McNamara (2000) find that the minimum for the median absolute error is attained for their data set at $k = 7$. Our analysis used an optimal value of $k^I_{opt} = 6$.

For TA and ROE criteria, one optimal choice for the number of peers is $k^{TA,ROE}_{opt} =$ 50. Employing less than 30 similar firms performs worse than our choice, while using more than 40 yields statistically equally precise valuation as that of our choice. It is worth mentioning that Cheng and McNamara (2000) choose $k^{TA,ROE} = 6$, a value far from the optimal one. These choices, most likely, have penalized the performance of the ROE based valuation in their study. In particular, for the P/E multiple, they find that ROE based valuation performed poorly, second only to TA. In contrast to their findings, we document superior performance for the pair (P/E, ROE). This pair is one of the five most accurate valuation methods.

For TA+I and ROE+I criteria, our optimal choice was $k_{opt}^{TA+I,ROE+I} = 12$. Using less than 10 peers perform strictly worse than our choice. The range starting at 10 produces valuations of the same accuracy as the method of our choice. Finally, for ROE+TA peer selection approach, one optimal choice is $k_{opt}^{ROE+TA} = 350$. For values of k smaller than 250, the valuation is strictly less precise than that corresponding to our choice, while the range that starts at 300 produces statistically equally precise valuation.

Before we conclude let us note that details about size of the set of peers for two of the the six peer selection criteria for which the k needs interpreting are given in section C of Appendix.

All results reported in the sequel are based on the optimal choice of the number of peers as explained in this section.

4.1.2 Summary statistics of $|E|$, the absolute valuation errors

Table 2 gives the medians and the 90% percentiles of the distributions of absolute relative errors (multiple estimation by the median and harmonic mean) of the thirty six (six multiples × six peer selection criteria) valuation approaches under study. To facilitate the discussion of results, the order is that of Table 3 which presents the dominance analysis. The values in the table can be directly compared to those in Alford (1992) and Cheng and McNamara (2000).

The measures of dispersion in Table 2 are larger than the similar ones in Alford (1992) and Cheng and McNamara (2000) reflecting most likely the fact that our sample is larger.

A hypothesis test of equal median absolute percentage shows that the first six

No.	Multiple	Criterion	Median				Harmonic Mean			
			25%-ile	Median	75%-ile	90%-ile	25%-ile	Median	75%-ile	90%-ile
1	P/B	ROE+TA	0.14	0.31	0.53	0.78	0.15	0.31	0.52	0.72
2	P/B	ROE	0.15	0.31	0.54	0.80	0.15	0.31	0.52	0.73
3	P/E	ROE+I	0.13	0.30	0.55	0.86	0.14	0.30	0.52	0.77
4	P/B	ROE+I	0.14	0.31	0.54	0.83	0.14	0.31	0.52	0.74
5	P/E	ROE	0.15	0.31	0.54	0.84	0.15	0.31	0.52	0.75
6	P/E	ROE+TA	0.14	0.31	0.54	0.81	0.14	0.31	0.53	0.76
7	P/E	I	0.14	0.32	0.59	0.89	0.14	0.31	0.56	0.83
8	P/E	I+TA	0.14	0.33	0.61	0.92	0.15	0.32	0.57	0.84
9	P/E	TA	0.16	0.35	0.61	0.90	0.16	0.34	0.58	0.84
10	P/B	I	0.17	0.37	0.63	1.04	0.17	0.36	0.59	0.83
11	P/B	I+TA	0.17	0.37	0.64	1.06	0.17	0.36	0.60	0.84
12	P/B	TA	0.19	0.39	0.65	1.01	0.18	0.38	0.62	0.84

Table 2: Descriptive statistics for the distributions of absolute relative errors of the 12 (2multiples $\times$ 6 peer selection methods) valuation approaches under study. The order is that of Table 3. The first six combinations of multiple and peer selection have median absolute errors that are statistically equal. All the other combinations have a significantly higher median absolute error.

combinations of multiple and peer selection in Table 2 have median absolute errors that are statistically equal. All the other combinations have a significantly higher median absolute value. While comparing median absolute error is a good starting point for performance evaluation, a more global criterion of comparison can be used. In the sequel we present the dominance analysis that compares the error distributions in their entirety, yielding a clear cut picture of the overall performance of different valuation methods.

4.1.3 Dominance analysis

Table 3 gives the dominance ordering of the valuation methods based on the P/E and P/B multiples.

Each entrance (i, j) of the Table 3 represents the value of the dominance measure $dm(F_j, F_i)$ as defined by equation (10) and explained in Table 5, where F_i is the i-th pair (multiple, peer selection approach) in the ranking given by the first two columns of the table. We note than, for a large majority of the pairwise comparisons, a relation of domination can be established. The results in Table 3 allows for a number of important and clear cut conclusions.

Findings regarding the valuation methods. The first main finding is that

	Multiple	Peers	A	B	C	4	5	6	7	8	9	10	11	12
A	P/B	ROE+TA	=											
B	P/B	ROE	=	=										
C	P/E	ROE+I	$\frac{2}{-1}$	$\frac{2}{-2}$	=									
4	P/B	ROE+I	1	=	$\frac{1}{-1}$	=								
5	P/E	ROE	2	1	2	=	=							
6	P/E	ROE+TA	2	2	1	=	=	=						
7	P/E	I	5	5	3	4	4	3	=					
8	P/E	I+TA	6	6	4	5	5	4	=	=				
9	P/E	TA	6	6	5	6	5	5	3	2	=			
10	P/B	I	7	8	8	7	7	7	5	5	3	=		
11	P/B	I+TA	8	8	8	8	7	7	6	5	3	=	=	
12	P/B	TA	10	10	10	10	10	10	8	8	5	3	3	=

Table 3: Dominance ordering of the valuation methods based on P/E and P/B multiples. The entrance (i, j) of the table represents the value of the dominance measure $dm(F_j, F_i)$ (see definition (10)) where F_i is the i-th pair (multiple, peer selection) in the ranking given by the first two columns of the table. The value 5 on the position (7,2) in the table indicates that the pair (P/B, ROE) (2) dominates the pair (P/E, I) (7) and that the difference between the estimated cdfs of the absolute relative pricing errors of the two methods is at most 5%, i.e. $0 \leq F_2(e) - F_7(e) \leq 5\%$. The value $\frac{2}{-1}$ in position (3,1) indicates that none of the methods (P/E, ROE+I) (3) and (P/B, ROE+TA) (1) dominates the other and that $-1\% \leq F_1(e) - F_3(e) \leq 2\%$.

six valuation methods[10] display an overall better precision than the rest of the approaches. Many of them are equally precise. When that is not the case, the differences are tiny: the biggest value of the dominance measure within this group is of 1%. Moreover, these six methods dominate all the others by a margin of at least 4% to 5%. Since the differences in performance are of little practical relevance[11], we believe there is no need to strictly classify them.

While in the sequel we comment on the relationship with other classifications in the literature[12], we would argue that, given the overall level of low accuracy of

[10] They are those with the best median absolute error in Table 2.

[11] Recall that the precision of the best multiple valuation method is poor: only 25% of the predicted firm stock prices are within 10% of the market price, more than 50% are priced with an error of at least 25% of the market price, more than 20% with an error of at least 50% of the market price.

[12] The articles concerned are Liu et al. (2002) and Cheng and McNamara (2000). Although Alford (1992) is one of the early references in the literature on the accuracy of multiple valuation methods, we believe that the relevance of his ranking is not as high as the importance of his seminal work due to the small size of the sample he analyzed. The dominance analysis we conducted

the multiple valuation, any one of these six valuation methods will produce results that are practically equally precise. Given the demonstrated relative superiority of these six methods, the following discussion will focus mostly on them.

Cheng and McNamara (2000) also find that (P/B, ROE+I) and (P/E, ROE+I) are the most precise methods, followed, in order, by (P/E, I), (P/E, I+TA), (P/B, ROE) and (P/E, ROE). Our results extend theirs[13]: the six mentioned methods are among the eight most precise ones according to Table 3. Two observations might explain the differences. The cited authors do not consider the peer selection method ROE+TA which yields two of our top valuation methods. The fact that we find valuation with peers selected by ROE criterion higher in the ranking might be explained by our findings in Section 4.1.1 of the paper. The number of peers selected on the basis of ROE proximity (k=6) in Cheng and McNamara (2000) is quite far from the optimal choice documented by our analysis ($k \geq 40$), resulting, most likely, in sub-optimal accuracy of their ROE based valuation. This finding could explain the poorer performance of the ROE based valuation that the authors document. Liu et al. (2002) consider only industry-based peer selection. Reducing the consideration to historical value drivers, our results extend theirs. Restricting the hierarchy in Table 3 to industry-based peer selection valuation methods yields the ordering in Liu et al. (2002).

Another fact worth noting is that the valuation based on the pair (P/E, I) is not among the most precise methods. The difference between this method and the top performing valuations is statistically significant. The dominance measure is of 5%.

Findings regarding the criteria for peer selection. The peer selection methods based on return on equity (ROE) criteria rank first. All six best performing valuations pairs use them. Second, selection of peers based on industry does not rank high. The results suggest that industry membership is not always an effective

---

restricting our sample to the years of his sample, i.e. 1978, 1982, and 1986, and to the firms with December fiscal year yielded precisely the same results as Alford (1992): the valuations based on Industry, I+TA, TA+ROE are equally accurate, i.e. the distribution of their absolute percentage errors are statistically indistinguishable, while valuation based on Industry is more accurate than that based on ROE or TA. As other later studies have shown, we document that a bigger sample allows for establishing a finer hierarchy between these methods.

[13] Recall that they consider only valuations based on P/E and P/B multiples.

criteria for selecting comparable firms and that one needs to use other criteria, for example, earnings growth, to direct the search for peers.

These results generalize those in Cheng and McNamara (2000) who find that both P/E and P/B multiples yield more precise valuations when paired with ROE+I peer-selection criteria than with industry alone[14].

As a corollary, comparing book value and earnings, the two popular accounting value drivers, we find that, when paired with appropriate peer selection methods, and over the whole sample, they are equally accurate.

4.1.4 The impact of the method of constructing the target multiple from peer multiple values

We conclude the discussion on the implementation and assessment of multiple valuation with a remark on the difference in accuracy between the use of the median or of the harmonic mean in the construction of the multiple. Figure 2 displays the typical relationship between the two error distributions. The error distribution corresponding to harmonic mean estimation dominates the one associated to median estimation of the multiple. The gain in precision, as measured by the dominance measure is, at roughly 2%, modest.

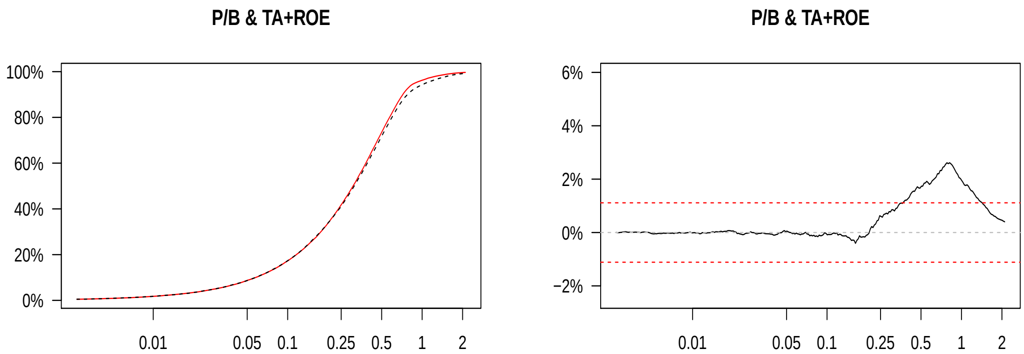


Figure 2: The estimated cdfs (left) and their differences (right) of $|E|$ (logarithmic scale) of the valuation pair (P/B, ROE+TA) when the multiple of the target firm is calculated as the median (dotted line, left-hand graph) or as the harmonic mean of the set of multiples of the comparable companies (full line, left-hand graph). The error distribution corresponding to harmonic mean estimation dominates the one associated to median estimation of the multiple. The gain in precision, as measured by the dominance measure is, at roughly 2%, rather small.

[14]They do not consider the TA+ROE peer selection criteria which yields according to our results better precision when paired to the P/B multiple. As they use a sub-optimal choice of the number of peers in the ROE based valuation, they also penalize the performance of these methods.

4.2 Relative valuation framework for value relevance of earnings and book values

The methodological developments in the previous sections lays the basis for conducting an accurate analysis of the time evolution of the aptness of relative valuation approach (based on the P/E and P/B multiples) to approximate observed market values. In the sequel, we argue that variations in the accuracy of relative firm valuation can be interpreted as indications of change in the value relevance of earnings and book value, respectively. In other words, we view the accuracy analysis in a (relative) valuation framework of value relevance of accounting information. In such a framework, changes in value relevance would be assessed by testing for shifts in the ability of intrinsic value derived from multiple valuation to approximate observed market values[15]. A decrease in the accuracy of the P/E and P/B valuations would be interpreted as a loss of value relevance of earnings and book values respectively.

Given a valuation After the systematic component is taken out, the residual variance (or equivalently residual standard deviation) directly gauges the unexplained noise level of the economic relation. As long as the economic relation does not change, variations in the dependent variable y can indeed be different in separate samples.

As many analysis in this section have a time dimension, most of the results will be reported as figures displaying four curves: the maximum (the upper solid line) and the minimum (the lower solid line) of the difference between the cdfs of absolute errors $|E|$ and the 99% confidence interval (see Appendix) (the dotted lines).

4.3 Scale effects

Following the developments in the distribution of relative errors to track changes in the value relevance assumes implicitly that pricing errors are proportional to the price, i.e. the valuation errors can be standardized by division with the price. However, Gu (2007) points out that, for price regressions, the proportionality as-

[15] This approach is consistent with interpretation 2 in Francis and Schipper (1999) which states that "financial information is value relevant if it contains the variables used in a valuation model or assists in predicting those variables".

sumption underlying the standardized pricing errors is invalid and that the relation between pricing errors and scale is nonlinear. To examine the validity of the proportionality assumption and how pricing errors are affected by scale changes in the case of multiple valuation and to make a comparison with what happens in the case of the price regressions, for each method, the absolute relative errors are pooled across years and sorted into deciles based on the values of prices. For each decile, a box plot is drawn. Figure 3 presents the results.

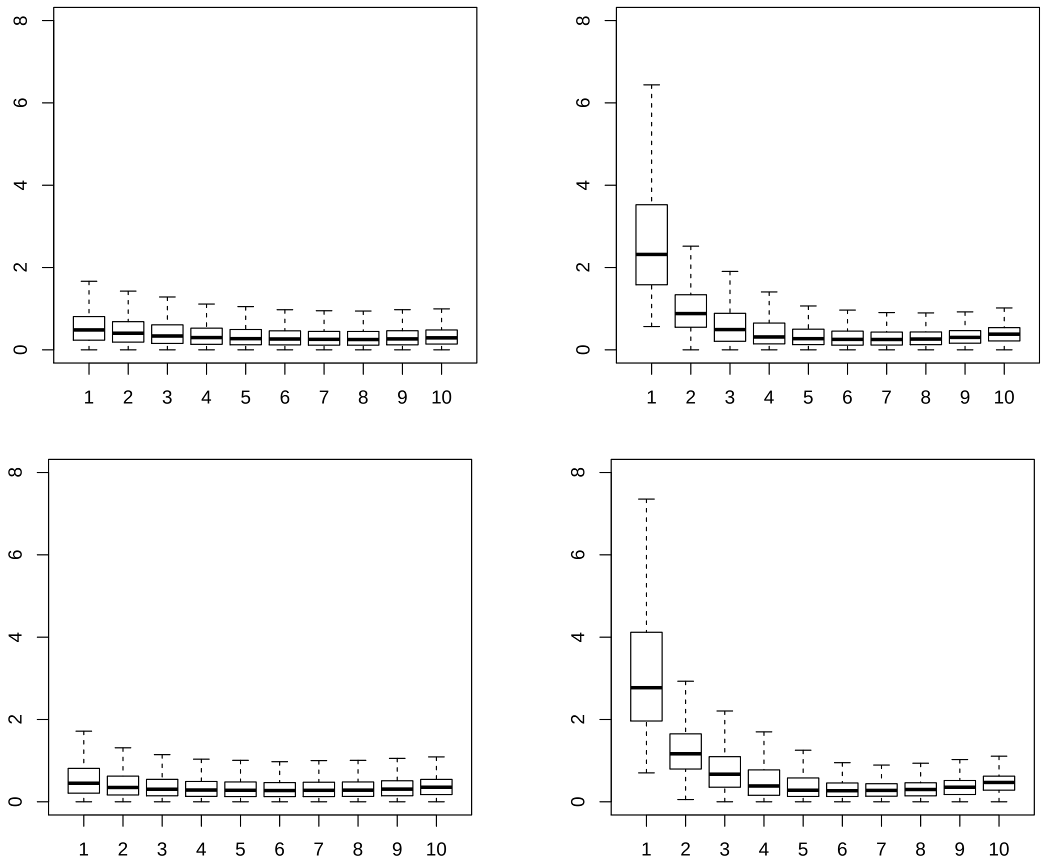


Figure 3: **Multiples vs. price regressions scaling.** The graphs show the box plots of the absolute relative errors from multiple valuation (*Left*) and price regressions (*Right*) conditional on the deciles of price. The price deciles go from low (1) to high (10). The graphs referring to the earnings are on top while those referring to book values are on the bottom. The graphs show that the price level seems to be the right scaling for multiple valuation errors but not for price regression errors.

Several points can be noted. First, as pointed in Gu (2007), the scale problem does exist for price regressions and is particularly severe at relatively low scale levels. For low prices, the absolute relative errors are much larger than for mid and high prices. Hence, larger relative errors might not signal a decrease of the value relevance of the accounting variable and could simply be due to a low level of the market price. The marked scale effects on the distribution of relative errors from price regressions suggest that the temporal patterns of variability of such errors are likely contaminated by the temporal scale changes.

Second, and more important, for multiple valuation the scale effect is almost nonexistent with relative errors corresponding to low prices (the first two or three decides) only slightly higher and slightly more variable than those for mid and high prices. As we will see in the sequel, the inferences for value relevance are robust to conditioning on the strength of the scale effect.

The practical absence of a scale effect on the distribution of relative errors from multiple valuations is reassuring evidence that the temporal patterns of variability of such errors reflect changes in the value relevance of the accounting variables and are not tainted by the temporal scale changes.

4.3.1 Time evolution of relative valuation accuracy in the sample

We begin with an analysis of changes in the accuracy of relative valuation across time. Figure 4 looks at the time evolution of the precision of P/E and P/B valuation (as measured by the dominance measure) by comparing the cdfs of the absolute errors $|E|$ and absolute errors adjusted for scale effects of, on one hand, the current year and, on the other hand, the first year in the sample. The implementation of the relative valuations uses the optimal choices of peer selection criteria and number of peers as described in the previous sections of the paper.

The figure displays the maximum (the upper solid/dotted lines) and the minimum (the lower solid/dotted lines) of the difference between the two cdfs. The solid line corresponds to the analysis made using errors not affected by scale effects, i.e. corresponding to the deciles 4 to 10 in Figure 3, while the dotted ones correspond to one based on all absolute errors $|E|$. The horizontal dotted lines mark the 99% confidence interval (see Appendix). A significant negative value identifies the years characterized by a decrease in the valuation precision while a significant positive value signals a year in which the valuation precision was superior to that in the

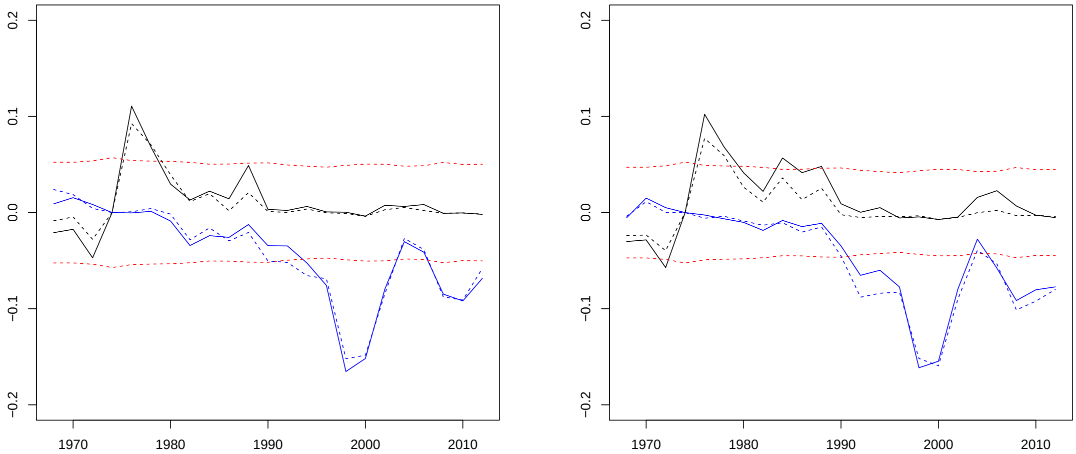


Figure 4: **Time evolution of the precision gap within the sample.** The accuracy gap is measured between valuations in the current year vs. the first year of the sample (1968). P/E valuation on the left, P/B valuation on the right. The upper (lower) solid (dotted) line represents the maximum (minimum) of the difference between the cdfs of the absolute errors not affected by scale effects, i.e. corresponding to the deciles 4 to 10 in Figure 3 (absolute errors $|E|$) of the current year and the first year in the sample. The dotted horizontal lines mark the 99% confidence interval (see Appendix). A significant negative value signals a decrease in the valuation precision.

beginning of the sample.

Figure 4 shows that the accuracy of the two valuations methods has significantly declined through the years, fact that, taken at face value, would be interpreted within the valuation framework as a decrease in value relevance of earnings and book value. The behavior of the two multiples is very similar indicating a comparable loss of relevance of both earnings and book values.

Moreover, the figure shows that inferences based on all relative errors are qualitatively identical and quantitatively very close to those based only on errors that do not display any scale effect, i.e. the deciles from 4 to 10 in Figure 3. This finding underlines one of the strengths of the relative valuation framework: the natural standardization of errors by the price level.

To avoid cluttering of figures, the results conditional on the the scale effect are only displayed in Figure 4. For all other analysis, the inferences are identical and only the results based on all the returns are shown.

In the sequel we investigate possible sources for the documented decrease in the accuracy.

4.3.2 Impact of known factors that affect the value-relevance

Previous literature has identified a number of factors that affect the value-relevance of earnings and book values: intangible intensity[16], one-time items, and firm size. In this section we investigate how valuation accuracy of the P/E and P/B multiples vary with the strength of these factors. In the framework of relative valuation, we evaluate the impact of each of these items on the value-relevance of earnings and book values. We measure their effect globally, on the whole sample, and we investigate the time evolution of the strength of their impact on valuation precision.

	Level	
Item	P/E	P/B
One-time items	-6% (2%)	-4% (2%)
Intangibles	-8% (2%)	-6% (2%)
Market capitalization	10% (1%)	8% (1%)

Table 4: **Precision difference between firms with high/low factor values - all sample.** The table reports the differences in valuation accuracy (as measured by the dominance measure) between firms characterized by high/low values of the factors affecting value relevance. The numbers in parenthesis gives the half-length of the 99% confidence interval (see Appendix). Negative (positive) values signal situations when firms with high values of the factor are less (more) precisely valued than those with low values.

Table 4 reports the overall effect of the three mentioned factors on the precision of valuations based on the P/E and P/B multiples. It shows a significant effect of the three factors when the whole sample is considered. The presence of one-time items and high intangible intensity influences negatively the precision of valuation while higher market capitalization is associated with higher valuation accuracy. The one-time items seem to have the lowest impact while the market capitalization is the most important of the the three factors.

While the results in table 4 give a static, overall impression, it is important to understand if the impact of the three factors has varied or not across time.

Figures 5, 6, and 7 display the time evolution of the size of the impact of the three factors on valuation accuracy. For each year in the sample, the figures

[16]For studying this factor we followed the classification of high/low-technology industries in Francis and Schipper (1999).

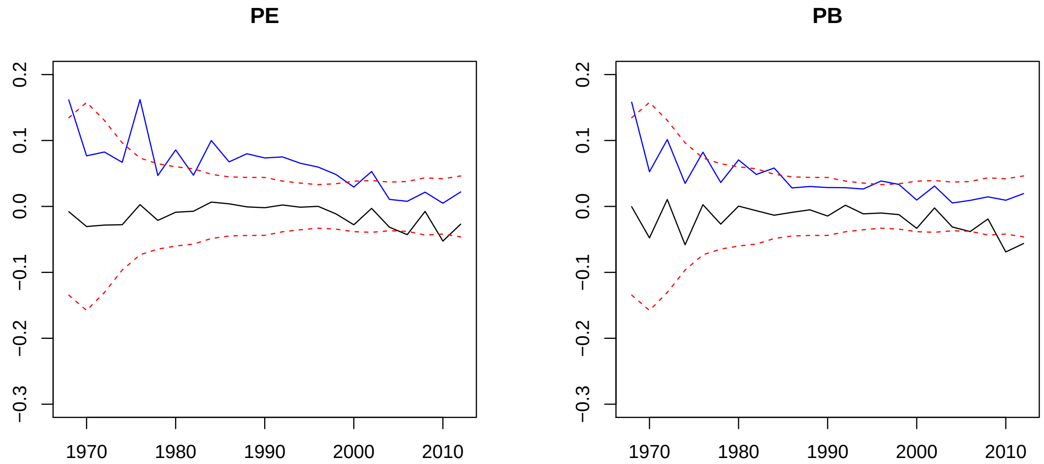


Figure 5: **Precision gap between valuations of firms with/without one-time items.** P/E valuation on the left, P/B valuation on the right. The upper (lower) solid line represents the maximum (minimum) of the difference between the cdf of the valuation errors of firms without one-time items and with one-time item. The dotted lines mark the 99% confidence interval (see Appendix). A significant positive value identifies the years with a more precise valuation of firms without one-time items.

compare the cdfs of the absolute errors when valuing firms with high and with low values of the corresponding factors. The graphs display the maximum (the upper solid line) and the minimum (the lower solid line) of the difference between the two cdfs. The dotted lines mark the 99% confidence interval (see Appendix). A significant negative value identifies the years characterized by a significantly less precise valuation of firms characterized by low values of the factor while a significant positive value indicates the years when valuation of firms with high values of the factor is more precise.

Figure 5 shows that the presence of one-time item has reduced the precision of valuation only temporally during two and a half decades for the P/E multiple and only occasionally for the P/B multiple.

Figure 6 shows that the significant negative impact of high intensity intangible factor reported in Table 4 is due only to the ten years preceding the dot-com market crash. Outside this period valuation accuracy is not significantly affected by the strength of the intangible factor.

Figure 7 shows that the only factor to continuously impact the valuation accuracy through the whole period under discussion is the market capitalization. The P/E valuation seems more sensitive with small firms being less accurately valued

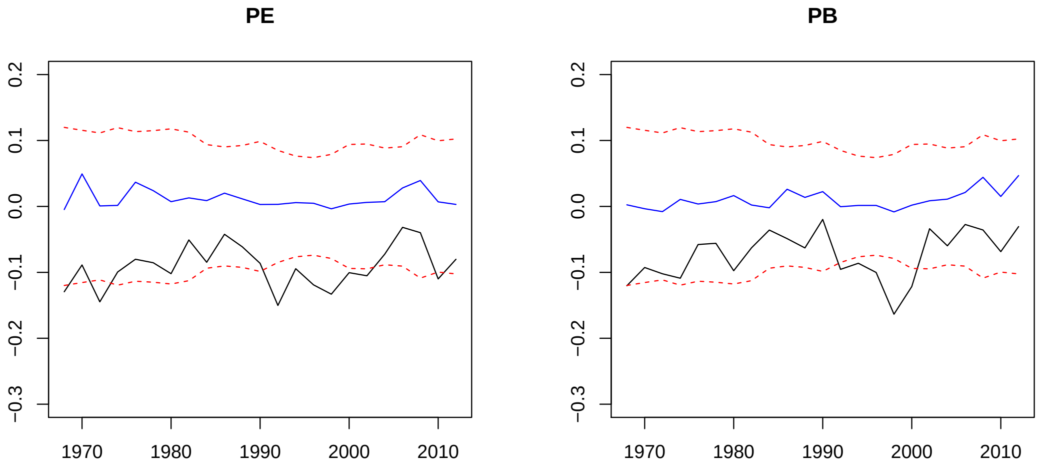


Figure 6: **Precision gap between valuations of firms with high/low intangibles.** P/E valuation on the left, P/B valuation on the right. The upper (lower) solid line represents the maximum (minimum) of the difference between the cdf of the valuation errors of high technology firms and low technology firms. The dotted lines mark the 99% confidence interval (see Appendix). A significant negative value identifies the years with a more precise valuation of firms with low investments in intangibles.

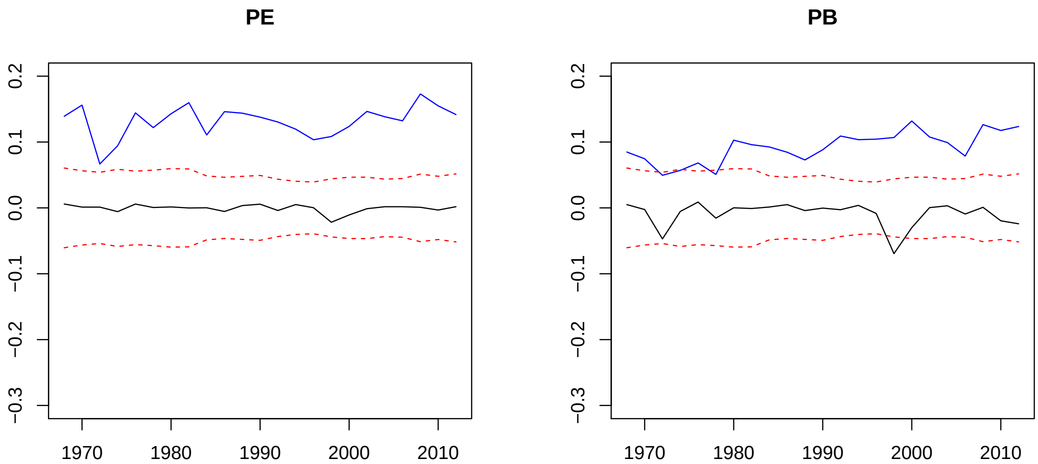


Figure 7: **Precision gap between valuations of large and small firms.** P/E valuation on the left, P/B valuation on the right. The upper (lower) solid line represents the maximum (minimum) of the difference between the cdf of the valuation errors of big firms and small firms. The dotted lines mark the 99% confidence interval (see Appendix). A significant positive value identifies the years with a more precise valuation of large firms.

with respect to the large firms.

We conclude that the only factor with a continuous and almost constant impact on the relative valuations under discussion is the market capitalization. The effect of the other two factors is weaker and, more importantly, is significant only temporary, i.e. during the decade ending the previous century for the intangible intensity and between 1975 and 2000 for the one-time items and only for the P/E valuation.

4.3.3 Sources of temporal changes in the value-relevance of earnings and book values

The developments in valuation precision in Figure 4 are the result of compounding effects of changes in, on one hand, the structure of the sample and, on the other hand, the precision with which firms with different characteristics are valued. For the clarity of the exposition we will focus on the market size factor. The prominent place of this factor already emphasized by the analysis in Section 4.3.2 is confirmed by the results of the subsequent analysis.

As large firms are more accurately valued than small firs, a lower percentage of large firms in the sample will reduce the valuation precision. If the accuracy of valuation of large firms does not evolve through time, the overall valuation precision decreases proportional to the respective change of the percentages of small/large firms in the sample. If, on the other hand, the accuracy of valuation of large goes down through time, the overall valuation decrease is more pronounced.

We will investigate now the time evolution of the two elements in the previous discussion impacting on the overall sample valuation precision: the structure of the sample and the relative accuracy of valuation of firms characterized by different strength of the factors that that affect value relevance.

Figure 8 displays the yearly proportion of small/large firms. It shows that the structure of the sample has strongly changed through time. The proportion of large firms has evolved in a non-linear fashion: from around 60% in the period 1960-1985, it went down to 30% in 2000 and it went up again to 60% in 2012.

Figure 9 shows that the valuation precision of large firms significantly decreased after 1990 with respect to the beginning of the sample. The decrease in the valuation precision of small firms although significant is less dramatic. Hence, the contribution of this factor to the decrease of the valuation precision in Figure 4 is the result of the compounding effect of the decrease/increase of the proportion of large firms and (mostly) the decreasing precision of large firm valuation.

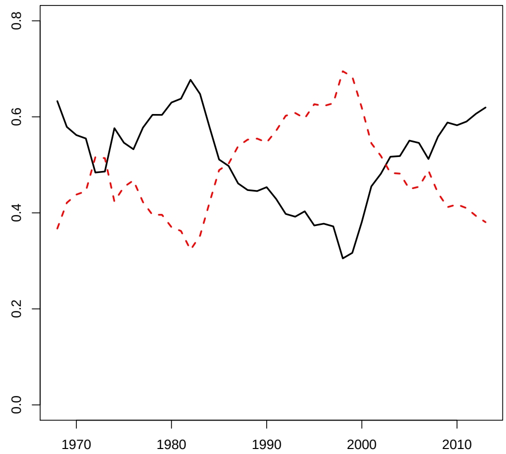


Figure 8: **Evolution of the proportion of small/large firms in the sample.** The full (dotted) line displays the proportion of large (small) firms in the sample. The size is measured by the market capitalization adjusted for the overall level of the market. The adjustment divides the market capitalization of a firm by a market-capitalization-weighted equity index. A firm is considered to be large if its adjusted market capitalization is higher than the median adjusted market capitalization over the whole sample. The proportion of large/small firms has varied widely across time.

