Assessing the price–earnings relation via machine learning

Catalin Starica* and Jan Marton†

Abstract

The relation between stock prices and accounting earnings has been a central theme in accounting research for more than half a century. By almost exclusively emphasizing a linear, parametric approach, the literature has been unable to convincingly overcome a number of modeling issues, including the effects of non-linearity and lack of fit (Holthausen and Watts [2001]).

We show that a non-linear, non-parametric approach based on recent advances in statistics and machine learning successfully addresses these modeling issues. Our methodology is validated in three ways: 1) residuals meet the orthogonality property, i.e., the estimated relation fits, 2) the firm-specific dependence of price on earnings agrees with theoretical predictions in the literature, and 3) our empirical earnings response coefficients yield reasonable cost of capital estimates.

Keywords: Earnings; price–earnings relation; earnings response coefficient; price level regression; machine learning; non-linear association; non-parametric regression

JEL classification: G10, G30, M41

1 Introduction

While capital market research has been a central theme in the accounting literature for several decades (Ball and Brown [1968]; Kothari and Wasley [2019]), the field is not without methodological challenges. Holthausen and Watts [2001] argue that the valuation models used in accounting research—specifically models that link stock prices or returns to accounting

*School of Economics and Business, University of Neuchâtel, Switzerland.

†School of Business, Economics and Law, University of Gothenburg, Sweden.

We thank Mary Barth, Stephan Hollander, Philip Joos, Robin Litjens, Stephen Penman, Emmeli Runesson, Kenth Skogsvik, Jacob Thomas, and seminar participants at Stanford University, Tilburg University, and University of Gothenburg for helpful suggestions and comments that have considerably improved the paper.

earnings—are not sufficiently elaborate to reflect knowledge about structure and relations in descriptive accounting theory. In particular, they emphasize that "most of the estimated [valuation] models are linear, when there is both ample theory and empirical evidence to support the notion that the relation between the variables in the models and value are non-linear." Consequently, they state that an "area for future research is to advance the valuation models used in the literature by explicitly considering rents, growth and abandonment options [aspects of descriptive accounting theory] and the resulting non-linear relations" (p. 66). Although these statements by Holthausen and Watts [2001] are 20 years old, they are still highly topical, as subsequent literature presents no apparent solution to the issues identified. Instead, the criticized methodology continues to be used (e.g., Balachandran and Mohanram [2011]; Barth et al. [2012]; Srivastava [2014]).

We argue that recent developments in statistics and machine learning (Hastie et al. [2009]; James et al. [2013]) allow for a solution to the issues identified by Holthausen and Watts [2001] that was not technically achievable at the time of their paper. More precisely, we propose and test a non-linear and non-parametric approach to mapping the relation between stock prices and earnings. The method stems from a paradigm shift within statistical science, which is likely to have a significant impact on accounting research. Breiman [2001b]—an important contributor to machine learning methods—discusses the emergence of a new culture in the use of statistical modeling. Traditionally, researchers in statistics assume that data is generated by a given parametric model (of which linear regression is the most prominent example). The new culture, using algorithmic modeling, is based on the general assumption that the data is drawn from an unknown multivariate distribution $(\mathbf{x}, y)$. The goal is to find an algorithm $f(\mathbf{x})$ such that for a given $\mathbf{x}$, $f(\mathbf{x})$ will be a "good" predictor of y. Through its scarcity of assumptions on the structure of the relation tying $\mathbf{x}$ to$y$, this approach is often said "to let the data speak." It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets.

Another way to think of the paradigm shift in statistics is in terms of the distinction between confirmatory and exploratory data analysis. For a confirmatory analysis, the form of the statistical model is determined before the data are considered. In the accounting literature, the model is often the ubiquitous linear regression model. Then, all the researcher needs to do is to estimate the values of some key parameters.[1] In an exploratory analysis, there is no

[1]The distinction is somewhat vague because in practice, few statistical analyses are truly confirmatory. Most

pre-determined model. Instead, statistical tools are used to empirically determine the relation of interest from the data. However, exploratory analysis is not akin to theoretical data mining. Identification of relations can be informed by existing theoretical considerations, which might provide valuable criteria for the choice of estimation algorithm. Moreover, known theoretical regularities can be used to validate the empirical results of the exploratory analysis (cf. Karolyi and Van Nieuwerburgh [2020]).

In the two decades that have passed since Breiman [2001b], algorithmic modeling has become readily available.[2] Although not yet commonly used in accounting research, machine learning is applied in the related field of finance (e.g., Gu et al. [2020]). In this paper, we show how the new algorithmic approach, closely informed and guided by accounting theory, can enrich parametric modeling and address long-standing issues in capital market research in accounting.

Efforts have previously been made in the accounting literature to address concerns of non-linearity highlighted by the critique in Holthausen and Watts [2001]. To our knowledge, all studies on the price–earnings relation that attempt to mitigate the issue of a mis-specified linear association do so within a parametric linear regression framework. Approaches used range from allowing coefficients to vary cross-sectionally (Barth et al. [1998]) to allowing for interactions with, e.g., negative earnings indicator variable and industry (Core et al. [2003]; Balachandran and Mohanram [2011]). While these methods may be partially effective, they require an *ex ante* assumption of the cause and structure of non-linearity. As a consequence, after their deployment, whether there is additional non-linearity and what effect it might have on inference and testing is still unknown (cf. Gow et al. [2016]).

Linear regression is often used as an approximation of non-linear relations in accounting research, and this can—in many cases—be sufficient. However, we show empirically that using a linear approximation for the non-linear price–earnings relation results in an unreliable hypothesis testing for our sample (all listed US firms from 1970 to 2016). Small changes in the sample with no effect on the underlying economic link between earnings and prices give statistically different results which, taken at face value, would provide evidence of change in the economic relation. Sample variation due to changes in features that serve as proxies for risk

of the time, some important parts of the model are unknown before the data are examined. For example, from a large set of potential explanatory variables, a subset may be selected for the final model.

[2]Software like Stata, R, MATLAB, and SPSS have full implementations of machine learning algorithms.

or growth (cf. Easton and Zmijewski [1989]; Collins and Kothari [1989]; Kothari and Shanken [2003]) leads to additional unwanted variation in the estimates. Moreover, we document that significant non-linearity remains even after we include interactions used in prior studies to address non-linearities within a linear framework.

We complement the empirical evidence against linear modeling with a theoretical motivation for adopting a non-linear framework. First, we show that the linear specification of Ohlson [1995]—commonly used as a theoretical motivation for linear modeling of the price–earnings link—does not state that a linear regression is a good approximation of the economic relation between prices and earnings. In fact, Ohlson's specification is a linear regression only under very strict assumptions that are unlikely to hold empirically. Second, we show that the mapping of stock prices as a function of discounted future cash flows (or other measures of future performance) can be represented through a non-linear valuation regression model.

We use the random forest (RF) machine learning approach (Breiman [2001a]; Hastie et al. [2009]) for inferring the non-linear relation between earnings and stock prices. This approach has at least three intuitively appealing properties that make it particularly useful for our application. First, it requires no *ex ante* assumptions of the type of non-linearity in the economic relation under scrutiny. Instead it adapts to non-linearities in the data, thereby helping to mitigate many of the methodological weaknesses of the linear framework identified in the literature. Second, the RF performs a local estimation of the relation between prices and earnings, taking into consideration only firms with similar characteristics, as prescribed by valuation theory. Third, it allows for efficient testing of an exhaustive list of potentially relevant characteristics defining the peers with similar price–earnings relations to be used in the local estimation. We construct the list based on extant accounting and finance literature.

A strength of linear regression is its ease of interpretability (James et al. [2013]). The association of each explanatory variable with the dependent variable is indicated by one coefficient (and one t-statistic). The output from the RF method is not as easy to interpret. Since we model the sensitivity of the price–earnings relation to a number of relevant firm characteristics, the RF yields a multi-dimensional surface (describing the joint dependence of prices on earnings and the additional characteristics), which cannot be easily visualized.

We address the interpretability issue in several ways. First, to assess the overall fit of our valuation model, we use the average size of the absolute percent valuation error, a onedimensional measure of model pertinence. Second, we present graphs of the functional relation between prices and earnings. These graphs show the type of relation (e.g., step function, curvilinear, or locally linear) that exists between the two variables, thereby identifying how the relation differs by level of earnings. Third, we measure and display the importance of each of the proxies for firm characteristics that shape the price–earnings relation.

We validate our methodology in a number of ways, mostly with links to existing accounting theory (cf. Holthausen and Watts [2001]). First, we show that the estimated model provides a good approximation of the economic relation between earnings and prices. The mean of the residuals conditional on the level of firm characteristics that the literature associates with non-linearity is approximately zero for levels of earnings, i.e., the orthogonality property is met.

Second, we find that the variables that are empirically most important in shaping the price–earnings relation agree with findings in extant literature. Out of the 45 proxies we consider, three are more important than the others: price-to-book (P/B), size, and leverage. The relation between prices and earnings is expected to vary by risk, economic rent, growth opportunities, and accounting conservatism (Holthausen and Watts [2001]; Kothari and Shanken [2003]; Liu and Thomas [2000]; Biddle et al. [2001]). P/B is a proxy for risk (Fama and French [1992]; Penman et al. [2018]), economic rent, growth opportunities (Fama and French [1992]), and unconditional conservatism (Roychowdhury and Watts [2007]).[3] P/B is the single most important explanatory variable, after earnings. Size (as measured by market value of equity) and leverage are additional proxies for risk. In accounting research, industry is often used as a proxy for variation in economic and accounting factors. We include industry as a categorical explanatory variable, but it is only the fortieth most important variable of the 45 we considered. These findings suggest that firm-specific characteristics (such as P/B, size, and leverage) might be more appropriate proxies for factors determining the price–earnings relation than the commonly used industry classification.

Third, we relate the estimated functional relation between prices and earnings to existing accounting literature. We document that the shape of the association depends on the level of earnings. For negative earnings, the association is flat, consistent with equity investors' abandonment or liquidation option (Hayn [1995]; Holthausen and Watts [2001]), limited liability

[3] Roychowdhury and Watts [2007] state that unconditional conservatism is one component of P/B and economic rent is another; however, we do not distinguish between these components in the P/B measure.

(which explains the occurrence of the abandonment option (Fischer and Verrecchia [1997]), different levels of information uncertainty (Shin [2003], and the effect of debt covenants (Core and Schrand [1999]). For high levels of earnings, the association is concave, consistent with the existence of transitory earnings components (Freeman and Tse [1992]; Core and Schrand [1999]) and increasing uncertainty of earnings (Subramanyam [1996]).

Fourth, a common argument against linear regression inference based on the Ohlson [1995] model is that the estimated earnings response coefficients (ERCs) are ”too small (Holthausen and Watts [2001]). Theoretically, the ERC should be close to the inverted discount rate for equity, that is, between 8 and 20 depending on the time period (Kothari [2001]). Instead, researchers (see, for example, Kothari and Shanken [2003]) report estimated ERCs of four on average, suggesting a cost of capital of around 25%. The inverted discount rate is theoretically correct if earnings are seen as permanent, and should not apply to transitory earnings. The functional relation between prices and earnings that we document is (close to) flat for extreme earnings (both negative and high earnings), consistent with investors seeing such earnings as transitory (see above). For non-extreme earnings, roughly between the third and the ninth deciles (corresponding to the range of more permanent earnings), the relation is approximately linear. We use this part of the curve to estimate the non-linear ERCs. We show that the non-linear ERCs have a ”reasonable” size, in the sense that they are the inverse of reasonable cost of capital values. For 47 annual cross-sections (1970–2016), we obtain an average non-linear ERC of 8.01, compared to 2.91 for a linear estimate. As the ERC is a proxy for the inverted discount rate, firm-specific ERCs should be associated with proxies for risk. As expected, we find that ERCs are strongly associated with risk-free interest rates, price-to-book ratios, firm size, and leverage. Further, consistent with risk being associated with future returns, we find that ERCs are associated with one-year-ahead returns even after controlling for the risk factors mentioned.

This paper makes an important methodological contribution. We present a statistically sound method for assessing the relation between stock prices and earnings. Our approach mitigates several of the econometric issues identified in prior research. It can be applied to other measures of performance in financial statements besides earnings, for example cash flows or components of earnings and cash flows. It yields a bias-free research design in which the relative usefulness of earnings (or other performance measures) can be assessed using the average size of the absolute percent error of a pricing informed only by the level of the performance measure. Our approach investigates the price–earnings relation with as few formal modeling assumptions as possible, with the stated purpose of revealing the relation unconstrained by hypothesized structures. By ”letting the data speak for itself” (Breiman [2001b]) we present empirical evidence to support the propositions of analytical studies such as Fischer and Verrecchia [1997] and Subramanyam [1996] and generalize prior empirical work based on parametric modeling of non-linearity (e.g., Hayn [1995]; Freeman and Tse [1992]).

The paper is structured as follows. In section 2 we present common linear models used in existing research and explain our focus on stock prices and not stock returns. Section 3 discusses the sample and gives descriptive statistics for the most important variables. In section 4 we test the fit of linear specifications to the price–earnings relation and show that such models give unreliable results when testing hypotheses. Section 5 theoretically justifies the non-linear approach, details how accounting theory guides and informs the empirical specification, and explains how the RF method works. The results of the non-linear modeling, including validation with links to descriptive accounting theory, are presented in section 6. Section 7 gives a specific example of how the proposed methodology can be used to develop and refine descriptive theory of accounting, based on a suggestion in Holthausen and Watts [2001]. Section 8 concludes.

2 Modeling the economic relation between prices and earnings

In this section, we present model specifications commonly used to study the relation between stock market and accounting variables. The specifications vary by stock market measure used and by variables included to reflect non-linearity. Both stock prices and returns are used to measure the relation between stock markets and earnings. While return specifications are commonly preferred, in this study we focus on the relation between stock prices and earnings. Kothari and Zimmerman [1995] argue theoretically that slope coefficients from price models are less biased than those from return models, since the price specification avoids the errors-in-variables problem of the return specification. Their empirical results support the conclusion that a price specification "gives more economically sensible earnings response coefficients."[4] Further, the empirical evidence in Kothari and Zimmerman [1995], where price and return specifications give very different results, suggests that they are not equivalent[5] (see also Christie [1987]), as often assumed.

2.1 Linear representation of the price–earnings relation

The extant accounting literature specifies the price–earnings relation in a linear regression framework. The specifications are estimated cross-sectionally, with separate regressions for each year.[6] Examples of basic models used are:

$$P_{i,0} = \beta_{0,0} + \beta_{1,0} NI_{i,0} + \varepsilon_{i,0}, \tag{1}$$

$$P_{i,0} = \beta_{0,0} + \beta_{1,0} NI_{i,0} + \beta_{2,0} B_{i,0} + \varepsilon_{i,0}, \tag{2}$$

where $P_{i,0}$ is the share price of firm i at time (year) 0, NI is earnings per share, and B is per-share book value of equity.

Specification (2) has been particularly popular and is often theoretically justified as an implementation of the Ohlson [1995] model (for an overview of this model, see section 9.4 in the appendix).

To address possible scale effects, some studies replace the per-share data by unscaled variables or scale the market value and accounting variables with, e.g., book value of equity, number of employees, or sales. However, the literature seems to have arrived at a certain consensus that deflating the variables by the number of shares gives results that are more consistent with formal benchmarks (Barth and Kallapur [1996]; Kothari and Shanken [2003]; Barth and Clinch [2009]; Aledo Martinez et al. [2019]).

[4]The median ERC from the return specification over 38 cross-sectional slope estimates corresponding to the period 1952–1989 was 1.64, corresponding to a cost of equity of 61%.

[5]For example, dividing the change in price by the previous period's price to obtain returns is often explained as a scaling factor introduced to improve the econometric conditions. Since previous period's price is a random variable rather than a constant, its effect in the equation *cannot* be understood as that of a scaling factor. After dividing by the previous period's price, the original variable, i.e., the price change, becomes a new variable.

[6]Some studies combine multiple years in a single regression, even if there is no theoretical justification for doing so.

2.2 Non-linearity within a linear framework

The linear models 1 and 2 are often amended to reflect predictions about non-linearity in the price–earnings association. For example, non-linearity caused by differences between profitable and loss-making firms or between firms in different industries can be captured with the following models:

$$P_{i,0} = \beta_{0,0} + \beta_{1,0}NI_{i,0} + \beta_{2,0}Loss_{i,0} * NI_{i,0} + \beta_{3,0}B_{i,0} + \varepsilon_{i,0}, \tag{3}$$

$$\begin{aligned} P_{i,0} = &\sum_{j=1}^{I-1} \beta_{j,0}^{(0)} Ind_j + \sum_{j=1}^{I} \beta_{j,0}^{(1)} Ind_j * NI_{i,0} + \sum_{j=1}^{I-2} \beta_{j,0}^{(2)} Ind_j * Loss_{i,0} * NI_{i,0} \\ &+ \sum_{j=1}^{I} \beta_{j,0}^{(3)} Ind_j * B_{i,0} + \sum_{j=1}^{I-2} \beta_{j,0}^{(4)} Ind_j * Loss_{i,0} * B_{i,0} + \varepsilon_{i,0}, \end{aligned} \tag{4}$$

where *Loss* indicates firm years with negative earnings and *Ind* is the industry to which the firm belongs. Specification (3) (see Core et al. [2003]) allows for different slope coefficients on profit and loss firms and addresses the differential in the information content of earnings of opposite signs. Losses tend to be less informative than profits (Hayn [1995]).

While, theoretically, the price–earnings relation has a substantial firm- or industry-specific component, specification (2) forces coefficients on earnings and book values to be the same for all firms (and industries) in a cross-section. Specification (4) allows for the intercepts and the coefficients on earnings and book values to vary across industries [7] (Balachandran and Mohanram [2011]).

Specifications (3) and (4) are examples of how particular types of non-linearity can be considered within a linear framework. However, the framework is too rigid for successful inference of complex non-linear relations such as, for example, those between prices and earnings. Both industry and firm-specific characteristics affect the price-earnings relation. Differences in economic rent, growth opportunities, and accounting conservatism cause non-linearity (Biddle et al. [2001]; Holthausen and Watts [2001]). Moreover, Kothari and Shanken [2003] show that aggregate growth and discount rate effects can influence slope coefficients of regression (2). Successful modeling of non-linearity in a linear framework requires that the structure of non-linearity is identified *ex ante*, an often difficult if not impossible task. In addition, once implemented, the researcher does not know whether additional non-linearity remains, or the effect

[7] We use the Fama and French [1997] classification of SIC codes into 12 industry groupings.

on inference of any such non-linearity unaccounted for. The tests in section 4.2 show that, while improving over (1) and (2), specifications (3) and (4) do not fit the economic price–earnings relation.

3 Variables definitions and sample

Apart from share price and earnings, the main variables used are price-to-book, market value, and leverage, as these three variables are empirically the most important for the inference of the price–earnings relation (additional variables are discussed and defined in the appendix, section 9.8). The sample is obtained from Compustat (accounting information) and CRSP (prices). It includes all firm-year observations for listed US firms over the period 1970–2016 with non-missing values for the main variables. Consistent with previous association studies, financial firms (in SIC codes 6000–6999) are excluded because their financial reporting differs from that of non-financial firms. The sample consists of 176,335 firm-year observations and 16,336 distinct firms. Figure 13 in section 9.1 in the appendix displays the number of firms per year, with a minimum of 1,666 in 1970 and a maximum of 5,622 in 1997.

Table 1 displays the summary statistics of the main variables[8], i.e., share price, earnings, the ratio of market value to book value of equity, market value, and leverage. We define earnings (*NI*) as income before extraordinary items (Compustat *IB*), divided by the number of shares. The number of common shares outstanding is the Compustat variable *CSHO*, while the book value of equity is the Compustat variable *CEQ*. The share price, obtained from CRSP, is equal to the monthly closing price collected at the beginning of the fourth month after the fiscal year end. Leverage is defined as total liabilities (Compustat *LT*) divided by the market value of common equity.

Table 2 displays the values of Pearson (upper triangle) and Spearman (lower triangle) correlations between the main variables in the study.

[8]The variables are winsorized at the 1% level, i.e., the smallest 0.5% are set to the 0.5 percentile and the largest 0.5% are set to the 99.5 percentile.

Variable	Mean	SD	10%	25%	50%	75%	90%
Share price	18.46	19.95	1.81	4.62	12.12	25.30	42.88
NI	0.82	1.98	-0.90	-0.09	0.59	1.65	3.02
P/B	2.58	4.69	0.55	0.95	1.66	3.00	5.51
log(Mktv)	4.93	2.35	1.99	3.17	4.77	6.55	8.09
Lev	1.25	2.21	0.09	0.23	0.58	1.36	2.79

Table 1: **Descriptive statistics.**

	Share price	NI	P/B	log(Mktv)	Lev
Share price	1.00	0.58	0.16	0.69	-0.20
NI	0.69	1.00	-0.01	0.36	-0.23
P/B	0.35	0.06	1.00	0.16	-0.17
log(Mktv)	0.77	0.41	0.39	1.00	-0.23
Lev	-0.21	-0.01	-0.61	-0.22	1.00

Table 2: **Correlation between variables.** *Upper triangle:* Pearson correlation, *Lower triangle:* Spearman correlation.