4.3.4 Explaining the temporal change in the value-relevance of earnings and book values

This section aims at understanding the importance of each one of the two mentioned compounding effects by separating their impact on the valuation accuracy.

To better understand their interaction, Figure 10 displays the estimated densities of the absolute pricing errors (in log scale) for large (thick lines, *Second row, left*) and small firms (thin lines, *Second row, right*) in 1968 (*First row, left*) and in 2000 (*First row, right*). In all four graphs a domination relationship can be established as shown in Figure 11. In both graphs on the first row, the valuation of large firms is more precise than that of small firms. The dominance measure is around 12% in both years. The second row clearly shows a decline in the precision of valuation from 1968 to 2000: the (dotted line) densities of absolute valuation errors in the year 2000 (for both small and large firms) are positioned to the right of the densities of the valuation errors in 2000 (full lines). The dominance measure for large firms is, at 22% much bigger than that for small firms (12%).

A change in the structure of the sample from one year to another corresponds

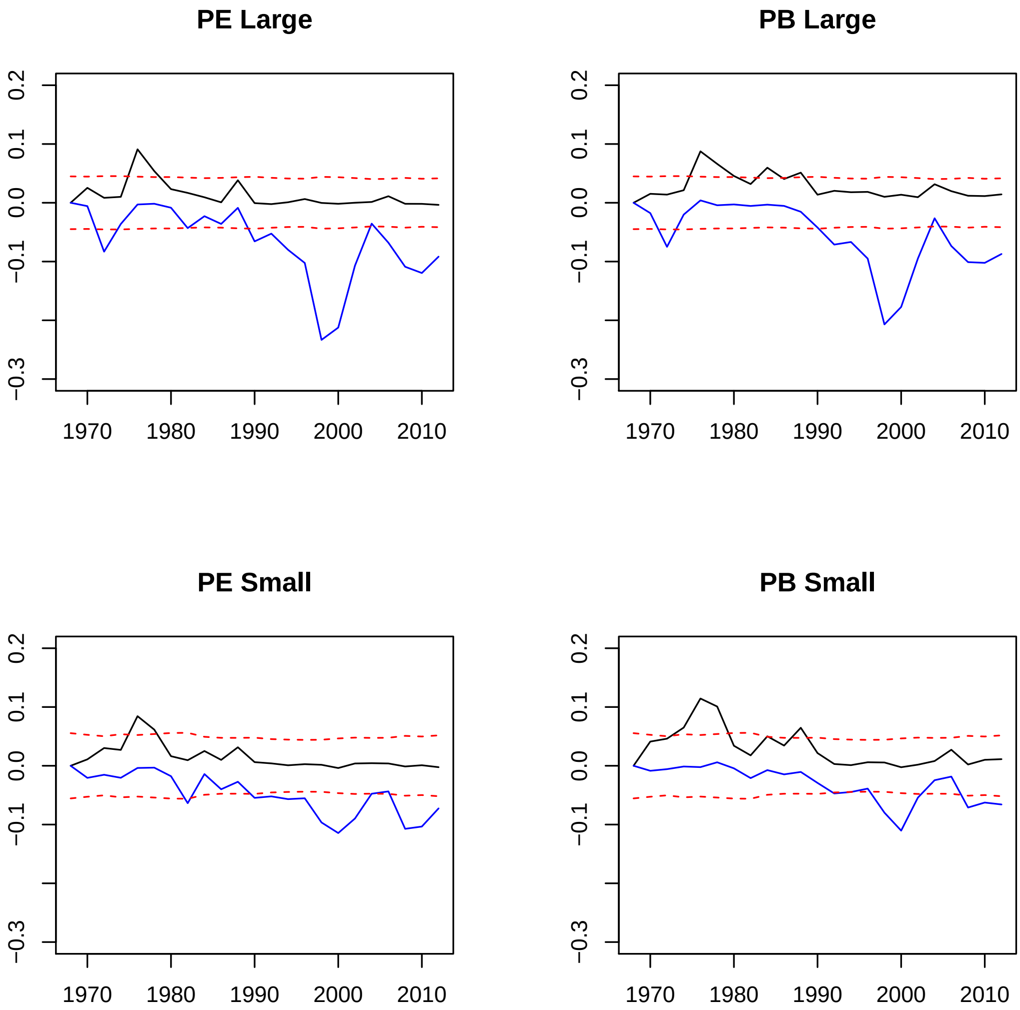


Figure 9: **Time evolution of the precision gap within the sample for large/small firms.** The accuracy gap is measured between valuations in the current year vs. the first year of the sample (1968). P/E valuation on the left, P/B valuation on the right. The upper (lower) solid line represents the maximum (minimum) of the difference between the cdfs of the absolute errors $|E|$ of the current year and the first year in the sample. The dotted lines mark the 99% confidence interval (see Appendix). A significant negative value signals a decrease in the valuation precision.

to observations moving, in the graphs on the first row, from thick to thin line densities (or viceversa). Suppose the percentage of large firms is lower in 2000 than in 1968. That means the proportion of observations described by the thick dotted density (the more precise one) is smaller than if the structure of the sample had not changed. Overall precision of the valuation is hence lower due to the structure change. The decrease in the precision with respect to 1968 is in fact amplified by the fact that the valuation for both large and small firms has declined from 1968 to 2000. Not only that more observations come from the density of small firms (that

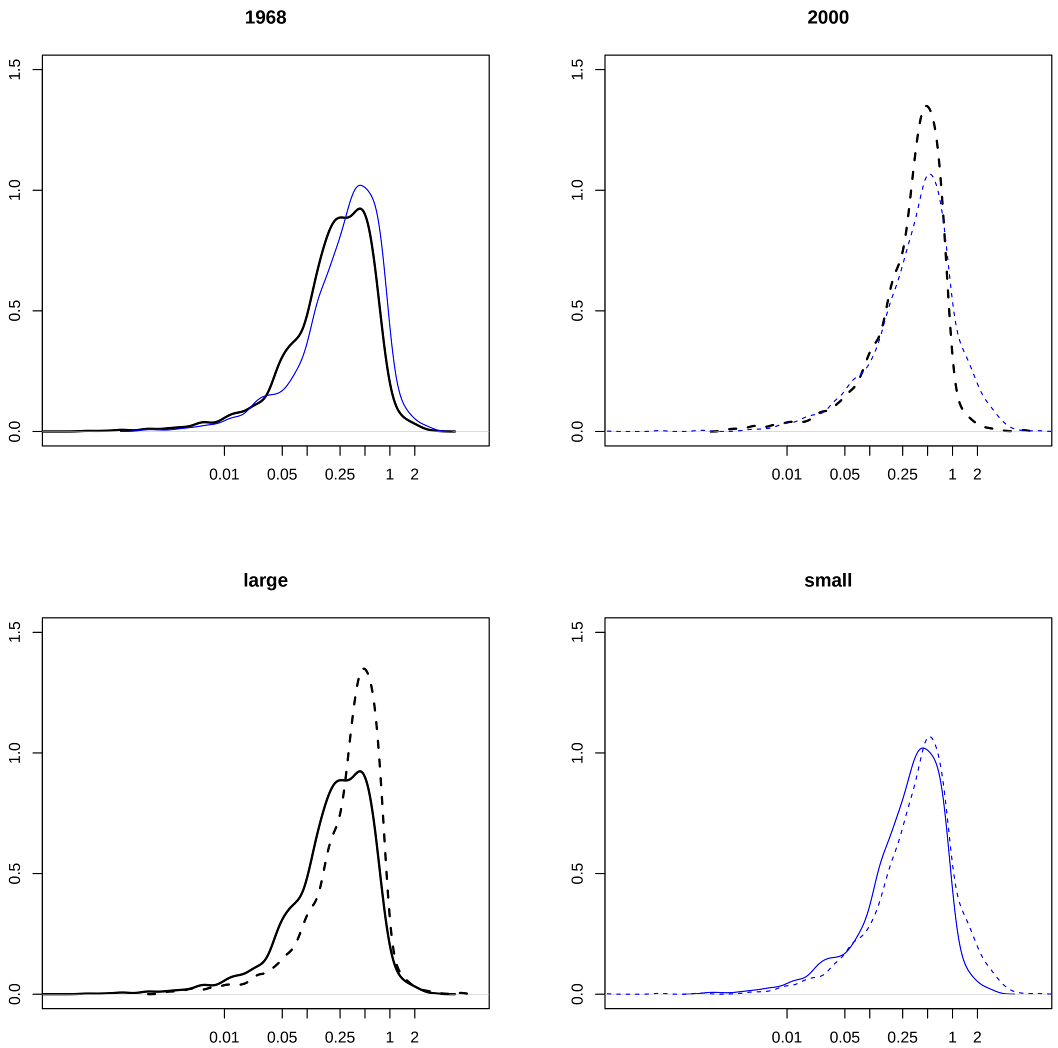


Figure 10: **Estimated densities of absolute valuation errors for big and small firms (1968 and 2000).** The absolute errors are on log-scale. Thick (thin) lines are used for the densities of large (small) firms while full (dotted) lines draw the densities of 1968 (2000). *Top, left:* 1968. *Top, right:* 2000. *Bottom, left:* large firms. *Bottom, right:* small firms.

is less precise than that of large firms) but this same density has drifted to the right with respect to its position in 1968.

This decomposition indicates one way of disentangling the two effects. Suppose we want to isolate the impact of the change in the proportion of small/large firms between 1968 and 2000. In 1968, 60% of the firms in the sample were large while in 2000 the proportion dropped to 30%. To see what the overall precision of valuation would have been if the accuracy of valuation in 2000 had stayed the same as that in 1968, we will simulate a sample of the size and structure of our sample in 2000 but where the valuation errors correspond to the precision of 1968. Concretely,

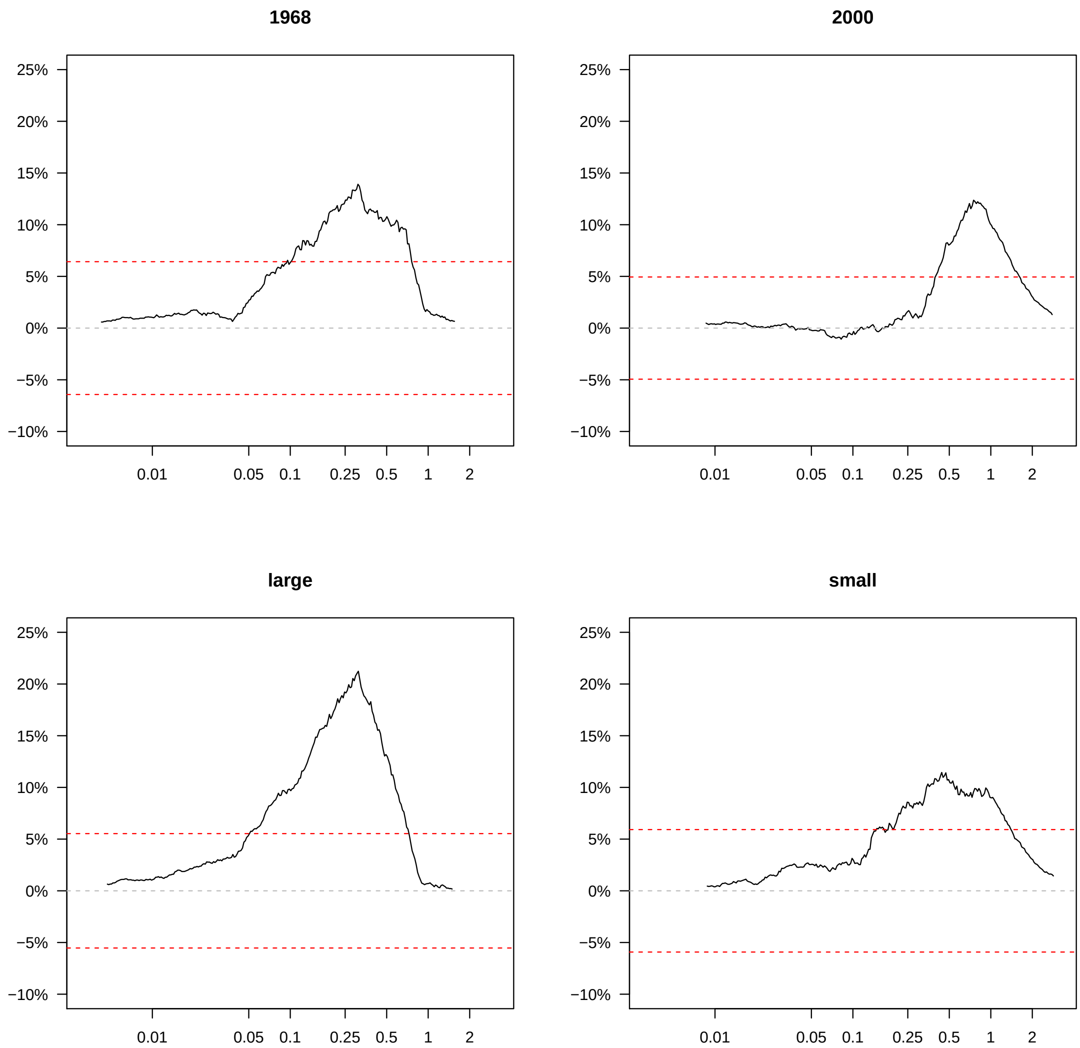


Figure 11: **Differences between estimated cdfs of absolute valuation errors for big and small firms (1968 and 2000).** The absolute errors are on log-scale. Thick (thin) lines are used for the densities of large (small) firms while full (dotted) lines draw the densities of 1968 (2000). *Top, left:* large firm error cdf minus small firm error cdf in 1968 valuations. *Top, right:* large firm error cdf minus small firm error cdf in 2000 valuations. *Bottom, left:* large firm error cdf in 1968 minus large firm error cdf in 2000. *Bottom, right:* large firm error cdf in 1968 minus large firm error cdf in 2000. The dotted lines mark the 99% confidence interval (see Appendix). A significant positive value signals a better performance of the valuation in the first place of the difference. Larger firms are more precisely priced both in 1968 and in 2000 (with a dominance measure of circa 12%). All firms were more precisely valued in 1968 than in 2000 (with a dominance measure of 12% for the small firms and of 22% for the large firms).

we will draw a number equal to 30% (70%) of the 2000 sample size from the 1968 distribution of errors for large (small) firms (continuous, thick, respectively thin curve in top, left graph in Figure 10). Comparing the resulting simulated distribution of errors with the real distribution from 1968 would give a measure of the loss of valuation precision due to the changes in the structure of the sample. If more factors were involved, the corresponding conditional distributions would be used in simulations.

Conversely, if interested in measuring the effect of changes in the precision with which firms with different characteristics are valued, we would simulate a sample of the same structure as that of 1968 with valuation errors corresponding to the precision of year 2000. Concretely, we will draw a number equal to 60% (40%) of the 2000 sample size from the 2000 distribution of errors for large (small) firms (continuous, thick, respectively thin curve in top, right graph in Figure 10). Contrasting the resulting simulated distribution of errors with the actual distribution from year 1968 would give a measure of the loss of valuation accuracy due to the changes in the precision with which firms of different sizes are valued.

Figure 12 displays the results of the two simulations explained above. The graphs on the top row isolate the impact on relative accuracy of the change in the proportion of small/large firms while those on the bottom row quantify the effect of changes in the precision with which firms of different sizes are valued. Each graph displays the maximum and the minimum of the differences between the cdfs of, on one hand, simulated valuation errors of the current year keeping the structure/precision constant and, on the other hand, actual absolute valuation errors from the first year in the sample (1968). It shows that the impact of changes in the structure of the sample is next to negligible while the effect of temporal changes in valuation accuracy is dominant.

The similarity to Figure 4 suggests that the explanation behind the time evolution of valuation accuracy in the overall sample is a decline in the valuation precision of both multiples after 1990. The decline affects mainly the large firms. While particularly low around the dot-com bubble, multiple pricing accuracy briefly improves in the middle of the first decade of the millennium to only decline again in the most recent years. In the relative valuation framework this finding is consistent with a decrease in value relevance of earnings and book values. This decrease is more pronounced after 1990 and affects earnings slightly more than book values.

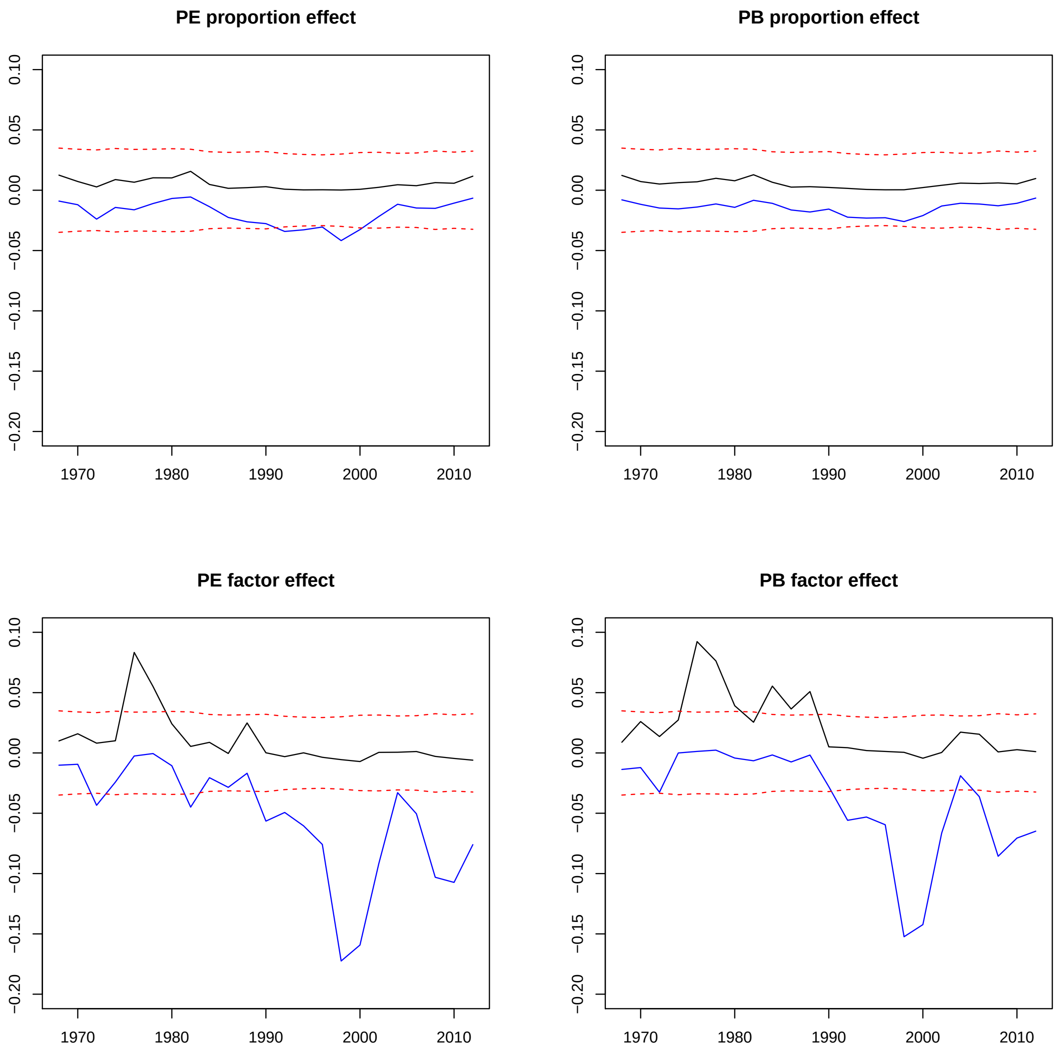


Figure 12: **Time evolution of the precision gap within the simulated samples.** The accuracy gap is measured between valuations in the current year vs. the first year of the sample (1968). *Top:* The sample analyzed was obtained by preserving the valuation precision to the levels of 1968 and drawing large and small firm valuation errors following the actual structure of the yearly samples. It isolates the impact of the change in the proportion of small/large firms between 1968 and the current year from that of the valuation precision. *Bottom:* The sample analyzed was obtained by preserving the structure of the yearly samples to the levels of 1968 but drawing large and small firm valuation errors from the actual error distributions. It isolates the impact of the change in the valuation precision between 1968 and the current year from that of the sample composition. P/E valuation on the left, P/B valuation on the right. The upper (lower) solid line represents the maximum (minimum) of the difference between the cdfs of the absolute errors —E— of the current year and the first year in the simulated sample. The dotted lines mark the 99% confidence interval (see Appendix). A significant negative value signals a decrease in the valuation precision.

5 Conclusions

We demonstrate how a relative valuation framework can be used to study the evolution of value relevance of earnings and book values across time. Its successful operationalization hinges on a number of methodological improvements in the implementation of multiple valuation method. In particular, extending the existing literature that compares accuracy using particular statistics of the distribution of valuation errors, we develop a framework for judging accuracy based on the notion of stochastic dominance that allows for an overall evaluation of performance.

The relative valuation framework we propose offers an alternative to the linear regression approach impaired by a number of econometric weaknesses when employed in the frame of value relevance studies (see Brown et al. (1999), Gu (2007)). In such a framework the measure of relevance of earnings or book values is the precision of P/E, respectively P/B, valuation. A loss of relevance of the accounting variable would substantiate in a decrease of the corresponding multiple valuation precision across time.

Using this framework, we document a significant reduction of the value relevance of earnings and book values mostly for large firms in the last twenty five years.

Appendix

A Empirical dominance measure

For a given sample of errors $(e_1, e_2, \ldots, e_{n_X})$, the estimator of F_X, the cdfs of the absolute relative error $|E_X|$ of method X, is the *empirical cumulative distribution function*:

$$\widehat{F}_{X,n_X}(x) := \frac{1}{n_X}\sum_{i=1}^{n_X} I_{(-\infty,x]}(|e_i|) = \frac{\#\text{ of } |\text{ errors }| \leq x}{n_X}, \tag{8}$$

where $I_A(x)$ is the indicator function:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

The statistical estimation error is described by the asymptotic distribution of the two sample Kolmogorov-Smirnov statistic:

$$D_{n_X,n_Y} := \sup_x |\widehat{F}_{X,n_X}(x) - \widehat{F}_{Y,n_Y}(x)|. \tag{9}$$

Under the null hypothesis that $F_X = F_Y$,

$$D_{n_X,n_Y} \leq c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$

with probability $1 - \alpha$[17].

To summarize, if $\widehat{F}_X$ and $\widehat{F}_Y$ denote the two estimated cdfs of the absolute errors $|E_X|$ and $|E_Y|$ corresponding to the valuation methods X and Y, respectively, we

[17]Given the large size of our sample we will use $\alpha = 0.01$ and $c(0.01) = 1.63$.

define the following measure of dominance:

$$\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y) \stackrel{def}{=} \begin{cases} \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS} \\ \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS} \text{ and} \\ & \quad \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS} \\ 0, & \text{if } -c_{KS} \leq \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \text{ and} \\ & \quad \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS} \\ ?, & \text{if } \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}, \end{cases} \tag{10}$$

where $c_{KS} = c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$, with $c(\alpha)$ as above, n_X and n_Y, the sample sizes used to estimate the two cdfs $\widehat{F}_X$ and $\widehat{F}_Y$, respectively.

Table 5 gives the interpretation of the measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$ and will be referred to in the sequel whenever comparing the precision of the competing valuation methods.

In the case that neither cdf dominates, the relationship between corresponding percentiles depends on and it changes with the order of the percentile. For example, median absolute error of X might be smaller than that of Y, while the 25%-ile of X might be greater than that of Y.

$\text{dm}(\widehat{F}_X, \widehat{F}_Y)$	When	Meaning
> 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X dominates F_Y or Method X is more precise than Y
< 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	F_Y dominates F_X or Method X is less precise than Y
$= 0$	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X is equal to F_Y or The 2 methods are equally precise
?	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	Neither cdf dominates or The 2 methods cannot be compared

Table 5: Interpretation of and necessary and sufficient conditions for the measure of dominance $\text{dm}(\widehat{F}_X, \widehat{F}_Y)$. The notations $\succ$, $\prec$ and $\asymp$ correspond to (in)equalities which are statistically true, i.e. are not rejected by the corresponding Kolmogorov-Smirnov hypothesis test. $\widehat{F}_X$ and $\widehat{F}_Y$ denote here the estimated cdfs (see definition (8)) of the absolute relative valuation errors $|E|$ of methods X and Y, respectively.

B Optimal number of peers

B.1 Operational version of the definition

When the cdfs of the errors are estimated from the data, the definition of the optimal number of peers in equation (7) can be rewritten as (see Table 5):

$$\begin{cases} \max\limits_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \succeq 0 & \text{and} \\ \min\limits_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \asymp 0 & \qquad \text{for all } k \neq k_{opt}, \end{cases}$$

where notations $\succeq$ and $\asymp$ correspond to (in)equalities which are statistically true, i.e. are not rejected by the corresponding Kolmogorov-Smirnov hypothesis test. In particular,

$$\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \succ 0 \text{ and } \min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \asymp 0 \tag{11}$$

for all k that yield a valuation accuracy that is strictly worse than that of k_{opt} and

$$\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \asymp 0 \text{ and } \min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \asymp 0 \tag{12}$$

for those k for which the valuation accuracy equals that of k_{opt}. Since k_{opt} does not depend on the multiple (as shown by the graphs in Figure 1), the statistics needed to classify the k's as strictly sub-optimal or optimal are defined as:

$$\begin{aligned} M^c_k &:= \max_m \left(\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \right) \\ N^c_k &:= \min_m \left(\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \right) \\ O^c_k &:= \min_m \left(\min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_k(e)) \right). \end{aligned} \tag{13}$$

A k for which $N^c_k \succ 0$ and $O^c_k \asymp 0$ will be sub-optimal for all the multiples, i.e. for any of the multiples under study, the valuation using k comparable firms for constructing the multiple c is strictly less accurate than that based on k_{opt} peers. A k for which $M^c_k \asymp 0$ and $O^c_k \asymp 0$ yields, for all multiples, a valuation accuracy that is statistically indistinguishable from the best one.

B.2 Statistical motivation of the optimal choice

Figure 13 gives the statistical motivation of the optimal choice. It displays the dominance measure of the methods using the optimal size of the peer set over those that make other size choices. For the choice of optimal number of peers in Figure 1, Figure 13 displays the dependency of the three statistics in definition (13) on k, the number of peers used in the construction of the multiple.

The three curves in each graph correspond in decreasing order to M^c_k, N^c_k, and P^c_k, respectively. The number constructing the curves identifies the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics (see equation (13)) is attained. The coding of the multiples is as follows : 1 - P/E, 2 - P/B. The dotted

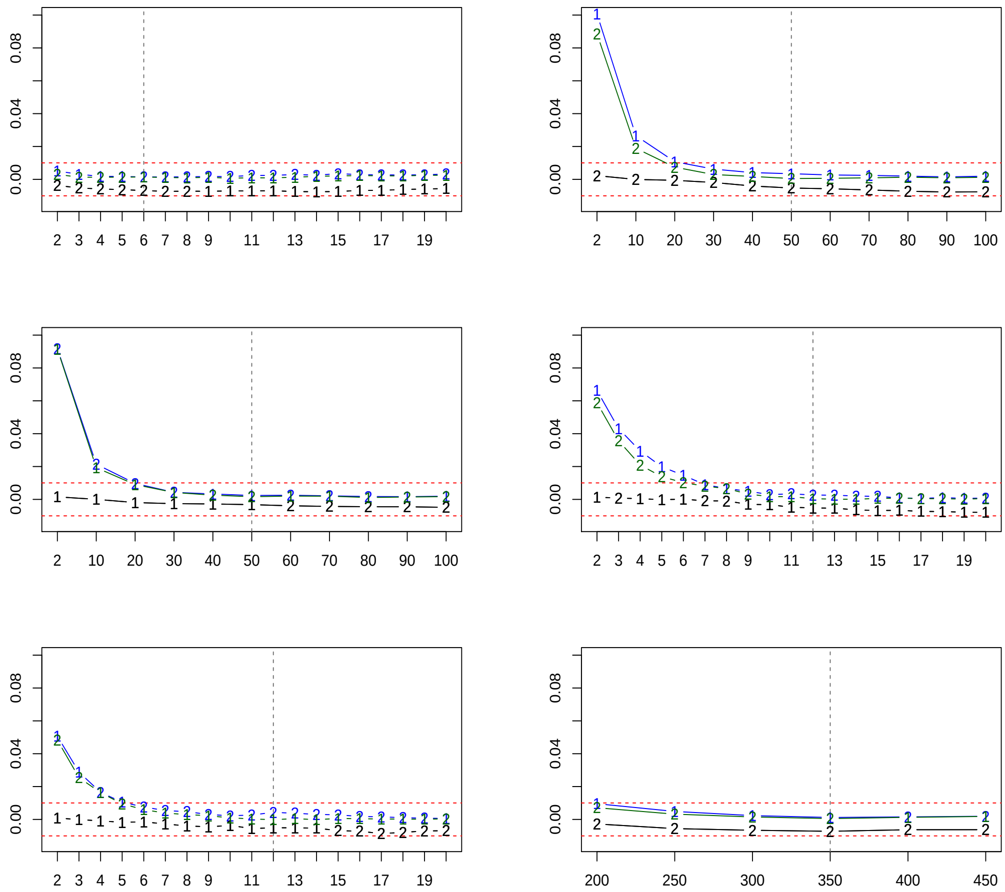


Figure 13: M, N, and P statistics (see definition (13)) as a function of k, the number of peers used in the construction of the multiple, for the choice of k_{opt} in Figure 1. Each graph reports three curves corresponding (in decreasing order) to M_k^c, N_k^c, and P_k^c, respectively, where c is the method of peer selection in the title of the graph. The curves are functions of k, the size of the set of comparable firms. The number constructing the curves indicate the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics is attained. The coding of the multiples is as follows : 1 - P/E, 2 - P/B. The dotted lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers.

lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. Note that $O_k^p \asymp 0$ for all k and all c and hence the values of the statistics M and N only will determine the relationship of optimality of a given k to k_{opt}. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers. It is worth noting that we are able to establish the relationship of optimality for all the k's in the range under discussion.

For the Industry criterion of peer selection, the shape of the functions is convex. While the minimal value is attained for $k = 6$ for all curves, the value of the functions does not change much with the number of peers. All the k's between 3 and 9 yield equally good accuracy while a k greater than 10 is sub-optimal.

C Optimal number of peers for I and ROE+TA criteria

Table 6 gives some details on the size of the set of comparable companies when selected on the base of the industry.

k	Mean	Median	Min	1st Q	3rd Q	Max
2	38	18	2	9	51	577
4	42	22	4	10	53	589
6	**45**	**25**	**6**	**11**	**56**	**689**
10	61	37	10	18	74	866
15	83	52	15	29	97	1186
20	98	59	20	34	113	1220

Table 6: Descriptive statistics for industry-based peer selection approach.The values corresponding to the optimal choice of the number of peers in our analysis are in bold.

Recall that, for the Industry criterion, for a given k, the set of peers includes all firms matched on the basis of four-digit SIC codes if the resulting industry contains k firms. Otherwise the definition of a firm's industry is determined by fewer SIC digits until at least k other firms are identified. As the cardinal of the set of comparable companies varies with the firm, the table gives the descriptive statistics of the size of the set of comparable companies as a function of k. The optimal choice of $k = 6$ corresponds hence to estimating the multiple from a set that averages 45 comparable companies.

Table 7 gives the descriptive statistics of the size of the set of peers as a function of k when the comparable companies are selected on the base of Total Assets and Return On Equity.

K	Mean	Median	Min	1st Q	3rd Q	Max
200	19	18	1	14	23	115
250	30	28	2	22	36	163
300	43	40	7	32	51	220
350	**58**	**54**	**14**	**43**	**69**	**279**
400	76	71	18	57	90	348
450	96	90	27	72	113	435

Table 7: Descriptive statistics for TA+ROE-based peer selection approach. The values corresponding to the optimal choice of the number of peers in our analysis are in bold.

Recall that, for this criterion, for a given k, the set of peers is obtained by taking the intersection of the sets of the k firms with the closest TA and ROE, respectively. The optimal choice of $k = 350$ corresponds hence to estimating the multiple from a set of roughly 50 comparable companies.

D Variables and multiples

The variables used to construct the common multiples under consideration were Common Shares Outstanding (CSHO), EPS excluding extraordinary items (EPSPX), Sales (S), Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA), Common Equity (CEQ), Net Income (NI), Long Term Debt (DLTT), Current Liabilities (LCT), Preferred Stock (UPSTK) and Preferred Dividends in Arrears (DVPA). The variables used for defining comparable companies (the peers) were Standard Industrial Classification code (SIC), Total Assets (TA) and Return on Equity (ROE).
The share Closing Price (P) was obtained from CRSP database, and the sum of DLTT (deflated by CSHO), LCT (deflated by CSHO), UPSTK (deflated by CSHO) and DVPA (deflated by CSHO), and P is defined as the Enterprise Value per share (EV). As the financial statements are publicly available a few months after the end of the accounting reporting period, the share closing price corresponding to the third month after the end of the reporting period were used as P and calculating EV.

E The sample

Table 8 shows in detail the loss of observations at each step of the construction of the sample.

Sample construction	Size
All annual reports in CRSP/Compustat (1968-2012)	372,647
1.Exclude annual reports with SIC codes between 6000 and 6799	-76,270
	296,377
2.Exclude annual reports with missing values	-139,053
	157,324
3.Exclude annual reports with infinite multiples	-2,345
	154,979
4.Exclude annual reports with negative multiples	-44,528
	110,451
5.Exclude annual reports with largest or smallest 1% multiple values	-7,889
	102,562
6.Exclude industries (1st digit SIC) with less than 50 peers in one year	-5,964
Final sample size	96,598

Table 8: Construction of the sample.

While the exclusion rules 1, 2, 3 and 5 are easy to understand, an explanation is needed for the other two. Shortly, the rule 4 yields the largest sample of company/year instances for which *both valuation methods* provide a market value and which allows for a fair comparison of the distribution of the errors produced by the different multiple based valuation methods. In more detail, negative multiples are generated by negative denominators. Recall that the multiple-based predicted market value of a company is obtained by multiplying the the peer-based predicted multiple with the value taken by the variable in the denominator of the multiple for the respective company in the given year. As the peer-based predicted multiple is, in the case of our analysis, always positive, a negative denominator implies a negative market value for the company. Hence, the fourth exclusion rule removes from the sample the company/year instances for which at least one of the valuation methods fails to give a value to the company by producing a negative market value. The sixth exclusion rule is set for ensuring enough comparable companies when the criterion of choice is industry.

Table 9 presents the sample size detailed for each year of the study period 1968-2012.

Year	1968	1969	1970	1971	1972	1973	1974	1975	1976
Size	1388	1518	1588	1665	1763	1647	1275	1517	1583
Year	1977	1978	1979	1980	1981	1982	1983	1984	1985
Size	1577	1608	1481	1548	1459	1439	1754	2082	2238
Year	1986	1987	1988	1989	1990	1991	1992	1993	1994
Size	2265	2461	2334	2148	2091	2162	2658	2985	3271
Year	1995	1996	1997	1998	1999	2000	2001	2002	2003
Size	3322	3520	3498	3145	2857	2572	2143	2226	2421
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012
Size	2698	2664	2642	2467	1904	1999	2311	2236	468

Table 9: Size of the sample by year.

Table 10 gives a few descriptive statistics of the common multiples under investigation.

Multiple	Mean	Median	Min.	1st Qu.	3rd Qu.	Max.	S.D.
P/E	23.33	15.80	3.22	10.53	24.68	301.60	27.92
P/B	2.40	1.80	0.36	1.17	2.91	15.89	1.98

Table 10: Description statistics of the multiples

References

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, *30*(1), pp. 94-108.

Baker, M., & Ruback, R. (1999). Estimating industry multiples. *Working paper*, Harvard University.

Beisland, L. A. (2009). A review of the value relevance literature. *The Open Business Journal*, *2*, pp. 7-27.

Brimble, M., & Hodgson, A. (2007). On the intertemporal value relevance of conventional financial accounting in australia. *Accounting & Finance*, *47*(4), pp. 599-622.

Brown, S., Lo, K., & Lys, T. (1999). Use of R^2 in accounting research: measuring changes in value relevance over the last four decades. *Journal of Accounting and Economics*, *28*(2), 83 - 115.

Chang, J. J. (1998). The decline in value relevance of earnings and book values. *Working paper*, University of Pennsylvania.

Cheng, C., & McNamara, R. (2000). The valuation accuracy of the price-earnings and price-book benchmark valuation methods. *Review of Quantitative Finance and Accounting*, *15*(4), pp. 349-370.

Collins, D. W., Maydew, E. L., & Weiss, I. S. (1997). Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics*, *24*(1), pp. 39-67.

Dontoh, A., Radhakrishnan, S., & Ronen, J. (2004). The declining value-relevance of accounting information and non-information-based trading: An empirical analysis. *Contemporary Accounting Research*, *21*(4), pp. 795-812.

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

Ely, K., & Waymire, G. (1999). Accounting standard-setting organizations and earnings relevance: Longitudinal evidence from nyse common stocks, 1927-93. *Journal of Accounting Research*, *37*, pp. 293-317.

Francis, J., & Schipper, K. (1999). Have financial statements lost their relevance?

Journal of Accounting Research, *37*(2), pp. 319-352.

Gilson, S. C., Hotchkiss, E. S., & Ruback, R. S. (2000). Valuation of bankrupt firms. *The Review of Financial Studies*, *13*(1), pp. 43-74.

Gu, Z. (2007). Across-sample incomparability of R^2s and additional evidence on value relevance changes over time. *Journal of Business Finance & Accounting*, *34*(7-8), pp. 1073–1098.

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The Journal of Finance*, *50*(4), pp. 1059-1093.

Kim, M., & Kross, W. (2005). The ability of earnings to predict future operating cash flows has been increasingnot decreasing. *Journal of Accounting Research*, *43*(5), pp. 753-780.

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics*, *53*(3), pp. 409-437.

Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. *Journal of Accounting Research*, *37*(2), pp. 353-385.

Lie, H., & Lie, E. (2002). Multiples used to estimate corporate value. *Financial Analysts Journal*, *58*(2), pp. 44-54.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research*, *40*(1), pp. 135-172.

Liu, J., Nissim, D., & Thomas, J. (2007). Is cash flow king in valuations? *Financial Analysts Journal*, *63*(2), pp. 56-68.

2.2 Paper II: Is P/E always more accurate than EV/EBITDA?

Is P/E always more accurate than EV/EBITDA ?

Jian Kang[1], Alain Schatt[2], and Catalin Starica[1]

[1] *School of Economics and Business, University of Neuchâtel*

[2] *HEC Lausanne, University of Lausanne*

Abstract

We compare the accuracy of firm valuation based on EV/EBITDA, a multiple that gained in popularity with the practitioners during the last decade, with that of the traditional multiple P/E. Our detailed analysis of a large sample of US companies, over a period of 45 years, shows that, when aggregating over all firms, P/E is more accurate than EV/EBITDA . Hence, we confirm the results of Liu, Nissim, and Thomas (2002). Our finding is robust to the implementation of the valuation method (characteristics and number of peers, aggregation of peer multiples), the type of data in the construction of multiples (reported numbers or forecasts), the period (recent years or not), the industry type (capital intensive industry or not). When different firm characteristics are taken into account, the relation becomes more intricate. For companies with low debt, EV/EBITDA is at least as precise as P/E. Moreover, EV/EBITDA is significantly more accurate than P/E when valuing firms that report largely negative special items and/or non-operating items. This finding is in line with the idea that not all items convey information to the investors. Overall, we conclude that EV/EBITDA leads to more accurate valuation in some specific cases.

Keywords: Valuation, multiples, EV/EBITDA, P/E, accuracy, stochastic dominance.

1 Introduction

Multiples are largely used by practitioners to evaluate listed firms (Demirakos, Strong, & Walker, 2004; Asquith, Mikhail, & Au, 2005; Damodaran, 2006; Deloof, De Maeseneire, & Inghelbrecht, 2009; Roosenboom, 2012). Traditionally, the most common valuation multiple has been by far the Price-to-Earnings (P/E) ratio. However, recent years have witnessed a change in the preferences of practitioners highlighted by recent surveys on valuation practice in the United States (Block, 2010) and in Europe (Bancel & Mittoo, 2014). Analyzing 1,209 responses by US financial analysts, Block (2010) finds that 41.7% use the price-to-earnings ratio as their primary metric, while 36.2% prefer EV/EBITDA. More importantly, the survey participants predict that the latter metric will become the primary measuring tool in the future. While P/E is the most used multiple for valuation of firms in a number of industries (Materials, Consumer Discretionary, Energy, Health Care, Utilities), EV/EBITDA takes the first place in other industries like Industrials, Telecommunication Services, Consumer Staples, and Technology. In further questioning of respondents, Block (2010) finds that the increasing popularity of EV/EBITDA stems from dissatisfaction with GAAP-related income and accounting standards. Bancel and Mittoo (2014) show similar results in Europe. A survey of 356 valuation experts across eight European countries finds that EV/EBITDA is the most popular multiple used by 83% of experts, followed by P/E favored by 68% of respondents.

In this paper, we analyze the relative accuracy of these two multiples. Our analysis is motivated not only by the mentioned evolution in analysts' preferences but also by the recent changes in the composition and the interpretation of the net income. Recent research shows that the number of firms reporting special (or non-recurring) items, as well as their size, has increased over the last decades (Bradshaw & Sloan, 2002;Johnson, Lopez, & Sanchez, 2011). This development is likely to have an impact on analysts forecasts as well as on the market value of the firms (Burgstahler, Jiambalvo, & Shevlin, 2002; Cready, Lopez, & Sisneros, 2012; Dechow & Ge, 2006; McVay, 2006; Riedl & Srinivasan, 2010). In particular, it is plausible that such items affect differently the accuracy of the two multiples, as they are only taken into account in the calculation of the net income and do not affect the size of EBITDA.

Analyzing multiples' accuracy is not an easy task because of the lack of consensus on the implementation of two essential steps of the valuation. First, a set of comparable companies (or peers) needs to be defined. Then, a ratio should be computed for the target firm by aggregating peers' multiples. For neither of the two steps a consensus exists on the manner to proceed. This fact explains why analysts usually use their discretion to select strategically the set of peers (De Franco, Hope, & Larocque, 2015; Paleari, Signori, & Vismara, 2014). Moreover, while the median of the peers' multiples is commonly employed to predict the target firm's multiple, other statistics are also used. Since the implementation issues significantly affect valuation accuracy (as shown by e.g. Alford, 1992; Bhojraj & Lee, 2002; Liu et al., 2002; Liu, Nissim, & Thomas, 2007), a thorough analysis of multiple accuracy should take all these issues into account.

In this paper, we perform a systematic investigation of the impact of different implementation choices on the valuation based on the two multiples. Our analysis is based on a sample covering a 45 year period (1968-2012) containing all US companies (for which the necessary data were available) listed in Compustat and CRSP data bases.

In a first step, we consider the impact on accuracy of the two main aspects of selection of the set of comparable firms: the characteristics of peers and their number. For the definition of peers, in line with the existing literature, we use six criteria: industry (proxied by the SIC code), size (proxied by the total assets), quality of growth (proxied by the ROE/ROIC), industry and size, industry and growth quality, and size and growth quality. Moreover, for each multiple we determine the optimal number of peers, i.e. the size of the peer set which yields the lowest pricing errors. Next, we consider the issue of aggregation of peer ratios into a prediction of the target firm's multiple confronting the use of the median to that of the harmonic mean. We use a comprehensive performance measure related to the notion of stochastic dominance to compare the overall accuracy of the two multiples. Although many researchers refer to it in other areas, especially for portfolios selection and performance comparison (Agliardi, Agliardi, Pinar, Stengos, & Topaloglouf, 2012; Belghitar, Clark, & Deshmukh, 2014; Kuosmanen, 2004), to the best of our knowledge, the concept of stochastic dominance has not been used in research related to multiples accuracy.

We find that, when aggregating over all firms in the sample, the P/E is more accurate than EV/EBITDA and that this result is robust to the implementation of the multiple method. It holds independently of the way the peers are selected (although we find that the number of peers greatly affects the valuation accuracy, and that the selection based on criteria containing ROE/ROIC consistently provides the most accurate valuation), the aggregation method of peer multiples, or the horizon used to compute the valuation accuracy (i.e. the errors are calculated with respect to the market price 3 months, 6 months or 12 months after the end of firm's financial year). The accuracy ranking is stable through time (although we document a reduction of the precision gap in the last decades) and over industries. Finally, the use of financial analysts forecasts instead of accounting (historical) data does not modify the order (although employing the forecasts increases the accuracy of both multiples). Overall, our results confirm those of Liu et al. (2002).

In a second step, we analyze the accuracy disparity between the valuations as a function of the two terms at the origin of the difference between the definitions of the two multiples: EV-P or debt (which relates the numerators of the two multiples) and EBITDA-E (which relates the two denominators). This decomposition allows for a refinement of the previous result: when different firm characteristics are taken into account, the relation becomes more intricate. We find that, while P/E is more precise for medium and high level of debt, EV/EBITDA dominates P/E for firms with low debt and a high gap between EBITDA and E. We provide evidence suggesting that largely negative values of special item and non-operating item components of EBITDA are at the origin of the mentioned superiority of the EV/EBITDA. They lower the performance of P/E valuation while not affecting that of EV/EBITDA . This finding ties nicely with the literature on financial reporting and helps understand the increased tendency of the analyst tracking services to report earnings before negative special items reported in Bradshaw & Sloan, 2002.

We interpret it as evidence for the fact that these items (when of the 'right' sign and size) do not provide information to investors. Finally, we document a strong similarity in the time patterns of the debt level and of the precision gap between the multiples. We infer that the documented reduction of the accuracy gap is consistent with a reduction of the level of debt assumed by firms.

We therefore conclude that P/E provides the most accurate valuation, except for firms with low levels of debt and largely negative special items and/or non-operating items. As the overall performance is an average of accuracies over different parts of the sample, the all-inclusive superiority of the P/E documented in the first part of the analysis is due to its better performance on a large section of the sample, that of firms with medium and high levels of debt.

Our paper contributes to the relatively scarce literature on multiples' accuracy. To the best of our knowledge only two other papers compare the accuracy of P/E and EV/EBITDA (Liu et al., 2002; Lie & Lie, 2002). Theirs analysis do not address all the implementation issues mentioned previously. In particular, both papers consider only one criterion for the selection of peers (industry), and neither of them takes into account the sensitivity of the valuation accuracy to the size of the set of peers. Moreover, they cover a shorter period and use incomplete statistics to compare the distribution of errors. In this paper, we conduct a systematic investigation of all the implementation issues mentioned by considering, for each of our 98,756 firm-year observations, different characteristics for peer selection, number of peers, methods of aggregation of the peer ratios. In addition, the use of stochastic dominance, which takes into account the whole distribution of errors, allows a comprehensive comparison and unifies the different approaches in the literature.

We also contribute to the financial reporting literature regarding the relevance of non-operating items, non-recurring items, or special items. As such items have increased in frequency and size over the last decades, a burgeoning literature discusses the impact of such changes on capital markets participants (Burgstahler et al., 2002; Cready et al., 2012; Dechow & Ge, 2006;McVay, 2006; Riedl & Srinivasan, 2010). We find that, for firms with low levels of debt, largely negative values of the special items (mainly) and the non-operating items (to a lesser extent) do not seem to convey information that is incorporated in the prices. For these firms, the accuracy of P/E is lower compared to that of EV/EBITDA. It is worth mentioning that our approach directly related to prices that compares the overall distribution of errors via stochastic dominance brings a fresh methodological perspective to this literature that focuses mainly on the analysis of returns via classical regression.

The rest of the paper is organized as follows. Section 2 presents our empirical design. In section 3, we report the results of the stochastic dominance analysis of P/E and EV/EBITDA valuation errors. Section 4 discusses the results of our accuracy decomposition with respect to the two main difference terms (EV-P and EBITDA-E) between the two multiples. Section 5 concludes the paper.

2 Empirical design

The multiple method estimates a firm's stock price by capitalizing company's current value of the accounting variable in the denominator of the multiple at the aggregate multiple computed for a set of comparable firms or peers. The construction of a set of comparable firms is therefore the main issue for the implementation of the two multiples under scrutiny in this study (P/E and EV/EBITDA).

2.1 The predicted price

A multiple m_i for the firm i is defined as the ratio between P_i, the stock price and Acc_i, the value of an accounting variable of the firm (at end of fiscal year t):

$$m_i := P_i/Acc_i. \tag{1}$$

For the multiple m, the predicted price $\widehat{P}_i(m, C)$ of the firm i is a function of the set C of peers and is given by:

$$\widehat{P}_i(m, C) = Acc_i \times \widehat{m}_i(C), \tag{2}$$

where $\widehat{m}_i(C)$ is the aggregate multiple of the firm estimated on the set C of comparable firms (peers). This notation emphasizes that a multiple valuation method is the result not only of a choice of a multiple m but also of a set of peers C. In the sequel, the valuation approaches will be referred by the pair (multiple, set of peers) that defines them.

In practice, the aggregate multiple $\widehat{m}_i(C)$ is often estimated as the median of the multiples of the peers. However, since the academic literature (see Baker and Ruback (1999), Liu et al. (2002)) documents accuracy gains from the use of the harmonic mean as an alternative to the median, we consider both definitions in this paper[1].

The multiples evaluated in our empirical analysis are: EV/EBITDA , defined as Enterprise Value per share divided by EBITDA deflated by the number of Common Shares Outstanding CSHO, and P/E , defined as the share price (P) divided by Earnings Per Share (EPS)[2]. For EV/EBITDA , the debt is removed from the predicted firm value to obtain the predicted price.

2.2 Multiples' accuracy

2.2.1 Relative and absolute errors

Since the set of comparable firms has a non-negligible impact on the value of the predicted price, the *relative error*[3] of the multiple of the firm i (for a given year t) is defined as a function of the multiple m and of the set of comparable firms C:

$$E_i(m, C) := \frac{\widehat{P}_i(m, C) - P_i}{P_i}, \tag{3}$$

where P_i is the actual stock price for the target firm i in year t, and $\widehat{P}_i(m, C)$ is the predicted stock price based on the pair (m, C) of the multiple m and the set of comparable firms C. The benchmark stock price for the target firm i, P_i, used to evaluate the accuracy of a multiple is the stock price of the firm i three months

[1] Section 3.2.2 contains a discussion on the size of the accuracy gain when using the harmonic mean. While the results we present are based on pricing that applies the harmonic mean definition of the multiple (as being more precise), our findings are robust to the choice of the aggregation method for the multiple.

[2] The earnings exclude the Extraordinary items.

[3] To avoid clutter in the displays, the graphs and the tables always show the absolute errors while the discussion of results uses mostly the more intuitive percent error terminology. For example, a relative error of 0.5 corresponds to a percent error of 50%.

after the end of the financial year[4].

2.2.2 An overall accuracy comparison measure

The academic literature on multiple valuation analyzes the accuracy of a valuation method (m, C) by looking at its pricing relative errors $E(m, C)$ (as defined in (3)) or at their absolute values $|E|$. The use of absolute errors places equal weight on positive and negative errors. The comparisons of precision are based on specific statistics of the distribution of errors. For example, Liu et al. (2002) and Liu et al. (2007) look at the distribution of the relative errors E and focus on "the interquartile range as the primary measure of dispersion". Alternatively, Alford (1992) states that "the accuracy of the different methods [ ] is assumed to be captured by the median and 90th percentile of the distribution" of absolute errors $|E|$. Cheng and McNamara (2000) uses the same two quantiles for their comparisons while Lie and Lie (2002) focuses on the percentage of firms within 20% of the actual price. Kaplan and Ruback (1995), Kim and Ritter (1999), Gilson, Hotchkiss, and Ruback (2000), Lie and Lie (2002) consider the "fraction of errors that are less than 15 percent", i.e. the 15th percentile of the distribution of $|E|$. We note that all comparisons make use of percentiles of the distribution of errors.

In this paper, we argue that the concept of stochastic dominance gives the natural set-up for comparing accuracy of competing valuation approaches. It encompasses all the criteria mentioned above and allows for precise statistical testing. While stochastic dominance has been extensively used in finance research, especially for the comparison of portfolios' performance (e.g., Agliardi et al., 2012; Belghitar et al., 2014; Kuosmanen, 2004), it has not yet been applied for the analysis of multiples accuracy.

The dominance criterion we advocate is both intuitively appealing and comprehensive. We say that valuation approach X dominates pricing method Y if the proportion of firms in the sample valued within a specified error is higher for method X than for method Y, and that happens for all error levels.

This definition can be stated formally making use of the cumulative distribu-

[4]The results remain qualitatively unchanged when we consider the stock price of the firm six, nine or twelve months after the end of the financial year t.

tion functions (CDF) F_X and F_Y of the absolute errors $|E|$ of methods X and Y, respectively. The valuation method X dominates[5] valuation method Y (we write $X \geq Y$), if

$$F_X(e) \geq F_Y(e), \tag{4}$$

for all errors e. In this case, we will prefer method X to method Y.

If e denotes a specific absolute error, say 20%, then the inequality $F_X(e) \geq F_Y(e)$ in the definition means that the percentage of firms valued by method X within 20% of the actual price is greater than, or equal to, the percentage of such firms valued by method Y. If method X dominates method Y, then whatever error level we may choose, there is always more precision delivered by method X than by Y.

Inequality (4) implies a clear relationship between the corresponding percentiles of the two error distributions. Since for a $p \in [0,1]$, the $(p * 100)\%$ percentile of the distribution X is defined as $F_X^{-1}(p)$, if method X dominates method Y, then

$$(p * 100)\% \text{ percentile of distribution } X \leq \ (p * 100)\% \text{ percentile of distribution } Y \tag{5}$$

for all $p \in [0,1]$. In words, all percentiles of X are smaller than the corresponding Y-percentiles. In particular, the median absolute error of method X is smaller than that of method Y. Since most of the comparisons in the literature are based on percentiles, the use of stochastic dominance generalizes the existing approaches.

In the sequel, the strength of the dominance relation between two valuation approaches X and Y (when established) will be measured by the *dominance measure*, denoted by $dm(X,Y)$ and defined in Table 1.

In practice, before performing a comparison of the accuracy of two valuation methods X and Y, F_X and F_Y, the CDF of the corresponding absolute errors, need to be estimated, and the statistical error needs to be taken into account when establishing a performance relationship between the valuation approaches. For details about how this is done rigorously, see section A in the Appendix.

[5]For our purpose, we had to adjust the common definition of *stochastic dominance* which states that *distribution Y dominates distribution X stochastically* at first order if, for any argument e, $F_X(e) \geq F_Y(e)$ where F_X stands for the cumulative distribution function (CDF) of the distribution X. This definition fits the case where smaller probabilities of low values are desirable, like in the study of poverty. In our case small values mean higher precision and are hence desirable.

Relation	$dm(X,Y) :=$	Meaning
$X > Y$ or $F_X(e) > F_Y(e)$	$\sup_e(F_X(e) - F_Y(e)) > 0$	Method X is more precise than method Y
$X < Y$ or $F_X(e) < F_Y(e)$	$\inf_e(F_X(e) - F_Y(e)) < 0$	Method Y is more precise than method X
$X = Y$ or $F_X(e) = F_Y(e)$	0	Method X is as precise as method Y
Neither method dominates the other	not defined	The 2 methods cannot be compared

Table 1: **Dominance measure.** The table defines the dominance measure $dm(X,Y)$ between two valuation methods X and Y. F_X and F_Y denote here the CDF of the absolute valuation errors $|E|$ of methods X and Y, respectively.

To summarize, the notion of dominance defined in Table 1 yields the most exhaustive criterion for comparing multiples' accuracy. It directly generalizes the approaches of Alford (1992), Cheng and McNamara (2000), Kaplan and Ruback (1995), Kim and Ritter (1999), Gilson et al. (2000), Lie and Lie (2002). While these studies compare a few particular percentiles of the distribution of absolute errors of competing multiples, the stochastic dominance approach draws its conclusions by a comparison of *all* the percentiles of the competing methods' error distribution.

2.3 The set of comparable firms

The definition of the group of peers C depends on the selection criteria of comparable companies as well as on the size of the set.

2.3.1 Characteristics of peers

The valuation literature states that the comparable companies should be chosen to be similar to the target firm along the dimensions that determine the multiples: risk, growth and cash flows (see for example, Damodaran (2006), McKinsey, Koller, Goedhart, and Wessels (2015)). Analysts as well as academics (Liu et al. (2002), Lie and Lie (2002)) often define comparable firms to be other companies in the target firms business. The implicit assumption made here is that firms in the same industry have similar risk, growth, and cash flow profiles and therefore can be compared with more legitimacy. An alternative to this practice consists in looking for firms that are similar in terms of valuation fundamentals. The specific measures of risk, growth, and cash flow generating potential to be used then depend on the multiple. For the P/E ratio, Alford (1992) and Cheng and McNamara (2000) compare the effectiveness of using peers chosen from the same industry with that of categorizations based upon fundamentals such as risk (proxied by total assets) and growth quality (measured by return on equity). With enterprise value multiples, the quality of growth is best captured by the return on invested capital (ROIC) (see Damodaran (2006), McKinsey et al. (2015)).

Following these recommandations, we use six methods of selecting comparable firms, based on three firm characteristics: industry membership, firm size (as measured by Total Assets), and growth quality (measured by Return on Equity or Return On Invested Capital, respectively)[6]. The six peer selection criteria are: Industry (I), Total Assets (TA), Return on Equity (ROE)/Return On Invested Capital (ROIC), Industry + TA, ROE/ROIC + Industry, ROE/ROIC + TA. For the Industry criterion, the peers are firms matched on the basis of four-digit SIC codes if the resulting industry contains enough many peers. Otherwise the definition of a firm's industry is progressively broadened until enough firms are identified. For TA and ROE/ROIC criteria, the peers are firms closest to the target company in terms of TA or ROE/ROIC. When Industry + TA or ROE/ROIC + Industry are used, the peers are firms in the set Industry (above) closest in terms of TA or ROE/ROIC. Finally, the peers chosen based on TA + ROE/ROIC criterion are obtained as the

[6] ROE is also a surrogate of earnings growth (Freeman, Ohlson, & Penman, 1982).

intersection of firms closest in terms of TA and of ROE/ROIC.

2.3.2 Optimal number of peers

The value of the multiple, and hence the predicted price, depends not only on the characteristics of the peers (industry, firm size and firm performance) but also on the size of the set of comparable companies. In the literature, no study provides a rigorous motivation for the choice of the size of the set of peers. For instance, Cheng and McNamara (2000) discuss valuation accuracy when the number of peers vary as a function of the number of digits of the SIC industry codes used to construct the set of comparable firms. However, they do not discuss the issue of the size of the set of peers for other peer selection criteria.

Thus, our paper extends the existing literature by determining an optimal number of peers for which the resulting valuation has the best *overall* accuracy relative to any other valuation based on a different number of peers. The natural definition of the optimal number of peers (k_{opt}) to be used in estimating the multiple reads as follows.

The optimal number of peers (k_{opt}) to be used in estimating the multiple is the number that yields a valuation *at least as accurate* as any other valuation (based on the same multiple and the same criterion of peer selection) with a different number of peers. The valuation accuracy of different choices of the number of peers is evaluated via the stochastic dominance and leads to the best *overall* accuracy.

More precisely, for a given pair (multiple, peer selection criterion), denote by F_k the CDF of the absolute errors $|E|$ when exactly k peers are used in the prediction of the multiple. Then the optimal number of comparable firms k_{opt} to be used in estimating the multiple, is defined by the condition:

$$F_{k_{opt}} \geq F_k \quad \text{for all} \quad k \neq k_{opt}, \tag{6}$$

where the inequality is in the sense of Table 1. In other words, the pricing using the k_{opt} number of peers is *overall* more precise than any other pricing using the same pair (multiple, peer selection criterion) but a different number of peers[7]. For

[7]Note that the optimal k might not be unique.

details about how to make this definition operational see Section B in Appendix.

2.4 Sample and variables

Our final sample comprises 96,598 firm-year observations, covering a period of 45 years between 1968 and 2012. The sample was constructed starting with all entries available in the CRSP/Compustat database and removing all the firm-year observations with

1. SIC code between 6000 and 6799 (no financial companies);
2. missing of at least one of the variables needed to construct the multiples and to define the comparable companies;
3. a negative value for at least one of the EBITDA or E;
4. at least one multiple that was among the largest/smallest 1%;
5. the 1st digit SIC industry set containing less than 50 peers. This last criteria is needed for ensuring enough comparable companies when the industry peer selection criterion is used.

A table that shows the size reduction corresponding to each of the mentioned steps can be found in section B.2 of the Appendix.

Table 2 provides descriptive statistics for the distribution of the two multiples under scrutiny. We note that both the position and the dispersion measures of the P/E ratio dominate the corresponding values of the EV/EBITDA multiple.

Multiple	Min	25%-ile	Mean	Median	75%-ile	Max	S.D.
P/E	3.22	10.44	23.52	15.77	24.82	301.60	28.37
EV/EBITDA	3.37	7.03	11.64	9.38	13.12	78.75	8.22

Table 2: **Descriptive statistics of the distribution of the** P/E **and** EV/EBITDA **multiples.** Both the position and the dispersion measures of the P/E ratio dominate the corresponding ones of the EV/EBITDA multiple.

The variables used to construct the multiples were Common Shares Outstanding (CSHO), EPS excluding Extraordinary Items (EPSPX), Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA), Common Equity (CEQ), Net Income (NI), Long Term Debt (DLTT), Current Liabilities (LCT), Preferred Stock (UPSTK) and Preferred Dividends in Arrears (DVPA). The variables used for defining comparable companies (the peers) were Standard Industrial Classification code (SIC), Total Assets (TA) and Return on Equity (ROE)/Return on Invested Capital (ROIC).

The share closing price (P) was obtained from CRSP database. The share closing price corresponds to the third month after the end of the reporting period. The Enterprise Value per share (EV) was defined as the sum of DLTT (deflated by CSHO), LCT (deflated by CSHO), UPSTK (deflated by CSHO) and DVPA (deflated by CSHO), and P.

3 Main results

We start this section by presenting the analysis that motivates the choice of the optimal number of comparable firms for the six peer selection criteria under investigation. These optimal choices are then used to produce the valuation errors which the rest of the paper analyzes in detail.

3.1 Accuracy sensitivity to the number of peers

Figure 1 shows the effect on valuation's accuracy of the number of peers used in predicting the target firm's multiple. Accuracy is measured here by the median absolute error. They provide preliminary motivation for the optimal choice of the number of peers that is made precise later in the section.

Each one of the six graphs in Figure 1 displays the two median absolute error curves ((1) for P/E and (2) for EV/EBITDA) corresponding to valuation using one of the six peer selection criteria under discussion. The vertical line in each graph indicates the optimal number of peers (k_{opt}) for each method of peer selection as explained in the sequel. The levels of the curves yield a consistent ordering of the valuation accuracy of the multiples *as measured by the median absolute error*. EV/EBITDA (curve 2) displays higher median absolute error for all the range

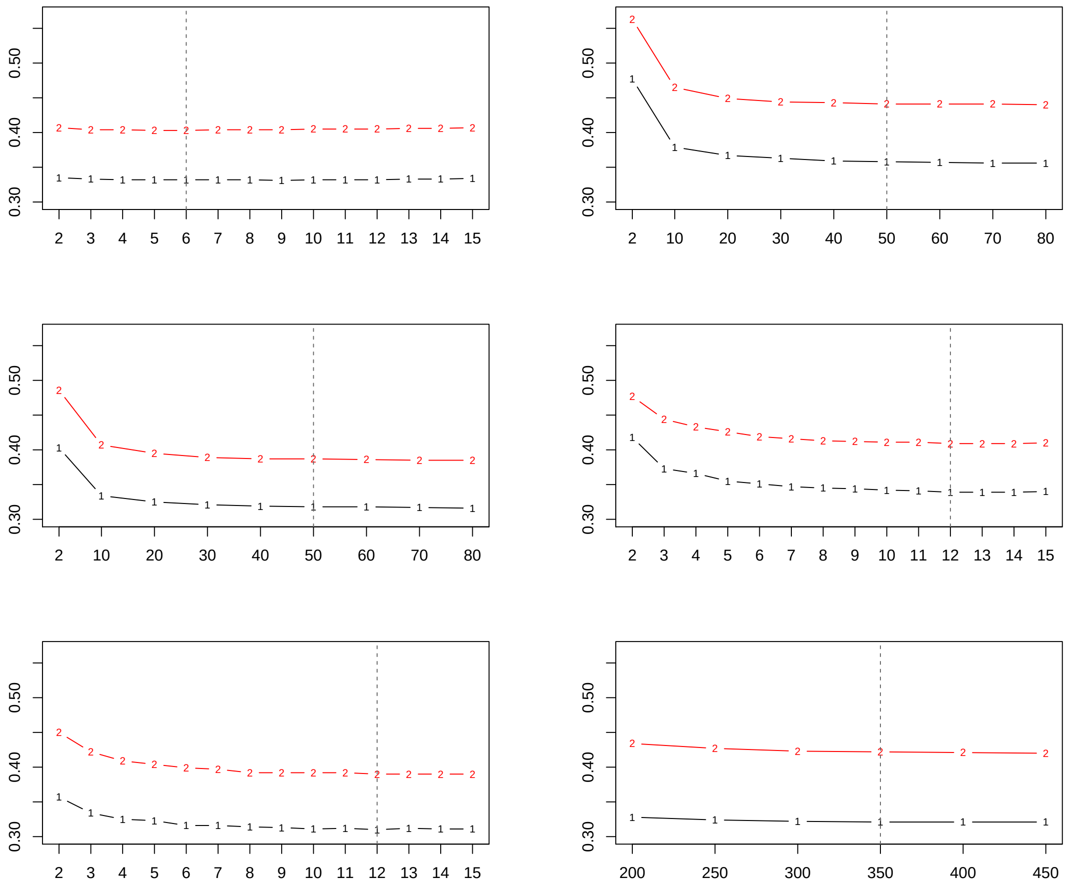


Figure 1: **Median absolute error $|E|$ as a function of the number of peers.** The graphs correspond to the six peer selection approaches: I and TA (*first row*), ROE/ROIC and I+TA (*second row*), I+ROE/ROIC and TA+ROE/ROIC (*third row*). Each curve corresponds to a multiple: (1) P/E, (2) EV/EBITDA. For each peer selection approach, the same pattern appears for both multiples. A consistent ordering of the multiples emerges: EV/EBITDA valuation displays constantly higher median absolute error. Peer selection based on ROE/ROIC criteria (left-second and left-third rows) seems most precise. The vertical lines indicate the optimal number of peers k_{opt} as defined in (6).

of k and for all peer selection criteria. Hence, the graphs provide preliminary evidence of the fact that valuations based on the multiple P/E are more accurate, independently of the peer selection criterion.