4 Linear regression and the economic relation between prices and earnings

In this section, we present empirical evidence that the linear specifications in section 2 are poor approximations of the non-linear price–earnings relation. For parsimony, and because it is by far the most used linear specification of the price–earnings regression, we use model (2) to illustrate consequences of the poor fit of linear specifications for the testing of hypotheses on the relation. Such tests give unreliable results. Even small changes in the sample composition—with no effect on the underlying economic relation—yield statistically different results. Therefore, although samples are expressions of the same data generating process, inference based on a linear regression model rejects the hypothesis of an identical economic relation. Consequently, although samples are expressions of the same data-generating process, inference based on a linear regression model rejects the hypothesis of an identical economic relation. Although our detailed investigation focuses on specification (2) based on Ohlson [1995], the consequences of a lack of fit are equally severe for the other specifications. Consequently, the implicit assumption that linear regression is an approximation of the non-linear relation between prices and earnings gives unreliable results for the sample we study (listed US firms during the period

1970–2016).

4.1 Conditional expectation—an econometric summary

For clarity of exposition and to establish the notation needed for the discussion in the remainder of the paper, we begin this section by presenting facts about conditional means, which are the primary focus of regression analysis.

Conditional means reduce the complexity of distributions to a single summary measure, which facilitates comparisons. For a pair of random variables (X,Y), the conditional mean (expectation) function is generically denoted by:

$$\mathbb{E}[Y|X=x]=m(x)$$

and is a function of x, the value taken by X, the variable we condition on. The function $x \to \mathbb{E}[Y|X=x]$ is, in general, a *non-linear* function of x.

The concept of conditional expectation is intuitively appealing, as we are used to thinking of an average as providing a representative value for a random variable. In that sense, the conditional expectation $\mathbb{E}[Y|X=x]$ expresses how Y varies with X by averaging the Ys of the pairs (X,Y) for which X takes values close to x.

By definition, the conditional expectation yields the following general decomposition. For *any* pair of random variables (X,Y), we can write:

$$Y=m(X)+\varepsilon, \quad \text{where} \quad m(x)=\mathbb{E}[Y|X=x], \tag{5}$$

that is, the expectation of Y given that the variable X takes a given value x, and where the error term ε satisfies the *orthogonality property*[9] an extra period:

$$\mathbb{E}[\varepsilon|X]=0. \tag{6}$$

It is important to keep in mind that the orthogonality property is not a restriction. It holds true by the definition of the conditional mean concept.

[9] Condition (6) implies $\mathbb{E}(h(X)\varepsilon)=0$ for any function h (which explains the term ”orthogonality”). In particular, for $h(x)=x$ it implies that $corr(X,\varepsilon)=0$ since $\mathbb{E}[\varepsilon|X]=0$ also implies that $\mathbb{E}(\varepsilon)=0$.

The conditional expectation function $m(x)$ is of primary interest for regression analysis because $\mathbb{E}[Y|X=x]$ is the best predictor of Y in the sense that it has the lowest mean squared error among all predictors based on X.

Linear regression can be used as an approximation of a non-linear relation. In empirical research, it is common to estimate the linear model (via ordinary least squares [OLS])

$$Y = \beta_0 + \beta_1 X + \varepsilon \tag{7}$$

even when the conditional mean $\mathbb{E}[Y|X]$ is a non-linear function of X. The linear model (7) is empirically unlikely to be accurate unless X is discrete and low-dimensional.[10] Consequently, in most cases, it is more realistic to view the linear specification (7) as an approximation.

This interpretation is motivated by the fact that the estimation of the linear equation (7) yields the best linear approximation of the possibly non-linear conditional mean function $x \to \mathbb{E}[Y|X=x]$. The error term satisfies the *projection property*:

$$\mathbb{E}[X\varepsilon] = 0, \quad \text{(which implies that } corr(X\varepsilon) = 0, \tag{8}$$

but not necessarily the orthogonality property (6). The error term of the linear model (7) satisfies the orthogonality property only if the conditional mean function $m(x)$ is linear. We emphasize that the estimation result is not necessarily a conditional mean, nor a parameter of a structural economic model. only if the conditional mean function $m(x)$ is linear To summarize, the answer to the question: "When does the linear model (7) represent the underlying economic dependence in the pair (X,Y)?" is: "When its error term satisfies the *orthogonality property*, i.e., $\mathbb{E}[\varepsilon|X] = 0$."

We use this fact to substantiate two statements. First, we argue that the relation of prices to earnings, and of the pair earnings and book value, are not linear, i.e., the specifications (1) and (2) are not expressions of the economic relation of prices to earnings. Second, we claim that the attempts to address the non-linearity in the price–earnings relation within a linear framework by including interactions, as proposed by specifications (3) and (4), are not fully successful. Significant non-linearity remains unexplained.

[10] A linear model only fully mimics a non-linear model when regressors are discrete and there is interaction between all regressors. Because of an exponential increase of interaction variables with more regressors, this approach is only usable with few regressors.

If the linear specifications represent the underlying economic relation, the OLS-estimated residuals of specifications (1) to (4) should approximately verify the orthogonality property (6). Section 4.2 presents evidence that this is not the case.

4.2 Approximation of the underlying economic relation of prices to earnings by a linear model

In this section we highlight theoretical and empirical issues concerning the commonly used approximation of the underlying economic relation of prices to earnings by a linear model.

4.2.1 The Ohlson model provides no theoretical justification for the price–earnings regression

The price–earnings regression (specification (2)) is often theoretically justified by the linear representation of the residual income model introduced in Ohlson [1995]. However, such a motivation is not supported by the main theoretical result of that paper. Ohlson [1995] does not state that a linear regression of prices on book value and earnings provides a good empirical approximation of the underlying economic relation of prices to earnings. In particular, we show (see section 9.5 in the appendix for a formal argument) that, without additional restrictive assumptions unlikely to hold empirically, the Ohlson [1995] model is not a good approximation of the relation between prices and earnings, i.e., the residuals of the model (2) do not satisfy the *orthogonality property* (6). In other words, there is no theoretical reason to assume that Ohlson's linear representation captures the cross-sectional economic relation between prices and earnings.

4.2.2 Empirical evidence of the fit of linear models

In this section, we empirically assess how well the linear models in Section 2.1 approximate the underlying economic relation of prices to earnings, i.e., we test whether the residuals of the estimated linear specifications satisfy the orthogonality property (6).

Figure 1 displays summary information about the distribution of residuals of specifications (1) through (4) conditional on earnings belonging to each of the 20 ventiles of NI (each segment/box belongs to the ventile specified on the x-axis). The use of ventiles is an opera-

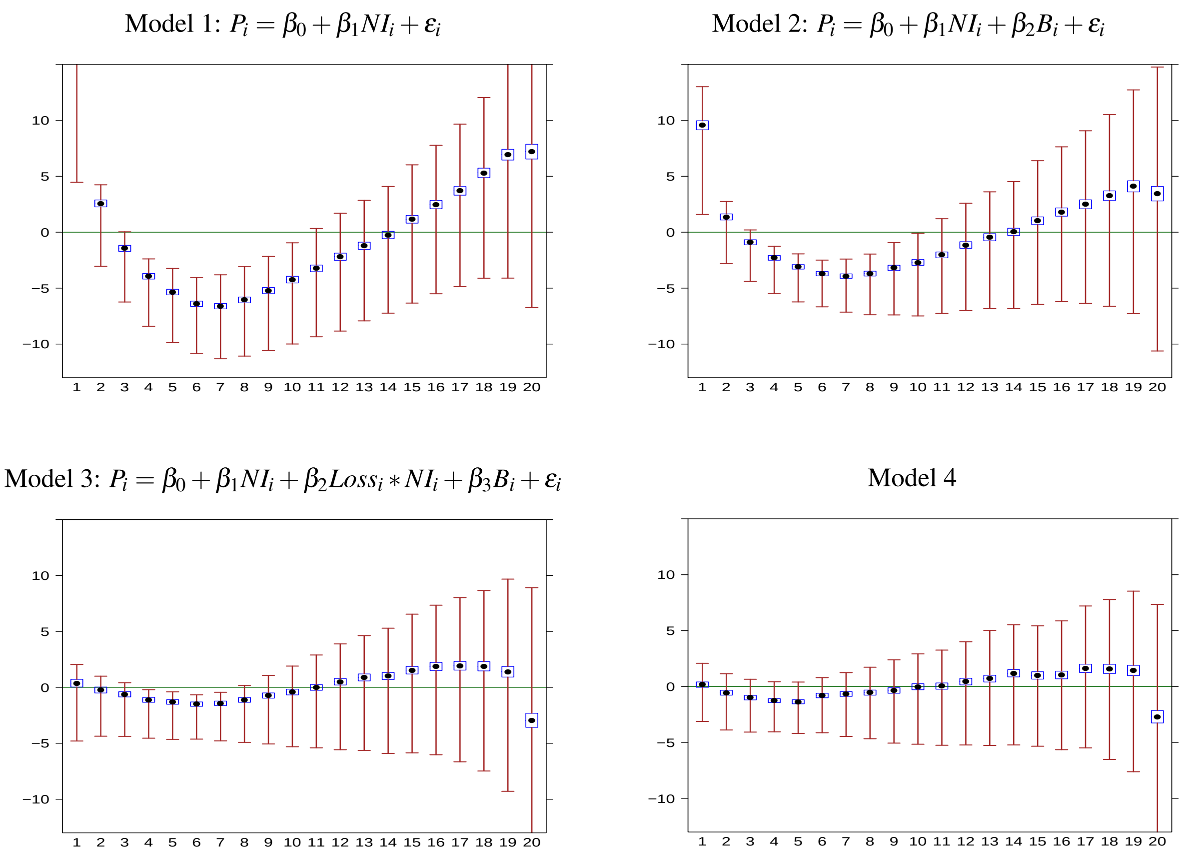


Figure 1: **Distribution of residuals of models (1) through (4), conditional on the level of NI.** The graphs display the 25th and 75th percentiles (the lower and upper ends of the whiskers, respectively), together with the mean (dot) and its 95% confidence interval (box) of the cross-sectional estimated residuals of the linear model (y-axis) conditional on ventiles of the level of NI (x-axis). The linear models are estimated cross-sectionally. For a given ventile, all the corresponding cross-section residuals are used to construct the three statistics (25th and 75th percentiles and mean) displayed above the ventile number. If the relations between the dependent and the independent variables in the models are linear, the 20 conditional means should be approximately zero.

tionalization of X being ”close to” x (see section 4.1). For a suitable model of the economic price–earnings relation, each of the 20 conditional means should be (approximately) equal to zero. None of the four models is close to satisfying this condition. While allowing for numerous interactions (for example, specification (4) infers 55 parameters) reduces the distance of the conditional expectations to zero, the gap remains (strongly) significant.

Intuitively, the graphs in Figure 1 are evidence of the fact that the lines fit by the models do not pass through the ”core” of the data; for some levels of earnings the prices are mostly above the line fitted by the models, while for other levels they are mostly below.

The fact that a linear model provides a poor approximation to the of interest (as the graphs in Figure 1 highlight for the case of the price–earnings relation) can lead to unreliable hypothesis testing (using the linear model). In section 4.3 we give concrete examples of such problems, which, for the sake of simplicity, are constructed using the Ohlson [1995] specification (2). Note that similar conclusions remain true for all models whose residuals violate the orthogonality property (6).

4.3 Econometric consequences of lack of fit of price–earnings linear models

A linear model fit to a non-linear relation runs the risk of being econometrically ineffective. Even small variations in the sample composition, without any changes in the economic relation, may cause the coefficients of the model to move significantly. In such a case, taking the linear inference at face value would result in a (false) rejection of the hypothesis of equal economic relation. This is because the linear projection may be a poor approximation to the conditional expectation function.

In this section, we test the stability of the models by making small changes to the sample. We remove 2.5% of observations, which could simulate an effect of using different sources of data (e.g., different databases), or of limiting the sample due to including control variables with data restrictions, which is common in empirical research.[11]

In a model that correctly specifies the economic relation, the errors are, for all levels of earnings, equally positive and negative. In such a case, removing a part of the sample does not affect the stability of the line, i.e., the slope does not change. In contrast, for a model where residuals are as shown in Figure 1, removing a part of the sample in which the errors are mostly on one side of the line will considerably tilt the regression line and significantly impact the estimation. In other words, the slope of the model will be different, although the economic relation is unchanged.

For example, assume that we split the data into parts—part 1 and part 2—which have different marginal distributions for the regressor X: part 1 has a lower mean value of X than part 2. For clarity of exposition, the example is graphically presented in Figure 2. If the shape of the conditional expectation function is sufficiently different between the ranges of X covered by the two halves of the data, the two linear projections estimated on the two halves will provide distinctly different approximations to the conditional expectation function. That is, fitting a linear model on a sample consisting of observations with the X variable mainly in part 1 will yield

[11] We want to emphasize that we are not removing outliers, i.e., pairs that do not represent the economic relation, but "normal" observations that are expressions of the relation to be estimated.

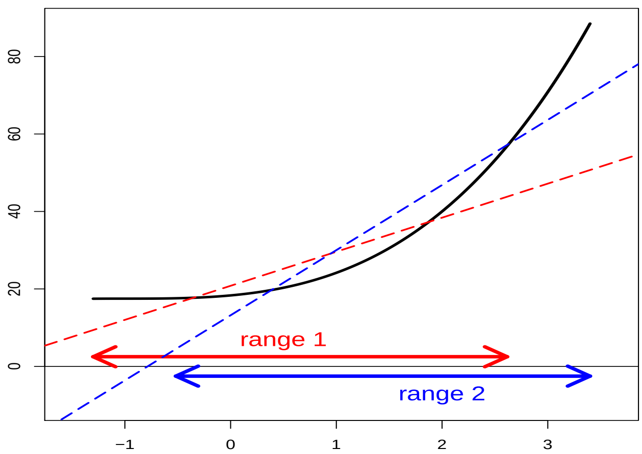


Figure 2: **Linear regression may be a poor approximation to the conditional expectation function.** The black curve is the conditional expectation function, the true economic relation between X and Y. We consider two sub-samples: the (X,Y) pairs with X in range 1 (marked with solid-line red arrows) and in range 2 (solid-line blue arrows). The shapes of the conditional expectation function in the two ranges of X are sufficiently different. Two linear projections are estimated on two subsets yielding distinctly different approximations to the conditional expectation function.

a model that is statistically different from the result of a linear estimation based on a sample of observations with the X variable mainly in part 2.

Taking the linear regression at face value, differences in the marginal distribution of the regressor X will be interpreted (due to the poor approximation of the linear model to the economic relation) as differences in the economic relation between the components of the pair (X, Y). The conclusion is incorrect because, in fact, there is no difference in the conditional mean function. The apparent difference is a by-product of a (poor) linear approximation to a non-linear conditional mean, combined with different marginal distributions for the conditioning variables. In the sequel we give examples of the econometric impact of non-linearity on testing the price-earnings economic relation using linear model approximations. Section 9.2 in the appendix further discusses how non-linearity can interact with omitted determinants to give incorrect results in hypothesis testing based on linear models.

Consequences of using linear models in the presence of non-linearity. Next we show how small changes in the sample induce significant changes in the coefficients of a linear model used to approximate a non-linear relation.[12] Accepting these coefficients at face value would lead to the erroneous conclusion that two samples issued from the same data generating process

[12]The same changes in a sample that expresses a linear relation would yield statistically identical coefficients.

are expressions of different economic relations.

Figure 3 presents an example of such a situation.[13] The graphs plot prices against *NI* (on the *x*-axis) for two sub-samples from the year 2015.[14] The two sub-samples were constructed as follows. First, we sorted the pairs in the cross-section in increasing order of the *NI* variable. Then, the first (second) sub-sample was obtained by removing every fourth observation among the 10% of pairs with lowest (highest) *NI*. The removed pairs are marked in the figure with squares, while the pairs in the sub-samples are marked with circles. The squares hence mark the pairs with the first, fifth, ninth, etc. largest (smallest) *NI* (down/up to the 10% threshold) in the 2015 cross-section. The marked sets each represent 2.5% of the sample. The two sub-samples are expressions of the *same economic relation*. A statistical model that correctly reflects this fact should not reject the hypothesis of no structural difference between the two sub-samples.

2015: First sub-sample: $\hat{\beta} = 6.82$ (0.18) Second sub-sample: $\hat{\beta} = 5.36$ (0.17)

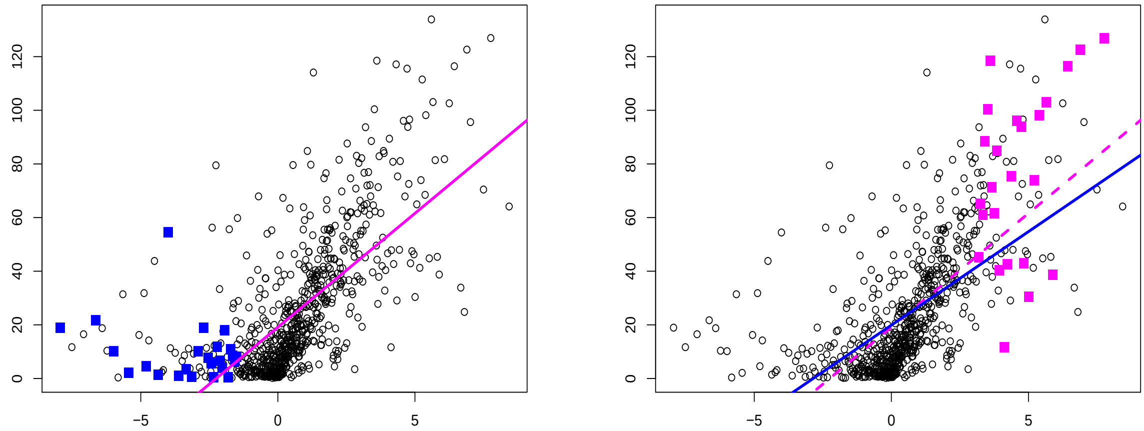


Figure 3: **Non-linearity confounds the linear regression.** The graphs display prices (on the *y*-axis) against *NI* (on the *x*-axis) for two sub-samples from the year 2015 constructed as follows. First, we sorted the pairs in the cross-section in increasing order of their *NI* variable. Then, the first (second) sub-sample was obtained by removing every fourth observation among the pairs with 10% lowest (highest) *NI*. The removed pairs are marked in the graphs with blue (red) squares. The circles mark the pairs in each of the two sub-samples. The marked sets each represent 2.5% of the sample. The two sub-samples are expressions of the same economic relation. The price–earnings relation of the estimated linear model (2) is displayed on each of the sub-samples. The dotted line in the second graph is the regression line estimated on the first sub-sample and differs significantly (see the estimates reported in the titles of the graphs) from the relation estimated on the second sub-sample. The standard error is shown in parenthesis after the values of β.

We estimate the linear price–earnings regression (2) on each of the sub-samples and display the two regression lines (the second graph displays both of them). Although the two sub-

[13]Since studies of the price–earnings relation involve cross-sectional estimation, we show examples of results for individual years. As noted later in this section, the years shown are representative of a large subset of the years in the period 1970–2016 (see Figure 4).

[14]We have already removed the most extreme 2% of observations (1% of each sign) for both the dependent and independent variables.

samples are expressions of the same economic relation, the two estimated linear models are statistically different. The t-statistic testing the hypothesis of equal slope is 2.83. Note also that the result is not due to outliers as we keep three fourths of the more extreme observations in each of the two sub-samples and that we have already removed 2% of the most extreme observations.

This example is pertinent to a situation where two researchers analyze the same cross-section but end up with slightly different sub-samples of firms due to different manipulations of the data used (e.g., through inclusion of different control variables). The differences we induce in the example in Figure 3, while innocuous when the economic relation between price and earnings is linear, interact with the strong non-linear pattern in the data and cause the mis-specified linear model to reject the true hypothesis.

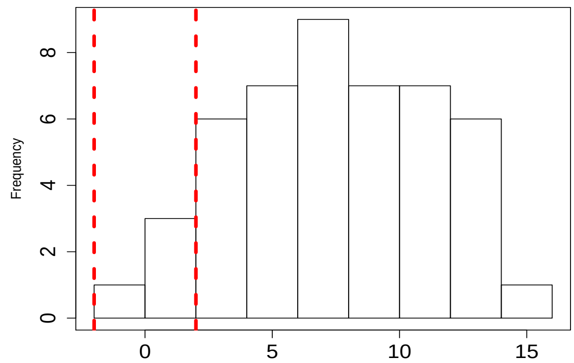


Figure 4: **Non-linearity confounds the linear regression—all cross-sections.** The graphs display the histograms of the t-statistics (on the x-axis) that test the equality of the regressions' earnings response coefficients from the two regressions (2) performed on pairs of yearly cross-sectional sub-samples (from 1970 to 2016) constructed according to the description in the caption of Figure 3. Values of the t-statistic outside the interval $[-2,2]$ correspond to rejection of the null of equal economic relation at a 95% confidence level. The graphs indicate that for a large number of cross-sections, a test based on the linear representation of the Ohlson model would wrongly reject the null.

We extend the analysis in Figure 3 to all yearly cross-sections in the sample and display the results in Figure 4. This figure presents the histograms of the t-statistics that test the equality of ERCs for two linear levels price–earnings regressions (2) performed on yearly cross-sectional sub-samples that are constructed according to the description above. Values of the t-statistic outside the interval $[-2,2]$ correspond to rejection of the null of equal economic relation at a 95% confidence level. The figure indicates that, for a large number of cross-sections, a test based on the linear representation of the Ohlson [1995] model would wrongly reject the null hypothesis (it is clear from the figure that 2015 is an average year). The graphs in Figure

4 clearly indicate that even small changes in the sample cause significant instability in the statistics that are usually used for testing hypotheses about the economic relation between prices and earnings.