To motivate our choice of the optimal number of peers defined by equation (6), the analysis on the valuation precision as measured by the median error in Figure 1 needs to be refined through the use of the notion of stochastic dominance[8]. Since presenting the details of such an analysis would lengthen the presentation of the

[8] Recall that stochastic dominance compares the error distributions corresponding to different sizes of the set of peers *in their totality* and not only through one particular statistic of the distribution.

results we decided to include them in section B.3 of the Appendix and to discuss here only the main findings.

We find that the optimal number of comparable firms depends only on the peer selection approach and not on the multiple used. For the Industry criterion (I), all the number of peers between 3 and 15 yield equally precise valuations while a k greater than 16 is sub-optimal. Alford (1992) uses $k = 6$, while Cheng and McNamara (2000) find that the minimum median absolute error is attained at $k = 7$. Our analysis will use an optimal number[9] of peers equal to 6: $k_{opt}^{I} = 6$.

For TA and ROE/ROIC criteria, our optimal choice for the number of peers is $k_{opt}^{TA,ROE/ROIC} = 50$. Values smaller than 30 perform worse than our choice, while values higher then 40 yield valuations that are statistically as precise as that using the optimal number of 50. It is worth mentioning that Cheng and McNamara (2000) choose $k^{TA,ROE} = 6$, a value far from the optimal one (see also the graph on second row, first column of Figure 1). This choice, most likely, has penalized the accuracy of valuation with peers selected on the ROE and TA criteria. In particular, for the P/E multiple, they find that ROE based valuation performed poorly, second only to TA. In contrast to their findings, we document superior performance for the pair (P/E , ROE). This pair is the second most accurate among the valuation methods under investigation.

The optimal choice for I+TA and I+ROE/ROIC criteria is $k_{opt}^{I+TA,I+ROE/ROIC} = 12$ while for TA+ROE/ROIC peer selection approach, one optimal choice[10] is $k_{opt}^{TA+ROE/ROIC} = 350$. Values smaller than 10 (for I+TA and I+ROE/ROIC) or 250 (for TA+ROE/ROIC), respectively, perform strictly worse than our choices. Values larger than 10 (for I+TA and I+ROE/ROIC) and 300 (for the TA+ROE/ROIC) respectively, produce valuations statistically equally accurate as the valuations of our choice.

[9]Due to the definition of the Industry peer selection (see Section 2.3.1), the number of comparable companies varies with the firm. The corresponding descriptive statistics are: mean=45, median=25, min=6, 1st quartile=11, 3rd quartile=56, max=689. The optimal choice of $k = 6$ corresponds hence to estimating the multiple from a set that averages 45 comparable companies.

[10]Since for a given k the peers are the intersection of the sets of the k firms with the closest TA and ROE/ROIC, respectively, the number of comparable companies varies with the firm. The corresponding descriptive statistics are: mean=58, median=54, min=14, 1st quartile=43, 3rd quartile=69, max=279. The optimal choice of $k = 350$ corresponds hence to estimating the multiple from a set that averages 58 comparable companies.

To sum-up, industry aside, the optimal number of comparable companies to be used in valuation is in fact a threshold. Valuations using any value above the threshold produce statistically equally accurate results. The optimal threshold is peer selection criterion specific. When the peers are chosen from a particularly large pool of candidates, using a low number of comparable companies to estimate the target multiple yields larger-than-needed errors.

3.2 Analysis of valuation errors

This section compares the absolute errors $|E|$ from valuations based on the two multiples under investigation. The errors correspond to valuations where, for every peer selection criterion, the optimal number of peers (obtained in the previous section) was used to define the set of comparable companies. First, we take the perspective common to the valuation literature and evaluate accuracy through a few particular percentiles of the distribution of errors. These results are comparable with the ones in the related literature. Then, in section 3.2.2 we report the outcomes of the encompassing analysis (based on the stochastic dominance notion) that compares *all percentiles* of two competing distributions of errors. This results are novel and more general than the existing ones.

3.2.1 Summary statistics

Table 3 shows the medians and the 25%, 75% and the 90% percentiles of the distributions of absolute errors (multiple prediction by the median-left hand half, and harmonic mean-right hand half of the table) of the (2 multiples) × (6 peer selection methods) valuation approaches under study. The methods are ordered by the median of the absolute error, from the smallest to the largest. The values in the table can be directly compared to those in Alford (1992) and Cheng and McNamara (2000).

The measures of dispersion in Table 3 are larger than the similar ones in Alford (1992) and Cheng and McNamara (2000) reflecting, most likely, the fact that our sample is much larger. A hypothesis test on the difference between medians of absolute error shows that the first three combinations of multiple and peer selection in

No.	Multiple	Criterion	Median				Harmonic Mean			
			25%-ile	Median	75%-ile	90%-ile	25%-ile	Median	75%-ile	90%-ile
1	P/E	I+ROE	0.13	0.30	0.55	0.86	0.14	0.30	0.52	0.77
2	P/E	ROE	0.15	0.31	0.54	0.84	0.15	0.31	0.52	0.75
3	P/E	TA+ROE	0.14	0.31	0.54	0.81	0.14	0.31	0.53	0.76
4	P/E	I	0.14	0.32	0.59	0.89	0.14	0.32	0.56	0.83
5	P/E	I+TA	0.14	0.33	0.61	0.92	0.15	0.32	0.57	0.84
6	P/E	TA	0.16	0.35	0.61	0.90	0.16	0.34	0.58	0.84
7	EV/EBITDA	I	0.17	0.38	0.68	1.12	0.17	0.37	0.66	1.04
8	EV/EBITDA	I+ROIC	0.17	0.38	0.68	1.15	0.17	0.38	0.66	1.05
9	EV/EBITDA	I+TA	0.18	0.39	0.70	1.17	0.18	0.38	0.67	1.07
10	EV/EBITDA	TA+ROIC	0.19	0.40	0.70	1.16	0.19	0.40	0.69	1.09
11	EV/EBITDA	ROIC	0.20	0.41	0.72	1.19	0.19	0.41	0.70	1.11
12	EV/EBITDA	TA	0.20	0.42	0.73	1.19	0.20	0.41	0.71	1.11

Table 3: **Descriptive statistics for the distributions of absolute errors.**

Table 3 have statistically equal median absolute errors. All the other combinations have significantly higher median absolute errors, and EV/EBITDA produces the highest median errors for all peer selection criteria.

3.2.2 Dominance analysis

Table 4 presents the dominance ordering of the valuations based on the two multiples under investigation. It displays the dominance measure defined in Table 1 calculated for all possible pairs of valuation approaches.

The (i, j) entry of the table represents the value of the dominance measure $dm(F_j, F_i)$ as explained in Table 1, where F_i is the i-th pair (multiple, peer selection criterion) (given by the first two columns of the table) in the ranking. We note than, in all the pairwise comparisons, a relation of domination can be established. The results in Table 4 allows for a number of important and clear cut conclusions.

Any valuation based on the P/E multiple is overall more accurate than any of the EV/EBITDA valuations. In other words, the most accurate EV/EBITDA valuation approach is less accurate than the least precise P/E-based valuation (with a strongly significant domination measure of 7%).

	Multiple	Peers	1	2	3	4	5	6
1	P/E	ROE+I	=					
2	P/E	ROE (+TA)	1	=				
3	P/E	I (+TA)	3	2	=			
4	P/E	TA	6	5	3	=		
5	EV/EBITDA	ROIC+I	8	7	6	4	=	
6	EV/EBITDA	I, ROIC, TA, ROIC+TA	10	9	7	4	3	=

Table 4: **Dominance ordering of the two multiples** EV/EBITDA **and** P/E**.** The (i, j) entry represents the value of the dominance measure $dm(F_j, F_i)$ (see Table 1 and the definition (9)) where F_i is the i-th pair (multiple, peer selection) (given by the first two columns of the table) in the ranking. The value 8 on the position (5,1) indicates that the pair (P/E , ROE+I) (1) dominates the pair (EV/EBITDA , ROIC+I) (5) and that the difference between the estimated cdfs of the absolute relative pricing errors of the two methods is at most 8%, i.e. $0 \leq F_1(e) - F_5(e) \leq 8\%$.

The EV/EBITDA multiple is significantly less precise[11] than the P/E : for any level of error $e = 0.01, 0.10, 0.25, \dots$ the proportion of firms priced by the most accurate P/E valuation with a precision better than e is larger than the proportion of firms valued by the most exact EV/EBITDA approach within the same precision. Moreover, for some error level, the difference between the two proportions can be as large as 11%.

Another conclusion is that, in the frame of P/E valuation, the TA criterion is the least interesting for the choice of peers. First, enhancing the Industry or ROE based selection of comparable firms by including the TA information does not improve valuation's precision (the second and third entries in the table). Second, the pair (P/E , TA) yields the least accurate pricing for the P/E multiple.

To give more intuition and some visual support to judging the magnitude of the differences in accuracy in Table 4, Figure 2 displays and compares the estimated CDFs of errors from, on one hand, the most accurate P/E valuation (i.e. P/E, ROE+I), and, on the other hand, the best performing EV/EBITDA valuation, i.e. (EV/EBITDA , ROIC+I). The left-hand side graph shows the two CDFs, while the graph on the right shows their difference together with the 99% confidence band (for statistical details, see section A in the Appendix.)

[11]Formally this reads as: the difference between the estimated CDF of the absolute pricing errors of the most accurate P/E valuation and the most precise EV/EBITDA pricing is positive and at most 11%, i.e. for all level of error e, $0 \leq F_{\mathrm{P/E}}(e) - F_{\mathrm{EV/EBITDA}}(e) \leq 11\%$.

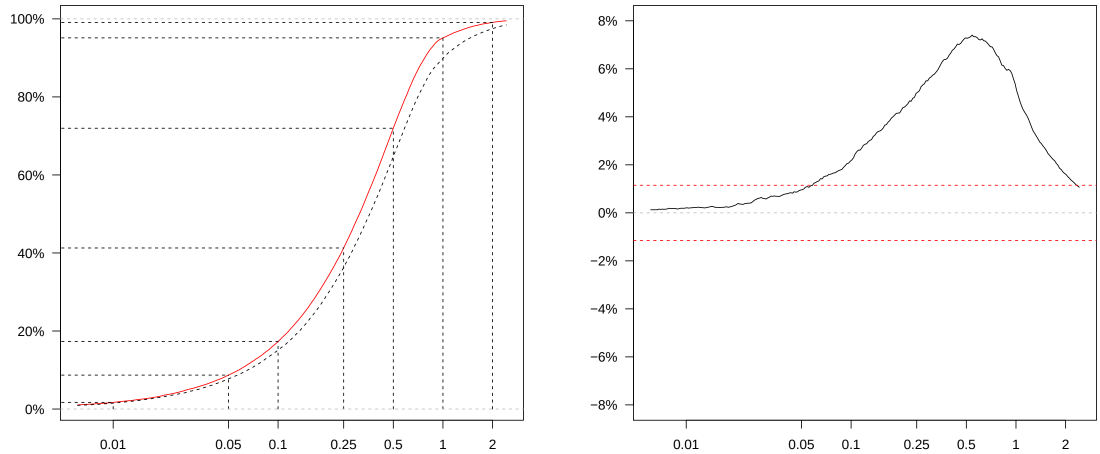


Figure 2: **The estimated CDF of $|E|$ (logarithmic scale) of selected valuation pairs (multiple, peer selection).** The graph on the left displays the cdfs of errors corresponding to valuation by the pair (P/E, ROE+I) (full line) and (EV/EBITDA, ROIC+I). On the right, the difference between the two CDFs is displayed together with the 99% confidence band based on the Kolmogorov-Smirnov statistic (see Section A in the Appendix).

3.3 Additional results

In this section, we present additional results regarding the difference in accuracy between P/E and EV/EBITDA: (1) when the median is used instead of the harmonic mean to aggregate the peers multiples; (2) over time and by industry; (3) when forward looking data is used;

3.3.1 Median vs. harmonic mean

We first discus the difference in accuracy between the use of the median or of the harmonic mean to aggregate peers' multiples. The error distribution corresponding to harmonic mean estimation stochastically dominates the one associated to median estimation of the multiple. The gain in precision, as measured by the dominance measure is, at roughly 2%, rather small.

3.3.2 Time evolution of accuracy

Since EV/EBITDA multiple gained in popularity with the practitioners during the last decade, it is interesting to analyze the accuracy of this multiple through the years. This section looks at the time evolution of the accuracy difference between P/E and EV/EBITDA, first on the whole sample, then on eight 1 digit SIC

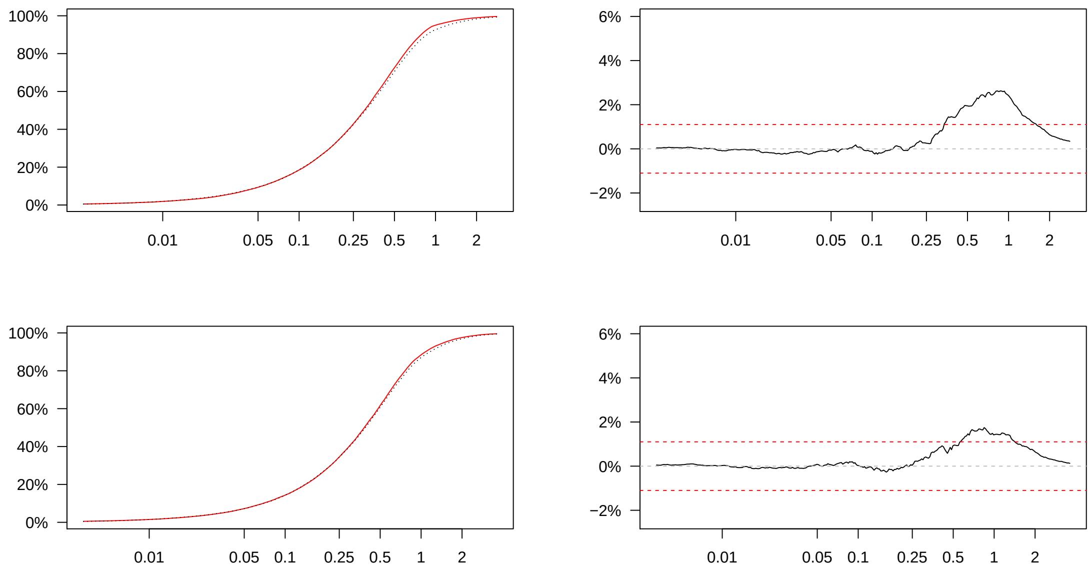


Figure 3: **Median vs. harmonic mean: the estimated CDFs of $|E|$ (left) and their differences (right) (logarithmic scale).** For each of the valuations (P/E , ROE+I) (top) and (EV/EBITDA , ROIC+I) (bottom), the figure compares the estimated CDFs of the errors when the multiple of the target firm is calculated as the median (dotted line, left-hand graph) or as the harmonic mean (full line, left-hand graph) of the set of multiples of comparable companies. The graphs on the right show the differences between the two CDFs with the 99% confidence band (see Section A for details). The error distribution corresponding to harmonic mean estimation dominates the one associated to median estimation of the multiple. The gain in precision, as measured by the dominance measure is, at roughly 2%, rather small.

industries.

We divided the original sample in nine intervals of five years and we repeated the analysis detailed in Section 3.2.2 on each one of them. The left-hand side graph in Figure 4 presents the more salient findings of this analysis. It displays the evolution of the difference in accuracy between, on one hand, the top methods in Table 4 (all P/E -based valuations) plus that of the best performing EV/EBITDA -based approach and, on the other hand, the best performing method of each five-year period. It is worth noting that the pair (P/E , ROE+I) has yielded consistently the best performing valuation in all nine 5-year periods. While the gap to the best performing P/E valuation decreased through time, the best EV/EBITDA - based approach remains strongly inferior. The smallest difference to the best P/E performing valuation corresponds to the most recent 5 year period and is as high as 7%.

For P/E valuation, it is worth noticing that, while in the beginning of the sample, the ROE+I peer selection criterion yielded more precise valuations, in the last 25 years the top three peer selection criteria produce equally accurate pricing showing a reduction of the importance of the choice of peer selection criterion.

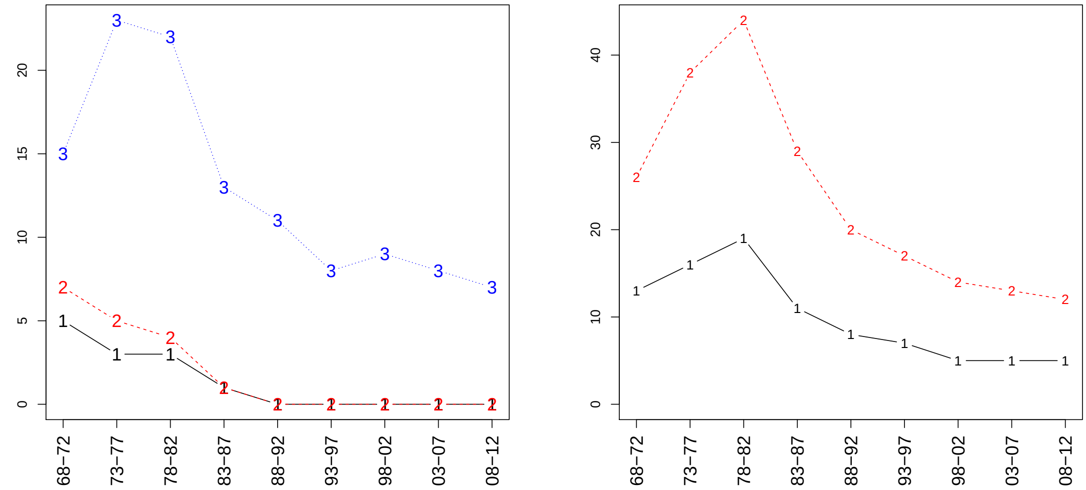


Figure 4: **Time evolution of the accuracy gap.** (*Left - Overall sample*): Time evolution of the accuracy gap between, on one hand, the second and third most accurate P/E valuations, i.e. (P/E , ROE) (curve 1), (P/E , I) (curve 2), and the best performing EV/EBITDA -based approach (EV/EBITDA , ROIC+I) (curve 3) on one hand, and the best P/E-based valuation of each sub-period, on the other hand. The accuracy gap is measured by the dominance measure. While the difference to the best performing P/E valuation is decreasing in time, the best EV/EBITDA - based approach remains strongly inferior. The smallest gap to the best performing valuation corresponds to the most recent 5 year period and is as high as 7%. (*Right - 1 digit SIC industries*): Time evolution of the dominance measure of the best performing P/E valuation over the best performing EV/EBITDA approach for the two industries for which the gap is lowest (curve 1), highest (curve 2), respectively. The lower curve is that of Manufacturing (SIC 2) while the upper curve corresponds to Transportation, Communications, and Utilities (SIC 4). The performance gap of the EV/EBITDA valuation for the other industries lies in between the two curves. The smallest gap is of 5% (15%) for the industry for which the EV/EBITDA based approach works the best (worst) and is attained in the most recent interval.

As noted in the introduction, EV/EBITDA is analysts' multiple of choice for certain industries (Block, 2010). Such a preference seems to imply a better relative accuracy of this multiple in some industries. The second half of the section looks at this issue. Since the mentioned preference is of a relatively recent date, our analysis is performed through time.

For each of the nine five year intervals and for each of the eight[12] industries defined by the 1 digit SIC code[13], we identified the best performing P/E and

[12]We removed the Financial and Public Administration, the last one for lack of data.

[13]Due to data availability limitations, we could not perform a more detailed analysis, where industries are defined by a 2 digit or higher SIC code.

EV/EBITDA valuations. We found that, independent of the industry and despite a narrowing gap, the best performing P/E valuation of each of the seven year intervals dominates the corresponding best performing EV/EBITDA.

The most salient results of the analysis are shown in the graph on the right of Figure 4. It shows the results for the most extreme two industries: Transportation, Communications, and Utilities as defined by the SIC 4, the industry in which the EV/EBITDA valuation is affected by the wider precision gap with respect to the best P/E valuation and Manufacturing (SIC 2), the industry where the gap is at it's lowest. The performance gap of the EV/EBITDA valuation for the other industries lies hence in between the two curves. For the industry for which the EV/EBITDA based approach works the worst (best), the dominance gap shrunk with time from a high of 45% (20%) to a low of 15% (5%).

3.3.3 Forward looking multiples

This section presents the results of a dominance analysis similar to that in Sections 3.2.2 performed on a smaller, more recent sample of companies for which consensus forecasts of future earnings and EBITDA are available. The multiples under consideration are P/FE1 defined as the share price (P) divided by future Earnings Per Share one year ahead, P/FE2 defined as share price (P) divided by future Earnings Per Share (EPS) two years ahead, and EV/FEBITDA defined as Enterprise Value per share divided by future EBITDA one year ahead deflated by CSHO. The peer selection approaches are those discussed in Section 2.3.1 with the exception of the TA+ROE/ROIC[14]. The analysis of the optimal choice for the number of comparable companies (in the spirit of Section 2.3.2) is reported in Section B.4 of the Appendix. Table 5 displays the dominance structure of the valuation approaches obtained by combining the four forward looking multiples with the five methods of peer selection.

The forward looking multiples based on forecasted future earnings dominate in precision with the multiple P/FE2 being the most accurate, independent of the peer selection method used, followed by the multiple P/FE1. The forward looking

[14]The sample size is too small for a rigorous study of this last approach.

	Multiples	Peers	1	2	3	4	5	6
1	P/FE2	all criteria	=					
2	P/FE1	all criteria	6	=				
3	P/E	I+ROE	18	13	=			
4	EV/FEBITDA	I, TA I+TA, I+ROIC	23	17	7	=		
5	EV/FEBITDA	TA	25	20	9	1	=	
6	EV/FEBITDA	ROIC, TA+ROIC	28	22	11	7	2	=

Table 5: **Dominance ordering based on forward looking multiples.** The entry (i, j) in the table represents the estimated value of the dominance measure $dm(F_j, F_i)$ (see Table 1 and definition (9)) where F_i is the i-th pair (multiple, peer selection) in the ranking given by the first two columns of the table. The value 23 on the entry (4,1) in the table indicates that valuations using the ratio P/FE2 paired with any peer selection method (1) dominate the valuations based on the ration EV/FEBITDA paired with any of the four peer selection approaches: I, TA I+TA, I+ROIC, (4) and that the difference between the estimated CDF of the absolute relative pricing errors of the two groups of methods is at most 23%, i.e. $0 \leq F_1(e) - F_5(e) \leq 23\%$.

P/E multiples dominate the historical P/E valuation (by 18% and 13%, respectively). This confirms and extends the findings of Liu et al. (2002). The forward looking multiple based on the future EBITDA performed significantly worse: its dominance measure with respect to P/FE2 is -23%. Even the best performing historical P/E valuation, which for this sample was, again, the pair (P/E, ROE+I), is more accurate than EV/FEBITDA (dominance measure of at least 7%).

4 Impact of the differences between P/E and EV/EBITDA

To better understand the findings in the previous section, this section looks at the accuracy gap as a function of the two terms at the origin of the difference between the definitions of the two multiples: EV-P or debt (which relates the numerators of the two multiples) and EBITDA-E (which relates the two denominators).

4.1 Main results

When analyzing the impact on accuracy of the difference between the numerators of the two multiples, (EV-P), we try to understand in fact how debt affects the relative precision of the two multiples. For the analysis along this dimension, we consider three levels of debt: low, medium, and high. While the first difference corresponds to a simple item, debt, the difference between the denominators (EBITDA-E) includes many items: Depreciation and Amortization (D&A), Interest and Taxes (I&T), but also Non-Operating items (NOPI) and Special items (SPI). To understand how the level of the second difference (EBITDA-E) affects the relative accuracy of the multiples, we split the firms along this dimension into three groups: low, medium, and high difference between EBITDA and E.

The marks that separate the levels of the two differences correspond to the 33% and 66%-percentiles of the ratio of debt to enterprise value (which corresponds to 0.2 and 0.4 of EV), and the proportion of EBITDA represented by the EBITDA-E (which correspond to 0.5 and 0.65 of EBITDA). We performed a stochastic dominance analysis on the 9 sub-samples defined by different levels in the two dimensions of debt and EBITDA-E.

	EBITDA-E			
Debt	Low	Medium	High	Differences due to EBITDA-E
Low	0% (=)	-4% (4)	-16% (6)	-16%
Medium	13% (4)	15% (6)	12% (6)	1%
High	12% (6)	17% (6)	25% (6)	13%
Difference due to debt	12%	21%	41%	

Table 6: **Dominance ordering based on the level of EV-P and EBITDA-E.** The table reports the dominance measure of the most accurate P/E valuation over the best EV/EBITDA pricing for the given sub-sample. The numbers in parenthesis give the difference between the position of the method in the overall hierarchy (similar to Table 4). The sign '=' indicates that the two most precise valuations are equally accurate on the given sub-sample. Negative values signal situations when EV/EBITDA is more precise than P/E.

Table 7 shows that both differences have an impact on the accuracy gap between the two multiples. The effects are, however, different. On one hand, the impact of debt (EV-P) on the relative accuracy of EV/EBITDA is negative. The increase of the level of debt (from the group with low debt to the group with high debt), worsens the performance of EV/EBITDA (as measured by the dominance measure) by 12% (low difference in EBITDA-E), 21% (medium difference in EBITDA-E), and

	EBITDA-E			
Debt	Low	Medium	High	Differences due to EBITDA-E
Low	0% (=)	-5% (4)	-20% (6)	-20%
Medium	9% (4)	14% (6)	13% (6)	5%
High	7% (6)	12% (6)	24% (6)	17%
Difference due to debt	9%	19%	45%	

Table 7: **Dominance ordering based on the level of EV-P and EBITDA-E.** The table reports the dominance measure of the most accurate P/E valuation over the best EV/EBITDA pricing for the given sub-sample. The numbers in parenthesis give the difference between the position of the method in the overall hierarchy (similar to Table 4). The sign '=' indicates that the two most precise valuations are equally accurate on the given sub-sample. Negative values signal situations when EV/EBITDA is more precise than P/E.

41% (high difference in EBITDA-E), respectively. On the other hand, the impact of the second difference (EBITDA-E) is more subtle. More precisely, it depends on the level of debt. When the debt is high, passing from a low difference between EBITDA and E to a large one reduces the relative accuracy of EV/EBITDA by 13%. For medium levels of debt, going from low differences between EBITDA and E to large ones does not affect the relative accuracy of EV/EBITDA (the difference is equal to 1%). Finally, if the level of debt is low, a larger difference EBITDA-E improves the relative accuracy of EV/EBITDA by 16%. For two of the nine sub-samples, (Low debt, Medium EBITDA-E) firms and (Low debt, High EBITDA-E) firms, we find that EV/EBITDA dominates P/E.

We conclude that debt is the term that most consistently influences accuracy: whenever the debt term is sufficiently large (66% of the sample), the EV/EBITDA valuation is less precise than P/E valuation independent of the level of the difference between EBITDA-E. The accuracy of EV/EBITDA is negatively affected by the difference EBITDA-E only for higher levels of debt. When the debt is relatively low, i.e. on a third of the sample, the precision of EV/EBITDA is at least as good as that of P/E. For at least 10% of the firm-years under investigation the EV/EBITDA is (statistically significant) more precise than P/E .

4.2 Additional analysis of the impact of EBITDA-E

The results in the previous section raise the natural question of why large differences (EBITDA-E) are associated with improved relative accuracy of EBITDA (for companies with low debt). Our explanation, supported by the analysis in this section, is that not all items provide relevant information to market participants[15]. They include the special items and, possibly, the non-operating items. Our analysis ties to an interesting current of research in the financial reporting literature that suggest that the frequency and the size of non-recurring and special items have increased a lot over the last decades, and that still debates on how capital markets participants react to the disclosure of such items (Burgstahler et al., 2002; Cready et al., 2012; Dechow & Ge, 2006; Johnson et al., 2011; McVay, 2006; Riedl & Srinivasan, 2010).

Table 8 presents the results of an additional analysis of the group of low debt firm-years, which represent a third of the sample (first line in Table 7). It groups the firms in three sub-groups corresponding to the low, medium and high values of single (or combinations of) item(s) that compose the difference between EBITDA and Earnings. For every one of the three levels, we report the measure of dominance of the best performing P/E over the most accurate EV/EBITDA, as well as the difference in their ranks. For the rank difference, we find a maximum value of 6 (group with high difference between EBITDA and E), which corresponds (when the difference is positive) to all P/E valuations being better than the best EV/EBITDA valuation. Negative values signal situations when EV/EBITDA is more precise than P/E.

We analyze 5 groups of items (I&T, D&A, SPI, NOPI, SPI&NOPI) and interpret the numbers in table 8 as follows. For a given group of items and a given level, a negative (positive) value of the dominance measure shows that the precision of the EV/EBITDA (P/E) valuation is superior to that of P/E (EV/EBITDA) valuation. As the value of the item has an impact only on the P/E ratio, we interpret that as evidence that the item does not (does) conveys information to the investors. The price set by a valuation with a multiple that is not affected by the item (EV/EBITDA) is closer to the market price than that based on a multiple

[15]In particular, the analysts might not want to take them into account when calculating multiples.

	Level		
Item	Low	Medium	High
EBITDA-E	0% (=)	-5% (4)	-20% (6)
I&T	0% (=)	0% (=)	0% (=)
D&A	0% (=)	0% (=)	0% (=)
-SPI	0% (=)	0% (=)	-12% (6)
-NOPI	0% (=)	0% (=)	-5% (6)
-SPI-NOPI	3% (3)	0% (=)	-14% (6)

Table 8: **Dominance ordering based on groups of items in the difference EBITDA-E.** The table reports the dominance measure of the most accurate P/E valuation over the best EV/EBITDA pricing for the low debt sub-sample of firms-years. The numbers in parenthesis give the difference between the position of the method in the overall hierarchy (similar to Table 4). The sign '=' indicates that the two most precise valuation approaches are equally accurate on the given sub-sample. Negative values signal situations when EV/EBITDA is more precise than P/E.

that is informed by the item (P/E).

The results in Table 8 indicate clearly which groups of items and which levels do not seem to convey information to the market participants. They are, in order of relevance, largely negative SPI&NOPI, largely negative SPI and largely negative NOPI. EV/EVITDA strongly dominates P/E (-14%) when SPI & NOPI is negative and represent a large proportion of EBITDA. The dominance is almost as strong in the sub-sample of firms with largely negative SPI (-12%) and significantly weaker for the firms with largely negative NOPI (-5%). In the other cases, either P/E dominates EV/EBITDA (when the 5 groups of items represent a small proportion of EBITDA), or the accuracy of P/E and EV/EBITDA are identical (when the 5 groups of items represent a medium proportion of EBITDA).

4.3 The effect of restated earnings

The findings in the previous section suggest obvious adjustments to the net income that might render it more consistent with market prices. As before, we think of the items and levels for which the precision of the EV/EBITDA valuation is superior as not conveying relevant information to the investors: removing the item in cause from EBITDA to construct the net earnings yields a multiple valuation (P/E) that is less aligned with the market prices. As a consequence, a better measure of operating performance would add back the concerned item to net income. One would restate the GAAP earnings as earnings before the concerned item to produce an operating performance measure more aligned with the market prices. Since the item values concerned are negative, the restatements yield higher net earnings.

This section evaluates the pertinence of this hypothesis. More concretely, three types of restated earnings were constructed. For low debt firms-years with extreme negative values of SPI, NOPI or their sum, the earnings were restated as earnings before the concerned item (or sum of items). Then new valuations were performed using the six peer selection criteria and the optimal number of peers as explained in Section 2.

Multiples with restated earnings		
P/(E-SPI)	P/(E-NOPI)	P/(E-SPI-NOPI)
-13% (6)	0% (=)	-14% (6)

Table 9: **Dominance ordering between P/E and alternative valuations based on restated earnings.** The table reports the dominance measure of the most accurate P/E valuation over the best pricing based on Price-to-Earnings multiples with restated earnings. For low debt firms-years with extreme negative values of an item (or sum of items), the earnings are restated as earnings before the concerned item. The numbers in parenthesis give the difference between the position of the method in the overall hierarchy (similar to Table 4 with EV/EBITDA replaced by the new multiple). The sign '=' indicates that the two most precise valuation approaches are equally accurate. Negative values signal situations where Price-to-(restated Earnings) multiples are more precise than Price-to-(GAAP Earnings).

Table 9 presents the results. While restating the earnings for firm-years characterized by largely negative NOPI did not improve valuation accuracy of the Price-to-Earnings ratio, restating the earnings when affected by largely negative SPI values improved the P/E precision by 13%. When the restating was triggered by a largely negative sum SPI+NOPI, the gain in precision was of 14%.