The problems with linear models that we noted in section 4.3 are potentially further exacerbated if there is an omitted variables issue. As noted in section 1, existing research documents a cross-sectional relation between ERCs from and risk (discount rates), growth opportunities, and economic rents (Holthausen and Watts [2001]; Kothari and Shanken [2003]). Section 9.2 in the appendix documents a behavior of the linear model similar to that documented in this section also in the case even when the sample is affected by changes in the range of important determinants of the price-earnings relation, like P/B and market value (but not in the range of NI).

5 Machine learning meets accounting

In this section we explain how statistical inference methodology—specifically recently developed machine learning methods—can provide a consistent estimation of the price–earnings relation.

Our goal is to infer the association between share prices and disclosed earnings in cross-sections. For a given firm i in the cross-section (year) 0, the general decomposition in equation (5) (see section 4.1) applied to the pair $(P_{i,0}, NI_{i,0})$ guarantees the existence of a regression that relates prices $P_{i,0}$ to earnings $NI_{i,0}$:

$$P_{i,0} = f_{i,0}(NI_{i,0}) + \varepsilon_{i,0}, \quad \text{where} \quad f_{i,0}(x) := \mathbb{E}[P_{i,0}|NI_{i,0} = x], \tag{9}$$

and the adjustment ε verifies the orthogonality property (6). Consequently, the non-linear regression function

$$x \to f_{i,0}(x) := \mathbb{E}[P_{i,0}|x],$$

can be inferred without bias (using methods from the non-linear regression estimation literature; see Hastie et al. [2009]) provided we have many similar observations of the price–earnings relation of which the pair $(NI_{i,0}, P_{i,0})$ is an expression.

However, it is intuitively clear that not all pairs $(NI_{j,0}, P_{j,0})$ in the cross-section 0 will be an expression of the relation of price to earnings realized in the specific pair $(NI_{i,0}, P_{i,0})$. The fact that the function f in the regression (9) bears the index $(i,0)$ formalizes the observation that the relation of prices to earnings depends on specific firm characteristics (e.g., in terms of risk, growth opportunities, and economic rents; see section 1).Hence, we cannot estimate the regression (9), as it is, on the entire cross-section.

Nonetheless, an unbiased estimation of the price–earnings relation for the firm i is possible, conditional on identifying a subset of firms j in the cross-section displaying a similar relation of prices to earnings, that is, the subset

$$S_i = \left\{\text{all firms } j \text{ in cross-section 0 for which } f_{j,0} \approx f_{i,0}\right\} \tag{10}$$

S_i includes all firms j for which the regression function $f_{j,0}$ in equation (9), describing the relation of prices to earnings, is similar to the regression function of the entity i. Once we construct the set S, we estimate the non-linear regression in equation (9) only on the sub-sample of pairs $(P_{j,0}, NI_{j,0})$ that correspond to the firms j in the set S_i. All these pairs are expressions of the same price–earnings relation, which can be estimated without bias. All other pairs are expressions of different price–earnings relations and including them in the sub-sample on which we perform the estimation would be equivalent to mis-specifying the model. In sections 5.1, 5.2, and 5.3 we show how we use the existing accounting and finance literature to identify the subset S. Section 5.4 explains how we operationalize their construction.

5.1 From a valuation model to a statistical regression specification

Assume that prices can be expressed as sums of discounted expectations (formed at time 0) of functions of future earnings (or other measures of performance, e.g., cash flows) and the price of equity risk:

$$P_{i,0} = \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0})]}{(1+r_{i,0})^t} \tag{11}$$

where $\mathbf{NI}_{i,t} := (NI_{i,1}, NI_{i,2}, \dots, NI_{i,t})$, $\mathbf{O}_{i,t}$ represents variables other than earnings, and $r_{i,0}$ denotes the risk-adjusted expected return on equity for firm i at time 0 while $\mathbb{E}_0$ is the market's expectation conditional on all information available at time 0. Concrete examples of valuation models include the residual income (RI) model and the Ohlson and Juettner-Nauroth (OJ) model (see section 9.4 in the appendix for details).

5.2 The (non-linear) price–earnings regression function

The modeling assumption in equation (11) allows us to make the general non-linear function in equation (9) more specific. More precisely, we show (see the details in section 9.7 in the appendix) that:

$$f_{i,0}(x) = \sum_{t=1}^{\infty} \frac{\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_{i,0}) \,|\, NI_{i,0} = x]}{(1 + r_{i,0})^t}.$$

Let us compare this expression to the model assumption (11). The two expressions are structurally the same and differ only through the information available to the investor when forming expectations about the future income stream. In model (11), we condition with all information available at time 0 (the price is a sum of conditional expectations of the type $\mathbb{E}_0[\cdot]$), while in the expression above, the conditioning set is restricted to the level of disclosed earnings of the firm at time 0 (the function $f_{i,0}$ is a sum of conditional expectations of the type $\mathbb{E}[\cdot \,|\, NI_i = x]$). It follows that the expression above (which is the regression function in (9)) is also a valuation. More concretely, $f_{i,0}$ is a valuation incorporating expectations shaped *only* by the current level of earnings of firm i.

Consequently, we can restate the relation (9). For any firm i, the price at time 0 can be written as:

$$P_{i,0} = \sum_{t=1}^{\infty} \frac{\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_{i,0}) \,|\, NI_{i,0}]}{(1 + r_{i,0})^t} + \varepsilon_{i,0}, \tag{12}$$

where $\mathbb{E}[\varepsilon | NI] = 0$, i.e., price is the sum of

- a valuation based on expectations of future earnings informed only by the current level of earnings of the firm (the value of earnings):

  $$\sum_{t=1}^{\infty} \frac{\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_{i,0}) \,|\, NI_{i,0}]}{(1 + r_{i,0})^t}$$

  and

- an orthogonal adjustment by investors that reflects all information (other than the level of earnings) available to investors:$\varepsilon_{i,0}$.

5.3 The determinants of the regression function $f_{i,0}$

As pointed out in section 5, the decomposition in equation (12) is firm-specific. The function

$$x \to \mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_i)\,|\,x], \tag{13}$$

appearing in the denominator of the summand in (12) is the best prediction of (or function of) the future earnings stream based on knowledge of the current levels of earnings. In other words, it describes how the current level of earnings projects the future earnings (or alternative measures such as cash flows) stream. The function (13) is shaped by the levels of the main determinants of the firm-specific relation between prices and earnings: risk, growth opportunities, economic rents and accounting conservatism. In this section, we take an exploratory and broad approach, as we attempt to identify, based on existing accounting and finance literature, classes of variables that potentially determine the form of function (13) and hence the price–earnings relation. The variables are grouped into three categories: risk and growth, industrial organization (IO), and accounting. In section 1, we noted that the firm-specific relation between prices and earnings depends on risk, growth opportunities, economic rents, and accounting conservatism, which are proxied by the main variables price-to-book, market value, and leverage. In this section, we take an exploratory and broad approach, as we attempt (based on extant accounting and finance literature) to identify a large number of variables that potentially affect the price–earnings relation. In total, 45 variables are identified; they are introduced and justified in section 9.8 in the appendix. In section 6, we present empirical results on the importance of these variables.

Risk and growth. According to the valuation literature, the main determinants of the evolution of future earnings streams and, hence, of the shape of the function (13), are a firm's cost of equity $r_{i,0}$ and growth $g_{i,0}$. In addition to well-known risk and growth factors such as price-to-book ratio and size, payout policy and financing variables are also associated with the risk and growth profile of a firm (Beaver and Ryan [2005]; Modigliani and Miller [1958]; Taggart [1991]; Skogsvik et al. [2012]).

More recently, the capital markets literature has emphasized the importance of investment and profitability proxies for valuation and market performance (Chen et al. [2011]; Novy-Marx [2013]; Fama and French [2015]; Ball et al. [2016]). Fama and French [2015] argue that "each stock's relevant expected return is determined by its price-to-book ratio and expectations of its future profitability and investment" (p. 2).[15]

Industrial organization framework. The economics and strategic management literature has studied how industry structure determines the profit-generating processes of firms. It provides a theoretical framework for the evolution of profits and explains its variation through differences in the competitive environment.

Accounting. The quality of financial statements—for example accruals quality—influences the ability of accounting numbers to serve as proxies for the economic concepts identified above.

To summarize, we identify three classes of (not necessarily disjoint) determinants[16] of the regression function f_i in (12): risk and growth determinants, IO (or economic) determinants, and accounting determinants.

Extant literature often uses industry as a comprehensive proxy for the risk and growth profile of a firm (which is clearly related to industry-specific factors: industry concentration, barriers to entry, product type, market share, etc.).[17] Our aim goes beyond that as we attempt to base our analysis on (potentially firm-specific) factors that explain differences within industries.

In total, we consider 45 variables that we individually justify and present in section 9.8 in the appendix. While the list of pertinent variables is based on extant accounting and finance literature, the choice of the most appropriate proxies is, in the end, an empirical question, which we address in section 6.1 and in section 9.8 in the appendix.

5.4 How to incorporate the constraint S into statistical estimation

Based on the discussion in section 5.3, we assume that firms with similar risk (r) and growth (g), IO/economic (E), and accounting (A) determinants have similar regression functions (9).

[15]If variables not explicitly linked to the decomposition of the market-to-book ratio as a function of expected return, expected profitability, and expected investment (such as size) help forecast returns, "they must do so by implicitly improving forecasts of profitability and investment or by capturing horizon effects in the term structure of expected returns" (Fama and French [2015], p. 2).

[16]The risk and growth profile of a firm is, for example, clearly related to industry-specific factors (industry concentration, barriers to entry, and product type).

[17]In section 4.2 we evaluate a linear model where slopes of earnings and book values are allowed to vary by industry.

Consequently, a possible specification of the set S in equation (10) for a firm i belonging to the cross-section (year) 0 is given by:

$$S_i = \{\text{all firms } j \text{ s.t. } r_{j,0} \approx r_{i,0},\ g_{j,0} \approx g_{i,0},\ E_{j,0} \approx E_{i,0},\ A_{j,0} \approx A_{i,0}\}.$$

As explained in the beginning of section 5, to consistently estimate the price–earnings relation for firm i, we need to perform regression (9) locally, on the set S_i. Our method of choice for the estimation of the relation between stock prices and earnings, the random forest (RF) algorithm, groups observations with similar determinants, implicitly constructing the sets S_i, and then use their dependent variable to locally infer the regression function. Below, we give a more detailed explanation of exactly how this estimation method operationalizes the construction of the sets S_i. An RF aggregates the estimates of a collection of classification and regression trees (CART) (for details, see Hastie et al. [2009]; James et al. [2013]). Tree-based methods (such as CART and RF) have a number of important advantages. First, their functioning is consistent with the discussion in section 5.3. They empirically group observations that behave similarly to each other. Consequently, the relation between stock prices and earnings for firm i is estimated based only on those firms that are similar (to the firm i) in terms of relevant determinants (the other explanatory variables). Second, they can approximate potentially severe nonlinearities. In particular, these methods naturally incorporate multiway predictor interactions, yielding a machine learning solution to a long-standing issue difficult to address with linear models.

Trees are fully non-parametric structures constructed with a logic that departs significantly from that of traditional parametric regressions. A tree grows sequentially, with a new branch added at each step. Each new branch separates the data remaining from the previous step into parts, based on one of the independent variables. As a result, the space of the determinants is partitioned into rectangles. The method approximates the unknown regression function (relating the independent variables to the dependent one) in each partition by the average of the outcome variables within each rectangle.

Figure 5[18] shows an example with two price determinants, say, $X_1 = NI$ and $X_2 = P/B$. The left graph shows how the tree assigns each observation to a partition based on its predictor values (X_1, X_2). First, observations are sorted on earnings per share: those with NI above t_1 are assigned to one partition, and those with NI below t_1 to another. Those with small earnings per

[18] The graphs in Figure 5are reproduced from Hastie et al. (2009).

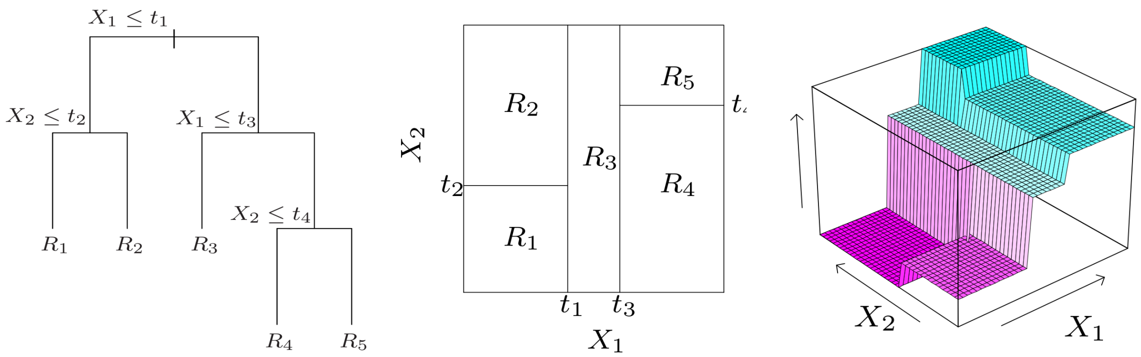


Figure 5: **Tree-based non-parametric estimation at work.** The left and center panels show the partition of a two-dimensional explanatory variable space by recursive binary splitting, as used in CART. The right panel shows a plot of the estimated regression function, which is locally constant on each of the five regions $R_1, R_2, \ldots, R_5$ of the explanatory variable space (reproduced from Hastie et al. (2009)).

share values ($X_1 \le t_1$) are then further sorted by their P/B. Observations with small earnings and $P/B \le t_2$ are assigned to region R_1, while those with small earnings and large P/B go into region R_2. The process continues similarly for the data points with large earnings ($X_1 > t_1$), which are further split based on the level of earnings. Those with earnings less than t_3 form the region R_3. The splitting continues for the observations with particularly large earnings ($X_1 > t_3$), yielding two regions R_4 ($X_1 > t_3$ and $X_2 \le t_4$) and R_5 ($X_1 > t_3$ and $X_2 > t_4$) according to the magnitude of their P/B. The choice of sorting variable from the set of predictors and of the corresponding split value maximizes the distance between average outcomes in each of the two new partitions. The newly constructed split best discriminates among the potential outcomes.

The graph in the middle shows the split of the predictors' space (determinants of the economic relation between prices and earnings) in the five regions produced by the sequence of binary partitions represented in the left graph. The graph on the right displays a (fictive) estimate of the non-linear function f representing the economic relation of the pair of determinants $(X_{1i} = NI_i,\ X_{2i} = P/B_i)$ with the dependent variable $Y_i = P_i$:

$$Y_i = f(X_{1i}, X_{2i}) + \varepsilon_i,$$

which generated the partition represented in the first two graphs. The vertical arrow shows Y (or P). Each region of the partition in the middle graph consists of firms with determinant pairs

$(NI,\ P/B)$ that relate similarly to the price P.Consequently, for $x \in R_k$, f is estimated as:

$$\hat{f}(x) := \frac{1}{n_k} \sum_{\{j\,:\,(X_{1j},\,X_{2j}) \in R_k\}} Y_j,$$

where n_k is the number of firms with the determinants $(X_1 = NI,\ X_2 = P/B)$ in region R_k. The graphs illustrate how interactions are modeled by trees. Let us consider the relation of $X_1 = NI$ to $Y = P$ for two different values (large and small) of $X_2 = P/B$. Graphically, we compare two sections (we vary X_1, while holding X_2 fixed) through the surface represented in the graph under discussion: one section near to the origin, the other at the extremity of the x_2-axis. We notice that in both cases, price is an increasing function of earnings. However, we can also see that the slope of the relation depends on the level of the P/B ratio: firms with higher P/B have higher slopes, an indication that high growth is positively impounded in prices.

Due to their flexibility, trees are prone to overfit. In our analysis, we address this concern by using RF, which combine estimates from many different trees. In concrete terms, our procedure draws a large number of bootstrap samples of the data, fits a CART regression tree to each of them, and then averages their estimates of the regression function. Trees for individual bootstrap samples tend to overfit, i.e., the estimated function varies excessively. Averaging over multiple estimations reduces this variation and eliminates the overfit.

Putting together the above explanation with the specification of the set S above, we estimate the price–earnings association by adding to the explanatory variable $NI_{i,0}$ in equation (9), proxies for firm's cost of equity $r_{i,0}$ and growth $g_{i,0}$, for its levels of economic $E_{i,0}$ and accounting $A_{i,0}$ determinants. In other words, we cross-sectionally estimate the non-linear regression:

$$P_{i,0} = f_0\Big(NI_{i,0},\ r_{i,0},\ g_{i,0},\ E_{i,0},\ A_{i,0}\Big) + \varepsilon_{i,0}, \tag{14}$$

where the error term $\varepsilon_{i,t}$ satisfies the orthogonality property

$$\mathbb{E}_t[\varepsilon_i \,|\, NI_i] = 0.$$

The functioning of the non-parametric regression explained above yields another intuitive interpretation of the regression function in (14), familiar to the reader used to the linear regression framework. Non-linearly regressing prices $P_{i,0}$ on the vector $(NI_{i,0},\ r_{i,0},\ g_{i,0},\ E_{i,0},\ A_{i,0})$

amounts to estimating the association between prices and earnings *while holding constant* the levels of risk, growth, economic determinants, and accounting determinants.

In other words, the non-parametric estimation infers the price–earnings association controlling for the levels of the factors that determine how the current earnings of a firm translate into future income. In addition, no *ex ante* knowledge of the relative importance of the explanatory variables is required, as the RF identifies those variables that lead to the largest reduction in mean squared error.

5.5 A non-linear research design alternative

The non-linear research design yields an estimate of the multidimensional economic relation between prices, on one hand, and earnings and the other determinants (risk, growth, economic and accounting proxies) of future earnings in model (11), on the other. This relation is expressed by the estimated regression function:

$$\big(NI,\, r,\, g,\, E,\, A\big) \rightarrow \widehat{f_0}\Big(NI,\, r,\, g,\, E,\, A\Big). \tag{15}$$

The expression in (15) is a function of five variables describing the joint dependence of prices on earnings and proxies for risk, growth, etc., which needs to be summarized to be interpretable. We solve the interpretability issue with the help of two measures: the *individual conditional expectation* (ICE) curves and the *partial dependence of prices on earnings* (PD). In section 7 we provide an example of how to apply RF, and there the results are interpreted by a third measure: relative valuation error.

In the case of a linear model, the price–earnings relation while the other determinants are kept constant is linear and hence described by one coefficient, a slope (plus an intercept). In the non-linear case, the same relation is described by a non-linear PD function. By definition, it is the average of the firm-specific price–earnings relation over all firms in the cross-section.

The firm-specific relation of prices to earnings is captured by the ICE curves. An ICE curve can be thought of as a simulation that shows what would happen to a given firm's stock price if only the level of earnings of the firm varied, while the other determinants remained constant (at the values they take for the given firm). The simulation is based on the estimated model (15). For more details about how ICE curves are constructed, see Goldstein et al. [2015].

Formally, the firm-specific association of earnings to prices for the firm i in cross-section 0 is given by the *ICE function*:

$$x \to \widehat{f}_0\Big(x;\, r_{i,0},\, g_{i,0},\, E_{i,0},\, A_{i,0}\Big), \tag{16}$$

where $r_{i,0},\, g_{i,0},\, E_{i,0}, and\, A_{i,0}$ are the values of firm i's determinants in cross-section 0. An ICE plot visualizes the dependence of price for each firm in a cross-section separately, by plotting the functions in (16),resulting in one line per firm.

Consequently, for a given level of earnings x, the dependence of price on earnings only is defined as the average of the ICE curves over all firms in the cross-section 0 yielding price's *partial dependence* function of earnings:

$$\widehat{f}_0^{(NI)}(x) := \frac{1}{n}\sum_{i=1}^{n} \widehat{f}_0(x;\, r_{i,0},\, g_{i,0},\, E_{i,0},\, A_{i,0}). \tag{17}$$

Examples of ICE plots are presented in section 6.3, where the ICE functions are used to define firm-specific ERCs. The partial dependence function appears prominently in section 6.4 and serves as a basis for the definition of a cross-sectional (non-linear) ERC.

6 Application of the non-linear research design

In this section, we apply the non-linear method presented in section 5 to the inference of the price–earnings relation. We begin by identifying which determinants have the largest effect on the relation. We validate the method by showing that the orthogonality property is met and present findings that agree with existing accounting theory in the literature.

6.1 Choice of proxies for $r_{i,0},\, g_{i,0},\, E_{i,0},\, A_{i,0}$

The first step in applying the non-linear research design is the choice of proxies for risk, growth, economic determinants, and accounting determinants. Informed by extant accounting and financial literature we put together an extensive list of proxies (described in section 9.8 in the appendix). The choice of the most appropriate variables is empirically determined in the non-parametric set-up. More precisely, we use RF to identify those variables that lead to the largest reduction in mean squared error.

Figure 17 in section 9.8 displays the importance of each of the 45 explanatory variables we consider and shows a clear ordering. Of importance is the percentile increase in mean square error when the values of a given variable are randomly shuffled, removing the association that possibly exists between that variable and the dependent variable. As expected, NI leads the ranking, with an importance of 34%. P/B is second, with an importance of 22%, followed by size ($Mktv$) at 18%. Thereafter, measures of financing and profitability are among the most relevant, while proxies for investment and accounting are the least relevant to the price–earnings association. However, we see that for most firm characteristics, the increase in the mean square error caused by their absence in explaining prices is rather modest (under 10%).

The literature suggests that the relation between prices and earnings varies by risk, economic rent, growth opportunities, and accounting conservatism (Holthausen and Watts [2001]; Kothari and Shanken [2003]; Liu and Thomas [2000]; Biddle et al. [2001]). As noted in section 1, the variables that are empirically shown to be most important—P/B, $Mktv$, and *Leverage*—are proxies for the expected determinants.

Based on the findings and for the sake of parsimony, we select those proxies with an importance greater than 10%, i.e., P/B, $Mktv$, and *Leverage*, as the firm characteristics that will inform the shape of the price–earnings relation. We do not include TA in the set as it has a (Spearman) correlation of 0.91 with $Mktv$. Results are qualitatively similar if we use the whole set of 45 variables (although the sample reduces drastically).

We recall that the role of the proxies is to assist in grouping the firms with similar determinants for the price–earnings relation. To emphasize this role and reduce their direct explanatory effect, we do not use the actual values of the proxies in the inference of regression (14). Instead, in every cross-section we order the firms according to their P/B ratio, $Mktv$, or *Leverage*, and use the centile to which the firm belongs, denoted by $rank(P/B)$, $rank(Mktv)$, and $rank(Leverage)$, instead of the actual values of the three firm characteristics. In this way, it is only the relative standing of the firms within the cross-section that informs the regression, and not the actual values of the firm characteristics.

We hence estimate the following non-linear regression:

$$P_{i,0} = f_0\Big(NI_{i,0};\ rank(P/B)_{i,0},\ rank(Mktv)_{i,-1},\ rank(Leverage)_{i,0}\Big) + \varepsilon_{i,0}, \tag{18}$$

where f is a non-linear function that we will estimate using the RF algorithm. The regression is the basis for all empirical results in this section 6 and in section 7.

6.2 Validation 1: Evidence of the fit of the non-linear estimation

In this section we present evidence that, in contrast to the linear model, the non-linear approach described in section 5 provides an exact description of the economic price–earnings relation. First, the mean of the residuals conditional on the independent variables being approximately zero (for the linear model, see section 4.2.2). Second, the model discriminates between differences in the marginal distribution and differences in the economic relation (for the linear model, see section 4.3). Estimation on sub-samples that differ in the range of earnings yields identical pictures of the price–earnings relation. Figure 6 displays the distribution of the estimated resid-

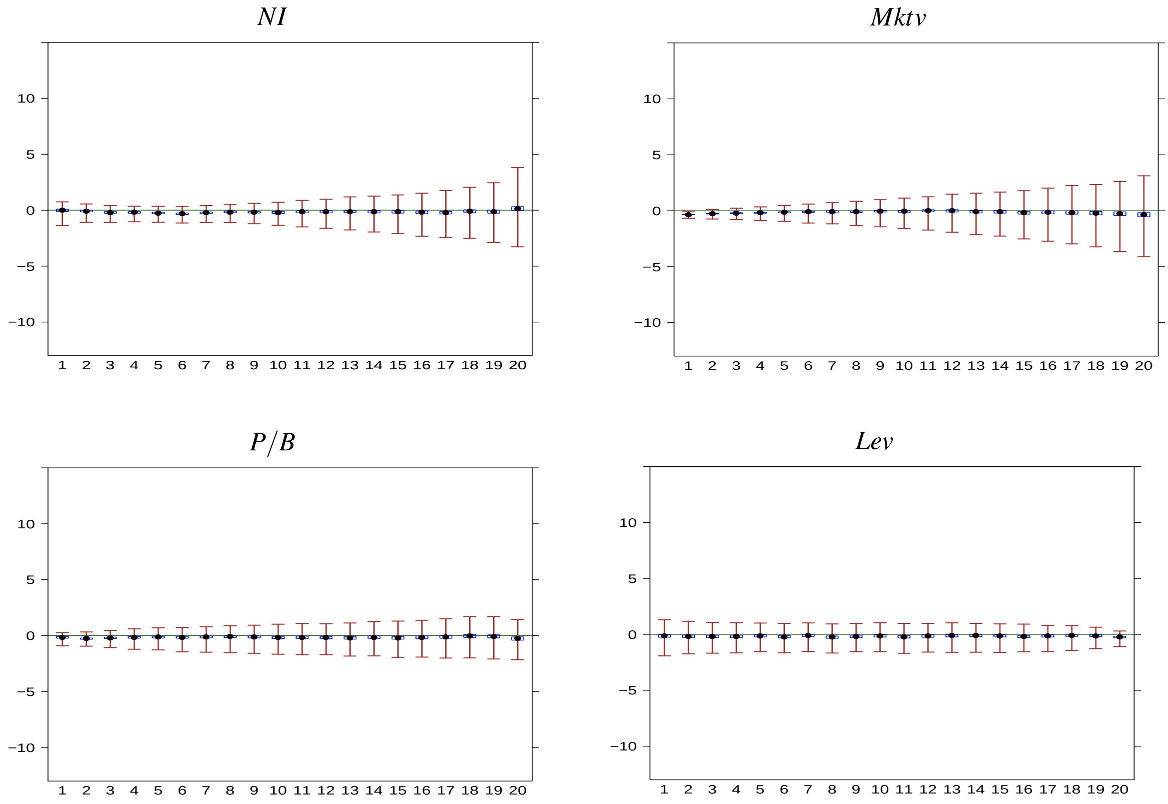


Figure 6: **Distribution of residuals of specification (18) conditional on the explanatory variables.** The graph displays the 25th and the 75th percentiles (the lower and upper ends of the whiskers, respectively) together with the mean (dot) and its 95% confidence interval (box) of the cross-sectional estimated residuals of the non-linear model (18) conditional on the ventiles of the level of *NI* (top left), *Mktv* (top right), *P/B* (bottom left), and *Leverage* (bottom right). The non-linear model (18) is estimated cross-sectionally and the firms are divided into 20 equal groups (ventiles) according to the level of each explanatory variable. For a given ventile of firms, all the corresponding cross-section residuals are used to construct the three statistics (25th and 75th percentiles and mean) displayed above the ventile number. If the underlying economic relation is well approximated by the model (18), all conditional means should be approximately zero. For ease of comparison, the scale of the graphs is the same as in Figure 1 (for the linear model).

uals of specification (18), conditional on the explanatory variables. The graphs show the 25th and 75th percentiles (the lower and upper ends of the whiskers, respectively) together with the mean (dot) and its 95% confidence interval (box) of the cross-sectional estimated residuals of the non-linear model (18), conditional on the ventiles of the level of *NI* (top left), *Mktv* (top right), *P*/*B* (bottom left), and *Leverage* (bottom right).

In contrast to Figure 1, the graphs in Figure 6 confirm the fit of the non-linear estimation. All four conditional distributions of the residual of the specification (18) satisfy, approximately, the orthogonality property (6).

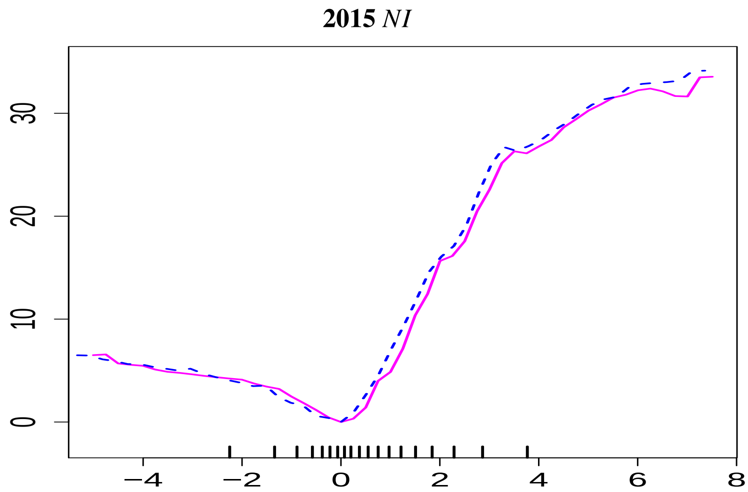


Figure 7: **The non-linear approach addresses the bias in the level regressions.** The graphs display the dependence of prices (on the *y*-axis) on *NI*, estimated by the approach described in section 5 for the pairs of sub-samples from the same cross-section analyzed using the linear models in figure 3. The pairs of sub-samples express the same economic relation between prices and earnings but differ in the ranges of values of *NI* (see the captions of figures 3 for more details). The solid (dotted) line corresponds to the samples obtained by removing firms with lower (higher) values of *NI*. The rug displayed on the *x*-axis corresponds to the ventiles, i.e., the 5th, 10th, ..., 95th percentiles, of the cross-section's distribution of *NI*. We see that the estimated dependence is very similar for the two sub-samples.

Figure 7 displays the partial dependence of price on earnings, $\widehat{f}_0^{(NI)}$, defined by (17), estimated non-linearly for the same pairs of sub-samples from the same cross-sections analyzed using the linear level regressions in section 4.3 (see figure 3). By construction, the pairs of sub-samples express the same economic relation between prices and earnings but differ in the ranges of values of the main variable *NI*. We see that the curves estimated on sub-samples that differ in the range of earnings are practically identical.

Section 9.3 in the appendix further shows that the non-linear model consistently estimates the economic relation also in the case when the sample is affected by changes in the range of important determinants of the price-earnings relation, like P/B and market value (but not in the range of NI).

6.3 Validation 2: Firm-specific dependence of price on *NI*

In this section, we discuss the *individual conditional expectation* (ICE) curves capturing the firm-specific association of earnings to prices, defined by the expression in (16). An ICE plot visualizes the dependence of price on earnings for each firm in a cross-section separately, resulting in one line per firm. Each ICE curve is a simulation (based on the estimated model (18)) that shows what would happen to a given firm's stock price if only the level of earnings of the firm varied, while the other determinants remained constant at the values they take for the given firm.

Figure 8 displays the *ICE* curves for four different cross-sections: 1974, 1983, 1995, and 2013. Each curve depicts the association of prices with earnings for a specific firm in the given cross-section. The graphs in the figure reveal the locally linear structure of the relation between earnings and prices. They give clear evidence that the price–earnings relation is flat for low values of *NI*, concave for high values, and roughly linear in the middle. The ticks on the x-axis mark the deciles of earnings values. The flat initial part corresponds to negative earnings and stretches from the smallest value to around the third decile. From the third to (roughly) the ninth decile, the curves are typically linear, with curves at the bottom (lower prices, i.e., higher risk and lower growth) displaying lower slopes than the curves at the top. The curves finish with another relatively flat (or sometimes slightly increasing) section, with the inflection point often above the ninth decile and stretching to the largest value. The two regions of no or slowly upward sloping shape correspond to the range of earnings with a large transitory component. In this range a one-unit increase in earnings is associated with a small (or no) increase in share price. The thick black line (bordered by yellow) is the mean of all curves (firms) in a given year (see the definition in (17)).

The firm-specific price–earnings relation visible in Figure 8 is consistent with the transitory nature of extreme earnings. The share price association with extreme earnings is smaller than the association with non-extreme earnings. One interpretation of this finding is that the market does not expect extreme earnings to be permanent, so the price adjustment over non-extreme earnings is smaller. When extreme positive earnings are not simply a result of one-time, large gains, the competition in the product market makes a sustained high level of profitability un-

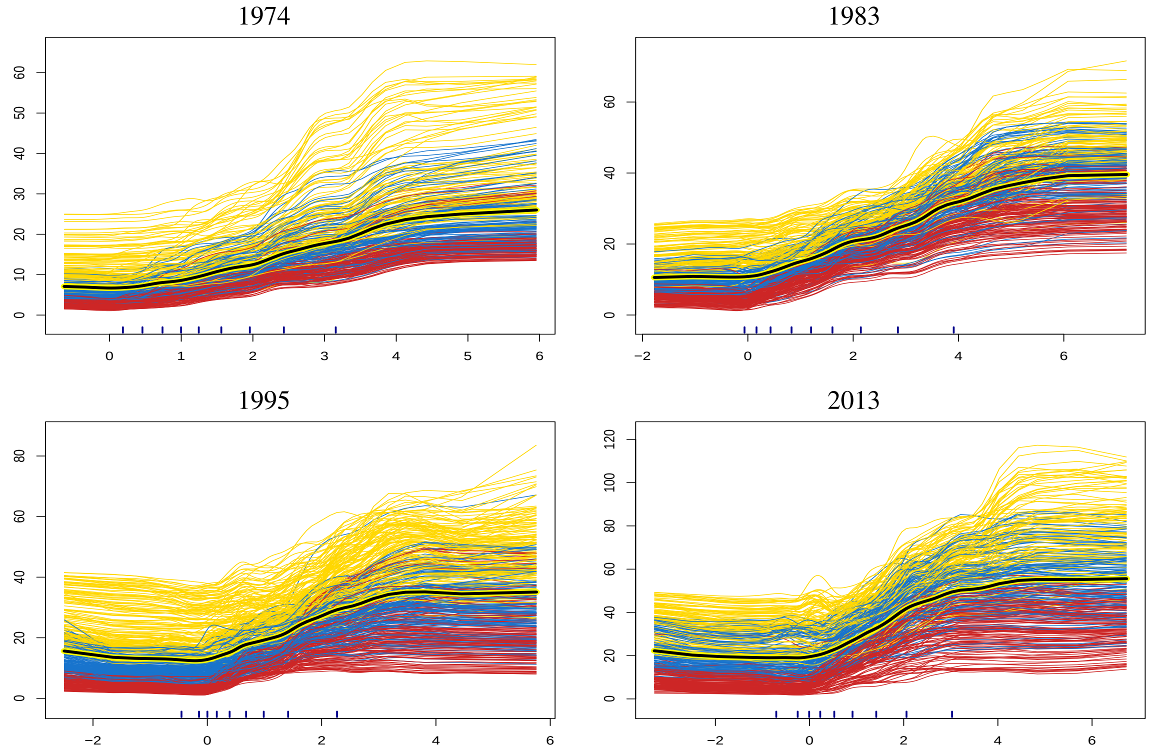


Figure 8: **Individual conditional expectation (ICE) curves.** Each graph displays the estimated functional relation between earnings (NI on the $x-axis$) and prices (on the $y-axis$) for all the firms in the cross-section specified. More precisely, each ICE curve shows what would happen to prices if only the level of earnings of a particular firm in the cross-section varied, while the other determinants remained constant (at their value for the given firm). The ticks on the x-axis mark the deciles. The thick black line is the mean of all the curves and gives a summary of the cross-section. The firms with size below (above) the 33rd (66th) percentile of the size variable in the cross-section are shown in red (yellow). The firms with size in the mid-range of the cross-section are shown in blue. The graphs give clear evidence of the common structure of the functional form of the price–earnings relation for all firms: flat at the beginning and end of the range and roughly linear in the middle. They also demonstrate a clear positive dependence between price level and size (for the same value of disclosed earnings).

likely. At the lower end of the earnings range, it can be argued that losses are likely to be temporary as shareholders can always liquidate the firm rather than suffer indefinite losses (Hayn [1995]; Fischer and Verrecchia [1997]).

The shape of the non-linearity presented by our analysis is consistent with the put option view (Hayn [1995]). Equity holders of loss-incurring firms "have a put option on the future cash flows of the firm whereby they can sell their shares at a price commensurate with the market value of the net assets of the firm" (p. 126). Consequently, only firms expected to improve will continue to operate, implying that observed losses should be temporary. Moreover, in this set-up, the value of the firm's equity is the largest of the present value of its expected earnings and its liquidation value, and negative earnings will not therefore be associated with a decline in share price that is proportional to the size of the loss.

The abandonment option framework in Hayn [1995] yields a simple model in which the firm's value, determined by its earnings, is constantly equal to L, the liquidation value of the firm, up to the level of expected earnings below which the liquidation option is triggered. Above that level, the function of X is linear, with the slope $k * X$, as the earnings are expected to continue in perpetuity. The slope[19] of the linear function k is the earnings response coefficient (ERC).

Consistent with these considerations, the graphs in Figure 8 suggest a straightforward extension of Hayn [1995]'s abandonment option model, which allows for a concave inflection at the positive extreme end of the earnings range, as implied by the assumption of high earnings being temporary (Freeman and Tse [1992]; Core and Schrand [1999]). The functional form of the depicted dependence can be well-approximated by a local linear function with three components (instead of two as in the cited model). Two of these, the first and the third[20], are flat (or close to flat), while the second component has a slope which is firm-specific. Formally, the extension of the abandonment option model looks like:

$$\hat{P}_{i,0}(X) = \begin{cases} L_{i,0}, & X \leq 0, \\ L_{i,0} + k_{i,0} \times X, & 0 \leq X \leq U_{i,0}, \\ L_{i,0} + k_{i,0} \times U_{i,0} + \tilde{k}_{i,0} \times (X - U_{i,0}), & U_{i,0} \leq X, \end{cases} \tag{19}$$

where $X = \mathbb{E}_0[NI_{i,1}]$, the expected earnings next period, and $\tilde{k}_{i,0} \leq k_{i,0}$. The constants $L_{i,0}$ and $U_{i,0}$ and the slopes $k_{i,0}$ and $\tilde{k}_{i,0}$ are functions of growth and risk and are firm-specific. The slope $k_{i,0}$ is the *non-linear firm-specific earnings response coefficient*. The slope $\tilde{k}_{i,0}$, which models the price–earnings relation for extreme earnings, is often equal to zero. There are two main differences with the abandonment option model: first, the expression in (19) takes into account the transitory nature of extreme positive earnings; second, it acknowledges that the parameters of the model are firm-specific (and not cross-sectionally constant as in Hayn [1995]).

While the non-linearity discussed—with lower slopes for both extremely low and high earnings—is well known in the literature, the RF allows two important contributions. First, we can empirically identify the extent of the middle linear component with no *ex ante* assump-

[19]When the true relation is as described above, a linear regression inferred on the entire range of X and constrained to have a single set of coefficients would result in an estimated ERC that under-estimates the true parameter k, with a lower explanatory power than that of the unrestricted regression. Moreover, L and k are firm-specific. As such, even estimating a linear regression that allows the coefficients to vary across regions of X would yield meaningless coefficients.

[20]The third component may be a linear relation with a slope that is smaller than that of the second component.

tions. Second, we can estimate firm-specific slopes.

The thick black line (bordered by yellow) in the graphs in Figure 8 is the partial dependence of price on earnings, $\widehat{f}_0^{(NI)}$, defined in (17). It gives the synthesis, over all firms in a cross-section, of the local linear relation that ties prices to earnings in the central range of earnings.

The firms (curves) in Figure 8 are color-coded. The firms with size below (above) the 33rd (66th) percentile of the size variable in the cross-section are shown in red (yellow). The firms with size in the mid-range of the cross-section's size values are shown in blue. We note that smaller firms have lower valuations for the same level of earnings (the red curves tend to be in the lower part of the graph, while the yellow ones tend to be higher up). To the extent that size is a proxy for risk, the positive association between firm size and the valuation based on expectations informed by the level of earnings is consistent with risk negatively affecting value.

The next two sections establish the relation between the $k_{i,0}$, the non-linear firm-specific ERC, and firm characteristics that proxy for risk and growth (section 6.3.1) and for future returns (section 6.3.2).

6.3.1 The relation between $1/k_{i,0}$ and relevant firm characteristics

The inverse of the firm-specific ERC $1/k_{i,0}$ is a measure of the risk of the firm i (Kothari and Shanken [2003]). As such, we expect it to be negatively correlated with size[21] and price-to-book ratio[22] (since growth is risky), and positively associated with leverage (the higher the leverage, the larger the risk) and the level of the risk-free interest rate.

Variable	Estimate	Cluster s.e.	t-value	$P(>\|t\|)$
Risk-free (1y)	0.5655	0.0110	51.20	<2e-16 ***
P/B	-0.0023	0.0001	-14.12	<2e-16 ***
Mktv(log)	-0.0112	0.0003	-29.28	<2e-16 ***
Leverage	0.0270	0.0005	47.98	<2e-16 ***
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'

Table 3: **Panel regression of $1/k_{i,t}$ on firm characteristics.** The estimated regression is $1/k_{i,t} = \alpha_0 + \alpha_1 rf_t + \alpha_2 P/B_{i,t} + \alpha_3 log(Mktv_{i,t}) + \alpha_4 ROE_{i,t} + \alpha_5 Lev_{i,t} + \varepsilon$. The inclusion of the risk-free rate takes care of the time-fixed effect. The standard errors are clustered by firm (Petersen, 2008).

[21] Differences in the price–earnings relation related to firm size may reflect differences in information environment (Collins and Kothari [1989]).

[22] The market-to-book value of equity is used as a proxy for a firm's economic growth opportunities.

Table 3 presents the results of a panel regression (with both time- and firm-fixed effects) of the inverse of the firm-specific ERC, $1/k_{i,0}$ on P/B, Mktv, Leverage and the rate of the one-year treasury bond. The inclusion of the risk-free rate takes care of the time-fixed effect. The firm-fixed effect is needed as the ERC is possibly determined by firm-specific factors other than, but correlated with, the independent variables considered. The standard errors are clustered by firm (Petersen [2008]). All the proxies are strongly associated with the risk measure in the expected direction, which supports ERCs as relevant measures of risk.