To summarize, we find that the prices of firms characterized by largely negative SPI or SPI+NOPI seem to better reflect an operating performance measured by GAAP earnings before SPI or before SPI and NOPI, respectively. These findings help to understand why there has been an increased tendency of the analyst tracking services to report earnings before negative special items (Bradshaw & Sloan, 2002). If negative special items do not convey information to the investors, the net income before special items is a better measure of operating performance.

4.4 Who's responsible for the closing of the accuracy gap?

Tables 7, 8 and 9 show that the precision gap (whose time evolution is shown in Figure 4) is the compounding result of two contrasting effects. In a sample, firms with medium and high debt will influence negatively the precision of EV/EBITDA while the presence of firms with low debt and largely negative values of the sum SPI+NOPI (and hence understated GAAP earnings) will penalize the accuracy of P/E valuation. The sign and the size of the resulting accuracy gap depends on the proportions of firms with the contrasting characteristics. To visualize this interaction, Figure 5 displays the time evolution of the proportion of firms with characteristics that determine the precision gap.

More precisely, we display on the same graph the time evolution of the percentage represented by the firms with medium and high debt in the sample, the evolution of the precision gap as measured by the dominance measure defined in Table 1 (curve 3 in the graph on the left in Figure 4) (dashed line), the percentages represented by firms with largely negative SPI+NOPI in the sample (dash-dotted) and in the sub-sample of low debt firms (dashed line). The graph documents a strong similarity in the pattern of the debt level and that of the precision gap. The reduction of the accuracy gap for the last three decades is consistent with a reduction of the level of debt assumed by the firms.

Although smaller than in the 70's and 80's, the precision gap between the two multiples remains significant. The explanation seems to be that, despite the gradual increase in the proportion of firms with low level of debt and largely negative SPI+NOPI, i.e. firms better priced by the EV/EBITDA multiple (from next to 0% in the beginning of the sample to 18% in the end), the percentage of firms with medium and high debt, i.e. firms more precisely valued by P/E multiple, has been, in the recent decades, roughly three times bigger.

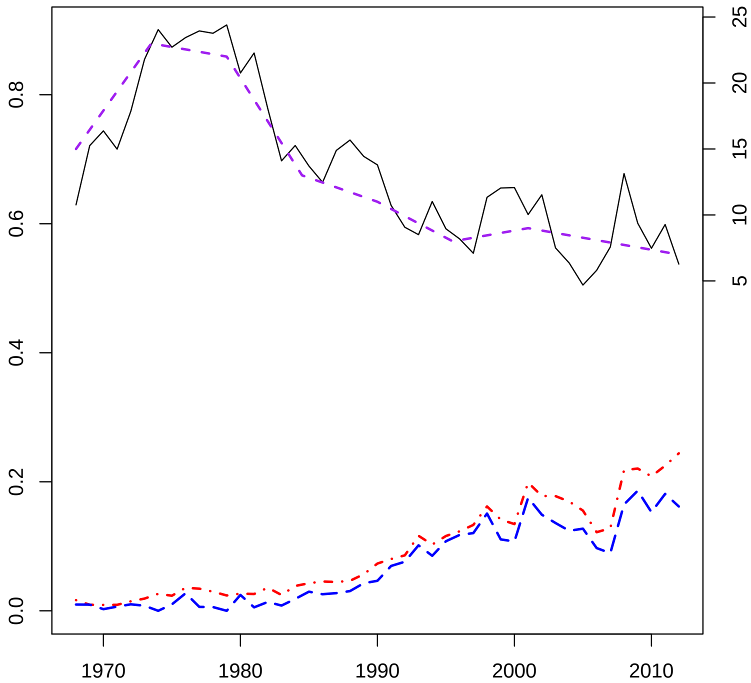


Figure 5: **Time evolution of proportions of firms with characteristics determining the precision gap between the two multiples.** *Top:* Yearly percentage represented by the firms with medium and high debt in the sample (full line). Superimposed, the evolution of the precision gap as measured by the dominance measure defined in Table 1 (curve 3 in the graph on the left in Figure 4) (dashed line). The scale for the precision gap in on the right vertical axis. *Bottom:* Yearly percentage represented by firms with largely negative SPI or SPI+NOPI in the sample (dash-dotted) and in the sub-sample of low debt firms (dashed line). We see a very good fit between the evolution of the precision gap and that of debt.

5 Conclusions

In this paper, we compare the accuracy of EV/EBITDA, a multiple that gained in popularity with the practitioners during the last decade, with that of the traditional P/E multiple. We systematically consider the impact of many implementation issues on accuracy, especially the comparable firm selection criteria, the size of the set of peers used in the construction of the multiple, and the aggregation of the target firms multiple. Extending the existing literature that compares multiples' precision using particular statistics of the distribution of valuation errors, we make use of the notion of stochastic dominance to compare the overall accuracy of the two multiples.

The results of our analysis suggest that, when aggregating over all firms in the sample, valuations based on EV/EBITDA are less accurate than these based on P/E whatever the implementation of the valuation method (i.e. criteria used for the selection of the peers, number of peers used). Our results are robust for aggregation of the peers multiples, forward-looking multiples, through time (over 45 years) and for various industries.

Our additional analysis of the impact on accuracy of the two terms (EV-P) and (EBITDA-E) at the origin of the difference between the two multiples refines the accuracy relation and puts forth two new results. First, we show that the domination of P/E decreases with the level of debt. Second, we document that EV/EBITDA is more accurate than P/E when valuing firms that have low level of debt and report largely negative values of special and non-operating items. This finding is in line with the idea that market participants, when valuing firms, care more about recurring items than about non-recurring and non-operating items. These new results are obviously interesting for investors and for accounting standard-setters.

Appendix

A Statistical details

For a given sample of errors $(e_1, e_2, \dots, e_{n_X})$, the estimator of F_X, the CDF of the absolute error $|E_X|$ of method X, is the *empirical cumulative distribution function*:

$$\widehat{F}_{X,n_X}(x) := \frac{1}{n_X}\sum_{i=1}^{n_X} I_{(-\infty,x]}(|e_i|) = \frac{\#\text{ of } |\text{ errors }| \le x}{n_X}, \tag{7}$$

where $I_A(x)$ is the indicator function:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases}$$

The statistical estimation error is described by the asymptotic distribution of the two sample Kolmogorov-Smirnov statistic:

$$D_{n_X,n_Y} := \sup_x |\widehat{F}_{X,n_X}(x) - \widehat{F}_{Y,n_Y}(x)|. \tag{8}$$

Under the null hypothesis that $F_X = F_Y$,

$$D_{n_X,n_Y} \le c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$

with probability $1-\alpha$[16].

To summarize, if $\widehat{F}_X$ and $\widehat{F}_Y$ denote the two estimated CDF of the absolute errors $|E_X|$ and $|E_Y|$ corresponding to the valuation methods X and Y, respectively, we

[16] Given the large size of our sample we will use $\alpha = 0.01$ and $c(0.01) = 1.63$.

define the following measure of dominance:

$$\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y) \stackrel{def}{=} \begin{cases} \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}, \\ \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS} \text{ and} \\ & \quad \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}, \\ 0, & \text{if } -c_{KS} \leq \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \text{ and} \\ & \quad \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}, \\ ?, & \text{if } \max\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min\limits_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}, \end{cases} \tag{9}$$

where $c_{KS} = c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$, with $c(\alpha)$ as above, n_X and n_Y, the sample sizes used to estimate the two cdfs $\widehat{F}_X$ and $\widehat{F}_Y$, respectively.

Table 10 gives the interpretation of the measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$ and will be referred to in the sequel whenever comparing the precision of the competing multiples.

B Optimal choice of number of peers by dominance analysis

B.1 Theoretical considerations

To make operational the choice of the optimal number of peers, let us denote by $F_k^{m,c}$ the CDF of the absolute errors $|E(m,C)|$ *when exactly* k *peers selected by the criteria* C *are used* in the construction of the multiple m. In other words, the peer set C consists of the k closest peers according to the peer selection criteria c. Formally, the optimal number of comparable firms k_{opt}, to be used in estimating

$\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$	When	Meaning
> 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X dominates F_Y or Method X is more precise than Y
< 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	F_Y dominates F_X or Method X is less precise than Y
$= 0$	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X is equal to F_Y or The 2 methods are equally precise
?	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	Neither cdf dominates or The 2 methods cannot be compared

Table 10: **The measure of dominance dm$(\widehat{F}_X, \widehat{F}_Y)$.** Interpretation of and necessary and sufficient conditions for the measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$. The notations $\succ$, $\prec$ and $\asymp$ correspond to (in)equalities which are statistically true, i.e. are not rejected by the corresponding Kolmogorov-Smirnov hypothesis test. $\widehat{F}_X$ and $\widehat{F}_Y$ denote here the estimated CDF (see definition (7)) of the absolute relative valuation errors $|E|$ of methods X and Y, respectively.

the multiple m, is defined[17] by the condition:

$$F_{k_{opt}}^{m,c} \geq F_k^{m,c} \text{ for all } k \neq k_{opt}.$$

When starting from data, this condition translates to:

$$dm(F_{k_{opt}}^{m,c}, F_k^{m,c}) \geq 0 \text{ for all } k \neq k_{opt},$$

[17]Note that the optimal k might not be unique.

or equivalently (see Table 10) to:

$$\begin{cases} \max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \succeq 0 \quad \text{and} \\ \min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \asymp 0 \qquad \text{for all } k \neq k_{opt}, \end{cases}$$

where notations $\succeq$ and $\asymp$ correspond to (in)equalities which are statistically true, i.e. are not rejected by the corresponding Kolmogorov-Smirnov hypothesis test. In particular, we should have

$$\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \succ 0 \text{ and } \min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \asymp 0$$

for all k that yield a valuation accuracy that is strictly worse than that of k_{opt} (strictly sub-optimal) and that

$$\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \asymp 0 \text{ and } \min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \asymp 0$$

for those k for which the valuation accuracy equals that of k_{opt} (optimal).

Since k_{opt} does not depend on the multiple (as shown by the graphs in Figure 1), the statistics needed to classify the k's as strictly sub-optimal or optimal are defined as:

$$\begin{aligned} M^c_k &:= \max_m \left(\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \right) \\ N^c_k &:= \min_m \left(\max_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \right) \\ O^c_k &:= \min_m \left(\min_e(\widehat{F}^{m,c}_{k_{opt}}(e) - \widehat{F}^{m,c}_{k}(e)) \right). \end{aligned} \tag{10}$$

Concretely, a k for which $N^c_k \succ 0$ and $O^c_k \asymp 0$ will be sub-optimal for all the multiples (i.e. for any of the multiples under study), the valuation using k comparable firms for constructing the multiple c is strictly less accurate than that based on k_{opt} peers. A k for which $M^c_k \asymp 0$ and $O^c_k \asymp 0$ yields (for all multiples) a valuation accuracy that is statistically indistinguishable from the best one.

B.2 The sample

Table 11 shows the size reduction of the sample that corresponds to each of the enumerated steps.

Sample construction	Size
All firm-year observations in CRSP/Compustat (1968-2012)	372,647
Exclude observations with SIC codes between 6000 and 6799	-76,270
	296,377
Exclude observations with missing values	-139,053
	157,324
Exclude observations with negative/infinite multiples	-46,872
	110,452
Exclude observations with at least a multiple among the largest/smallest 1% of the values	-5,608
	104,44
Exclude industries (1st digit SIC) with less than 50 peers in a given year	-6,088
Final sample size	98,756

Table 11: **Sample construction.**

B.3 Empirical results - initial sample

Figure 6 displays the dependency of the three statistics in definition (10) on k, the number of peers used in the construction of the multiple.

The three curves in each graph correspond in decreasing order to M_k^c, N_k^c, and O_k^c, respectively. The number constructing the curves identifies the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics (see equation (10)) is attained. The coding of the multiples is as follows : (1) - P/E , (2) - P/B, (3) - EV/EBITDA . The dotted lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. Note that $O_k^p \asymp 0$ for all k and all c and hence the values of the statistics M and N only will determine the relationship of optimality of a given k to k_{opt}. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers. It is worth noting that we are able to establish the relationship of optimality for all the k's in the range under discussion. The graphs in Figure 6 show that an appropriate choice of the size of the set of comparable companies is essential for

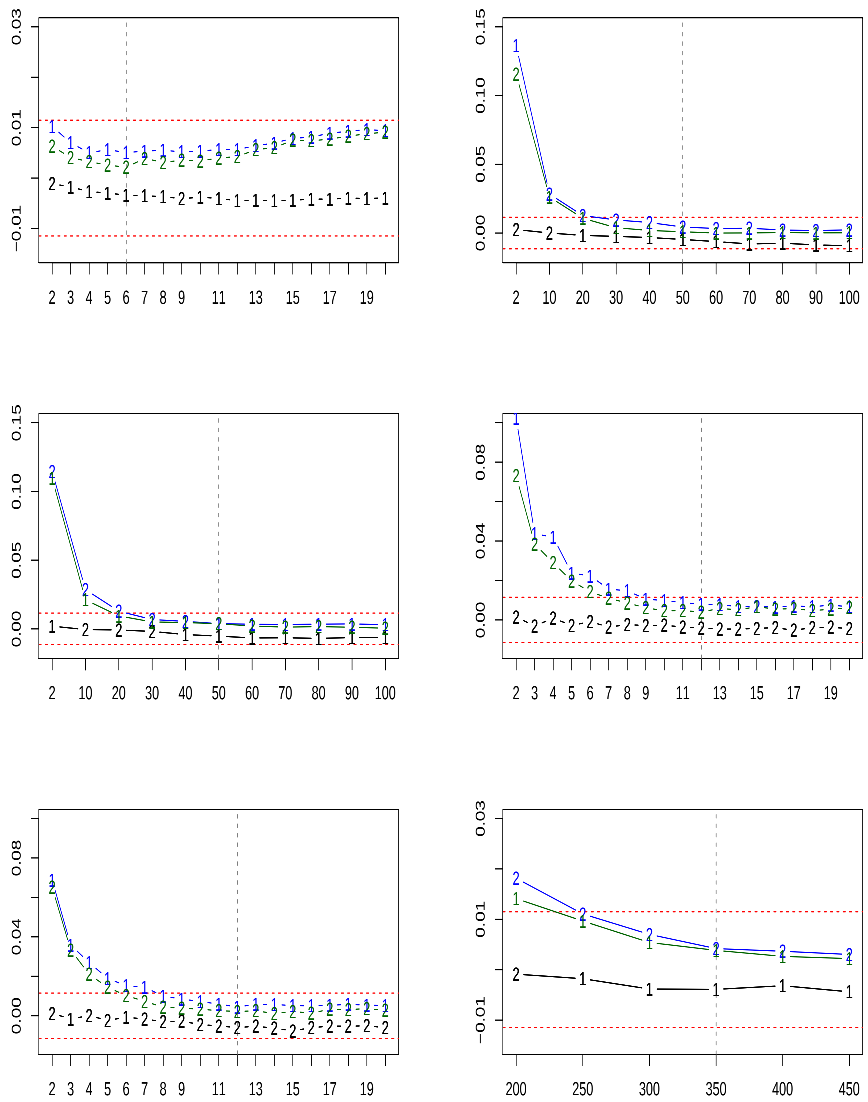


Figure 6: M**,** N**, and** O **statistics.** This figure shows the statistics (see definition (10)) as a function of k, the number of peers used in the construction of the multiple, for the choice of k_{opt} in Figure 1. Each graph reports three curves corresponding (in decreasing order) to M_k^c, N_k^c, and O_k^c, respectively, where c is the method of peer selection in the title of the graph. The curves are functions of k, the size of the set of comparable firms. The number constructing the curves indicate the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics is attained. The coding of the multiples is as follows : (1) P/E , (2) P/B, (3) EV/EBITDA . The dotted lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers.

a fair comparison of the performance of different approaches. In particular, a set of peers composed by a small number of comparable firms (up to 5 for the I+TA and I+ROE/ROIC criteria, or up to 15 in the case of ROE/ROIC and TA) yield valuation that are overall less precise than the optimal ones.

Overall, figures 1 and 6 show that an appropriate choice of the size of the set of comparable companies is essential for a fair comparison of the performance of different approaches.

B.4 Empirical results - the more recent sample

As discussed in Section B.4, Figures 7 and 8 present the results of the analysis of the effect of the number of peers in the construction of the multiple on valuation's accuracy. Figure 7 displays, for every multiple and every peer selection approach, the median absolute error as a function of the size of the set of comparable companies used in valuation. It gives an overall idea about the size of the optimal choice of the cardinal of the set of peers. Figure 8 provides statistical motivation for the optimal choice. It shows the dominance measure of the methods using the optimal size of the peer set over those that make other size choices. The graphs in Figures 7 and 8 show that an appropriate choice of the size of the set of comparable companies is essential for a fair comparison of the performance of different approaches.

Each graph in Figure 7 reports six median absolute error curves corresponding to valuation using the peer selection approach in the graph's title and the various multiples. Note that the shape of the function does not depend on the multiple. As a result the optimal number of comparable firms will depend only on the peer selection approach and not on the multiple used. The vertical line in each graph indicates the optimal number of peers for each method of peer selection as explained in the sequel.

For the choice of optimal number of peers in Figure 7, Figure 8 displays the dependency of the three statistics in definition (10) on k, the number of peers used in the construction of the multiple.

The three curves in each graph correspond in decreasing order to M_k^c, N_k^c, and P_k^c, respectively. The number constructing the curves identifies the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics (see equation (10)) is attained. The coding of the multiples is as follows : (1) P/E , (2) P/B, (3) EV/EBITDA , (4) P/FE1, (5) P/FE2, (6) P/FB, (7) EV/FEBITDA. The dotted

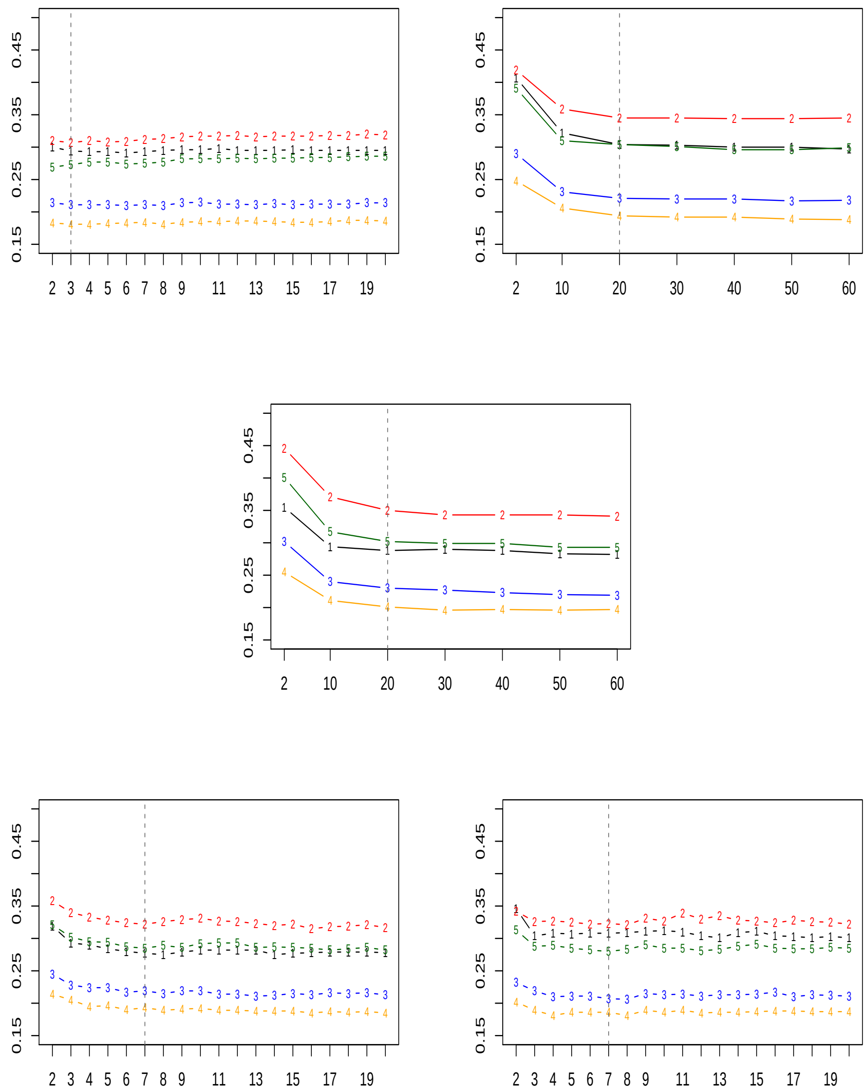


Figure 7: **Median absolute error $|E|$ as a function of the number of peers used in the construction of the multiple.** The graphs correspond to the six peer selection approaches: I and TA (*first row*), ROE/ROIC (*second row*), I+TA and I+ROE/ROIC (*third row*). The curves in each graph display the median absolute error corresponding to multiple valuation as follows : 1 - P/E , 2 - EV/EBITDA , 3 - P/FE1, 4 - P/FE2, 5 - EV/FEBITDA. The graphs show the same patterns for all the multiples (and a given peer selection approach) as well as a consistent ordering of the multiples (according to the median absolute error of the valuation) with enterprise multiples displaying constantly higher median absolute error. The vertical lines indicate the optimal number of peers as defined in (6)

.

lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. Note that $O_k^p \asymp 0$ for all k and all c and hence the values of the statistics M and N only will determine the relationship of optimality of a given k to k_{opt}. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve

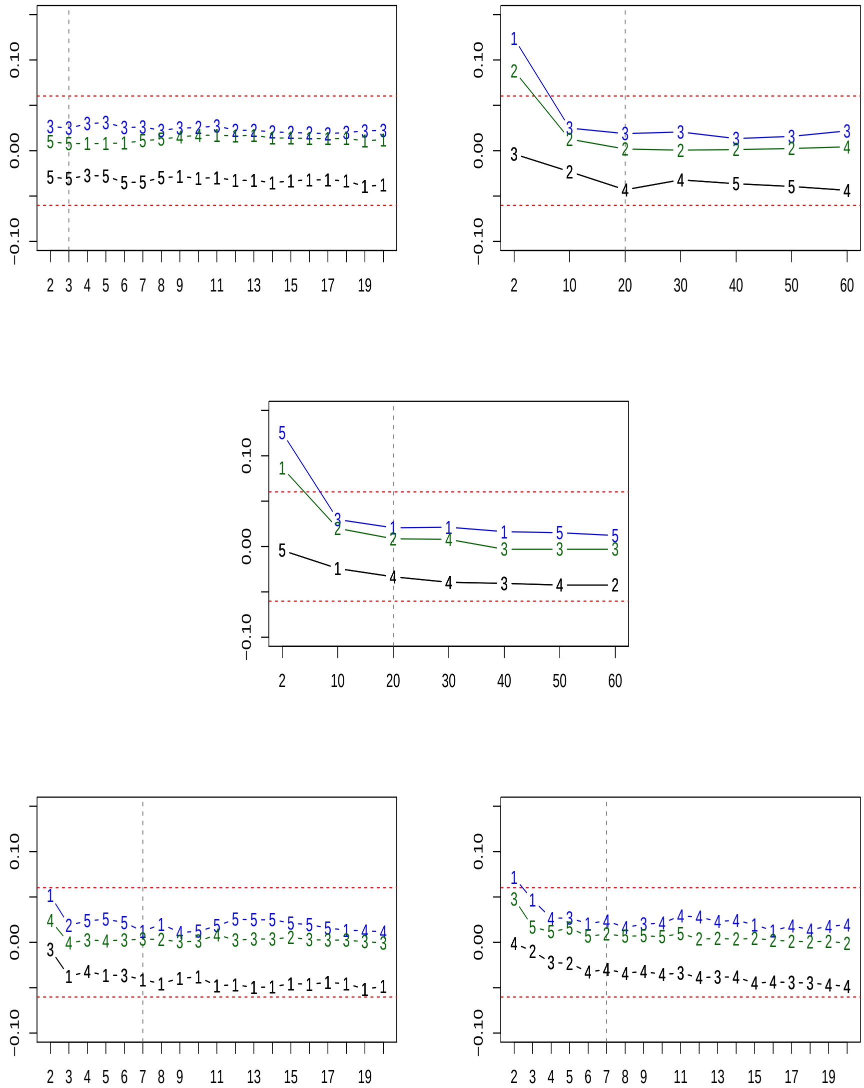


Figure 8: **M, N, and P statistics as a function of k, the number of peers used in the construction of the multiple, for the choice of k_{opt} in Figure 1.** The graphs correspond to the six peer selection approaches: I and TA (*first row*), ROE/ROIC (*second row*), I+TA and I+ROE/ROIC (*third row*). Each graph reports three curves corresponding (in decreasing order) to M_k^c, N_k^c, and P_k^c, respectively, where c is the method of peer selection in the title of the graph. The curves are functions of k, the size of the set of comparable firms. The number constructing the curves indicate the multiple for which $\max_m$ or $\min_m$ in the definition of the three statistics is attained. The coding of the multiples is as follows : (1) - P/E , (2) - EV/EBITDA , (3) - P/FE1, (4) - P/FE2, (5) - EV/FEBITDA. The dotted lines correspond to the 1% confidence band of the Kolmogorov-Smirnov statistic. For a given method of peer selection c, a value of the higher (M) (middle (N)) curve above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers.

above (below) the upper limit of the confidence band in position k indicates that valuations using the method of peer selection c and a set of k comparable firms is strictly less precise than (as precise as) that using k_{opt} peers.

References

Agliardi, E., Agliardi, R., Pinar, M., Stengos, T., & Topaloglouf, N. (2012). A new country risk index for emerging markets: A stochastic dominance approach. *Journal of Empirical Finance*, *19*(5), 741 - 761.

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, *30*(1), 94-108.

Asquith, P., Mikhail, M., & Au, A. (2005). Information content of equity analyst reports. *Journal of Financial Economics*, *75*(2), 245-282.

Baker, M., & Ruback, R. (1999). Estimating industry multiples. *Working paper*, Harvard University.

Bancel, F., & Mittoo, U. R. (2014). The gap between theory and practice of firm valuation: Survey of european valuation experts. *Available at SSRN 2420380*.

Belghitar, Y., Clark, E., & Deshmukh, N. (2014). Does it pay to be ethical? evidence from the ftse4good. *Journal of Banking and Finance*, *47*, 54 - 62.

Bhojraj, S., & Lee, C. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, *40*(2), 407-439.

Block, S. (2010). Methods of valuation: Myths vs. reality. *The Journal of Investing*, *19*(4), 7-14.

Bradshaw, M. T., & Sloan, R. G. (2002). Gaap versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research*, *40*(1), 41–66.

Burgstahler, D., Jiambalvo, J., & Shevlin, T. (2002). Do stock prices fully reflect the implications of special items for future earnings? *Journal of Accounting Research*, *40*(3), 585-612.

Cheng, C., & McNamara, R. (2000). The valuation accuracy of the price-earnings and price-book benchmark valuation methods. *Review of Quantitative Finance and Accounting*, *15*(4), 349-370.

Cready, W. M., Lopez, T. J., & Sisneros, C. A. (2012). Negative special items and future earnings: Expense transfer or real improvements? *The Accounting Review*, *87*(4), 1165-1195.

Damodaran, A. (2006). *Damodaran on valuation: Security analysis for investment and corporate finance* (2nd ed.). John Wiley, Inc.

Dechow, P., & Ge, W. (2006). The persistence of earnings and cash flows and the role of special items: Implications for the accrual anomaly. *Review of Accounting Studies*, *11*(2-3), 253-296.

De Franco, G., Hope, O.-K., & Larocque, S. (2015). Analysts choice of peer companies. *Review of Accounting Studies*, *20*(1), 82-109.

Deloof, M., De Maeseneire, W., & Inghelbrecht, K. (2009). How do investment banks value initial public offerings (IPOs)? *Journal of Business Finance and Accounting*, *36*(1-2), 130-160.

Demirakos, E. G., Strong, N. C., & Walker, M. (2004). What valuation models do analysts use? *Accounting Horizons*, *18*(4), 221-240.

Freeman, R. N., Ohlson, J. A., & Penman, S. H. (1982). Book rate-of-return and prediction of earnings changes: An empirical investigation. *Journal of Accounting Research*, *20*(2), 639-653.

Gilson, S. C., Hotchkiss, E. S., & Ruback, R. S. (2000). Valuation of bankrupt firms. *The Review of Financial Studies*, *13*(1), 43-74.

Johnson, P. M., Lopez, T. J., & Sanchez, J. M. (2011). Special items: A descriptive analysis. *Accounting Horizons*, *25*(3), 511-536.

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The Journal of Finance*, *50*(4), 1059-1093.

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics*, *53*(3), 409 - 437.

Kuosmanen, T. (2004). Efficient diversification according to stochastic dominance criteria. *Management Science*, *50*(10), 1390-1406.

Lie, H., & Lie, E. (2002). Multiples used to estimate corporate value. *Financial Analysts Journal*, *58*(2), 44-54.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal* *of Accounting Research*, *40*(1), 135-172.

Liu, J., Nissim, D., & Thomas, J. (2007). Is cash flow king in valuations? *Financial Analysts Journal*, *63*(2), 56-68.

McKinsey, C., Koller, T., Goedhart, M., & Wessels, D. (2015). *Valuation: Measuring and managing the value of companies* (6th ed.). John Wiley, Inc.

McVay, S. E. (2006). Earnings management using classification shifting: An examination of core earnings and special items. *The Accounting Review*, *81*(3), 501-531.

Paleari, S., Signori, A., & Vismara, S. (2014). How do underwriters select peers when valuing IPOs? *Financial Management*, *43*(4), 731–755.

Riedl, E. J., & Srinivasan, S. (2010). Signaling firm performance through financial statement presentation: An analysis using special items. *Contemporary Accounting Research*, *27*(1), 289–332.

Roosenboom, P. (2012). Valuing and pricing IPOs. *Journal of Banking and Finance*, *36*(6), 1653-1664.

2.3 Paper III: Price-to-earnings valuation - imperfect heuristic? Maybe, but then a precise one!

Price-to-earnings valuation, an imperfect heuristic? May be, but then a precise one!

Jian Kang and Catalin Starica

School of Economics and Business, University of Neuchâtel

Abstract

We show that following the prescription of classical valuation models and allowing for non-linearities in the construction of the multiple greatly improves the precision of the price-to-earnings valuation: pricing errors are within 15 percent of stock prices for about 60 percent of our sample. Although regarded sometimes as an imperfect heuristic, the P/E valuation, enhanced by the non-linear feature, seems to be more precise than more comprehensive valuation methodologies based on detailed projections of future as reported in the accounting literature.

Keywords: Valuation, price-to-earnings multiple, accuracy, stochastic dominance.

1 Introduction

Multiples are largely used by practitioners (Demirakos, Strong, and Walker (2004); Asquith, Mikhail, and Au (2005); Damodaran (2006)). Their main appeal is simplicity as they bypass explicit projections and present value calculations required by more comprehensive approaches. Does this simplicity come at the cost of a lesser precision? The pricing error analysis we conduct in this study shows that it is not always the case provided that the prescriptions of the accounting valuation models are heeded.

The valuation approach under investigation pairs the forward earnings-to-price ratio with industry peer selection criterion. Our choice is motivated by the formal interpretability of this multiple as well as by the documented accuracy of the pair among a large class of multiples and peer selection methods (Liu, Nissim, and Thomas (2002), Kang and Starica (2014)). We find that allowing for non-linearities in the prediction of the multiple, as prescribed by accounting valuation models, greatly improves valuation precision. Although not evaluated on the same sample, the accuracy of the enhanced price-to-earnings multiple pricing compares favorable with that of more comprehensive valuation methods as reported in the literature.

The pricing error analysis we conduct assumes market efficiency which implies, in particular, that the best value estimate is the one closest to investors' beliefs about intrinsic value[1]. In this setup, the measurable goal of a valuation approach is a distribution of the pricing error as tight around zero as possible.

The gains in valuation precision reported in this study come from better predicting the multiple. The key observation is that, for multiples ratios of price to value driver, the relative pricing error, i.e. predicted price minus actual price scaled by actual price, is equal to the relative prediction error for the multiple, i.e. predicted multiple less actual multiple scaled by actual multiple[2]. Hence, for multiples ratios of price to value driver, better pricing comes only through improved prediction of

[1]Another setting of interest assumes that efficient market hypothesis might not hold in the short run but that market price ultimately reverts to fundamentals. The best value estimate in this case comes from the method that comes closest to the value to which price reverts. This setup is the subject of parallel research to be reported elsewhere.

[2]This equivalence does not hold for valuations where the multiples is constructed as the ratio of Enterprise Value to value driver.

the multiple.

The focus on better predicting the multiple is not new in the accounting literature. In a paper closely related to our study, Bhojraj and Lee (2002) use linear regressions and firm profitability, risk and growth characteristics to predict future price-to-book and enterprise value-to-sales ratios. In a first step, they use the residual income valuation model to derive the fundamental determinants of the two common valuation multiples. Forecasts for valuation multiples are then constructed through linear regressions of observed valuation multiples on the fundamental determinants (although the model specifies a non-linear relation of these determinants to the multiples). The authors report sharp improvements in performance over the traditional technique: the median absolute relative pricing errors decreases from 0.55 to 0.35 for enterprise-value-to-sales and from 0.38 to 0.29 for price-to-book multiple, respectively.

In a discussion of the mentioned paper, Sloan (2002) raises a number of issues, the most important being: the use of a linear approach for modeling relations that are clearly non-linear, the role and use of previous year value of the multiple in forecasting the current value, and the heuristic nature of the multiple approach. He concludes by directly challenging the use of multiple valuation: instead of relying on 'imperfect heuristics' like valuation multiples, the practitioners should use 'more rigorous valuation methodologies'.

Our study builds on this criticism and brings a number of contributions to the valuation literature. First, it better puts into light the scope for non-linear prediction techniques in pricing. In the traditional multiple valuation approach, the price-to-earning ratio of the target company is commonly predicted to be equal to the median price-to-earning ratio of peers (in our case the industry multiple median). In Section 3 we argue that the relationship between the quantity to be predicted, the firm's price-to-earnings multiple, and the predictor, the industry median price-to-earning ratio, is non-linear. The empirical analysis in Section 5 confirms the model-based argument. It also shows that the departure from linearity affects mostly the industries with small/large median price-to-earnings.

Second, it demonstrates the practical impact of non-linear prediction techniques.

Allowing for non-linearities in predicting target firm's price-to-earnings, as prescribe by valuation accounting models, greatly enhances the pricing precision as shown in Section 6.1: the median absolute pricing error decreases from 0.21 to 0.12. We find (Section 6.2) that the gains in pricing precision benefit mostly the firms with small/large price-to-earnings, the firms in industries with a small price-to-earnings and the firms with high forecasted growth. We also find (Section 6.4) that direct non-linear prediction of prices is more precise than plain-vanilla multiple valuation. If non-linearities are taken into account, prices can be more accurately set without passing through the intermediate step of multiple prediction.

Third, we shed light on the role and use of multiple's previous year value in forecasting the current value. It was noted that "if practitioners are simply interested in generating the best possible forecast of a company's current or future valuation multiple, then the company's past valuation multiple would be the most useful forecasting variable" Sloan (2002). We show in Section 6.3 that previous year price-to-earnings ratios, while an useful forecasting variable on its own, cannot further improve the accuracy of predicting the multiple once the non-linearity is taken into account. Adding previous year price-to-earnings ratios to the industry median multiple in the non-linear prediction does not significantly decrease the relative pricing error.

Our findings reposition the multiple pricing with respect to the more comprehensive valuation approaches. On a sample of more than 22,000 firm-years, our non-linear enhancement of the price-to-earnings multiple method shows a median absolute pricing error of 10.6% (standard deviation of 16.5%). To our best knowledge, the most precise comprehensive valuation in the literature up to day is reported in Courteau, Kao, and Richardson (2001). It is based on the discounted cash flow model and employs *Value Line*-forecasted prices in the terminal value expressions. It documents a median absolute pricing error of 13.7% (standard deviation of 19.7%) on a sample ten times smaller than ours. It hence seems that, while price-to-earnings valuation might be referred to as an 'imperfect heuristic', when strengthened by the understanding of accounting valuation models, it becomes a precise pricing method.

To summarize, our study takes up the "broader question of why we care so much about the valuation multiples in the first place" posed in Sloan (2002). The answer we propose is simple: because they deliver (when properly enhanced by the insight of classical accounting models).

2 Prior research comparing pricing accuracy of multiples and comprehensive accounting models

To put into perspective the precision gains outlined by our study, this section reviews the literature that reports on the accuracy of the two classes of pricing: multiple valuation and model-based pricing. We start with the academic studies that report on the accuracy of the multiple valuation method.

Using price-to-earnings multiple to value 4698 firm-years, Alford (1992) reports an average median absolute error over the three years in the sample of 23.9% for peer selection based on industry and return on equity. On a larger sample of 30,310 firm-years, Cheng and McNamara (2000) record an average mean absolute error over the twenty years in the sample of 26.4% for the P/E valuation and of 24.1% for the P/B valuation. In both cases the peers were chosen on the intersection of industry and return on equity criteria.

Liu et al. (2002) compared the performance of a large number of multiples relative to current stock price for a sample of 26,613 observations from 1982 to 1999. They found that the forward earnings-to-price ratio performs best with a interquartile of the relative error of 0.30 (they do not report statistics of the absolute error). Bhojraj and Lee (2002) used warranted multiples to predict future price-to-book and price-to-sales ratios for a target firm given its value driver characteristics. They show sharp improvements in performance over the traditional techniques of matching firms on size and/or industry. The best performance they report is that of the price-to-book ratio with a median absolute error of 0.29. On a sample of 8,621 covering only the 1998 fiscal year, Lie and Lie (2002) report a median absolute error between 22.4 % and 34.4 % using ten different multiples for U.S. equity data.

Schreiner and Spremann (2007) investigated twenty seven multiples on a sample of 592 European firms for the period 1996 to 2005. They found that, for the historical multiples, the median absolute error ranged between 44% and 25.4% while for the forward multiples the range was from 43.8% to 21.5%. The best performance was that of the forward price-to-earnings multiple based on the two-year ahead analyst prediction of earnings.

In a direct comparison between model-based intrinsic valuation and multiple pricing, Courteau, Kao, O'Keefe, and Richardson (2003) test the hypothesis that errors from a valuation that converts *Value Line* predictions of relevant accounting variables and prices into an estimate of the intrinsic value of the firm are lower than those of three industry-multiplier approaches. One of the multiples used is a version of the forward earnings yield multiple discussed in our study. They predict lower errors for the direct valuation based on the belief that no two firms are alike. Given heterogeneity across firms within an industry, greater valuation accuracy can be expected when the value is computed by drawing on the analyst's detailed knowledge about the target firm (i.e., the direct method), as opposed to relying on inferences made from information about comparable firms within the same industry (i.e., industry-multiplier approaches). On a sample of quarterly data of 43,204 quarter-firms from 1990-2000 they report a median absolute valuation error of 16.8% for the direct valuation compared to 27.8%, 26.1% and 19.9%, respectively, for the three multiples. The forward price-to-earnings multiple performance (19.9%) is comparable to the value reported on our sample (21.8%).

Kang and Starica (2014) systematically investigated the impact of peer selection criteria, of the number of peers to be considered, and of the method of predicting the multiple for thirty-six pairs of multiple and peers selection criterion. On a sample of 96,598 firm-years covering the period from 1968 to 2010, they documented the superiority of the price-to-earnings and price-to-book multiples coupled to peer selection based on return on equity and total assets or return on equity and industry criteria. The median absolute pricing error of multiples based on historical data ranged between 30% and 64%. On a smaller, more recent sample of 3,672 firm-years between 2000 and 2010, the forward price-to-earnings was the best performing multiple with a median absolute pricing error of 21%.

To summarize, multiple valuation precision depends strongly on the multiple as well as on the peer selection approach: the median absolute error ranges between 20% and more than 60%. Most precise multiples with a mean absolute error around 20% are versions of the forward price-to-earnings ratio.

More comprehensive valuation approaches are based on accounting models, such as discounted dividend model (DDM), discounted cash flow (DCF) model, residual income model (RIM) or discounted abnormal earnings (AE) model. Unlike the multiple method, these approaches do not make us of the market value information. Instead they convert analysts' predictions of relevant accounting variables or prices into an estimate of the intrinsic value of the firm. Several recent studies have investigated the ability of one or more of these valuation methods to generate reasonable estimates of market values.

On a sample of 2,907 firm-year of publicly traded firms followed by *Value Line* during the period 1989-93, Francis, Olsson, and Oswald (2000) found that AE value estimates perform significantly better than DIV or DCF value estimates. For the most precise valuation in each class, the median absolute prediction error for the AE model was of 30%, that of the DCF approach of 41%. The DIV model yielded the less precise valuations with a median absolute error of 69%. On a sample of 36,532 firm-year observations over the period 1981-1998, Sougiannis and Yaekura (2001) document an absolute pricing error of the RIM of 47%.

Heinrichs, Hess, Homburg, Lorenz, and Sievers (2011) examined the performance DDM, DCF and RIM models for a sample of 15,658 firm-year observations of firms covered by Value Line from 1986 to 2006. The best performing DDM had an median absolute error of 81%, for the most precise DCF model the absolute pricing error was of 50% while the most accurate approach, the RIM model with growth of 2%, had a median absolute error of 32%. Courteau, Kao, and Tian (2015) also document the superiority of the RIM model with extended assumption on accrual over the DCF model on a smaller sample of 5,123 firm-year observations between 1990 to 2000. The reported mean absolute median valuation error lied between 32% and 44%.

Recent literature argues that, within each class of valuation models, the model that employs forecasted price in the terminal value expression generates the lowest prediction errors, compared with models that employ nonprice-based terminal value. Courteau et al. (2001) discussed RIM and DCF valuation and showed that using *Value Line* forecasted target price in the terminal value expression significantly improves pricing accuracy. On a sample of 2110 firm-years from 1992 to 1996, they document, for the non-price-based models, errors of 38% for the RIM and of 40% to 44% for the DCF approach. In contrast to that, the price-based approaches display significant accuracy improvements: absolute valuation error of both models is reduced to 14%. Although the increase in accuracy is sharp, the sample is restricted to 422 large firms due to the wealth of information needed. The price-based terminal value approach, as implemented in the paper, can be used only for firms covered by *Value Line*.

As a conclusion, the valuation literature seems to agree that the precision of non-price-based terminal value model-based valuation is low: the median absolute pricing error is larger than 30%. The accuracy can be sharply improved by the use of proxies of the theoretically ideal terminal value expressions in each model. The most precise model-based valuations in the literature (which is also the most precise approaches we are aware of) report an median absolute valuation error of 16.8% for a version of the AE model (Courteau et al. (2003)) and of 14% for the RIM and DCF models with price based terminal value (Courteau et al. (2001)). This precision comes at a cost: market's expected stock price at the horizon and the premium of that price over book value for a particular accounting system, i.e. the theoretically ideal terminal value expressions in each model, need to be inferred. Surrogates of these expressions are available for far less firms than covered by the usual analysts' earnings forecasts.

3 A non-linear relation

The following considerations give some formal arguments for the non-linear nature of the relationship between the forward multiple and the median industry earnings yield. They also provide a formal common background for the set of pricing approaches discussed in the empirical sections. Basic accounting valuation models suggest that past earnings-to-price ratios as well as current industry median earnings yield can be viewed as approximations of the unknown cost of equity capital (COEC) of the firm to be priced. They also postulate that the link between the forward price-to-earnings multiple and the rate of return on equity is non-linear. To make these models operational, simplifying assumptions are needed. The pricing approaches under investigation are obtained by plugging approximations of the unknown COEC of the firm to be priced in different operational versions of valuation models.

3.1 Approximation of cost of equity capital by earnings yield

In the sequel, we view the forward earnigs-to-price ratio as an estimate of the cost of equity capital for the firm. As Easton (2004) points out, use of the earnings yield to represent the COEC is based on the assumption that a single year of earnings is representative of the future stream of earnings. This interpretation is motivated by recent accounting research (Easton and Monahan (2005), (2010), Larocque and Lyle (2014)) which shows that the simple earnings-to-price ratio outperforms most of the commonly employed implied cost of capital measures in the literature. For the sake of the argument, we note that the interpretation we adopt is supported by the Residual Earnings valuation model (REM):

$$V_0^E = CSE_0 + \sum_{t=1}^{\infty} \frac{NI_t - r_E \times CSE_{t-1}}{(1+r_E)^t} \tag{1}$$

which, under the simplifying assumption of zero growth of the residual earnings after year 1, yields:

$$V_0^E = CSE_0 + (NI_t - r_E \times CSE_{t-1}) \times \sum_{t=1}^{\infty} \frac{1}{(1+r_E)^t} = CSE_0 + \frac{NI_1 - r_E \times CSE_0}{r_E} = \frac{NI_1}{r_E}. \tag{2}$$

Supposing that the current price reflects the intrinsic value of the firm, this implies that the forward looking ratio NI_1/P_0 equals firm's expected rate of return on equity (inverse engineering):

$$\frac{NI_1}{P_0} \approx r_E. \tag{3}$$

A similar interpretation holds for the ratio NI_2/P_0. With no growth after the k-th year ($k > 1$) and assuming the clean surplus relation holds:

$$CSE_1 = CSE_0 + NI_1 - D_1,$$

where D_1 denote the dividends, the REM model (1) simplifies to:

$$V_0^E = \sum_{t=1}^{k-1} \frac{D_t}{(1+r_E)^t} + \frac{NI_k}{r_E(1+r_E)^{k-1}} = \sum_{t=1}^{k-1} \frac{PR_t \times NI_t}{(1+r_E)^t} + \frac{NI_k}{r_E(1+r_E)^{k-1}}, \tag{4}$$

where PR stands for the pay-out ratio, i.e. the proportion of the net income payed as dividends. When $k = 2$, the model reads:

$$V_0^E = \frac{D_1}{1+r_E} + \frac{NI_2}{r_E(1+r_E)} = \frac{PR_1 \times NI_1}{1+r_E} + \frac{NI_2}{r_E(1+r_E)}. \tag{5}$$

Two observations help simplify the expression in (5) for most of firms. First, the first term in the sum $PR_1 \times NI_1/(1 + r_E)$ is dwarfed by the second term $NI_2/r_E(1+r_E)$. Canceling the $(1+r_E)$ factor, two consecutive earnings, normally close in size, are multiplied first, by PR_1, a value typically between 0 and 0.5, and second, by $1/r_E$, a value typically between 5 and 20. Second, $r_E(1 + r_E) \approx r_E$ for the range of values common to most firms ($r_E \in [0.05, 0.25]$). With this in mind, if current price correctly reflects the value of the firm, reverse engineering yields:

$$\frac{NI_2}{P_0} \approx r_E. \tag{6}$$

The discussion above motivates two operational approximations for the unknown COEC of the firm i to be priced base on past data. These are the past lags of the forward ratios:

$$r_{E,i} \approx \text{lagged}\left(\frac{NI_{1,i}}{P_{0,i}}\right) \quad \text{and} \quad r_{E,i} \approx \text{lagged}\left(\frac{NI_{2,i}}{P_{0,i}}\right). \tag{7}$$

Before we discuss another approximation of the COEC that is based on contemporaneous information, let us mention an operational form of the REM model in equation (4) that will be the basis of direct prediction of price discussed in Section 6.4. Assuming constant earnings' growth rate g_i for the periods 3 to k and constant pay-out PR_i, the value of the firm i can be expressed as:

$$\begin{aligned} V_{0,i}^{E} &= \frac{PR_i \times NI_{1,i}}{1+r_{E,i}} + \frac{PR_i \times NI_{2,i}}{(1+r_{E,i})^2} \times \left[\sum_{t=0}^{k-3} \left(\frac{1+g_i}{1+r_{E,i}} \right)^t + \frac{1}{r_{E,i}} \times \left(\frac{1+g_i}{1+r_{E,i}} \right)^{k-3} \right] \\ &= \tilde{f}\left(r_{E,i},\ NI_{1,i},\ NI_{2,i},\ g_i,\ PR_i\right), \end{aligned} \tag{8}$$

with $\tilde{f}$ a non-linear function.

3.2 Approximation of the COEC by industry's median earnings yield

The link between the forward ratio and the COEC in equations (3) and (6) can be used alternatively to construct an approximation of the unknown cost of equity capital of the firm i (to be priced) starting from the observed ratios of the other firms in the same industry. The precision of the approximation is based on the assumption that all firms in the industry share a common operating cost of capital.

More concretely, by the weighted average cost of capital formula (WACC), the expected return on equity of firms j in the same industry, $r_{E,j}$, can be decomposed as:

$$r_{E,j} = r_F + \frac{V_j^D}{V_j^E}(r_F - r_{D,j}), \tag{9}$$

i.e. in a common[3] part, represented by the operating cost of capital, r_F, and an idiosyncratic part represented by the product between the market leverage of the firm and the required return spread.

This decomposition and the approximations in equations (3) and (6) suggest that the forward earnings yield ratio of the firms in the same industry may be used to predict the unknown cost of equity capital of the firm i to be priced. In particular, applying the median in equation (9) yields an useful representation of

[3] We assume that r_F is common to all firms in the same industry.

the industry median earnings:

$$\operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}\left(\frac{NI_{\cdot,j}}{P_{0,j}}\right) \approx \operatorname*{median}_{\substack{j\in SIC\\ j\neq i}} r_{E,j} = r_F + \operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}\left(\frac{V_j^D}{V_j^E}\times(r_F - r_{D,j})\right).$$

This expression shows that the industry median earnings is a quantity closely related to

$$r_{E,i} = r_F + \frac{V_i^D}{V_i^E}(r_F - r_{D,i}),$$

the unknown COEC of the firm to be priced. It contains the common part, represented by the operating cost of capital, and it approximates the idiosyncratic component by the median idiosyncratic contribution of the firms in the industry.

This discussion motivates approximating, in the sequel, the unknown COEC of the firm to be priced $r_{E,i}$ by the industry median earnings yield:

$$r_{E,i} \approx \operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}\left(\frac{NI_{1,j}}{P_{0,j}}\right) \quad \text{and} \quad r_{E,i} \approx \operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}\left(\frac{NI_{2,j}}{P_{0,j}}\right). \tag{10}$$

3.3 Pricing based on different representations of the V/NI ratio

In this section we discuss the implications that different simplifying assumptions in the REM model have on the valuation methodology. The operational assumptions are those needed to match the information available in the sample of the empirical analysis in Section 6: predictions by the analysts of one- and two-year ahead earnings and of the future short-term growth rate in earnings. The set of valuation approaches under discussion is obtained by plugging the two approximations of the COEC of the firm to be priced in equations (7) and (10) in the operational versions of the accounting model presented by the expressions (2) and (4).

The simplest version of the REM model represented in equation (2) postulates the value-to-earnings ratio of the firm i to priced as

$$\frac{V_{0,i}^E}{NI_{1,i}} = \frac{1}{r_{E,i}}.$$

As a consequence, the intrinsic values for the firm i is given by:

$$V_{0,i}^E = NI_{1,i} \times \frac{1}{r_{E,i}} \simeq \begin{cases} NI_{1,i} \times \text{lagged}\left(\frac{NI_{\cdot,i}}{P_{0,i}}\right), \\ NI_{1,i} \times \underset{\substack{j \in SIC \\ j \neq i}}{\text{median}}\left(\frac{P_j}{NI_{\cdot,j}}\right), \end{cases} \tag{11}$$

depending on the approximation used for the unknown COEC of the firm to be priced. The last expression corresponds to the plain-vanilla multiple valuation which is our benchmark while the first two correspond to the scenario where the current multiple is approximated simply by its most recent past value. The precision of these approaches to valuation is empirically investigated in Section 6.3.

The more general form of the value model in equation (4) yields

$$\frac{V_{0,i}^E}{NI_{1,i}} = \sum_{t=1}^{k-1} \frac{PR_t}{(1+r_{E,i})^t} \times \frac{NI_{t,i}}{NI_{1,i}} + \frac{1}{r_{E,i}(1+r_{E,i})^{k-1}} \times \frac{NI_{k,i}}{NI_{1,i}}.$$

Assuming constant earnings' growth rate g for the periods 3 to k and constant pay-out, the value-to-earnings ratio of the firm i can be expressed as:

$$\begin{aligned} \frac{V_{0,i}^E}{NI_{1,i}} &= \frac{PR_i}{1+r_{E,i}} + \frac{PR_i}{(1+r_{E,i})^2} \times \frac{NI_{2,i}}{NI_{1,i}} \times \left[\sum_{t=0}^{k-3}\left(\frac{1+g}{1+r_{E,i}}\right)^t + \frac{1}{r_{E,i}} \times \left(\frac{1+g}{1+r_{E,i}}\right)^{k-3}\right] \\ &= \tilde{f}\left(r_{E,i},\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right), \end{aligned} \tag{12}$$

with $\tilde{f}$ a non-linear function. Similar expressions hold for the ratio $V_{0,i}^E/NI_{2,i}$.

This equation shows the non-linear[4] dependency of the value-to-earnings ratio on the expected rate of return on equity. It indicates that the value-to-earnings ratio can be approximated by a non-linear function of four variables, giving hints about other variables (besides the cost of equity capital) to be used in the prediction of the multiple.

Equation (12) yields the following expression for the intrinsic value of the firm i:

$$V_{0,i}^E \simeq NI_{\cdot,i} \times \tilde{f}\left(r_{E,i},\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right). \tag{13}$$

[4]It goes without saying that giving up on the simplifying assumptions that yield equation (5) will only accentuate the non-linear nature of the dependency of the forward multiple on the expected rate of return.

We conclude with two important remarks concerning the pricing approach based on this representation. The first one relates the new approach to the plain-vanilla valuation. To render operational the relationship in equation (13) we replace $r_{E,i}$, the unknown COEC of the firm to be priced, by one the approximations in equations (7) and (10) and we estimate the non-linear function $\tilde{f}$. When approximating $r_{E,i}$ by the industry median earnings yield as shown in equation (10), the expression of the value-to-earnings ratio in equation (12) becomes:

$$\begin{aligned}\frac{V_{0,i}^{E}}{NI_{\cdot,i}} &\simeq \tilde{f}\left(\underset{\substack{j\in SIC\\ j\neq i}}{\text{median}}\left(\frac{NI_{\cdot,j}}{P_{0,j}}\right),\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right)\\ &= f\left(\underset{\substack{j\in SIC\\ j\neq i}}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right),\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right)\end{aligned} \tag{14}$$

for some non-linear function f since

$$\underset{j}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right) = 1/\underset{j}{\text{median}}(NI_{\cdot,j}/P_{0,j}).$$

This shows that pricing based on the price-to-earnings representation in equation (12) generalizes the plain-vanilla multiple approach in equation (11): while classical multiple pricing constructs the multiple of the firm to be valued as the industry median price-to-earnings, in the approach proposed in this study the multiple is constructed as a non-linear function of the industry median (and of other variables).

It is worth noting that, if the unknown COEC of the firm to be priced is approximated by the industry mean earnings yield, i.e.

$$r_{E,i} \approx \underset{\substack{j\in SIC\\ j\neq i}}{\text{mean}}\left(NI_{1,j}/P_{0,j}\right) \quad \text{or} \quad r_{E,i} \approx \underset{\substack{j\in SIC\\ j\neq i}}{\text{mean}}\left(NI_{2,j}/P_{0,j}\right),$$

the expression of the value-to-earnings ratio in equation (12) yields:

$$\begin{aligned}\frac{V_{0,i}^{E}}{NI_{\cdot,i}} &\simeq \tilde{f}\left(\underset{\substack{j\in SIC\\ j\neq i}}{\text{mean}}\left(NI_{\cdot,j}/P_{0,j}\right),\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right)\\ &= f\left(\underset{\substack{j\in SIC\\ j\neq i}}{\text{harmonic mean}}\left(P_{0,j}/NI_{\cdot,j}\right),\ \frac{NI_{2,i}}{NI_{1,i}},\ g_i,\ PR_i\right)\end{aligned} \tag{15}$$

for some non-linear function f since

$$\operatorname*{mean}_j(P_{0,j}/NI_{\cdot,j}) = 1/\operatorname*{harmonic}_j \operatorname{mean}(NI_{\cdot,j}/P_{0,j}).$$

Hence, pricing based on the price-to-earnings representation in equation (15) generalizes the implementation of the plain-vanilla multiple valuation, common in the valuation literature, that replaces the multiple of the firm to be priced by the industry harmonic mean price-to-earnings: in the approach advocated by this study the multiple is constructed as non-linear function of the harmonic mean (and of other variables).

Second remark concerns the variables to be used in the prediction of the multiple when pricing based on the value representation in equation (13). The expression in the mentioned equation suggests the possibility of simultaneous use of different approximations for the unknown COEC. The details of the implementation are presented in the next section.

As a conclusion, this section presented formal considerations that motivate our approach and position it as a natural generalization of the multiple valuation. The plain-vanilla multiple value is obtained starting with the simplest formal expression of the value-to-earnings ratio and approximating the unknown COEC of the firm to be priced by the industry median earnings-to-price ratio. The non-linear approach to pricing proposed in this study plugs the same approximation of firm's unknown COEC in a more general representation of the value-to-earnings ratio as a non-linear function of four variables. As we will see in the sequel, using multiple approximations of the unknown COEC of the firm to be priced in the non-linear expression of the ratio enhances the precision of valuation.

4 Empirical design

To predicts a firm's stock price the multiple method multiplies a value driver (such as earnings) by the corresponding estimated multiple. The multiple is obtained from the ratio of stock price to the value driver for a group of comparable firms (or peers) most commonly chosen from the same industry.

4.1 Prediction of the multiple

To price the firm i in the year t, the multiple valuation approach first needs to predict the firm's multiple $(P/NI)_{t,i}$ for the year t. The predicted multiple $\widehat{(P/NI)}_{t,i,C}$ is a function of the multiples of a set C of comparable firms (peers). In a second step, the predicted price $\widehat{P}_{t,i,C}$ is obtained by multiplying the predicted multiple and the value of future earnings $NI_{t+1,i}$:

$$\widehat{P}_{t,i,C} = NI_{t+1,i} \times \widehat{\left(\frac{P}{NI}\right)}_{t,i,C}. \tag{16}$$

4.1.1 Plain-vanilla case

In plain-vanilla multiple valuation, the predicted multiple is commonly constructed as the median of the multiples of the peer companies and the comparable firms are defined based on various industry classifications yielding:

$$\widehat{\left(\frac{P}{NI}\right)}_{t,i} := \underset{\substack{j \in SIC \\ j \neq i}}{\text{median}} \left\{ \frac{P_{t,j}}{NI_{t+1,j}} \right\}. \tag{17}$$

The plain-vanilla multiple valuation based on the prediction in equation (17) is the first benchmark for our comparisons. For the firms for which previous year prices are available more benchmarks are considered. They include the valuation described by equation (11) based on the approximation of the multiple by its most recent past value and the trivial valuation that values the firm by its most recent past price value. These benchmarks are useful for gauging the real contribution of the plain-vanilla valuation with respect to trivial approaches based on past information.

4.1.2 Non-linear approach

As discussed in the previous section, we argue that the relationship between the multiple of a firm $P_{t,i}/NI_{t+1,i}$, its cost of equity capital, the ratio of future earnings and the earnings' growth rate is not linear and hence, the predicted multiple of the firm to be priced should be constructed as a non-linear function of the predicting variables. In particular, the median industry multiple or the lagged multiple of the firm, respectively should contribute non-linearly to the construction of the multiple.

To evaluate the importance of the different variables and CEOC approximations, we compare the pricing precision of valuations where the multiple is predicted as a non-linear function of the following subsets of the variables put forth in equation (12):

$$\widehat{\left(\frac{P}{NI}\right)}_{t,i} := \begin{cases} f(g), \\ f(NI_{t+2,i}/NI_{t+1,i}), \\ f(g, NI_{t+2,i}/NI_{t+1,i}), \\ f(\text{lagged}\,(NI_{\cdot,i}/P_{0,i})), \\ f(\operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}(P_{t,j}/NI_{t+\cdot,j})), \\ f(\operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}(P_{t,j}/NI_{t+\cdot,j})\,, \text{lagged}\,(NI_{\cdot,i}/P_{0,i})), \\ f(\operatorname*{median}_{\substack{j\in SIC1\\ j\neq i}}(P_{t,j}/NI_{t+\cdot,j})\,, \operatorname*{median}_{\substack{j\in SIC2\\ j\neq i}}(P_{t,j}/NI_{t+\cdot,j})\,, \operatorname*{median}_{\substack{j\in SIC3\\ j\neq i}}(P_{t,j}/NI_{t+\cdot,j})). \end{cases} \tag{18}$$

We removed the pay-out ratio from the predictor set for empirical considerations. The number of firms for which all four predictors in equation (12) are available is small relative to that for which the first three exist. Considering a set of four predictors would have drastically reduced the sample size[5].

The empirical analysis (the results of which are presented in Section 6.1.1) shows that the variables, NI_2/NI_1 and g do significantly improve the accuracy of the pricing (with respect to the plain-vanilla multiple) when used together in the non-linear framework. However, once an industry median multiple is used in forecasting, the pricing precision is not significantly increased by inclusion of these two variable among the predictors. Hence, the operational predicting equation simplifies to:

$$\widehat{\left(\frac{P}{NI}\right)}_{i,t} := f\left(\operatorname*{median}_{\substack{j\in SIC\\ j\neq i}}\left\{\frac{P_{t,j}}{NI_{t+1,j}}\right\}\right).$$

The non-linear function f is to be estimated. To address the issue of the choice of the industry depth we enlarge the predictor set to include different approximations

[5]On the smaller sample for which all predictors are available, the pay-out ratio does not seem to improve pricing precision when added to the industry median.

of the unknown COEC of the firm to be priced. These are the earnings yield medians calculated using peers with the same one, two and three-digit SIC codes. Hence the most parsimonious predictor specification (as we will see in the sequel) is given by:

$$\widehat{\left(\frac{P}{NI}\right)}_{i,t} := f\left(\underset{\substack{j\in SIC1\\ j\neq i}}{\text{median}}\left\{\frac{P_{t,j}}{NI_{t+1,j}}\right\},\ \underset{\substack{j\in SIC2\\ j\neq i}}{\text{median}}\left\{\frac{P_{t,j}}{NI_{t+1,j}}\right\},\ \underset{\substack{j\in SIC3\\ j\neq i}}{\text{median}}\left\{\frac{P_{t,j}}{NI_{t+1,j}}\right\}\right), \tag{19}$$

whith f a trivariate, non-linear function to be estimated.

4.2 Multiples' Accuracy

In this section we discuss the quantities to evaluate and the measures by which we asses the pricing precision of different valuation methods. The setup of our study that assumes market efficiency and, in particular, that the best value estimate is the one closest to investors' beliefs about intrinsic value entail the quantity to evaluate when comparing different valuation methods: the relative pricing error. It also implies that the measurable goal of a valuation approach is a distribution of the pricing error as tight around zero as possible or, alternatively, a distribution of the absolute pricing error with the lowest central measure and the less spread around it.

While current accounting literature emphasizes the use of specific statistics of relative and absolute pricing errors, we argue that a notion close to that of stochastic dominance gives a more complete picture in the precision comparisons. Moreover, the proposed measure encompasses all the criteria currently used in the literature and allows for precise statistical testing.

4.2.1 Relative pricing errors

We define the *relative (percentage) error*[6] of the multiple of the firm i in year t:

$$\%E_{i,t} := \frac{\widehat{P}_{t,i} - P_{t,i}}{P_{t,i}}, \tag{20}$$

where $P_{t,i}$ is the actual stock price for the target firm i in year t, and $\widehat{P}_{i,t}$ is the predicted stock price based on the set of comparable firms. For most of the paper[7], the benchmark stock price for the target firm i in year t, $P_{t,i}$ used to evaluate the precision of multiple based predictions, is the stock price of the firm i three months after the end of the financial year t.

The methodology proposed in the sequel focuses on improving the precision of predicting the price-to-earnings multiple. We note that the multiple relative error defined by:

$$\left[\widehat{\left(\frac{P_0}{NI_1}\right)}_{t,i} - \frac{P_{t,i}}{NI_{t+1,i}} \right] / \frac{P_{t,i}}{NI_{t+1,i}}, \tag{21}$$

and the price relative error defined in (20) are equal: the error (20) is obtained from the error in equation (21) by multiplying both the numerator and the denominator by $NI_{t+1,i}$, the future earnings of the target company. As a consequence, better multiple prediction translates in more accurate price prediction. More precisely, the precision of both multiple prediction and multiple based stock price prediction is described by one relative error.

4.2.2 An overall accuracy comparison measure

The academic literature on multiple valuation analyzes the accuracy of pricing by looking at its pricing relative errors $\%E$ as defined in (20) or at their absolute values $|\%E|$. The use of absolute prediction errors gives equal weight on negative and positive errors. In general, some specific statistics of one of these distributions are calculated and then used to compare different methods. For example, Liu et al.

[6] To avoid clutter in the displays, the graphs and the tables always show the relative errors while the discussion of results uses mostly the more intuitive percent error terminology. For example, a relative error of -0.1 corresponds to a percent error of -10% (i.e. an under-pricing of 10%), a relative error of 0.5 corresponds to a percent error of 50% (i.e. an overpricing of 50%).

[7] The results remain qualitatively unchanged when we consider the stock price of the firm six, nine or twelve months after the end of the financial year t.

(2002) and Liu, Nissim, and Thomas (2007) look at the distribution of the pricing relative errors $\%E$ and focus on "the interquartile range as the primary measure of dispersion". Alternatively, the precision of two pricing approaches is evaluated by comparing the same percentile of the two error distributions. Alford (1992) states that "the accuracy of the different methods [ ] is assumed to be captured by the median and 90th percentile of the distribution" of absolute errors $|\%E|$ (see also Cheng and McNamara (2000)). Lie and Lie (2002) considers the percentage of firms within 20% of the actual price, i.e. the 20th percentile of the distribution of $|\%E|$. Similarly, Kaplan and Ruback (1995), Kim and Ritter (1999), Gilson, Hotchkiss, and Ruback (2000), Lie and Lie (2002) focus on the "fraction of errors that are less than 15 percent", i.e. the 15th percentile of the distribution of $|\%E|$.

The comparisons in this study are based on the stochastic dominance approach introduced in Kang and Starica (2014). As discussed there, this approach encompasses all the criteria mentioned above and allows for precise statistical testing. For the sake of completeness we recall here the general frame. We say that valuation method X dominates valuation method Y (we write $X \geq Y$), if $F_X(e) \geq F_Y(e)$, for all errors e. F_X and F_Y denote here the cumulative distribution functions (CDF) of the absolute relative valuation errors $|\%E|$ of methods X and Y, respectively. In this case, we will prefer method X to method Y.