6.3.2 The relation between $1/k_{i,0}$ and future returns

As a measure of risk, the inverse of the firm-specific ERC $1/k_{i,0}$ should be positively associated with future returns (as higher risk should be associated with higher returns). We regress 12-month future realized returns, $ret.12$, on $1/k_{i,0}$, controlling for firm characteristics that are known to be associated with returns: P/B, size, and leverage. The regressions are performed as an unbalanced panel. The inclusion of the risk-free rate functions as time-fixed effects. The firm-fixed effect is needed as the returns are possibly determined by firm-specific factors other than, but correlated with, the independent variables considered. The standard errors are clustered by firm (Petersen [2008]).

Variable	Estimate	Cluster s.e.	t-value	$P(>\|t\|)$
ERC^{-1}	0.5377	0.0313	17.15	<2e-16 ***
Risk-free (1y)	-1.8070	0.0742	-24.33	<2e-16 ***
P/B	-0.0183	0.0011	-16.80	<2e-16 ***
Mktv(log)	-0.1281	0.0034	-37.07	<2e-16 ***
Leverage	0.04899	0.0030	16.31	<2e-16 ***
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'

Table 4: **Panel regression of one-year returns, $ret.12_{i,t}$, on $1/k_{i,t}$ and proxies for firm risk and growth.** The estimated regression is $ret.12_{i,t} = \alpha_0 + \alpha_1 rf_t + \alpha_2(1/k_{i,t}) + \alpha_3 P/B_{i,t} + \alpha_4 log(Mktv_{i,t}) + \alpha_5 Leverage_{i,t} + \varepsilon$.

Table 4 presents the results of a panel regression of the 12-month future returns on the inverse of the firm-specific ERC, $1/k_{i,0}$, P/B, Mktv, Leverage, and risk-free rate. The risk measure $1/k_{i,0}$ explains the future returns beyond the classical factors of size and value or growth.

6.4 Validation 3: The cross-section of non-linear ERC

In this section, following Kothari and Zimmerman [1995], we validate the novel research design introduced in section 5 by verifying that the non-linear regression ERCs have the "right" size and imply economically justifiable risk premium values.

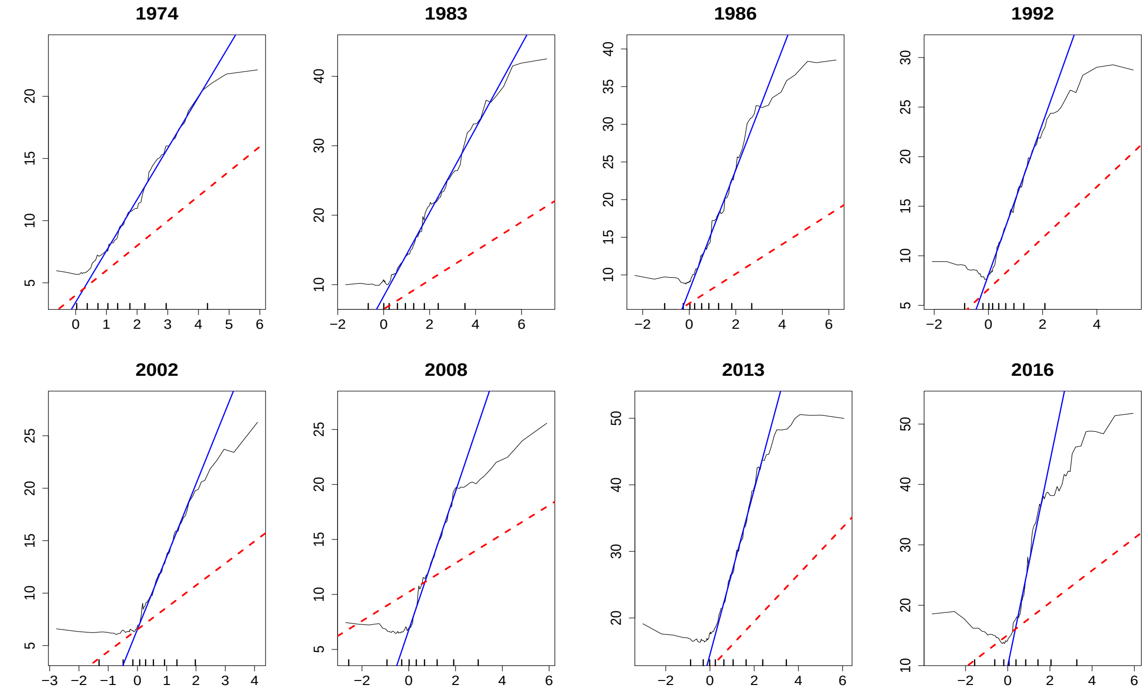


Figure 9: **Linear fit for the estimated functional relation between earnings and prices – selected years.** The graphs display the estimated partial dependence function (17) for all firms in the cross-section in the title of the graph (solid, black), the linear fit for the range of earnings where the dependence is approximately linear (roughly between the third and eighth deciles) (solid, blue), and the linear fit of the levels regression (2) (dotted, red). We see clearly that the linear regression earnings response coefficient (ERC) is much smaller than the corresponding non-linear ERC.

We define the cross-section's non-linear ERC, ERC_0, as the slope of the linear segment of the PD plot displaying the functional relation between earnings and share prices (the thick black curves in Figure 8). Figure 9 displays a number of cross-sections[23] of the estimation of the *non-linear regression ERC*, i.e., the linear fit of a regression line on the linear segment of the one-way PD plot. The figure also shows, for comparison, the price–earnings linear relation implied by the estimated linear regression in (2) (dotted, red). We see that the slope of the linear fit of the Ohlson [1995] linear specification of the RI model is much smaller than the corresponding slope of the local linear fit of the non-linear research design.

[23]Due to space constraints here, we show all 47 cross-sections in section 9.9 in the appendix.

ERC_0 improves over the market price–earnings ratio

$$\frac{(\text{average price})_0}{(\text{average } NI)_0}$$

that is commonly used in the literature to gauge the size of the ERCs (see section 6.4.2, and section 9.10 in the appendix).

6.4.1 The non-linear levels earnings response coefficient has the "right" size

The size of the ERCs has been an ongoing puzzle since the beginning of the 1980s (Beaver et al. [1980]). Empirical estimates of ERC magnitudes range from 1 to 3 (see, for example, Kormendi and Lipe [1987]; Easton and Zmijewski [1989]). Using price–earnings multiple as an estimate of the ERC, one expects a magnitude[24] of 6 to 20 (depending on the cross-section) (Kothari [2001]). Predictions based on the discount rate used by investors lead to expected values for the slope coefficient of 7 or higher (Hayn [1995]; Kothari and Zimmerman [1995]).

The relatively small magnitude of the ERC compared to its predicted value has resulted in at least four hypotheses: (a) prices leading earnings, (b) inefficient capital markets, (c) noise in earnings and deficient GAAP, and (d) transitory earnings (Kothari [2001]). In this section we introduce evidence supporting another explanation (although one related to transitory earnings): the linear model is mis-specified when fit to the non-linear economic price–return relation. This mis-specification is responsible for the gap between the linearly estimated ERCs and their expected magnitude. A correctly specified design yields ERCs that are economically reasonable.

Figure 10 displays histograms of the 47 values of the linear ERC obtained from the yearly cross-sectional estimation of the linear levels regression in (2) (on the left-hand side) and of the *non-linear level regression ERC* ERC_0 defined at the beginning of this section (right-hand side). The mean ERC is 2.91 for the linear model and 8.01 for the non-linear specification. While, as previously reported in the literature, the cross-sectional linear regression ERCs are relatively small (compared to their predicted values), the non-linear regression ERCs have economically reasonable size.

[24]Theaverage price–earnings ratio in our sample ranges between 6.6 and 21.8.

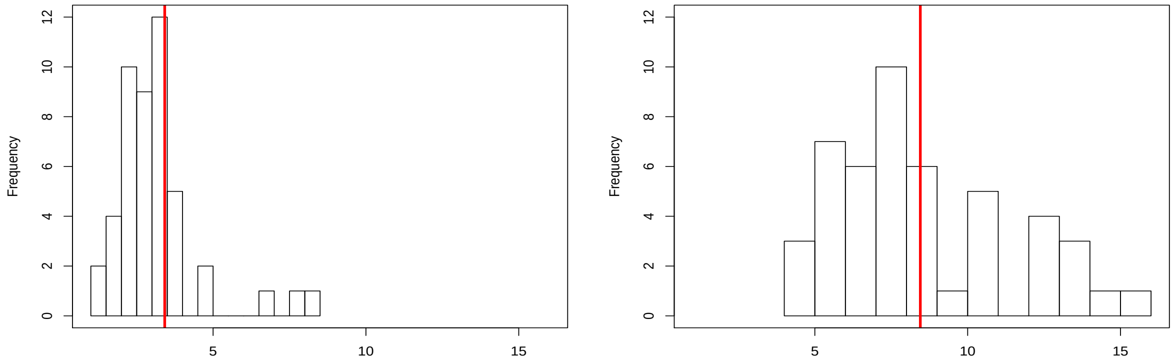


Figure 10: **The linear and non-linear earnings response coefficients (ERCs).** The graphs show the histograms of the 47 cross-sectional levels linear ERC, $\beta_{2,0}$, in (2) (left) and of the cross-sectional mean slope of the linear component (for all firms in a given year) in the non-linear design, a_0 (right). The vertical line marks the mean ERC: 2.91 for the linear model and 8.01 for the non-linear design. The standard deviations are 1.39 and 2.50, respectively.

6.4.2 Implied risk premium comparison

To better gauge the values of the ERCs estimated in the cross-sections, we construct the cross-sectional risk premium implied by the two types of ERCs. Barth and Kallapur [1996] argue that the ERC should equal $1/r$, where r is a discount rate that can be equated with the cost of capital. Following their arguments, we view estimated ERCs that yield unreasonably high cost of capital estimates as signs of mis-specification, i.e., failure of the modeling approach to capture the economic relation between prices and earnings.

The non-linear approach yields an estimate of the cross-sectional risk premium:

$$r_{e,0}^{(non-lin)} = 1/\widehat{ERC_0} - r_{f,0}, \tag{20}$$

where $r_{f,0}$ is the cross-sectional risk-free rate.[25] The linear approach based on Ohlson's linear specification (2) of the RI model yields another estimate:

$$r_{e,0}^{(lin)} = 1/\widehat{\beta}_{2,0} - r_{f,0}. \tag{21}$$

The magnitude of the values $r_{e,0}^{(lin)}$ and $r_{e,0}^{(non-lin)}$ are easier to interpret than that of the median slope ERC_0 or the $\beta_{2,0}$ coefficient.

Figure 11 displays the time series of the risk premiums implied by the linear research design (21) (top), as well as the non-linear research design estimate (20) (bottom). The linear approach

[25]We used the ten-year US government T bond rate.

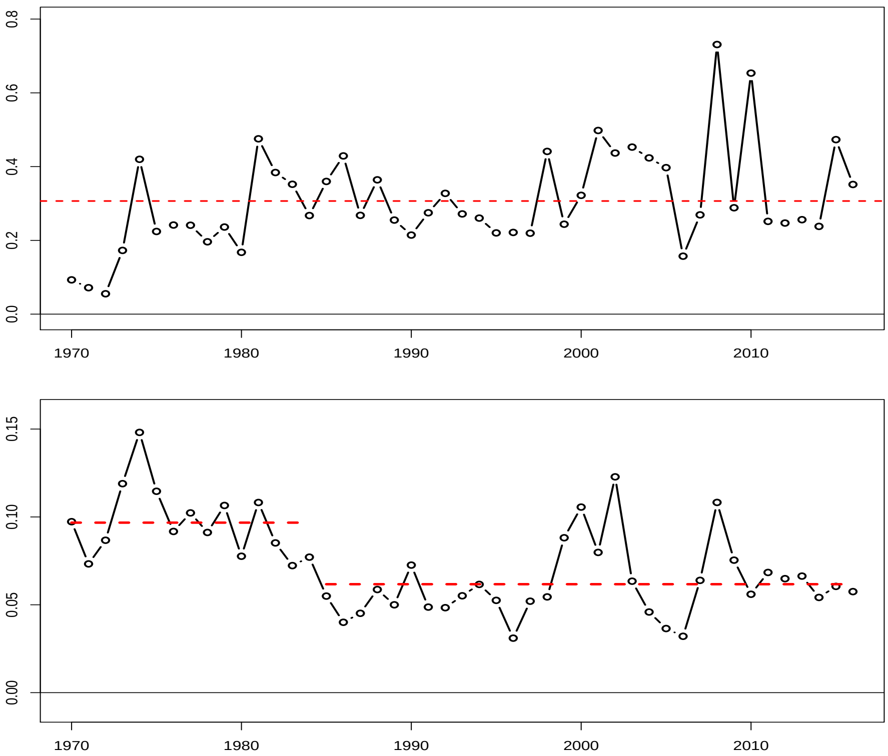


Figure 11: **The risk premium implied by the linear and non-linear earnings response coefficients (ERCs).** The graphs display the time series of risk premiums implied by the regression ERC β_2 in (2), given by the expression in (20) (top) and by the non-linear ERC (defined at the beginning of this section), defined in (21) (bottom). Risk premiums are on the $y-axis$ and years on the $x-axis$. The yearly risk premiums yielded by the non-linear estimation are economically justifiable (the mean implied market risk premium is 5.8% for the last three decades), while the premiums implied by the linear estimation are not (the mean implied market risk premium is 31%).

yields unreasonably large risk premium estimates. In contrast, the risk premium implied by the non-linear approach is economically justifiable and comes close to estimates produced through alternative approaches.

The criteria that we employ to validate the non-linear research design are neither unique nor the only ones that can be used. Different criteria will apply, depending upon research design and loss function. Commonly (see Kothari and Zimmerman [1995]; Kothari [2001]), the estimated ERCs are compared to the market price–earnings ratios; however, we avoid using this criterion for statistical reasons. The price–earnings ratio (or the ratio of average price to average earnings per share in cross-sections) is a one-point estimate of the slope parameter of a linear relation between earnings and prices.[26] As such, it is a very unreliable estimator, with a large variance. As further evidence of the lack of statistical reliability, we note that estimating

[26]Moreover, as we document in this paper, the price–earnings relation is not an over-all linear association and the contribution of the non-linear part biases the estimation.

the cross-sectional risk premium starting from the ratio of average price to average earnings per share in cross-sections yields negative risk premiums for most years t in the period from 1980 to 2010 (see section 9.10 in the appendix).

7 An example of how to use the non-linear research design

This section gives an example of how the methodology we introduced in this paper can be used to advance and refine accounting theory. It is based on a suggestion in Holthausen and Watts [2001] (p. 67). The authors emphasize the importance of addressing the "weaknesses in the current valuation models used in accounting research" (p. 66). They argue that such models should be able to reflect basic "descriptive accounting theory" and give an example of what they mean by that: "consider the relation between accounting earnings and stock prices. It seems plausible that an accounting theory might predict that the accounting earnings of firms with higher risk and growth measure future cash flows with greater error and bias [...]. The extent to which current earnings capture future cash flows is likely to be smaller for riskier and higher growth firms" (p. 67). A model of the relation between prices and earnings should reflect such differences. We show that the method we propose gives the expected results.

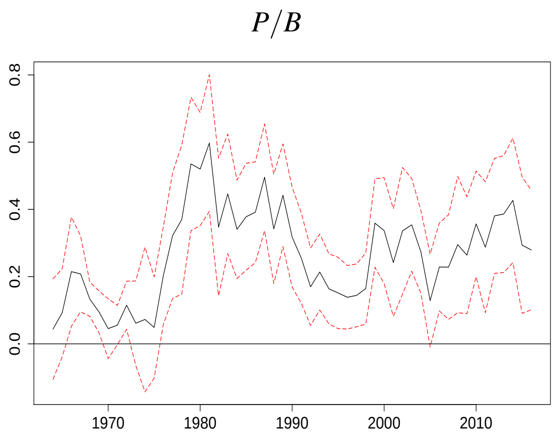
We formally test the hypothesis that the earnings of firms with higher risk and/or growth inform the expectation formation process of price setting less than those of firms with lower risk and/or growth. For this example, we choose two proxies for risk and/or growth. The first is the price-to-book ratio P/B, the second, the volatility of cash flows from operations. P/B is commonly used in extant literature as a proxy for growth as well as for the level of conservative accounting. Cash flow volatility is a measure of economic risk.

To introduce the measure of the pertinence of earnings to the price of a firm we note that regression (18) is a valuation: the estimated $\hat{f}_0$ is the value of the firm based on expectations about future cash flows (or earnings) informed only by the level of reported earnings. The measure of pertinence is defined as the absolute value of the relative valuation error:

$$|P_i - \widehat{f_i}|/P_i.$$

This represents the proportion of the price represented by investors' adjustment to a price informed only by the level of earnings. The higher this measure, the less pertinent the earnings numbers to the price expectation formation process.

We test the hypothesis for each cross-section. For a given year, we use each of the two proxies to divide the firms in the cross-section into halves. One half contains the firms with proxy values in the lower 50% of the range while the other half contains the rest. In other words, we create two sub-samples: one composed of firms with low levels of risk/growth (as measured by the proxy) and the other containing firms with high levels of risk/growth. The test statistic is the difference between the means of the pertinence measure of the high and low risk/growth sub-samples. The null hypothesis is that of equal pertinence of earnings for the two levels of risk/growth. As we are performing the (yearly) test 47 times, we present the results in the form of a graph (as opposed to a table). We display the test statistic together with a 95% confidence interval. If the test statistic is positive and this interval does not include zero we conclude that, for the year under study, the null hypothesis is rejected in favor of the alternative of lower earnings pertinence to price expectation formation for higher risk/growth firms.



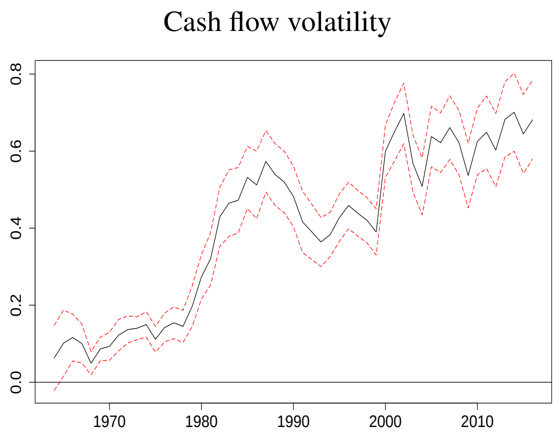


Figure 12: **Earnings of firms with different risk and growth profiles inform prices to different extents.** The graphs display the yearly difference between the mean relative valuation error of firms with high (top 50%) vs. low (bottom 50%) price-to-book (P/B) (left) and cash flow volatility (right). Mean relative valuation errors are on the $y-axis$ and years on the $x-axis$. The black line shows the mean by year, while the red lines show the 95% confidence interval of the mean. For the years where the bottom red line is above the $x-axis$ there is a statistically significant difference between firms with high and low values of P/B and cash flow volatility.

The results are presented in Figure 12 and show that, for most of the years, the earnings of firms with lower levels of risk/growth are significantly more pertinent to price formation than the earnings of firms with higher levels of risk/growth. This example provides evidence that our approach is a significant step in the desirable direction of development of valuation models proposed in Holthausen and Watts [2001] (p.67): "Links between the accounting numbers and valuation models will have to be specified in a way that provides testable implications about accounting."

8 Conclusions

Many of the seminal papers investigating the relation of prices to earnings were published before the recent exponential developments of non-linear modeling and prediction methodology in the fields of statistics and machine learning. Since the accounting literature is highly consensual on the fact that the relation of prices to earnings is non-linear (Holthausen and Watts [2001], advances in non-linear modeling offer a promising avenue for capital market research in accounting. Our paper is a step in this direction.

We show with concrete examples how non-linearities affect a linear model. Even small changes in a sample, which do not affect the economic relation expressed, cause the coefficients of a linear model to change significantly. Taking the linear inference at face value would lead to the false rejection of the hypothesis of equal economic relation.

We outline a non-linear research design in which existing accounting literature informs modern statistical inference methodology to address the issue of consistent estimation of the non-linear relation between prices and earnings. We start with a general valuation model and show how to deduce a non-linear regression specification of the economic relation in the model.

We estimate the non-linear regression using random forests (RFs). In addition to the level of earnings, the specification contains (as independent variables) proxies for determinants of a firm's future cash flow stream. The non-parametric feature of our design does not require any *ex ante* assumptions about the structure of the price–earnings relation.

For any level of earnings, the RF method approximates the price–earnings relation of a given firm with a local average of the prices of firms with characteristics similar to those of the given firm. Consequently, the RF has a property that corresponds to the intuition that similar firms have a similar price–earnings relations. Further, the RF identifies those variables that are important to the relation. Empirically, we find that the most important variables are the ratio of price to book value of equity, size, and total liabilities to market value of equity. The remaining (numbering more than 40) variables we investigate are of marginal importance.

This paper makes two main contributions. First, we find—absent any *ex ante* assumptions about the structure of the relation between earnings and price—empirical support for theoretical predictions in the literature. The estimated price–earnings relation is consistent with limited liability and the option view of equity (Hayn [1995]; Fischer and Verrecchia [1997]): equity price increases with earnings and is bounded below by zero, strictly convex for low levels of earnings, and strictly concave for high levels of earnings. Further, we bring evidence that non-linear regression ERCs have the ”right” size and yield economically justifiable risk premium values (cf. Kothari and Zimmerman [1995]).