If e denotes an absolute prediction error, say 20%, then the inequality in the definition, $F_X(e) \geq F_Y(e)$, means that the percentage of firms valued by method X within 20% of the actual price is greater than or equal to the percentage of such firms valued by method Y. In other words, there is at least as high a proportion of precisely valued firms by method X as by method Y (precision here means an absolute error smaller than e). If method X dominates method Y, then whatever error level we may choose, there is always more precision delivered by method X than by Y.

The definition above implies a clear relationship between the corresponding percentiles of the two distributions. Since for a $p \in [0, 1]$, the $(p * 100)\%$ percentile of the distribution X is defined as $F_X^{-1}(p)$, if method X dominates method Y, then

$$F_X^{-1}(p) \leq F_Y^{-1}(p), \tag{22}$$

for all $p \in [0,1]$, i.e. all percentiles of X are smaller than the corresponding Y percentiles. In particular, the median absolute error of method X is smaller than that of method Y.

The strength of the relation of dominance between two valuation measures X and Y (when established) will be measured by the *dominance measure*, denoted by $dm(X,Y)$. Table 1 gives an equivalent formulation for the definition of the relation of dominance that can be implemented statistically and introduces the definition of the dominance measure.

Relation	When	Dominance measure	Meaning
$X > Y$ or $F_X(e) > F_Y(e)$	$\sup_e(F_X(e) - F_Y(e)) > 0$ and $\inf_e(F_X(e) - F_Y(e)) = 0$	$dm(X,Y) := \sup_e(F_X(e) - F_Y(e)) > 0$	Method X is more precise than method Y
$X < Y$ or $F_X(e) < F_Y(e)$	$\sup_e(F_X(e) - F_Y(e)) = 0$ and $\inf_e(F_X(e) - F_Y(e)) < 0$	$dm(X,Y) := \inf_e(F_X(e) - F_Y(e)) < 0$	Method Y is more precise than method X
$X = Y$ $F_X(e) = F_Y(e)$	$F_X(e) = F_Y(e)$ for all errors e	$dm(X,Y) := 0$	The 2 methods are equally precise
Neither method dominates the other	$\sup_e(F_X(e) - F_Y(e)) > 0$ and $\inf_e(F_X(e) - F_Y(e)) < 0$	$dm(X,Y)$ is not defined	The 2 methods cannot be compared

Table 1: **Dominance measure.** This table propose definitions of the dominance measure $dm(X,Y)$ and interpretations of and necessary and sufficient conditions for the relation of dominance between two valuation methods X and Y. F_X and F_Y denote here the CDF of the absolute relative valuation errors $|\%E|$ of methods X and Y, respectively.

The notion of dominance (as defined in Table 1) applied to the distribution of absolute errors $|\%E|$ yields the most exhaustive comparison of the accuracy of competing valuation methods[8].

In practice, before performing a comparison of the accuracy of two valuation methods X and Y, the CDF of the corresponding absolute error $|\%E|$, F_X and F_Y,

[8]The notion of dominance directly generalizes the approaches in Alford (1992), Cheng and McNamara (2000), Kaplan and Ruback (1995), Kim and Ritter (1999), Gilson et al. (2000), Lie and Lie (2002), which compare quantiles of the distribution of absolute errors of competing approaches. If method X dominates method Y, all percentiles of X are smaller than those of Y. In particular, the median, the 15th or the 90th percentiles of the absolute percentage error of method X are smaller than those of method Y.

need to be estimated and the statistical error needs to be taken into account when establishing a performance relationship between the valuation approaches, as well as when calculating the dominance measure $dm(\widehat{F}_X, \widehat{F}_Y)$ between the estimated cdfs, $\widehat{F}_X, \widehat{F}_Y$. For details about how this is done rigorously, see section A in the Appendix.

4.3 Sample and variables

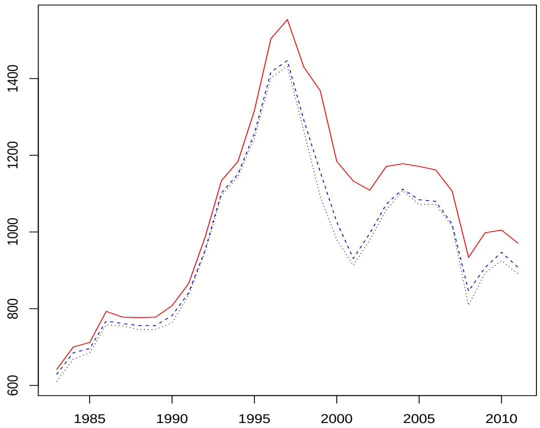
Our sample comprises 27,901 firm-year observations, covering the period 1983-2011 (29 years). It includes all data in CRSP/Compustat for which I/B/E/S analysts' forecast earnings and growth were available and for which the forward multiples under study were positive. More precisely, we started with all firm-years in the CRSP/Compustat and removed all the firm-year observations with

1. missing or negative values in the variables used constructing the multiples, i.e. price, NI_1, NI_2, g;
2. SIC code between 6000 to 6799, i.e. companies from the financial sector;
3. extreme (upper and lower) 1% values of the multiple;
4. no industry peers (the same first digit SIC code) in one year.

(For more detail on the steps in the construction of the sample, see Tables 9 and 10 in the Appendix.)

Figure 1 shows the time evolution of the number of firms in the yearly samples. It peaked in the middle of the 90's around a size of 1500 per year. The first years of the sample are characterized by small yearly samples of six to seven hundreds firms. The more recent years show also a decrease in the number of firms per year.

It is worth noting that our valuation exercise covers a significantly larger number of companies than Courteau et al. (2001), the study that reports the most accurate model-based valuation. The size of our sample is more than double in 1992 and more than four times bigger in 1996. We recall that the cited study covers the period between 1992 and 1996 and follows 442 firms. Our sample augments from 920 (?) firms in 1992 to 1460 (?) firms in 1996.



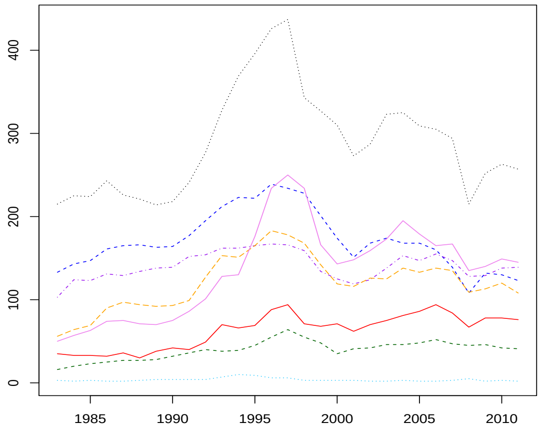


Figure 1: **Time evolution of the number of firms in the sample.** (*Left*): Detailed construction of the yearly samples. The curves (from top to bottom) represent the yearly numbers of companies: 1. after excluding the sector of Finance, Insurance, and Real Estate industry (SIC codes 6000-6799) and those companies with missing values of price, NI_1, NI_2 and g (top), 2. with positive, finite forward multiples P_0/NI_1, P_0/NI_2 (middle line), 3. final sample which excludes the companies with 1% most extreme multiples as well as those without 1-digit SIC peers in the year (bottom). (*Right*): Number of firms in the yearly samples by industry (1-digit SIC code). From top to bottom: SIC 3 (Manufacturing), SIC 2 (Manufacturing), SIC 4 (Transportation, Communications, and Utilities), SIC 5 (Wholesale and Retail Trade), SIC 1 (Mineral and Construction Industries), SIC 7 (Service Industries), SIC 8 (Service Industries), SIC 9 (Public administration).

The drivers used to construct the forward multiples under consideration are the mean of I/B/E/S one year forward earnings per share forecast (NI_1) and the mean of I/B/E/S two year forward earnings per share forecast (NI_2). The long term growth forecast (g) is defined as the mean of I/B/E/S analysts' growth forecasts. The share closing price (P) is obtained from CRSP database and represents the stock price at the end of the third month after the end of the reporting period.

Table 2 gives a few descriptive statistics of the forward multiples under investigation.

Multiple	Mean	Median	Min.	1st Qu.	3rd Qu.	Max.	S.D.
NI_1/P_0	0.069	0.065	0.004	0.047	0.086	0.222	0.033
NI_2/P_0	0.084	0.078	0.013	0.059	0.101	0.269	0.036

Table 2: **Descriptive statistics of the multiples.**

4.4 Statistical implementation

The algorithm described next is applied sequentially to the firms in each year of the sample. To get the industry median earnings yield approximation of the COEC we adopt a hold-out procedure under which the industry multiple is obtained from the multiple attributes of the target firm's industry peers in the given year. More concretely, for each industry-year, we remove one firm observation at a time and calculate the industry multiple using the remaining firms in that industry-year. We put the firm back into the sample and repeat the operation until all the firms within the industry-year are covered. Then we move to the next industry until the year subsample is exhausted.

Once we calculated the multiples for all the firms in the year subsample, we apply classification and regression tree (CART) based methods to estimate the nonlinear relation between $(P/NI)_{i,t}$ and the explanatory variables in equation (14) in a first step, and to forecast the multiple in a second step. Concretely, every year, the subsample is randomly partitioned in twenty subsets of the same size. The data in the nineteen subsets that do not contain the target firm is used to estimate the non-linear function f in (14). In the second step, the estimated f is applied to the predictor values of the target firm yielding the predicted multiple $(\widehat{P/NI})_{i,t}$. The predicted price for the firm and the year under consideration is obtained as in equation (16) by multiplying the predicted multiple by the corresponding analysts' future earnings forecasts. Once the prices and hence the pricing errors for all the firms in one year are calculated, the algorithm is repeated on the firms in the next year until the whole sample is covered.

We note that the CART approach is insensitive to monotone transformations and deals elegantly with irrelevant inputs as well as with the issue of multicollinearity. A more formal presentation of the CART approach employed is available in Section C of the Appendix.

Since the actual stock price for the target firm has not been previously used in the estimation, the valuation errors approximate those that arise when the analyst is valuing a private firm using the multiplier approach.

5 Empirical evidence of a non-linear relation

While Section 3 gave some formal reasoning for the non-linear nature of the relation between the forward multiple and the median industry earnings-to-price ratio, this section looks at this relation through the data. Two aspects of the empirical relationship are considered. First, we directly estimate the relationship between firm's earnings-to-price ratio and median industry ratio and confirm its non-linear nature. Second, we examine the estimated forecasting non-linear function used in the prediction exercise. We find that the overall look of the forecasting function is similar to the estimated relationship. While the reported similarity in shape offers informal confirmation of the appropriateness of our approach, statistical evidence of the fit of the prediction function is given in Sections 6.1 and 6.3 where the precision of the pricing is discussed in detail.

5.1 Direct estimation

The argument for a non-linear relationship outlined in Section 3 is strengthen by the direct estimation results presented in Figure 2. The graph displays the estimated relationship between firm's earnings-to-price ratio on the y-axis and median industry ratio on the x-axis. The four curves correspond to the estimated relationship between the two forward ratios NI_1/P_0 (dotted) and NI_2/P_0 (full) and two industry medians estimated on two-digit SIC industry peers (top two lines, one full and one dotted) and three-digit SIC industry peers (the other two lines), respectively. They have a similar shape and reveal a complex relationship. While for values ranging between 0.05 and 0.10 the industry median is a relatively good approximation of the firm ratio, for industries with a low (high) median the average firm has a higher (lower) ratio.

5.2 Non-linear forecasting function

The importance of the non-linearity in *forecasting* the multiple is highlighted by the graphs in Figure 3 that show the partial dependency functions corresponding to each one of the three industry medians. Since the estimated prediction function

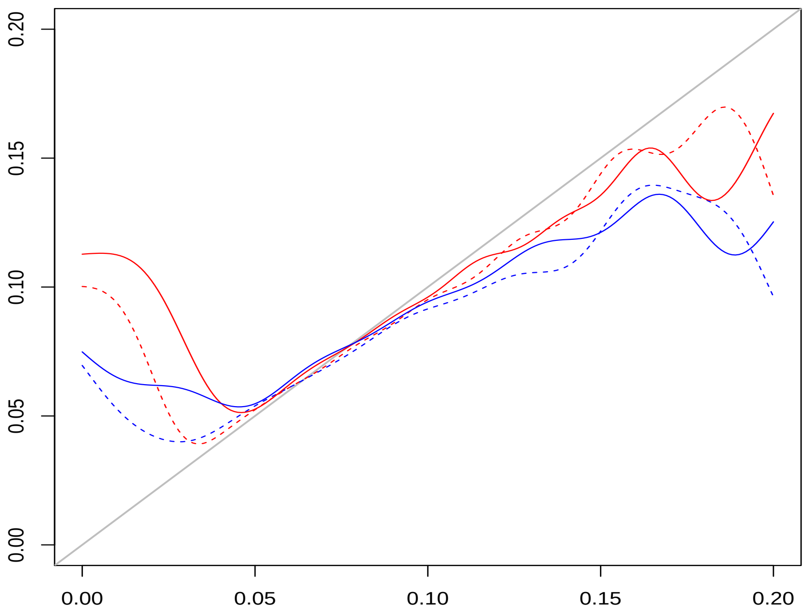


Figure 2: **Non-linear relationship.** The curves display the estimated relationship between the industry median forward ratio in the abscissa and firm's forward ratio, in the ordinate. The dotted lines are used for the NI_1/P_0 multiple while the full ones correspond to the NI_2/P_0 ratio. The top two lines (one full and one dotted) correspond to median multiple estimated on two-digit industry peers while the other two correspond to the three-digit industry peer median. The graph shows non-linearities in the dependency.

f defined in equation (19) is trivariate, in order to isolate the contribution of each of the prediction variables, we need to fix the values of the other two. These variables are given the values available in the sample and then the median forecast is calculated yielding the partial dependency function.

More concretely, for a given x, $pd_i(x)$, the partial dependency function at x of the predictor i, is the median prediction for all the triplets that have x on the i-th position and data values for the other entries. For the median forward ratio of the firms with common first SIC digit, for example, the partial dependency function pd_1 is defined as:

$$pd_1(x) = \underset{(m_2,m_3)}{\text{median}} f(x, m_2, m_3), \tag{23}$$

where f is the estimated non-linear trivariate function in equation (19) and the median is taken over all the pairs (m_2, m_3) of median forward ratio of firms with common first two, respectively three SIC digits.

The graphs in Figure 3 display the median forecasted value of the forward

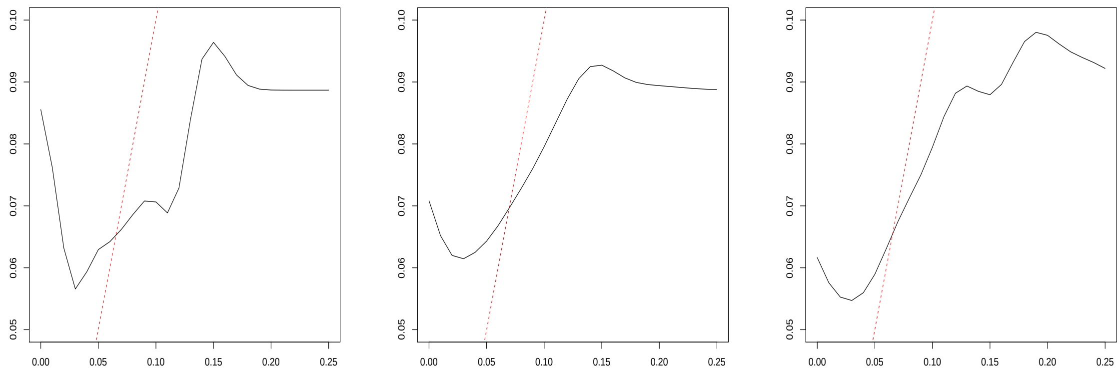


Figure 3: **Non-linear forecasting functions.** Partial dependence plots of the one-digit (*Left*), two-digit (*Center*), and three-digit (*Right*) SIC industry median, respectively. The predicted ratio is NI_1/P_0. The plots give an idea of the dependency of the predicted value of the multiple on each one of the three industry medians used in the prediction. The shape of the three partial dependency functions, quantifying the forecasting relationship, is similar to the shape of the directly estimated relationship in Figure 2 confirming the appropriateness of our forecasting approach.

earnings-to-price ratio[9] NI_1/P_0 on the y-axis, while the x-axis corresponds to the values of each one of the three median industry earnings-to-price. The plots give an idea of the dependency of the predicted value of the multiple on each one of the three industry medians used in the prediction. We note that the shape of the three partial dependency functions, quantifying the forecasting relationship, is similar to the shape of the directly estimated relationship displayed in Figure 2 giving informal confirmation of the appropriateness of our forecasting approach. More formal evidence of the fit of our prediction set-up is given in Sections 6.1 and 6.3 where the precision of the pricing is discussed in detail.

6 Results

This section analyses in detail the pricing performance of the non-linear approach at the center of this study. We begin by an evaluation of its accuracy. We compare its precision with that of the plain-vanilla price-to-earnings valuation. We find that the non-linear enhancement proposed in this article dominates by a good margin the plain-vanilla valuation. Next, we investigate the profile of the firms whose valuations gain in precision through the non-linear enhancement of the multiple

[9] To keep the scale on the two axis comparable we chose to display on the y-axis the ratio NI/P, i.e. the inverse of the multiple.

method and identify a number of their characteristics: high/low earnings-to-price ratio, membership to industries with low median ratio, high short-term growth forecast by the analysts. Finally, we investigate the role that lagged multiple value could play in predicting the current multiple. We find that plain-vanilla multiple valuation does not seem to improve over the trivial approach that predicts the current multiple by its lagged value. When accounting for non-linearity, the predicting contribution of the lagged value of the multiple over that of the industry median multiple is practically equal to zero.

6.1 Analysis of errors

In this section we compare the accuracy of the plain-vanilla price-to-earnings valuation with that of the non-linear enhancement proposed in this article. After we report the main statistics of the relative and absolute pricing errors, we study the dominance relation of the two approaches. We find that the non-linear approach dominates the plain-vanilla multiple valuation and that the dominance measure is equal to 25%, i.e. a quarter of the sample.

6.1.1 Summary statistics

For ease of comparison with the results in the literature on the firm valuation we present first the main statistics of the relative (Table 3) and absolute (Table 4) pricing errors. The errors correspond to pricing when the multiple is predicted as a non-linear function of subsets of the prediction variables in equation (14). First line of each multiple predicting formula refers to the P_0/NI_1 multiple while the second line corresponds to the P_0/NI_2 ratio. The benchmark is the plain-vanilla multiple valuation (first entry in the table). The asterisk marks the methods that are statistically more precise than the plain-vanilla based on the Kolmogorov-Smirnov test presented in Section A of the Appendix (see equation (29)).

We note that the precision of pricing when the multiple is constructed as a non-linear function of analysts' growth rate or the ratio of the predicted earnings only is comparable with that of plain-vanilla multiple valuation. If both variables are used, the pricing is statistically equally accurate.

Predicted multiple formula	Mean	S.D.	Median	75%-25%	90%-10%	95%-5%
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = \underset{j\in SIC2}{\text{median}}\left(\frac{P_{0,j}}{NI_{1,j}}\right)$	0.000	0.469	0.000	0.499	1.068	1.490
(plain-vanilla)	0.001	0.420	0.001	0.441	0.948	1.322
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(g_i)$	-0.043	0.523	-0.043	0.513	1.105	1.566
	-0.048	0.458	-0.048	0.474	1.013	1.401
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\frac{NI_{2,i}}{NI_{1,i}}\right)$	-0.040	0.502	-0.040	0.519	1.097	1.531
	-0.039	0.501	-0.039	0.519	1.100	1.531
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(g_i, \frac{NI_{2,i}}{NI_{1,i}}\right)$	-0.055	0.423	-0.055	0.441	0.958	1.335
	-0.053	0.423	-0.053	0.442	0.957	1.336
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\underset{j\in SIC}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right)\right)^*$	-0.013	0.309	-0.013	0.351	0.685	0.946
	-0.016	0.268	-0.016	0.309	0.577	0.797
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\frac{NI_{2,i}}{NI_{1,i}}, g_i, \underset{j\in SIC2}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right)\right)^*$	-0.046	0.343	-0.046	0.335	0.775	1.109
	-0.044	0.333	-0.044	0.315	0.742	1.061
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(\underset{j\in SIC1}{\text{median}}, \underset{j\in SIC2}{\text{median}}, \underset{j\in SIC3}{\text{median}})^*$	-0.018	0.268	-0.018	0.268	0.579	0.861
	-0.020	0.237	-0.020	0.238	0.500	0.723

Table 3: **Statistics for pricing errors.** First line of each predicting multiple formula refers to the P_0/NI_1 multiple while the second line corresponds to the P_0/NI_2 ratio. The asterisk marks the methods that are statistically more precise than the plain-vanilla based on the Kolmogorov-Smirnov test presented in Appendix (see equation (29)). The figures show substantial gains in precision for the non-linear approaches that use the industry median multiple as a predictor.

We also not that, once the median earnings yield is used, the accuracy of the non-linear approach becomes strictly better. Adding the analysts' growth rate and/or the ratio of the predicted earnings does not improve the precision of the pricing. For this reason, these variables will not be used as predictors in the sequel.

The last entry in the table presents the most precise pricing that corresponds to a non-linear prediction of the multiple using three different approximations of the unknown COEC of the firm to be priced. These are the median earnings yields calculated for the three definitions of industry that correspond to one, two and respectively three first common digits in the SIC code[10]. The use of this set of

[10]Adding the forth approximation corresponding to the median earnings yield of the industry

predictors, theoretically motivated in Section 3, addresses also the implementation issue of the choice of peers based on industry criterion.

The figures in Table 3 show substantial gains in precision for the non-linear approach: the standard deviation of the pricing errors decreases by 42%, the interquartile range by 45%.

Predicted multiple formula	Mean	25%-ile	Median	75%-ile	90%-ile
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = \underset{j\in SIC2}{\text{median}}\left(\frac{P_{0,j}}{NI_{1,j}}\right)$	0.331	0.110	0.245	0.452	0.701
(plain-vanilla multiple valuation)	0.298	0.098	0.218	0.400	0.630
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(g_i)$	0.368	0.116	0.253	0.459	0.721
	0.322	0.112	0.244	0.429	0.659
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\frac{NI_{2,i}}{NI_{1,i}}\right)$	0.350	0.120	0.261	0.461	0.708
	0.335	0.114	0.251	0.441	0.679
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(g_i, \frac{NI_{2,i}}{NI_{1,i}}\right)$	0.302	0.105	0.228	0.407	0.627
	0.304	0.106	0.229	0.411	0.630
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\underset{j\in SIC}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right)\right)^*$	0.226	0.086	0.175	0.302	0.477
	0.195	0.076	0.155	0.258	0.395
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f\left(\frac{NI_{2,i}}{NI_{1,i}}, g_i, \underset{j\in SIC2}{\text{median}}\left(\frac{P_{0,j}}{NI_{\cdot,j}}\right)\right)^*$	0.245	0.077	0.177	0.336	0.540
	0.233	0.073	0.165	0.320	0.514
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(\underset{j\in SIC1}{\text{median}}, \underset{j\in SIC2}{\text{median}}, \underset{j\in SIC3}{\text{median}})^*$	0.188	0.060	0.131	0.248	0.430
	0.166	0.056	0.120	0.219	0.359

Table 4: **Statistics for absolute pricing errors.** First line of each predicting multiple formula refers to the P_0/NI_1 multiple while the second line corresponds to the P_0/NI_2 multiple. The asterisk marks the methods that are statistically more precise than the plain-vanilla based on the Kolmogorov-Smirnov test presented in Appendix (see equation (29)). The figures show substantial gains in precision for the non-linear approaches that use the industry median multiple as a predictor.

The improvement in precision reported in Table 4 is considerable: a reduction of 47% (44%) of the median absolute error for the P_0/NI_1 (P_0/NI_2) multiple, of 45% for the 75%-ile, and of 39% (42%) for the 90%-ile of the absolute pricing error

defined by all four digits of the SIC code does not improve the precision. The specification we propose is hence parsimonious.

for the P_0/NI_1 (P_0/NI_2) multiple.

6.1.2 Dominance analysis

This section presents the more complete performance comparison that uses the notion of dominance defined in Section 4.2.2. Based on the preliminary results in the previous section we focus our discussion on the better performing valuation approaches, i.e. those based on predicting the multiple as a non-linear function of approximations of the COEC of the firm to be priced by one and three industry medians, respectively. The results of the analysis are displayed in Figure 4 and in Table 5.

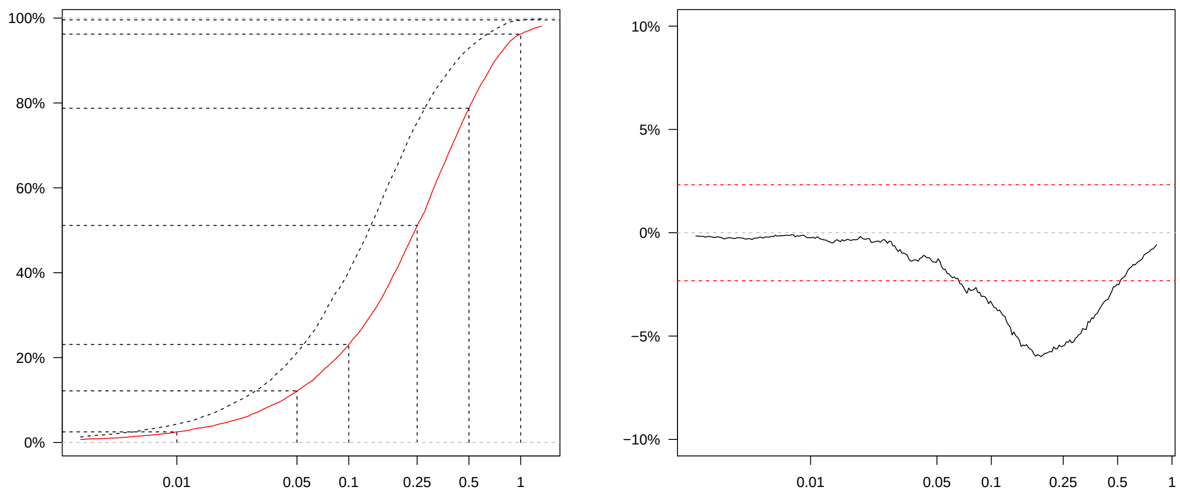


Figure 4: **Dominance plots.** (*Left*): Estimated cdfs of the absolute relative pricing errors of the plain-vanilla (continuous line) and non-linearly enhanced P_0/NI_1 valuation (non-linear prediction of the multiple using three industry medians). (*Right*): The difference between the estimated cdfs of the absolute relative pricing errors of the non-linearly enhanced P_0/NI_1 and P_0/NI_2 valuation. A negative value outside the confidence bands indicates a statistically significant superior performance for the P_0/NI_2 valuation. The proportion of the sample priced with an error of at most 15% of the true price by the P_0/NI_2 approach is 5% larger than the proportion priced with the same error by the P/NI_1 valuation.

The left hand-side graph in Figure 4 displays the estimated cdfs of the absolute relative pricing errors of the plain-vanilla (continuous line) and the non-linearly enhanced P_0/NI_1 valuation. It shows that accounting for the non-linearities in predicting the multiple produces significant gains in precision for all level of error. For any level of error e, the percentage of firms valued by the non-linear approach within the prescribed error e of the actual price is strictly greater than the percentage of such firms valued by the classical approach. The difference in precision between the two methods is most pronounced for $e = 0.25$: while the plain-vanilla values 53% of the firms with an error of at most 0.25 of the actual price, for the non-linearly enhanced approach the proportion valued with the same precision is as high as 78%, i.e. an increase of a quarter of the sample.

The right hand-side graph in Figure 4 displays the difference between the estimated cdfs of the absolute relative pricing errors of the non-linearly enhanced P_0/NI_1 and P_0/NI_2 valuations. The dotted lines mark the limits of the confidence band corresponding to the Kolmogorov-Smirnov statistic (see Section A in the Appendix for details). The graph shows that the P_0/NI_2 pricing is statistically more precise than the P_0/NI_1 with a maximal gain of almost 7% for the error of 0.15, i.e. the proportion of the sample priced with an error of at most 15% of the true price by the P_0/NI_2 approach is almost 7% larger than the proportion priced with the same error by the P_0/NI_1 valuation.

Table 5 gives more detailed information about the dominance relations between different multiple prediction approaches. The two forward multiples are predicted as the industry (SIC 2) median multiple (with the firm to be valued removed) (”plain-vanilla”), as a non-linear function of the median industry multiple (”non-linear 1 median”) and as a non-linear function of three industry medians (SIC 1, SIC 2, SIC 3) (”non-linear 3 medians”).

The table shows that generalizing the plain-vanilla prediction by allowing for non-linearity in the prediction of the multiple brings an improvement in the dominance measure of 14% for both forward ratios (lines 3 vs. 5 and 4 vs. 6 in the Table 5) while using three industry medians in the non-linear forecasting brings an extra improvement of 11% (lines 1 vs. 3 and 2 vs. 4). The overall improvement is, for both multiples, of 25%. In other words, the non-linearly enhanced approach is better than the plain-vanilla for any level of precision. For the error level where the gain of precision is maximal, e_{max}, the proportion of firms valued with a precision better than e_{max} augments by a quarter of the sample for the non-linear approach.

	Multiple	Method	1	2	3	4	5	6
1	P_0/NI_2	non-lin 3 medians	=					
2	P_0/NI_1		5	=				
3	P_0/NI_2	non-lin 1 median	11	7	=			
4	P_0/NI_1		16	11	7	=		
5	P_0/NI_2	plain-vanilla	25	20	18	12	=	
6	P_0/NI_1		30	25	24	17	6	=

Table 5: **Dominance ordering of different multiple prediction approaches.** The entry (i, j) in the table represents the estimated value of the dominance measure $dm(F_j, F_i)$ (see Table 1 and definition (30)) where F_i is the i-th pair (multiple, prediction method) in the ranking given by the first two columns of the table. We predict the multiple using plain-vanilla one industry (SIC 2) median ('plain-vanilla') and non-linear approaches with one industry (SIC 2) median predictor ('non-lin 1 median') and three industries (SIC 1, SIC 2, SIC 3) medians predictors. The value 25 on the entry (5,1) in the table indicates that valuations based on a non-linear prediction of multiple P/NI_2 using three industry medians dominate the plain-vanilla valuations using the same ratio and that the difference between the estimated CDF of the absolute relative pricing errors of the two approaches is at most 25%, i.e. $0 \leq F_1(e) - F_5(e) \leq 25\%$ for all e.

6.2 Who gains in accuracy?

In the sequel we investigate the profile of the firms whose valuations gain in precision through the non-linear enhancement of the multiple method. We find that the firms that benefit most are those with high/low earnings-to-price ratio, those in industries with a low earning-to-price median, as well as those with a high short-term growth forecast by the analysts.

Since we suspect that the relationship between firm's characteristics and the gain in precision is non-linear we chose to not perform a linear regression. Instead, we divided the range of the given feature into twenty subintervals and calculated the conditional mean of the precision gain given the value of the characteristic under investigation. The gain in precision is measured as the relative change in the absolute pricing error:

$$\frac{|\%E|_{\text{non-linear}} - |\%E|_{\text{plain-vanilla}}}{|\%E|_{\text{plain-vanilla}}}. \tag{24}$$

Figure 5 shows the results of the analysis. On the x-axis the graphs display the values of the characteristic under investigation while the y-axis shows the relative reduction of absolute pricing error. The firm characteristics under scrutiny are earnings-to-price ratio (*Upper-left*), industry median earnings-to-price (*Upper-right*), and short-term analyst forecasted earnings' growth (*Lower-center*). The

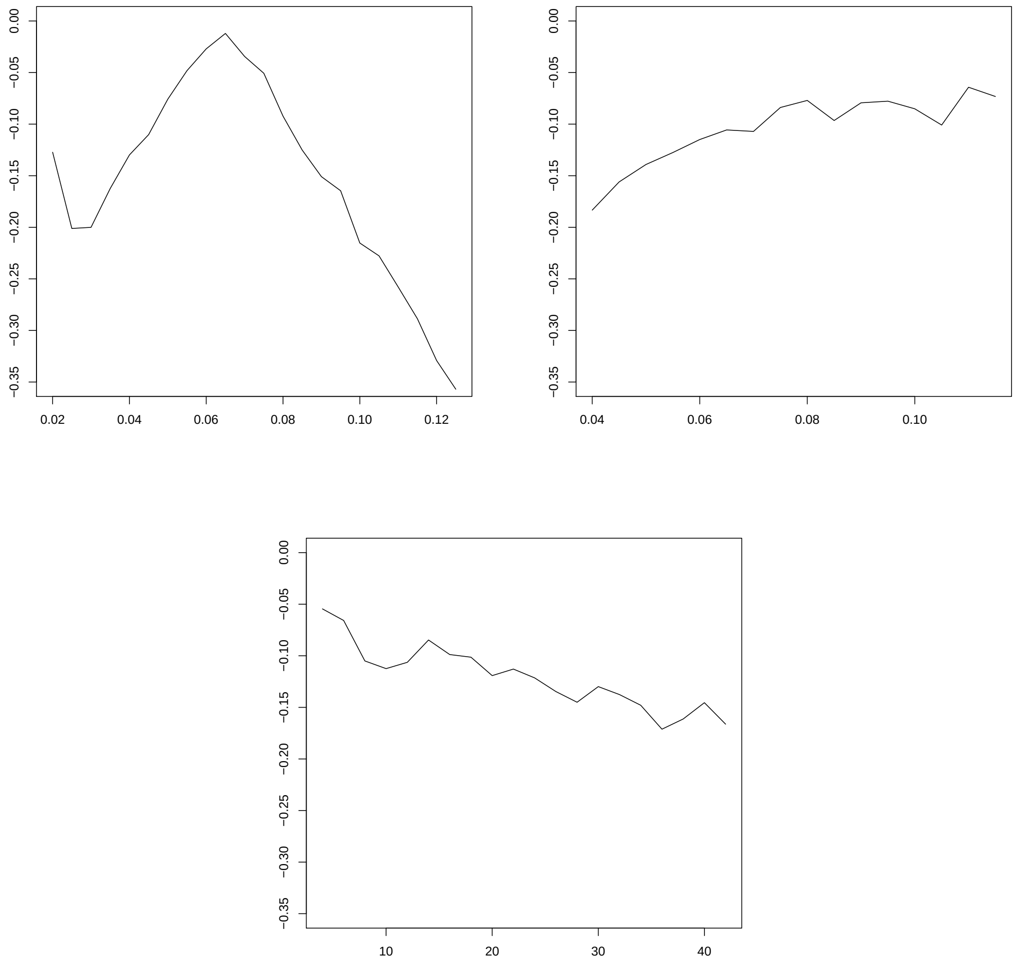


Figure 5: **Error improvement.** The graphs display on the x-axis the values of the characteristic under investigation and on the y-axis the relative reduction on the absolute pricing error. The characteristics under scrutiny are firm's NI/P (*Upper left*), industry median NI/P (*Upper right*), and short-term analyst forecasted earnings growth (*Lower center*). The graphs show a non-linear relationship for the first two characteristics. The firms with higher reduction of pricing error are those with high/low earnings-to-price ratio, those in industries with a low earning-to-price median, as well as those with a high short-term growth forecast by the analysts.

upper-left graph displays the strongest dependency and indicates that valuation errors for firms with low earnings-to-price ratios decrease by up to 20% while firms with high earnings-to-price ratios see an accuracy improvement of up to 35%. Firms in the central value range gain little in valuation precision. The upper-right display shows a more even reduction of the pricing error as a function of industry median ratio: firms in industries with low earning-to-price ratio gain as much as 18% in relative precision while for the rest of the firms the reduction of the pricing errors is of around 9%. The dependency of the valuation error on analysts' forecasted growth, displayed in the bottom row, is more linear and indicates a negative relationship: the valuation of low growth firms improves by 5% while the pricing error for firms with high forecasted short term growth decreases by up to 18%.

6.3 Multiple prediction using past valuation multiples

For traded firms, previous year price and previous year multiple values could help the valuation. It is believed for example, that "if practitioners are simply interested in generating the best possible forecast of a company's current or future valuation multiple, then the company's past valuation multiple would be the most useful forecasting variable" (Sloan (2002)). In this section we take up this statement and investigate the relevance of lagged price and lagged multiple values in pricing. Besides possibly improving the precision, employing these variables in simple forecasting schemes provide us with useful benchmarks for accuracy.

6.3.1 Plain-vanilla vs. past lag multiple prediction

More concretely, it is of interest to evaluate the accuracy of the most trivial prediction set-up, in which the past year multiple is used as a forecast of the current value. The predicted price is obtained by simply multiplying past year's multiple by current year's earnings. For ease of comparison with the results in the literature on firm valuation we present first the main statistics of the absolute pricing errors (Table 6) and in a second step a more detailed dominance analysis (Figure 6). Since the sample we investigate changes slightly[11] (not all firms used in the previous analysis have lagged information) we report also the valuation precision of plain-vanilla and non-linear enhanced multiple approaches.

Table 6 collects the main statistics that illustrate the findings to be discussed in this section. The entries relevant to the theme of this subsection are contained in the first panel, i.e. the first two lines of the table.

The statistics in the first panel of the table indicate that the plain-vanilla multiple valuation does not seem to improve over the trivial approach that predicts the current multiple by its lagged value. In fact, the percentiles up to the 75%th

[11]The sample size decreases to 22,610 firm-years. Due to the smaller sample the precision improves slightly but with no statistical significance.

Formula of multiple prediction	25%-ile	Median	75%-ile	90%-ile
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = \underset{j \in SIC2}{\text{median}} \left(\frac{P_{0,j}}{NI_{\cdot,j}} \right)$	0.104	0.237	0.436	0.682
(plain-vanilla multiple valuation)	0.092	0.206	0.381	0.606
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = \text{lagged} \left(\frac{P_{0,j}}{NI_{\cdot,j}} \right)$	0.099	0.216	0.416	0.771
	0.091	0.198	0.365	0.615
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f \left(\text{lagged} \left(\frac{P_{0,j}}{NI_{\cdot,j}} \right) \right)$	0.095	0.212	0.389	0.633
	0.088	0.190	0.345	0.544
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f \left(\underset{j \in SIC2}{\text{median}} \left(\frac{P_{0,j}}{NI_{\cdot,j}} \right) \right)^*$	0.081	0.168	0.290	0.457
	0.071	0.147	0.248	0.388
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(\underset{j \in SIC1}{\text{median}}, \underset{j \in SIC2}{\text{median}}, \underset{j \in SIC3}{\text{median}})^*$	0.058	0.127	0.240	0.411
	0.052	0.115	0.210	0.356
$\widehat{\frac{P_{0,i}}{NI_{\cdot,i}}} = f(\underset{j \in SIC1}{\text{median}}, \underset{j \in SIC2}{\text{median}}, \underset{j \in SIC3}{\text{median}},$	0.051	0.117	0.233	0.408
$\text{lagged} \left(\frac{NI_{\cdot,i}}{P_{0,i}} \right))^*$	0.047	0.106	0.207	0.351

Table 6: **Statistics for absolute pricing errors when lagged variables are used.** The first line of each entry refers to pricing with the NI_1/P_0 multiple while the second line concerns NI_2/P_0 valuation. *First panel:* The first entry refers to plain-vanilla valuation while the second one concerns the trivial prediction of the multiple by its past value. The figures show little if any gain in precision from using multiple valuation with respect to trivial prediction using past multiple value. Predicting the price by its previous year value works worse than pricing based on predicting the multiple by its lagged value. *Second panel:* The first entry refers to the non-linear enhancement of the plain-vanilla valuation, i.e. the multiple is predicted as a non-linear function of the SIC2 industry median, while the second one concerns the non-linear prediction of the multiple based on its lagged value. The figures show substantial gains in precision for the non-linear approach only in the case of the industry median predictor. *Third panel:* The first entry refers to non-linear prediction of the multiple using three industry medians while the second one concerns the non-linear prediction that adds the lagged multiple as the forth predictor. Adding the lagged multiple does not significantly improve the precision of the pricing.

are lower for the simple-minded forecast for both P/NI multiples. Figure 6 sheds more light on the overall relation between the two distributions. In the case of the P_0/NI_1 multiple neither of the approaches dominates. The percentiles up to the 80%th are lower for the simple-minded forecast. The multiple valuation produces fewer larger (than more than 50% of the actual price) errors. In the case of the P_0/NI_2 multiple the two approaches are equally precise: the difference between the estimated cdfs is inside the 95% confidence bands. Considering these results, the statement in the opening of the section seems to hold: lagged multiple value approach performs as well as the plain-vanilla multiple valuation.

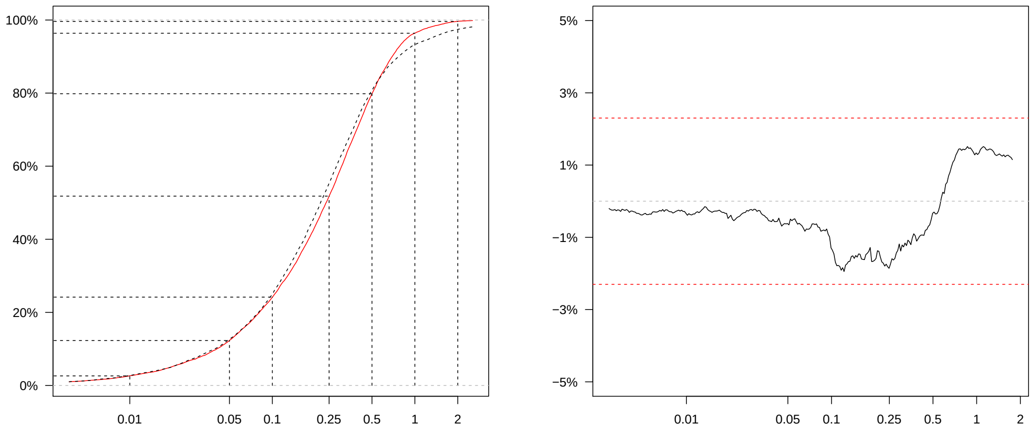


Figure 6: **Pricing error distributions: plain-vanilla vs. lagged multiple.** *Left:* Estimated cdfs of the absolute pricing errors of trivial prediction of the multiple by its lagged value (dotted line) and of plain-vanilla P_0/NI_1 multiple valuation. Neither of the approaches dominates. The percentiles up to the 80%th are lower for the simple-minded forecast. The multiple valuation produces less larger (than 50% of the actual price) errors. *Right:* The difference between the estimated cdfs of the absolute relative pricing errors of trivial prediction of the multiple by its lagged value and of the plain-vanilla P_0/NI_2 multiple valuation. Values inside the confidence bands indicates that the two methods are a statistically equally precise.

Before moving further, it is worth recalling the size of the median absolute pricing errors for different comprehensive approaches based on accounting models reported in the literature (see Section 2). Most of them are significantly higher than that of the approach which trivially forecasts the current multiple by the previous year multiple.

6.3.2 Non-linear prediction with lagged multiple

Next we investigate the effect of allowing for non-linearities in predicting the multiple when lagged multiple values are available. The pricing set-up of the previous section is extended in two ways. First, we stick to one predictor, i.e. industry SIC 2 median and lagged multiple, respectively, and examine the impact on accuracy of the more general non-linear prediction approach. Second, we investigate the influence on prediction's precision of adding the lagged multiple (as a forth predictor) to the three industry medians. We find that non-linear prediction of the multiple does not enhance the precision when the predicting variable is the lagged price while its impact is significant when applied to the industry median multiple predictor. We interpret this finding as further support of the argument for non-linear prediction methods outlined in Section 3 of the paper. We also find that, while adding the lagged multiple as a forth predictor improves the accuracy of pricing, the gain is close to being statistically not significant. We conclude that, contrary to the opening statement of the section, the company's past valuation multiple is not a very useful forecasting variable. When accounting for non-linearity, its predicting contribution over that of the industry median multiple is of little value. These conclusions are motivated by the results in the two lower panels of Table 6 as well as by the graphs in Figure 7.

The second panel (lines three and four) of Table 6 display basic statistics of the distribution of absolute pricing errors for the non-linear approach when the prediction variables are the industry SIC 2 median in one case and lagged multiple in the other. The figures show that the non-linear enhancement hardly improves the pricing precision when the predicting variable is the lagged price while its impact is significant when applied to the industry median multiple predictor.

The last panel (the last two lines) of the Table 6 contains basic statistics of the distribution of absolute pricing errors for the non-linear approach when the prediction variables are the three industry SIC 2 medians in one case and the medians and the lagged multiple in the other. The figures show that adding the lagged multiple seems to improve very little if at all the precision of the pricing.

The results of the statistical test of the significance of the contribution of the lagged value variable are displayed in Figure 7. The graphs show the difference between the estimated cdfs of the absolute relative pricing errors when predicting the multiple using first, the three medians and second, the medians together with the lagged value of the multiple. The confidence bands (dotted lines) are those of the Kolmogorov-Smirnov statistic in equation (29) in Section A the Appendix. Values of the difference inside the confidence bands indicate that the two methods are (statistically) equally precise. For both multiples we find that the contribution of the extra predictor is practically not significant (it is barely significant for a small range of errors in the case on P_0/NI_2). In the light of these results we can

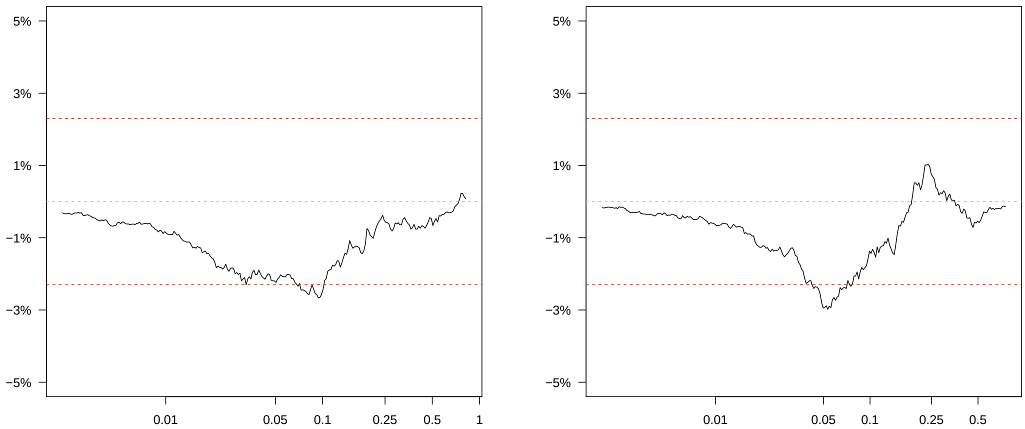


Figure 7: **Pricing error distributions: multiple prediction with three medians vs. three medians plus lagged multiple.** Differences between the estimated cdfs of the absolute relative pricing errors when predicting the multiple with first, three industry medians and second, three industry medians plus the lagged multiple value (*Left:* P_0/NI_1 and *Right:* P_0/NI_2). The confidence bands (dotted lines) are those of the Kolmogorov-Smirnov statistic in equation (29) in the Appendix. Values inside the confidence bands indicates that the two methods are a statistically equally precise. The contribution of the extra predictor is close to being statistically not significant.

conclude that past value of the multiple is not a very useful forecasting variable. When accounting for non-linearity, its predicting contribution over that of industry median multiples is practically equal to zero.

6.4 Non-linear direct prediction of prices

This section investigates the scope for direct non-linear prediction of price. Challenging the usefulness of multiples in valuation, one could try to apply the non-linear prediction approach illustrated in the case of multiple valuation directly to prices. In the sequel we discuss how this approach can be operationalized started form the expression in equation (8).

The value of the firm to be priced can be represented as a non-linear function of the unknown COEC, future earnings and earnings' growth rate[12]:

$$V_{0,i}^{E} = \tilde{f}\left(r_{E,i},\ NI_{1,i},\ NI_{2,i},\ g_i\right), \tag{25}$$

with $\tilde{f}$ a non-linear function. The first specification considered approximates $r_{E,i}$, the unknown COEC, by the industry median earning-to-price ratios as in equation

[12]The pay-out ratio has been dropped from the predictors' list due to limited availability.

(10). In the second specification, the lagged earnings yield as in equation (7) are added to the explanatory variables. The unknown non-linear function $\tilde{f}$ is estimated following the statistical procedure discussed in Section 4.4. For completeness, we also evaluate the trivial approach that uses last year price as the forecast of the current one. We consider two benchmarks: the best plain-vanilla multiple valuation and the best non-linear multiple approach.

Formula for price prediction	25%-ile	Median	75%-ile	90%-ile
$\widehat{P_{0,i}} = \text{lagged } P_{0,i}$	0.113	0.238	0.431	0.745
$\widehat{P_{0,i}} = f(\text{medians},\ NI_{1,i},\ NI_{2,i},\ g_i)$	0.069	0.166	0.341	0.638
$\widehat{P_{0,i}} = f(\text{medians, lagged} \left(\frac{NI_{\cdot,i}}{P_{0,i}}\right),\ NI_{1,i},\ NI_{2,i},\ g_i)$	0.064	0.151	0.307	0.573
Best plain-vanilla multiple valuation	0.092	0.206	0.381	0.606
Best non-linear multiple valuation	0.047	0.106	0.207	0.351

Table 7: **Statistics of absolute pricing errors for direct prediction of prices.** Prices are predicted by past year value (first line), by an estimated non-linear function of the variables in expression (25) with the unknown COEC approximated, first, by industry median earning-to-price ratios as in equation (10) (second line) and, second, by the same variables plus the lagged earnings yields as in equation (7) (third line). The figures suggest that non-linear multiple valuation is more precise than direct prediction of prices, which, at its turn, is more precise than plain-vanilla approach.

Table 7 reports basic statistics of absolute pricing errors for direct prediction of prices. It is worth noticing that the plain-vanilla multiple valuation is not much more precise (if at all) than trivial valuation that uses last year price as a forecast of the current one. The figures in Table 7 also suggest that non-linear direct prediction of prices is more precise than plain-vanilla approach but less accurate than the non-linear enhanced multiple valuation.

These conclusions are confirmed by the graphs in Figure 8 which displays the differences between the estimated cdfs of absolute errors of the best direct price prediction on one hand and the best plain-vanilla multiple valuation (left-hand side) and the best non-linear multiple valuation (right-hand side), respectively, on the other hand. They show that nonlinear multiple valuation is more precise than the direct prediction of prices, which, at its turn, is more precise than the plain-vanilla

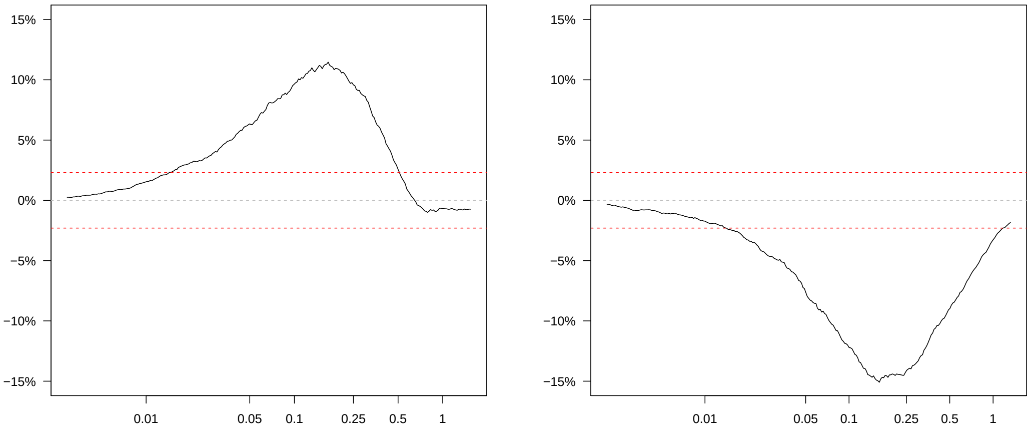


Figure 8: **Pricing error distributions relative to best direct price prediction.** Differences between the cdfs of absolute errors of best direct price prediction in Table 7 and best plain-vanilla multiple valuation (left-hand side) and between best direct price prediction and best non-linear multiple valuation (right-hand side). The confidence bands (dotted lines) are those of the Kolmogorov-Smirnov statistic in equation (29) in the Appendix. Values inside the confidence bands indicates that the two methods are a statistically equally precise. The positive (negative) values in the first (second) graph show the superiority (inferiority) of direct price prediction with respect to the plain-vanilla approach (non-linear enhancement proposed in this article). Non-linear multiple valuation is more precise than direct prediction of prices, which, at its turn, is more precise than plain-vanilla approach.

approach. As a conclusion, the direct prediction of prices can challenge the plain-vanilla multiple valuation but not the non-linearly enhanced multiple approach.

7 Conclusions

We revisit the price-to-earnings multiple valuation approach. Interpreting the median industry earning ratio as a quantity close to the expected return on equity and heeding the prescriptions of valuation models commonly used in the accounting literature, we motivate the need for a non-linear prediction of the multiple. We propose the use of multiple industry median ratios as predicting variables.

We show that non-linearly forecasting the multiple using three industry median multiple constructed according to different depth definition of industry (one, two or three SIC digits) sharply improves the accuracy of the prediction of the multiple. Since the relative pricing error for multiples that are ratios of prices to value driver is equal to the relative prediction error for the multiple, the non-linear prediction of the multiple greatly enhances pricing precision. We report a reduction of 47%

of the median absolute pricing error, of 45% for the 75%-ile, and of 40% for the 90%-ile of the absolute error with respect to plain-vanilla multiple pricing.

We find that the firms that benefit most from the increasing in precision are those with high/low earnings-to-price ratio, those in industries with a low earning-to-price median, as well as those with a high short-term growth forecast by the analysts.

For traded firms, previous year price and previous year multiple values could help the valuation. Somewhat surprisingly, we find that a trivial lagged multiple value approach performs as well as the plain-vanilla multiple valuation. We also find that, when accounting for non-linearity, the predicting contribution of the lagged value of the multiple over that of the industry median multiple is not statistically significant.

Our findings reposition the multiple pricing with respect to the more comprehensive valuation approaches. Although not evaluated on the same sample, the performance of the non-linearly enhanced price-to-earnings multiple pricing compares favorable with that of more comprehensive valuation methods as reported in the literature. On a sample ten times larger, our non-linear enhancement of the price-to-earnings multiple method shows a median absolute pricing error of 10.6% (standard deviation of 16.5%) compared to a median absolute pricing error of 13.7% (standard deviation of 19.7%) of the most precise comprehensive valuation in the literature (Courteau et al. (2001)) based on a discounted cash flow model. We argue that, while price-to-earnings valuation might be an 'imperfect heuristic', when strengthened by the understanding of accounting valuation models, it becomes a precise pricing method.

Summing up, we propose a simple answer to the "broader question of why we care so much about the valuation multiples in the first place" posed in Sloan (2002). That is: because they deliver (when properly enhanced by the insight of classical accounting models).

Appendix

A Statistical details

For a given sample of errors $(e_1, e_2, \ldots, e_{n_X})$, the estimator of F_X, the CDF of the absolute error $|\%E_X|$ of method X, is the *empirical cumulative distribution function*:

$$\widehat{F}_{X,n_X}(x) := \frac{1}{n_X}\sum_{i=1}^{n_X} I_{(-\infty,x]}(|e_i|) = \frac{\%\text{ of } |\text{ errors }| \leq x}{n_X}, \tag{26}$$

where $I_A(x)$ is the indicator function:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A. \end{cases} \tag{27}$$

The statistical estimation error is described by the asymptotic distribution of the two sample Kolmogorov-Smirnov statistic:

$$D_{n_X,n_Y} := \sup_x |\widehat{F}_{X,n_X}(x) - \widehat{F}_{Y,n_Y}(x)|. \tag{28}$$

Under the null hypothesis that $F_X = F_Y$,

$$D_{n_X,n_Y} \leq c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \tag{29}$$

with probability $1 - \alpha$[13].

To summarize, if $\widehat{F}_X$ and $\widehat{F}_Y$ denote the two estimated CDF of the absolute errors $|\%E_X|$ and $|\%E_Y|$ corresponding to the valuation methods X and Y, respectively,

[13] Given the large size of our sample we will use $\alpha = 0.01$ and $c(0.01) = 1.63$.

we define the following measure of dominance:

$$\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y) \stackrel{def}{=} \begin{cases} \max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS} \\ \\ \min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)), & \text{if } \min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS} \text{ and} \\ & \quad \max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS} \\ \\ 0, & \text{if } -c_{KS} \leq \min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \text{ and} \\ & \quad \max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS} \\ \\ ?, & \text{if } \max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS} \text{ and} \\ & \quad \min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}, \end{cases} \tag{30}$$

where $c_{KS} = c(\alpha)\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$, with $c(\alpha)$ as above, n_X and n_Y, the sample sizes used to estimate the two cdfs $\widehat{F}_X$ and $\widehat{F}_Y$, respectively.

Table 8 gives the interpretation of the measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$ and will be referred to in the sequel whenever comparing the precision of the competing multiples.

B Sample construction

Table 10 presents the sample size detailed for each year of the study period 1968-2012.

$\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$	When	Meaning
> 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X dominates F_Y or Method X is more precise than Y
< 0	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	F_Y dominates F_X or Method X is less precise than Y
$= 0$	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \asymp 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \leq c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \geq -c_{KS}$	F_X is equal to F_Y or The 2 methods are equally precise
?	$\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \succ 0$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) \prec 0$ $\max_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) > c_{KS}$ and $\min_e(\widehat{F}_X(e) - \widehat{F}_Y(e)) < -c_{KS}$	Neither cdf dominates or The 2 methods cannot be compared

Table 8: **The measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$.** Interpretation of and necessary and sufficient conditions for the measure of dominance $\mathrm{dm}(\widehat{F}_X, \widehat{F}_Y)$. The notations $\succ$, $\prec$ and $\asymp$ correspond to (in)equalities which are statistically true, i.e. are not rejected by the corresponding Kolmogorov-Smirnov hypothesis test. $\widehat{F}_X$ and $\widehat{F}_Y$ denote here the estimated CDF (see definition (26)) of the absolute relative valuation errors $|\%E|$ of methods X and Y, respectively.

C Classification and regression trees

The classification and regression tree (CART) approach was introduced in Breiman, Friedman, Stone, and Olshen (1984). The basic idea is rather intuitive. The predictor space, i.e. the range of the variable $\operatorname*{median}_{j \in C}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\}$ in our case, is partitioned into N regions, R_1, R_2,...,R_N. For each one of the regions the response function is constant c_n and equal to the mean value of the independent variable :

$$c_n := mean\left(\left(\frac{P}{NI}\right)_{i,t,C} \mid \operatorname*{median}_{j \in C}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\} \in R_n\right). \tag{31}$$

Sample construction	Size
All firm-year observations in CRSP/Compustat (1983-2011)	261,155
Exclude observations with missing values in price	-124,737
	136,418
Exclude observations with SIC codes between 6000 and 6799	-24,746
	111,672
Exclude observations with the price lower than \$3	-14,494
	97,178
Exclude observations with missing values in forecast Earnings and Growth	-66,725
	30,453
Exclude observations with negative multiple drivers	-2,075
	28,378
Exclude observations with largest or smallest 1% multiple values	-475
	27,903
Exclude industries (1st digit SIC) with less than 50 peers in one year	-2
Final sample size	27,901

Table 9: **Construction of sample.**

Year	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Size	611	668	685	758	755	745	746	763	835	947
Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Size	1095	1144	1245	1402	1432	1264	1092	980	913	979
Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	
Size	1057	1108	1072	1072	1015	809	893	925	891	

Table 10: **Size of the sample by year.**

Thus a tree model can be expressed as

$$\widehat{\left(\frac{P}{NI}\right)}_{i,t} := T\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right) = \sum_{n=1}^{N} c_n I\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\}\in R_n\right), \tag{32}$$

where I_A is the indicator function of the set A defined in (27) while the parameters $\Theta := \{R_n, c_n\}$.

To give an idea of the construction of the regions R_n, we describe the first step of the algorithm. It starts with all the data in one region and choses a prediction variable l and a split point s in its range of values. The two new regions R_1 and R_2 are defined as the pair of half-planes,

$$\begin{aligned} R_1(l,s) &= \left\{\underset{j\in C}{\text{median}}\left(\frac{P_{j,t}}{NI_{j,t+1}}\right) \mid \underset{j\in SIC_l}{\text{median}}\left(\frac{P_{j,t}}{NI_{j,t+1}}\right)\le s\right\} \text{ and} \\ R_2(l,s) &= \left\{\underset{j\in C}{\text{median}}\left(\frac{P_{j,t}}{NI_{j,t+1}}\right) \mid \underset{j\in SIC_l}{\text{median}}\left(\frac{P_{j,t}}{NI_{j,t+1}}\right) > s\right\}. \end{aligned} \tag{33}$$

The variable l and the cut point s l are the solutions of the following minimization problem:

$$\min_{l,s}\left\{\min_{c_1}\sum_{\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\}\in R_1(l,s)}\left(\left(\frac{P}{NI}\right)_{i,t,C}-c_1\right)^2+\min_{c_2}\sum_{\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\}\in R_2(l,s)}\left(\left(\frac{P}{NI}\right)_{i,t,C}-c_2\right)^2\right\}. \tag{34}$$

The algorithm continues then to split the existing regions in the same fashion.

While CART is insensitive to monotone transformations and deals elegantly with irrelevant inputs, its predictive power is not so good because of the lack of robustness. To answer this weakness, the method of boosting trees has been introduced by Friedman (1999). The boosted tree predictor is defined as a sum of trees in equation 35.

$$\widehat{\left(\frac{P}{NI}\right)}_{i,t} := b_K\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)=\sum_{k=1}^{K}T\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta_k\right). \tag{35}$$

The steps of the algorithm are as follow:

1. $b_0\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)=0.$

2. For k in 1 to K:

   (a) Compute
   $\hat{\Theta}_k=\arg\min_{\Theta_k}\sum_{i=1}^{N}L\left(\left(\frac{P}{NI}\right)_{i,t,C},b_{k-1}\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)+T\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)\right)$,
   $L(.)$ is the loss function.

   (b) $b_K\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)=b_{k-1}\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right)+T\left(\underset{j\in C}{\text{median}}\left\{\frac{P_{j,t}}{NI_{j,t+1}}\right\};\Theta\right).$

References

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research*, *30*(1), 94-108.

Asquith, P., Mikhail, M., & Au, A. (2005). Information content of equity analyst reports. *Journal of financial economics*, *75*(2), 245-282.

Baker, M., & Ruback, R. (1999). Estimating industry multiples. *Working paper*, Harvard University.

Bancel, F., & Mittoo, U. R. (2014). The gap between theory and practice of firm valuation: Survey of european valuation experts. *Available at SSRN 2420380*.

Bhojraj, S., & Lee, C. (2002). Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, *40*(2), 407-439.

Block, S. (2010). Methods of valuation: Myths vs. reality. *The Journal of Investing*, *19*(4), 7-14.

Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and regression trees*. Chapman & Hall.

Cheng, C., & McNamara, R. (2000). The valuation accuracy of the price-earnings and price-book benchmark valuation methods. *Review of Quantitative Finance and Accounting*, *15*(4), 349-370.

Courteau, L., Kao, J., O'Keefe, T., & Richardson, G. (2003). Gains to valuation accuracy of direct valuation over industry multiplier approaches. *Working paper*, *Available at SSRN 393120*.

Courteau, L., Kao, L., & Richardson, G. D. (2001). Equity valuation employing the ideal versus ad hoc terminal value expressions. *Contemporary Accounting Research*, *4*(Winter), 625-661.

Courteau, L., Kao, L., & Tian, Y. (2015). Does accrual management impair the performance of earnings-based valuation models? *Journal of Business Finance and Accounting*, *42*(1-2), 101-137.

Damodaran, A. (2006). Damodaran on valuation: Security analysis for investment and corporate finance-2/e.

Deloof, M., De Maeseneire, W., & Inghelbrecht, K. (2009). How do investment banks value initial public offerings (ipos)? *Journal of Business Finance & Accounting*, *36*(1-2), 130-160.

Demirakos, E. G., Strong, N. C., & Walker, M. (2004). What valuation models do analysts use? *Accounting Horizons*, *18*(4), 221-240.

Easton, D., P. (2004). Pe ratios, peg ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review*, *79*(1), 73-95.

Easton, D., P., & Monahan, S. (2005). An evaluation of accounting based measures of expected returns. *The Accounting Review*, *80*(2), 501-538.

Francis, J., Olsson, P., & Oswald, D. (2000). Comparing the accuracy and explainability of dividend, free cash flow, and abnormal earnings equity value estimates. *Journal of Accounting Research*, *38*(Spring), 45-70.

Friedman, J. (1999). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, *38*, 367–378.

Gilson, S. C., Hotchkiss, E. S., & Ruback, R. S. (2000). Valuation of bankrupt firms. *The Review of Financial Studies*, *13*(1), 43-74.

Heinrichs, N., Hess, D., Homburg, C., Lorenz, M., & Sievers, S. (2011). Extended dividend, cash flow and residual income valuation models - accounting for deviations from ideal conditions. *Working paper*, *Available at SSRN: http://ssrn.com/abstract=1145201*.

Kang, J., & Starica, C. (2014). Have earnings and book values lost relevance? a relative valuation framework. *Working paper*.

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The Journal of Finance*, *50*(4), 1059-1093.

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics*, *53*(3), 409 - 437.

Larocque, S., & Lyle, M. (2014). Implied cost of equity capital estimates as predictors of accounting returns. *Working paper*, *Available at SSRN 2370686*.

Lie, H., & Lie, E. (2002). Multiples used to estimate corporate value. *Financial Analysts Journal*, *58*(2), 44-54.

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal* *of Accounting Research*, *40*(1), 135-172.

Liu, J., Nissim, D., & Thomas, J. (2007). Is cash flow king in valuations? *Financial Analysts Journal*, *63*(2), 56-68.

Roosenboom, P. (2012). Valuing and pricing ipos. *Journal of Banking & Finance*, *36*(6), 1653-1664.

Schreiner, A., & Spremann, K. (2007). Multiples and their valuation accuracy in european equity markets. *Working paper*, *Available at SSRN: http://ssrn.com/abstract=957352*.

Sloan, R. (2002). Discussion of who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research*, *40*(2), 441-444.

Sougiannis, T., & Yaekura, T. (2001). The accuracy and bias of equity values inferred from analysts' earnings forecasts. *Journal of Accounting, Auditing and Finance*, *16*(4), 331-356.