Second, we propose a research design that gives a consistent estimation of the price–earnings relation. Focusing on average errors, the method is useful for comparing the information content for stock market valuation of different predictors. For example, it can be used to compare earnings with cash flows or to compare earnings based on different accounting standards. We also give an example—suggested by Holthausen and Watts [2001]—in which we show how the method can be used to test predictions made in accounting theory. In addition, the method can be used to identify empirical structures in the data, which allows for the application of linear regression in specific settings (cf. Gow et al. [2016]).

9 Appendix

9.1 The sample: size of the cross-section

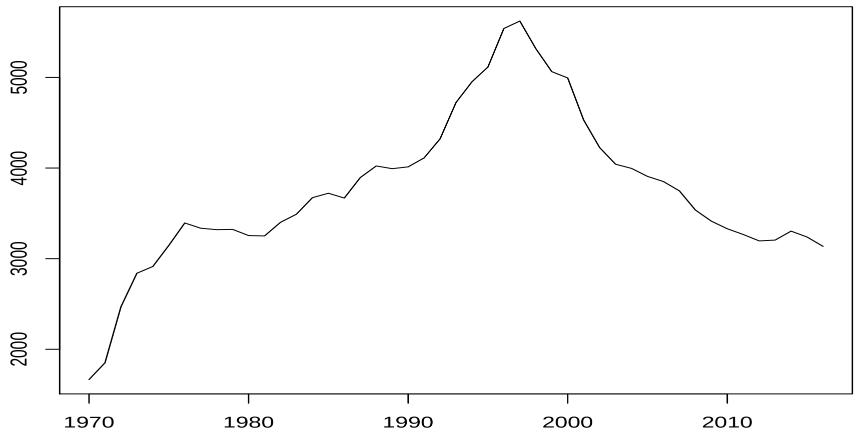


Figure 13: **Number of firms per year.** The cross-section size varies between a minimum of 1,666 (in 1970) and a maximum of 5,622 (in 1997).

Figure 13 displays the number of firms per year, with a minimum of 1,666 in 1970 and a maximum of 5,622 in 1997.

9.2 Interaction between non-linearity and omitted variables

The problems with linear models that we noted in section 4.3 are potentially further exacerbated if there is an omitted variables issue. As noted in section 1, existing research documents a cross-sectional relation between ERCs from returns specification and risk (discount rates), growth opportunities, and economic rents (Holthausen and Watts [2001]). Kothari and Shanken [2003] extends the findings to ERCs from price specifications.

Table 5 presents evidence, based on our sample, of the relation between ERCs from the price–earnings linear regression and two proxies for risk and growth opportunities, size ($Mktv$) and price-to-book (P/B). Given the non-linear nature of the underlying relations, we do not attach much confidence to these estimates. However, we note that the size of the differences between extreme deciles is very large.

The types of econometric issues related to the interaction of non-linearity and omitted determinants are exemplified in Figure 14. The graphs display pairs of samples that express the same economic relation between prices and earnings but differ in the value range of each of two proxies. The samples

Decile		1	2	3	4	5	6	7	8	9	10
Mktv	β_{NI}	0.40	1.20	1.72	2.52	3.03	3.44	3.87	4.49	4.98	6.77
	Std	0.10	0.16	0.20	0.25	0.28	0.31	0.35	0.43	0.48	0.65
P/B	β_{NI}	0.98	2.77	4.45	5.51	6.31	7.48	8.61	9.85	11.82	12.92
	Std	0.18	0.24	0.28	0.31	0.34	0.36	0.39	0.44	0.51	0.75

Table 5: **Estimated earnings response coefficients conditional on the level of risk and growth proxies.** We ordered the observations in each cross-section according to the level of the proxy ($Mktv$ and P/B) and ran ten regressions (2) on each of the ten sub-samples formed by the firms with the proxy in a given decile. We report the mean and the standard deviation of the distribution of the estimated coefficients for each of the deciles. We note a strong increasing relation between the level of the estimates and the decile of the proxy: the higher the proxy, the larger the estimated earnings coefficients, on average.

in the top (bottom) graphs (proxy: $Mktv$ – top, P/B – bottom) are sub-samples of the 1994 and 1996 cross-sections. The sub-samples in the left-hand-side (right-hand-side) graphs are missing every other firm with the values of the proxy in the lowest (highest) decile. The missing observations are marked with squares (blue on the left, magenta on the right). The solid line represents the regression line of the sample in the graph, while the dotted line is the regression line corresponding to the paired sample. We see that lower (higher) values of the proxies are associated with lower (higher) earnings and prices. Since the relation between prices and earnings is non-linear and depends on the level of earnings, a sample that is richer (poorer) in firms with higher proxy values will have a higher (lower) slope. The differences between the two slopes are statistically significant: in the case of the top (bottom) graphs, the t-statistic of the test of equal slopes is 2.84 (3.37). Although we know that the two sub-samples are expressions of the same economic relation between earnings and prices, the linear model rejects the hypothesis of equality.

This example is pertinent to a situation where two cross-sections are expressions of the same economic relation between price and earnings but differ in terms of the risk or growth profiles of the firms. The interaction between non-linearities and omitted determinants exemplified in Figure 14 would lead (based on the mis-specified linear model) to the rejection of the correct hypothesis.

We performed the analysis in Figure 14 on all yearly cross-sections in the sample. The results are presented in Figure 15, which displays the histograms of the t-statistics that test the equality of slopes for the pairs of regressions on yearly cross-sectional sub-samples constructed according to the description above. The pairs of sub-samples in each cross-section express the same economic relation between prices and earnings but differ in the values taken by each of two proxies. For the left (right) graphs, the proxy is $Mktv$ (P/B). Values of the t-statistic outside the interval $[-2, 2]$ correspond to rejection of the null of equal economic relation at a 95% confidence level. The graphs in the figure indicate that, for a large number of cross-sections, a test based on the linear representation of the Ohlson [1995] model

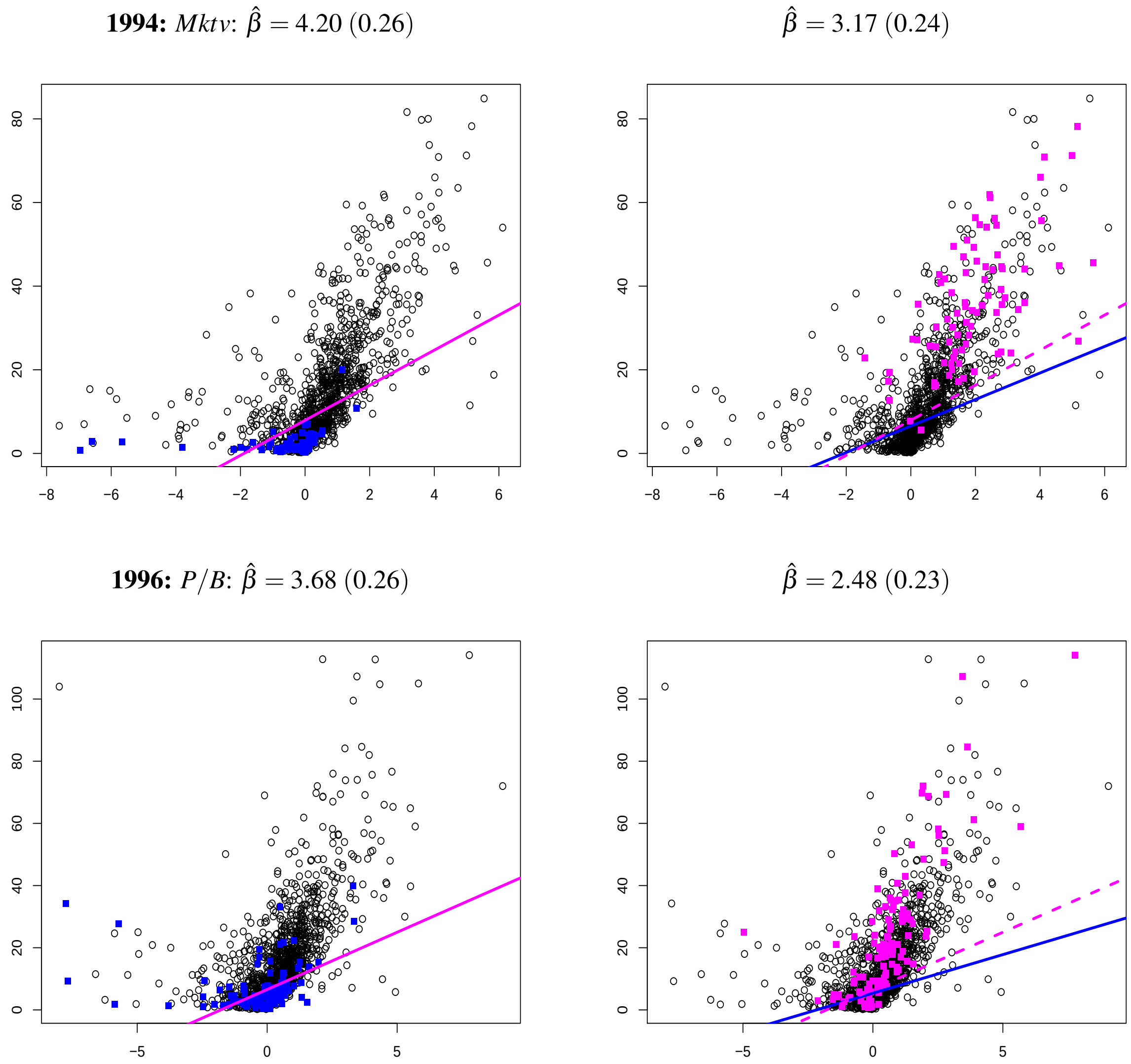


Figure 14: **Non-linearity and omitted determinants confound the linear regressions.** The graphs display prices (on the y-axis) against NI (on the x-axis) for two sub-samples from the same cross-section constructed as follows. The pairs of sub-samples express the same economic relation between prices and earnings but differ in the range of values of each of two proxies. For the top (bottom) graphs, the proxy is $Mktv$ (P/B). The sub-samples in the left-hand-side (right-hand-side) graphs are missing every other firm with the values of the proxy in the lowest (highest) decile. The missing observations are marked with squares (blue on the left, magenta on the right). The solid line represents the regression line of the sub-sample in the graph, while the dotted line is the regression line corresponding to the paired sub-sample. We see that lower (higher) values of the proxies are associated with lower (higher) earnings and prices. The size of the standard error is shown in parenthesis after the values of β.

would wrongly reject the null hypothesis. Even changes in the sample structure that are not directly related to prices or earnings, but to other determinants of their relation, can cause significant instability in the statistics that are commonly used for testing hypotheses about the economic relation between prices and earnings.

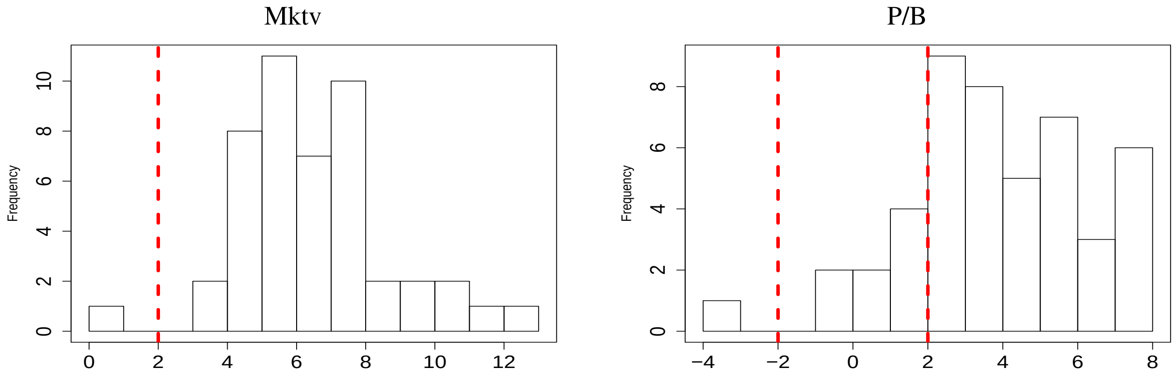


Figure 15: **Non-linearity and omitted variables confound the linear regressions - all cross-sections.** The graphs display the histograms of the t-statistics (on the x-axis) that test the equality of the regression earnings response coefficients, from the two regressions (2) performed on pairs of yearly cross-sectional sub-samples (from 1970 to 2016) constructed according to the description in the caption of Figure 14. The pairs of sub-samples in each cross-section express the same economic relation between prices and earnings but differ in the range of values of each of two proxies. For the left (right) graphs, the proxy is $Mktv$ (P/B). Values of the t-statistic outside the interval $[-2, 2]$ correspond to rejection of the null of equal economic relation at a 95% confidence level. The graphs in the figure indicate that for a majority of cross-sections a test based on the linear representation of the Ohlson model would wrongly reject the null of equal relation.

9.3 Evidence of the fit of the non-linear estimation

As already seen in section 6.2, the model discriminates between differences in the marginal distribution of earnings and differences in the economic relation (for the linear model, see section 4.3). In the sequel, we document that when the sub-samples differ in the range of determinants of the price–earnings relation (P/B or market value), the shape of the estimated economic relation is identical over most of the range of earnings (up to and above the ninth decile).

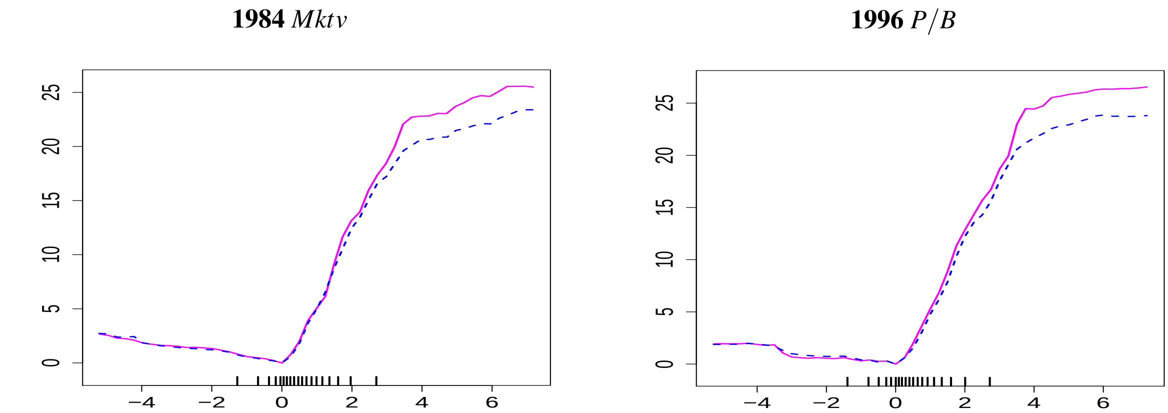


Figure 16: **The non-linear approach addresses the bias in the level regressions.** The graphs display the dependence of prices (on the y-axis) on NI, estimated by the approach described in section 5 for the pairs of sub-samples from the same cross-section analyzed using the linear models in figure 14. The pairs of sub-samples express the same economic relation between prices and earnings but differ in the ranges of values of $Mktv$ and P/B (see the caption of figure 3 for more details). The solid (dotted) line corresponds to the samples obtained by removing firms with lower (higher) values of NI or proxies. The rug displayed on the x-axis corresponds to the ventiles, i.e., the 5th, 10th, . . . , 95th percentiles, of the cross-section's distribution of NI. We see that the estimated dependence is very similar up to at least the 95th percentile.

Figure 16 displays the partial dependence of price on earnings, $\widehat{f}_0^{(NI)}$, defined by (17), estimated non-linearly for the same pairs of sub-samples from the same cross-sections analyzed using the linear level regressions in sections 9.2 (see figure 14). By construction, the pairs of sub-samples express the same economic relation between prices and earnings but differ in the ranges of values of the two main proxies for the determinants of the price–earnings relation, $Mktv$ (left) and P/B (right). We see that the curves estimated on sub-samples that differ in the range of the two proxies are almost identical over most of the range of earnings, i.e., at least until the 95th percentile. Moreover, the discrepancy noticeable for higher earnings is to be expected, as firms with the same level of earnings but higher values of the two proxies have significantly higher prices. As such, the samples obtained by removing firms with higher values of the two proxies will have a "thinner" upper part of the price distribution and hence a dependence function that tapers off earlier.

9.4 Ohlson and Ohlson Juettner-Nauroth models

The Ohlson model is a linear expression of the *RI* valuation relation (Preinreich [1936]; Edwards and Bell [1961]; Peasnell [1982],

$$\begin{aligned} P_{i,0} &= B_{i,0} + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[NI_{i,t} - r_{i,0} \times B_{i,t-1}]}{(1+r_{i,0})^t} = B_{i,0} + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[RI_{i,t}]}{(1+r_{i,0})^t} \\ &= B_{i,0}\left(1 + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[RI_{i,t}/B_i]}{(1+r_{i,0})^t}\right), \end{aligned} \tag{22}$$

which expresses the value of firm i at time 0 as the book value (B) plus discounted future expected abnormal earnings ($NI - r \times B_{-1}$) (r_0 denotes the price of equity risk at time 0 while $\mathbb{E}_0$ stands for market's expectation conditional on all information available at time 0).

To linearize the non-linear expression (22), Ohlson [1995] makes two additional assumptions on the dynamics of *RI*:

$$\begin{aligned} RI_{i,t} &= \omega_i RI_{i,t-1} + \nu_{i,t-1} + \varepsilon_{1,i,t}, \\ \nu_{i,t} &= \delta_i \nu_{i,t-1} + \varepsilon_{2,i,t}, \end{aligned} \tag{23}$$

where $0 \leq \omega_i, and \delta_i < 1$ are two constants determined exogenously by a firm's economic environment and its accounting practices, $k_{i,0} = \omega_i r_{i,0}/(1 + r_{i,0} - \omega_i)$ and $\alpha_{i,0} = 1 + r_{i,0}/(1 + r_{i,0} - \omega_i)(1 + r_{i,0} - \delta_i)$. ν_t is "information other than abnormal earnings" (i.e., events that have not affected current B_0 and NI_0), while $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are unpredictable (zero mean conditional on the information at time t: $\mathbb{E}_t(\varepsilon_{1,t+j}) = 0$, $\mathbb{E}_t(\varepsilon_{2,t+j}) = 0$) "disturbance" terms.

Under these assumptions, the non-linear relation (22) admits a linear representation that explains prices as a linear combination of current book values and residual earnings:

$$P_{i,0} = B_{i,0} + k_{i,0} r_{i,0}^{-1} \times RI_{i,0} + \alpha_{i,0} \times \nu_{i,0}. \tag{24}$$

However, this linear representation is not necessarily a regression and hence cannot justify linearly regressing prices on earnings and book values in cross-sections. A formal proof of this statement is presented in the next section.

Ohlson and Juettner-Nauroth [2005] provide an alternative model (the OJ model, also known as the AEG model) that allows for deviations from the clean surplus relation that are to be expected under current major accounting standards (US GAAP and IFRS). Equity value is the capitalized sum of (i) the next-period earnings and (ii) the present value of the expected changes in subsequent earnings adjusted for dividends, the so-called *abnormal earnings growth*:

$$P_0(OJ) = \frac{E_0[NI_{i,1}]}{r_{i,0}} + \sum_{t=1}^{\infty} \frac{E_0[NI_{i,t+1} + r_{i,0} dps_{i,t} - (1+r_{i,0}) NI_{i,t}]}{r_{i,0}(1+r_{i,0})^t}. \tag{25}$$

where dps represents dividends per share and all other variables are as defined above.

This model is more general than the residual income (RI) model as it does not require clean surplus accounting. It is based on expected earnings and expected earnings growth and can be seen as the M&M-consistent version of the Gordon model.

9.5 The Ohlson (1995) linear representation of the residual income model does not motivate the linear price–earnings regression

This section identifies the supplementary second order conditions on the dynamics of *RI* that are needed for Ohlson's linear representation of the RI model to be a good empirical approximation of the underlying economic association between earnings and prices. There is no particular reason to assume that similar conditions actually hold in the data.

Proposition 1. *Assume prices verify the RI valuation model (22) and Ohlson's assumptions (23). Assume that the sequence* $(\varepsilon_{1,t}, \varepsilon_{2,t})$ *is second-order stationary with the covariance structure described by* $\gamma_{\varepsilon_1,\varepsilon_2}(h) := Cov(\varepsilon_{1,t}\varepsilon_{2,t-h})$. *Denote by* $\gamma_X(h) := Cov(X_t, X_{t-h})$ *the auto-covariance function of the series* X_t.

Then the following condition on the second order structure of the innovations in the RI dynamics:

$$\sum_{i,j,k=0} \omega^k \delta^{i+j} \gamma_{\varepsilon_2}(k+1+j-i) + \sum_{i,j=0} \delta^i \omega^j \gamma_{\varepsilon_1,\varepsilon_2}(j-i) = 0$$

is necessary for the linear expression (24) to be a linear regression, i.e., for the orthogonality property (6) to hold.

Proof. Note that the linear dynamics assumptions (23) imply (we suppress the index making the quantities firm-specific) that:

$$RI_t = v_{t-1} + \omega v_{t-2} + \omega^2 v_{t-3} + \ldots + \varepsilon_{1,t} + \omega \varepsilon_{1,t-1} + \omega^2 \varepsilon_{1,t-2} + \ldots$$

and that

$$v_t = \varepsilon_{2,t} + \delta \varepsilon_{2,t-1} + \delta^2 \varepsilon_{2,t-2} + \ldots$$

for all t. Hence:

$$cov(v_t, v_{t-h}) := \gamma_v(h) = \sum_{i,j=0} \delta^{i+j} \gamma_{\varepsilon_2}(h+j-i)$$

Note that:

$$\begin{aligned} cov(v_t, RI_t) &= Cov(v_t, v_{t-1} + \omega v_{t-2} + \omega^2 v_{t-3} + \ldots) \\ &\quad + Cov(v_t, \varepsilon_{1,t} + \omega \varepsilon_{1,t-1} + \omega^2 \varepsilon_{1,t-2} + \ldots). \end{aligned}$$

The first term in the previous decomposition can be written as:

$$\begin{aligned} Cov(v_t, v_{t-1} + \omega v_{t-2} + \omega^2 v_{t-3} + \ldots) &= \sum_{k=0} \omega^k \gamma_v(k+1) \\ &= \sum_{i,j,k=0} \omega^k \delta^{i+j} \gamma_{\varepsilon_2}(k+1+j-i), \end{aligned}$$

while the second term can be written as:

$$\begin{aligned} Cov(v_t, \varepsilon_{1,t} + \omega \varepsilon_{1,t-1} + \omega^2 \varepsilon_{1,t-2} + \ldots) &= Cov\left(\sum_{i=0} \delta^i \varepsilon_{2,t-i}, \sum_{j=0} \omega^j \varepsilon_{1,t-j}\right) \\ &= \sum_{i,j=0} \delta^i \omega^j \gamma_{\varepsilon_1,\varepsilon_2}(j-i). \end{aligned}$$

Since $cov(v_t, RI_t) = 0$ is a necessary condition for the linear expression in (24) to be a linear regression, the statement holds. □

Unless binding assumptions on the second order structure of the disturbance terms and on their interaction are made, there is no reason to suppose that the covariance between ν_t and RI_t is equal to zero.

Hence we can state that:

Proposition 2. *The linear representation of the RI valuation model in (24) is not necessarily a linear regression. Without further binding assumptions on the second moments of the "disturbance" sequences* $(\varepsilon_{1,s}, \varepsilon_{2,t})_{s,t}$*,* $\mathbb{E}[\nu_0|RI_0]$ *is not necessarily equal to zero.*

This states that the economic relation of the *RI* valuation model cannot always be inferred empirically if expressed in the linear form of Ohlson's model. Worse still, the researcher cannot know when she has been successful. Consequently, the implications of the model cannot be verified empirically.

Based on the formal argument above, a short, intuitive explanation[27] of the result reads as follows. It seems reasonable to assume, for example, that the size of the contribution of "information other than abnormal earnings" to determining *RI* may be similar in consecutive years if the economic conditions in which the firm operates do not evolve much and the firm characteristics are stable, i.e., $corr(\nu_0, \nu_{-1}) > 0$. A large/small correction to previous *RI* to obtain current (this year's) *RI* is likely to be followed by a large/small correction to obtain next year's *RI*. Unless $corr(\nu_0, \omega RI_{-1} + \varepsilon_{1,0})$ is negative and exactly compensates $corr(\nu_0, \nu_{-1})$, condition $corr(\nu_0, RI_0) = 0$, and hence $\mathbb{E}[\nu_0|RI_0] = 0$, is not fulfilled. There is no reason why such a perfect compensation would always occur. In other words, without precise extra assumptions on the parameters of the information dynamics (more precisely, on its second order structure), the linear expression of the *RI* valuation model in (24) is not a linear regression associating prices with book values and *RI*.

9.6 Non-linear regression and its estimation

Definition 1. *The decomposition:*

$$Y = f(X) + \varepsilon$$

is called a regression *if*

$$f(X) = \mathbb{E}[Y|X]$$

or, equivalently, if the orthogonality property

$$\mathbb{E}[\varepsilon|X] = 0$$

[27]This argument also covers the case $\delta = 0$ discussed in this appendix. Since no restrictions on the variance and covariance of the disturbance terms are made, it is possible that $\delta = 0$ and $corr(\nu_0, \nu_{-1}) \neq 0$.

holds. For more details see, for example, Györfi et al. [2002] or Stock and Watson [2012].

The expected value of Y conditional on X is the best predictor of Y given X in a sense that is made precise by the following result (for more details on the notion of conditional expectation and the related results see, for example, Billingsley [1995]).

Proposition 1 (Prediction property)**.** *The function that minimizes*

$$\mathbb{E}(Y - m(X))^2$$

is $m(X) = E[Y|X]$.

The orthogonality property, essential to the definition of a regression, is *necessary*[28] for its consistent estimation (Stock and Watson [2012]; Györfi et al. [2002]). The following proposition states the inferability of the regressions as introduced in definition 1 and concludes the econometric considerations.

Proposition 2 (Consistent inference of $\mathbb{E}[Y\,|\,X]$)**.**

Given a sample (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n) *from a regression model (5), the non-parametric estimation literature provides a large choice*[29] *of approaches that yield consistent estimates of the regression function* $f(X)$ *(Györfi et al. [2002]; James et al. [2013]).*

For most of the approaches the rate of convergence of the estimates to the regression function are also known.

9.7 Price = valuation informed by earnings level + investors' adjustment

This section gives the theoretical motivation of the passage from the valuation model in (11) to a non-linear regression relation in (14).

In brief, our approach is intuitive and works as follows. We start with a set of predictors of future income (e.g., earnings or cash flows). Market expectation of future income at horizon t is broken up via the decomposition property in (5) into the best (possibly non-linear) projection of future income by the predictor set and a term that collects the corrections to the projection due to other information available when forming the expectation in addition to the current values of the predictors set. The correction

[28] In particular, it implies $corr(\varepsilon, X) = 0$, which, in the case of the linear regression, ensures that no omitted variables bias the estimation of the coefficients.

[29] The list of approaches includes (but it is not limited to) local averaging estimates (including kernel, partitioning, and nearest neighbor estimates), least squares estimates (using splines, neural networks, and radial basis function networks), penalized least squares estimates, local polynomial kernel estimates, and orthogonal series estimates.

term complements the information in the observed data. Formally, it is orthogonal in the sense of the definition (6) to the predictors set.

Consequently, the price, which is modeled as the sum of discounted future income, decomposes into the sum (over all time horizons) of discounted best forecasts of future income by the predictors set (the regression function) plus a pricing correction term due to other information available to the market (the error term). The pricing correction fulfills the condition for being a regression error term since it is the sum of discounted corrections to future income expectations, each complying with the orthogonality property. The resulting valuation specification is hence a non-linear regression that can be consistently estimated using techniques from the field of non-parametric statistics.

We denote by `PREDICT.NI` a set of predictors of future earnings available at time 0 (current earnings, earnings growth, risk profile information about the firm, etc.).

We are now ready to state our main result:

Proposition 3 (Existence of regression specifications of the accounting model)**.**

1. `There exist specifications of the valuation relation (11) that are regressions (in the sense of definition 1):`

   Suppose prices are given by equation (11) and let `PREDICT.NI` *be a set of predictors of future NI available at time* 0 *(as specified above). Then, there exist* $f_{i,0}$*, a possibly non-linear, firm- and predictor-specific function, and* ε_i*, an error term, such that:*

$$P_{i,0} = f_{i,0}(\texttt{PREDICT.NI}_{i,0};\ r_{i,0}) + \varepsilon_{i,0} \tag{26}$$

   where

$$\mathbb{E}[\varepsilon_0 \,|\, \texttt{PREDICT.NI}_0] = 0. \tag{27}$$

2. `The regression functions are valuations incorporating expectations shaped only by the current values of the predictors PREDICT.NI:`

   Moreover,

$$f_{i,0}(\mathbf{x}\,;\, r_{i,0}) := \sum_{t=1}^{\infty} \frac{\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\,;\, r_{i,0}) \,|\, \texttt{PREDICT.NI} = \mathbf{x}]}{(1 + r_{i,0})^t}$$

   i.e., the regression function in the decomposition (26) amounts to valuations incorporating expectations of future earnings formed only on the basis of the current values of the predictors `PREDICT.NI`.

3. `The error terms represent investor correction based on informationother than the levels of the predictor set available at time` 0`:`

   The error term in the decomposition (26), ε_i, amounts to investor correction (grounded in other available information) to a price set upon expectations shaped only by the observed values of the predictors `PREDICT.NI`.

4. `The size of the error terms is a measure of expectation formation pertinence of the predictors set:`

   The size of the absolute value $|\varepsilon_i|$ reflects the extent to which predictor levels inform expectations of future earnings incorporated in prices.Higher absolute values indicate lower contributions of the predictor levels to pricing.

Condition (27) guarantees that the non-linear regression specifications of the accounting valuation (11) in (26) can be consistently estimated. The issue of omitted variables bias is structurally ruled out.

The stated result is general and flexible. It forms the basis of a specific implementation of the general accounting model (11) for testing hypotheses about the relation between prices and any choice of accounting constructs associated with prices. For a given set of predictors of future income streams, the result states the existence of a specification of the general valuation relation (11) that is a regression and, hence, that can be consistently inferred on data. The regression function is specific to the set of predictors. To different sets of predictors correspond different functions f as the relation to value is variable-specific. For each firm, the estimation yields an error term whose size gives the measure of the contribution of predictors' observed values to market setting of firm price. Hypotheses on the association between prices and the set of predictors can be tested on the magnitude of the estimated errors.

The choice of the set of predictors `PREDICT.NI` depends on the research question. Examples include (but are by no means limited to):

$$\texttt{PREDICT.NI} := NI,$$

if the research concerns earnings;

$$\texttt{PREDICT.NI} := (NI, B),$$

when the research concerns the pertinence of bottom-line items; and

$$\texttt{PREDICT.NI} = (CFO, TACC),$$

for studies on the incremental relevance of the decomposition of earnings into cash flows and accruals (CFO stands for cash flow from operations, while $TACC$ denotes total accruals).

Proof of the main result. Although proposition 3 is a direct consequence of the decomposition property in 5, we explain its derivation in a few of steps that better convey the intuition behind the decomposition in (26).

The proposition is an expression of the modern view of capital market research on the relation between current financial statement data and firm value as a two-step process (Bernard [1995]). First, the current information is used to project future financial statement data. Second, the link between those forecasts and current value must be specified: in our case, by the valuation expression in (11). While observed levels of the predictors project future earnings, other information is available to investors. They use this ”other information” to correct projections into expectations of future income streams (first link).

Consequently, the prices, given by the valuation expression in (11), can be split into two terms: a sum of discounted best projections of future income by the predictor set plus a sum of discounted corrections due to ”other information” available at time 0. The size of the second term informs about the importance of the contribution of ”other information” to the process of expectation formation.

Step 1. Current values of the predictors set project future earnings.

This step is formalized by applying the decomposition property 5 to the pair $(\mathtt{PREDICT.NI}_i,\ h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0}))$. For the firm i, we write:

$$\begin{aligned} h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0}) &= \mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0}) \,|\, \mathtt{PREDICT}_{i,0}] + \delta_{i,0}(t) \\ &:= g_{i,0}(\mathtt{PREDICT.NI}_{i,0};\ t) + \delta_{i,0}(t). \end{aligned} \tag{28}$$

That is to say, we decompose the function of future *NI* t time units ahead, into its best projection[30] by the current values of the predictors, $g_{i,0}(\mathtt{PREDICT.NI}_{i,0};\ t)$, and a left-over piece, orthogonal to the projection. The orthogonality property reads:

$$\mathbb{E}[\delta_{i,0}(t) \,|\, \mathtt{PREDICT.NI}_{i,0}] = 0.$$

The left-over piece amounts to the part of the function of future NI at time t that cannot be foreseen knowing only the current levels of the predictors.

The functions $g_{i,0}(\cdot\ ;\ t)$, the projections of future NI by the current values of the predictors set, describe the value-creation growth expectations for the firm i given the level of the predictors at time 0. They model[31] the persistence of earnings and are an expression of the factors that determine abnormal

[30] Recall that, according to proposition 1, $\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0}) \,|\, \mathtt{PREDICT.NI}_{i,0}]$ is the best predictor of the function of future NI given the current values of the predictor set, generally a non-linear function of the predictors that we denote by $g_{i,0}(\cdot\ ;\ t)$.

[31] The empirical specification of the valuation regression relation will assume that they are roughly similar for

earnings: firm size, product-type, capital intensity, barriers to entry, and accounting practices (see also section 5.3, and section 9.8 in this appendix).

Step 2. Expectations about future NI integrate the projections by current values of the predictors with other available information. This "other information" is complementary (orthogonal) to that contained in the observed predictors.

To see the implication of the decomposition of the function of future *NI* in (28) for the process of expectation formation, we condition with the set of information available at time 0 to obtain market expectation of the function h_t:

$$\begin{aligned} \mathbb{E}_0[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0})] &= \mathbb{E}_0[g_{i,0}(\texttt{PREDICT.NI}_i;\ t)] + \mathbb{E}_0[\delta_{i,0}(t)] \\ &= g_{i,0}(\texttt{PREDICT.NI}_i;\ t) + \mathbb{E}_0[\delta_{i,0}(t)] \\ &= g_{i,0}(\texttt{PREDICT.NI}_{i,0};\ t) + \varepsilon_{i,0}(t), \end{aligned} \tag{29}$$

since $\mathbb{E}_0[g_{i,0}(\texttt{PREDICT.NI}_{i,0};\ t)] = g_{i,0}(\texttt{PREDICT.NI}_{i,0};\ t)$, and where we denote $\mathbb{E}_0[\delta_{i,0}(t)]$ by $\varepsilon_{i,0}(t)$.

This decomposition reflects the fact that expectations about future NI are formed by adjusting the projections by current values of the predictors with other information available at time 0. The correction term $\varepsilon_i(t)$ amounts exactly to the contribution of the latter. Its complementarity is reflected by its orthogonality to the information in the current levels of the predictor set. To see this, note that, according to the theorem of iterated expectations:

$$\begin{aligned} \mathbb{E}[\varepsilon_i(t)\,|\,\texttt{PREDICT.NI}_{i,0}] &= \mathbb{E}\left[\mathbb{E}_0[\delta_i(t)]\,\middle|\,\texttt{PREDICT.NI}_{i,0}\right] \\ &= \mathbb{E}[\delta_i(t)\,|\,\texttt{PREDICT.NI}_{i,0}] = 0. \end{aligned}$$

Consequently, $|\varepsilon_i|$ is a measure of the pertinence of observed values of the predictors to expectations formation and hence to price setting. The larger $|\varepsilon_i(t)|$, the smaller the contribution of the current values of the predictors to shaping expectations about future earnings at time t.

Step 3. Aggregation of discounted expectations about future NI yields a firm- and predictor-specific decomposition of price into two terms: one collecting the contributions of the observed predictors to expectation formation and another assembling the corrections brought about by other available information.

comparable firms.

Expression (11) together with the decomposition in equation (29) implies:

$$\begin{aligned} P_{i,0} &= \sum_{t=1}^{\infty} \frac{\mathbb{E}[h_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0}) \,|\, \mathtt{PREDICT}_{i,0}]}{(1+r_{i,0})^t} + \sum_{t=1}^{\infty} \frac{\varepsilon_0(t)}{(1+r_{i,0})^t} \\ &=: f_{i,0}(\mathtt{PREDICT.NI}_i;\ r_i) + \varepsilon_{i,0}, \end{aligned}$$

where the errors ε_0 verify:

$$\mathbb{E}[\varepsilon_i \,|\, \mathtt{PREDICT.NI}_i] = 0.$$

This ends the proof. $\square$

9.8 The choice of proxies

Implementing the specification (14) requires that we specify proxies for risk, growth, and the economic and accounting determinants discussed in section 5.3. While the list of pertinent proxies is constructed based on the extant accounting and finance literature, the choice of the most appropriate variables is an empirical question, which we address below.

9.8.1 Definition of the proxies

Risk and growth. According to the extant valuation literature, the main determinants of the evolution of future income streams and, hence, of the shape of the functions in (13), are firm's cost of equity $r_{i,0}$ and firm growth $g_{i,0}$. Risk and growth include both direct proxies (see table 6) and indirect proxies; indirect proxies include those for risk (investment, see table 8, and financing, see table 10) and for growth (profitability, see table 7, and payout policy, see table 9).

P/B, together with size, is one of the risk factors in the original Fama-French three-factor model (Fama and French [1992]). More recently, Penman et al. [2018] establish conditions under which P/B is a valid measure of equity investment risk. They show that the conditions under which P/B is a valid risk characteristic involve a particular form of accounting that resembles GAAP. They conclude that P/B indicates expected return because it forecasts expected earnings growth and the risk that the expectation may not be met.

We use two proxies for size: current total assets and previous-year market size. The two proxies are based on the relative total assets/market value of the firm in the cross-section and hence are cross-section-specific. The firms in a cross-section are sorted into ventiles (based on each of the two proxies) and the $Size_i$ variables are defined as the cross-section ventile to which the firm i belongs.

We use cash flow volatility as a measure of economic risk. This volatility is firm-specific and is calculated with at least six values over the previous eight years. Since cash flow volatility is correlated to level of cash flow and volatility of accruals (the latter is one of our accounting reporting measures), we separate the cash flow volatility component that is orthogonal on these last two variables by running a non-linear regression of cash flow volatility on cash flow and volatility of accruals. The firm's ventiles of the innovation term in the regression is our measure of cash flow volatility.

We consider the following direct measures of growth: past earnings ($NI.g$), total assets ($TA.g$), sales ($Revt.g$), and book ($B.g$) median growth.

Variable Name	Definition
P/B	Previous year price to book ratio
Size - TA	Firm's ventile of total assets in the cross-section
Size - Mktv	Firm's ventile of previous year $Mktv$ in the cross-section
$EPS.g$ - Earnings growth	Change in earnings per share (EPS)/lagged EPS
$Revt.g$ - Sales growth	Change in sales/lagged sales
$TA.g$ - Assets growth	Change in total assets (AT)/lagged AT
$B.g$ - Book growth	Change in common equity (ceq)/lagged equity
Economic risk	Firm's volatility of cash-flow per share

Table 6: **Direct proxies of risk and growth.** Growth rates are calculated as median rate over at least six of the eight previous years. Volatility is calculated with at least six of the eight previous years' values.

While the market-to-book ratio functions as an important determinant of expected returns (Fama and French [1992], more recently, the capital markets literature has emphasized the importance of investment and profitability proxies to valuation and market performance (Chen et al. [2011]; Novy-Marx [2013]; Fama and French [2015]; Ball et al. [2016]).

Profitability. Novy-Marx [2013] shows that profitability predicts gross profit growth, earnings growth, and free cash flow growth. We consider six proxies for profitability. We include the three measures based on gross profit in Novy-Marx [2013], Fama and French [2015], and Ball et al. [2016]. We also include two commonly used profitability measures based on operating income (Kahle and Stulz [2017]), as well as return on assets (ROA) and return on equity (ROE).

Investment. Chen et al. [2011] argue that the investment factor plays a similar role to that of the Fama and French [1993] value factor. Firms with higher valuation ratios have more opportunities for growth, invest more, and earn lower expected returns than firms with lower valuation ratios. Theoretically, firms invest more when their profitability is high (e.g., Fama and French [2006]). As such, controlling for profitability, investment should be negatively correlated with expected returns.

Variable Name	Definition
$Prof_1$ - Operating income-to-Assets	Operating income (OI) before depreciation (oibdp) minus interest (xint) minus taxes (txt)/lagged assets
$Prof_2$ - R&D-adjusted OI-to-Assets	OI/assets plus R&D/lagged assets
$Prof_3$ - Gross profit-to-Assets	Revenues (sale) minus cost of goods sold (cogs) divided by assets (Novy-Marx [2013])
$Prof_4$ - Gross profit-to-Book value	sale-cogs-SG&A expenses (xsga) divided by book equity (ceq) (Fama and French [2015])
$Prof_5$ - Operating profitability	(sale-cogs-xsga+ R&D expenses (xrd) divided by assets (Ball et al. [2016])
ROA - Return on assets	Earnings before extraordinary items (ib)/lagged assets
ROE - Return on equity	Earnings before extraordinary items/lagged book equity

Table 7: **Proxies for profitability.**

Following Chen et al. [2011], we include investment-to-assets (I/A) as the annual change in gross property, plant, and equipment (Compustat annual item "ppegt") plus the annual change in inventories ("invt") divided by the lagged book value of assets ("ta").[32] Fama and French [2015] define the investment factor based on change in total assets divided by last year's total assets. We consider a similar variable as one of our direct measures of growth.

Variable Name	Definition
Inv_1 - Capital Expenditures-to-Assets	Capital expenditures (capx) / lagged assets (at)
Inv_2 - R&D intensity	R&D (xrd) / lagged assets[33]
Inv_3 - Tangibility	Fixed assets (ppent) divided by assets
Inv_4 - Inventory-to-Assets	Inventory (invt) divided by assets
Inv_5 - Cash-to-Assets	Cash and marketable securities (che) divided by assets
Inv_6 - SG&A Investment component	Main SG&A-Maintenance Main SG&A (see Enache and Srivastava [2018])
Inv_8 - Investment-to-Assets	Change in gross property, plant, and equipment (ppegt) plus the change in inventories (invt), divided by the lagged assets (Chen et al. [2011])

Table 8: **Proxies for investment.**

Recent empirical literature has documented a steady increase in intangible capital accumulation over time. Firms with lower tangibility ratios have lower valuations, most likely due to higher risk. A predominance of lower tangibility assets reduces the amount a firm can pledge as collateral. As a con-

[32] Changes in property, plant, and equipment capture capital investment in long-lived assets used in operations over many years, such as buildings, machinery, furniture, and other equipment. Changes in inventories capture working capital investment in short-lived assets used in a normal operating cycle, such as merchandise, raw materials, supplies, and work in progress.

sequence, firms with lower asset tangibility are more vulnerable to adverse economic conditions (such as recessions and credit crunches, i.e., low liquidity). The difficulty in raising funds during such events might lead to costly external financing or impair the firm's ability to undertake profitable investments. Almeida and Campello [2007] show that tangibility has a significant impact on investment. More precisely, investment–cash flow sensitivities increase with tangibility. We use the tangibility ratio (*Tang*) (ppent/at) as a proxy for this firm characteristic.

Other measures of investment we consider include capital expenditure/assets, R&D intensity, and inventory/assets.

Since a large proportion of intangible investments is made in avenues other than R&D (Corrado et al. [2005], many studies use expenses reported in the SG&A category as a proxy for total intangible investments. However, many SG&A expenses support current, rather than future, operations. For this reason, we follow Enache and Srivastava [2018] and divide SG&A expenses based on whether an expense is intended to produce a current or a future benefit. The investment component of main SG&A is one of our measures of investment.

Payout policy. Beaver and Ryan [2005] find that payout policy is a significant predictor of future growth in sales and book value over a five-year horizon, whereas past growth in sales and book value does not have predictive power beyond two years. The payout ratio measures growth in terms of reinvestment: full payout, low growth, zero payout, and high growth. We use four different payout measures, which are defined in Table 9.

Variable Name	Definition
PO_1 - Dividends-to-Assets	Cash dividends on ordinary stock (dvc), divided by lagged assets
PO_2 - Total payout-to-Assets	dvc + prstkc, divided by lagged assets
PO_3 - Repurchases-to-Assets	Purchase of stock (prstkc) minus any decrease in preferred stock (pstk), divided by lagged assets

Table 9: **Proxies for payout policy.**

Financing. The positive relation between the equity cost of capital and financial leverage is a core tenet of financial economics going back to the seminal work by Modigliani and Miller (Modigliani and Miller [1958]; Taggart [1991]).[34] More precisely, while leverage in itself does not affect equity risk, a

[34]Empirical research has had difficulty documenting a leverage risk premium in stock returns.

firm's operating risk is amplified through financial leverage (Skogsvik et al. [2012]). We consider six leverage measures, defined in Table 10.

Variable Name	Definition
Fin_1 - Book leverage	Long term debt (dltt) plus debt in current liabilities (dlc), divided by assets (at)
Fin_2 - Market leverage	Long term debt (dltt) plus debt in current liabilities, divided by assets (at) minus book equity (ceq) plus the market value of common equity (csho*prcc)
Fin_3 - Net Leverage	Long-term debt (dltt) plus debt in current liabilities minus cash (che), divided by assets (at)
Fin_4 - Interest-to-Assets	Interest expense (xint) divided by lagged assets
Fin_5 - Total liabilities-to-Market	Total liabilities (lt) divided by the market value of common equity (csho*prcc)
Fin_6 - Net debt-to-Market	dltt+dlc+pstk+dvpa-che divided by the market value of common equity (csho*prcc) divided by book (seq)

Table 10: **Proxies for financing.**

Industrial organization (IO) (or economic). Complementary to each other, the economics literature and the strategic management literature have studied how industry structure determines the profit generating process of firms. This provides a theoretical framework for the evolution of profits and explains its variation through differences in the competitive environment. The well-established *structure-conduct–performance* relation (Bain [1956]) implies that low/high levels of industry concentration[35] should be associated with normal/abnormal profitability. The *SCP* relation implies that industry characteristics play a determinant role in explaining firm profits. Firms in the same industry should converge to a common industry profit rate.

Inter-industry measures. The IO literature identifies the inter-industry traits with an impact on profits persistence and hence on the evolution of future earnings. The most important ones are industry concentration, barriers to entry, and product type.

We use the Fama-French 48 industry classification. We assume that the industry classification serves as a proxy for the product type and barriers to entry industry structure dimensions. We measure the concentration with the Herfindahl-Hirschman Index (HHI), which is based on the relative sales of the

[35] A low level of industry concentration reflects the presence of a substantial number of similar firms and no substantial barriers to entry.

firms in a given industry.[36] If we denote by $Total.SALE_I$ the sum of the sales of all firms in the industry I, and by $Mkt.sh_i$ the market share of the firm i,

$$Mkt.sh_i = \frac{SALE_i}{Total.SALE_I} \tag{30}$$

and the HHI is defined as:

$$HHI = \sum_{i=1}^{N} (Mkt.sh_i)^2 , \tag{31}$$

where N is the number of firms in industry I.

Intra-industry (firm-specific) measures. While the traditional IO literature uses the industry as the fundamental unit of analysis, intra-industry efficiency differences are pervasive. Within the same industry, efficient firms obtain large market shares and earn abnormal economic rents. Market share is one effect of scale-related efficiency. It is also a measure of market power related to quality differences, patents, and price discrimination. Other related intra-industry measures include firm-size and capital intensity (Lev [1983]; Baginski et al. [1999]).

The proxy for market share, $Mkt.Sh_i$, is defined in (30). The proxy for capital intensity, $Cap.Int$, is calculated by dividing the total assets of a company by its sales. It is reciprocal of total asset turnover ratio.

Variable Name	Definition
Industry	The Fama-French 48 industry classification
Industry concentration	The HHI index defined in (31)
Market share	The ratio of firm's sale over the sum of the sales of all firms in the industry (30)
Capital intensity	The ratio of total assets of a company to its sales

Table 11: **Industrial organization variables.**

Accounting. The structure of GAAP influences the ability of accounting numbers to serve as proxies for the economic concepts identified above. In particular, accounting standards and practice have an effect on the persistence of abnormal earnings (Feltham and Ohlson [1995, 1996]; Zhang [2000]; Cheng [2005]). While under unbiased accounting and perfect competition a firm's residual earnings equals its cost of equity, if the competition is imperfect, the firm can charge prices higher than its costs, resulting in economic rents and abnormal ROE that is no longer zero. Under conservative accounting,

[36] Sales provide, possibly, a better measure of the real activity of a firm than total assets or earnings as they are less influenced by accounting manipulation.

accounting measures depart from economic measures and a firm's abnormal ROE can be different from zero even when the firm operates under perfect competition. The level of conservative accounting is determined by both industry- and firm-specific factors. While industry characteristics play an important role in determining the level of non-discretionary or unconditional conservatism, managerial preferences are an obvious firm-specific driver. Extant accounting literature documents that aspects of the account-

Variable Name	Definition
SPI	The ratio of special items of a company to its assets
$C.score$	The score constructed in Penman and Zhang [2002].
$R\&D$	The ratio of $R\&D$ expenses of a company to sales
$Lifo$	The ratio of $Lifo$ expenses of a company to sales
Abnormal accruals - $Jones$	The abnormal accruals form Jones-type model in the specification of Chen et al. [2011]
Abnormal accruals - $D-D$	The absolute value of the residual of the model in (Dechow and Dichev [2002])
Size of accruals - $Acc\ size$	The median (over at least 6 of the last 8 years) of the accruals scaled by average of total assets.
Volatility of accruals - $Acc\ vol$	The volatility of accruals per share (over at least 6 of the last 8 years)
Correlation - $cor.CFO.ACC$	The correlation between cash flows and accruals calculated over at least 6 of the last 8 years)

Table 12: **Proxies for accounting.**

ing recognition process other than conservatism have an impact on the prediction of future earnings. Burgstahler et al. [2002] show that the presence and the sign of special items significantly affect the persistence of earnings. Earnings of firms reporting non-zero special items are less persistent, with the magnitude of the persistence coefficient for earnings of firms with positive special items being statistically higher than that for firms with negative special items. Furthermore, evidence in Cready et al. [2012] suggests that negative special items signal real future performance improvements. Alford and Berger [1999] document that special items impair analysts' ability to predict future earnings.

We use accruals volatility as a measure of accounting reporting uncertainty. This volatility is firm-specific and is calculated with at least six values over the previous eight years. Since accruals volatility is correlated with the level and the volatility of cash flows (the latter is one of our risk measures), we separate the accruals volatility component that is orthogonal on these last two variables by running a non-linear regression of accruals volatility on cash flow and its volatility. The firm's ventiles of the innovation term in the regression is our measure of accruals volatility.

The proxies used in the empirical estimation of the non-linear specification in (14) are neither unique nor the only ones that can be used. Other variables can extend or amend the set of proxies. In robustness tests, we experimented with firm's beta as a proxy for risk, as well as with the conservatism index in Penman and Zhang [2002] as a proxy for conservatism. The results are qualitatively equal to the ones we present below.

9.8.2 The importance of the proxies

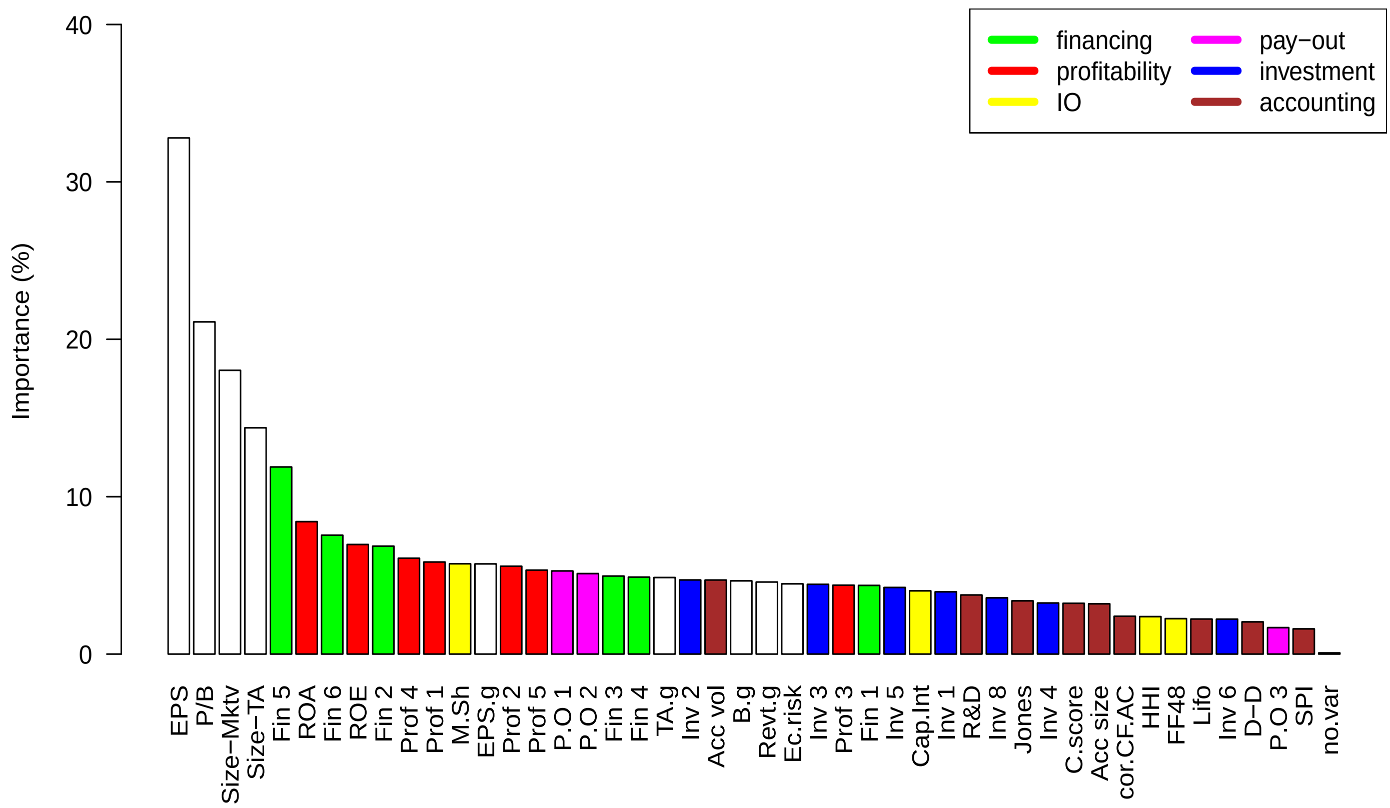


Figure 17: **Average importance of the explanatory variables.** The graph displays the average importance of each of the 45 explanatory variables we consider and shows a clear ordering. Measures of financing and profitability are among the most relevant to the price–earnings association, while proxies for investment and accounting uncertainty are the least relevant.

The importance of an explanatory variable is defined as the (percentage) reduction in the mean squared error when the given variable is removed from the set of explanatory variables. Figure 17 shows the average importance of each of the 45 explanatory variables defined above.

9.9 Slope estimation

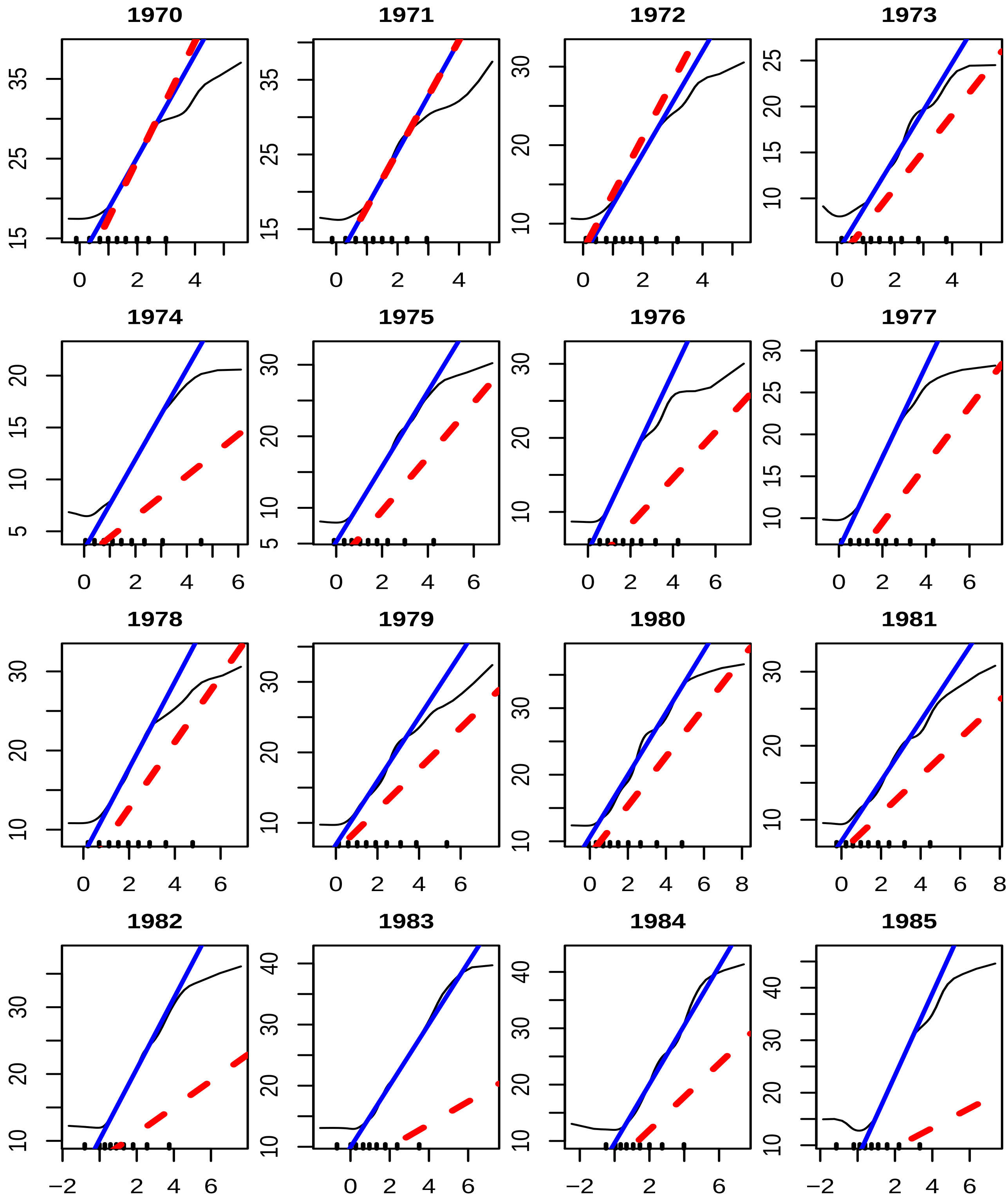


Figure 18: **Linear fit for the estimated functional relation between earnings and prices: 1970–1985.** The graphs display the estimated partial dependence function (17) for all firms in the cross-section in the title of the graph (solid, black), the linear fit for the range of earnings where the dependence is approximately linear (roughly between the third and eighth deciles) (solid, blue), and the linear fit of the levels regression (2) (dotted, red). We see clearly that the linear regression earnings response coefficient (ERC) is much smaller than the corresponding non-linear ERC.

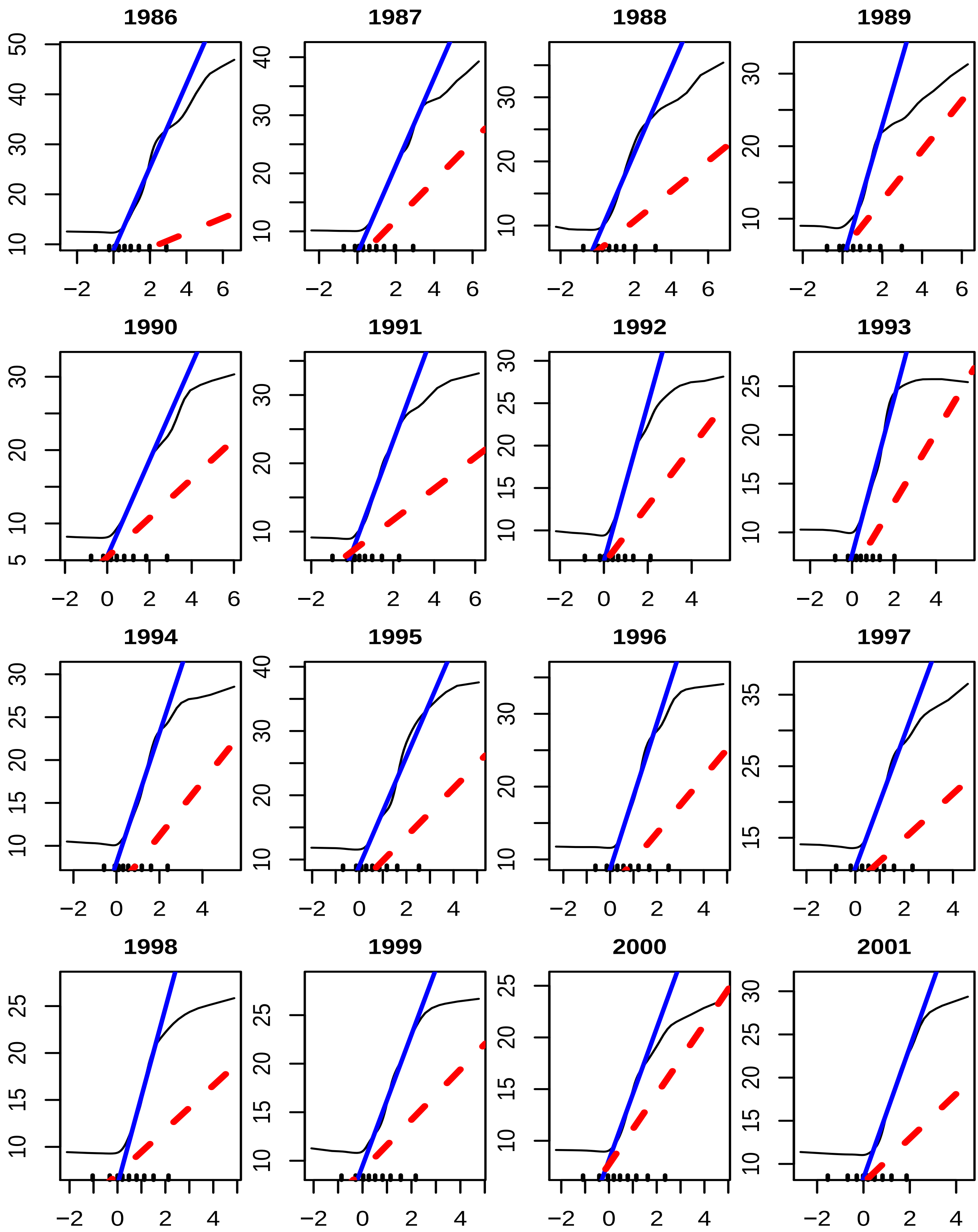


Figure 19: **Linear fit for the estimated functional relation between earnings and prices: 1986–2001.** The graphs display the estimated partial dependence function (17) for all firms in the cross-section in the title of the graph (solid, black), the linear fit for the range of earnings where the dependence is approximately linear (roughly between the third and eighth deciles) (solid, blue), and the linear fit of the regression (2) (dotted, red). We see clearly that the linear regression earnings response coefficient (ERC) is much smaller than the corresponding non-linear ERC.

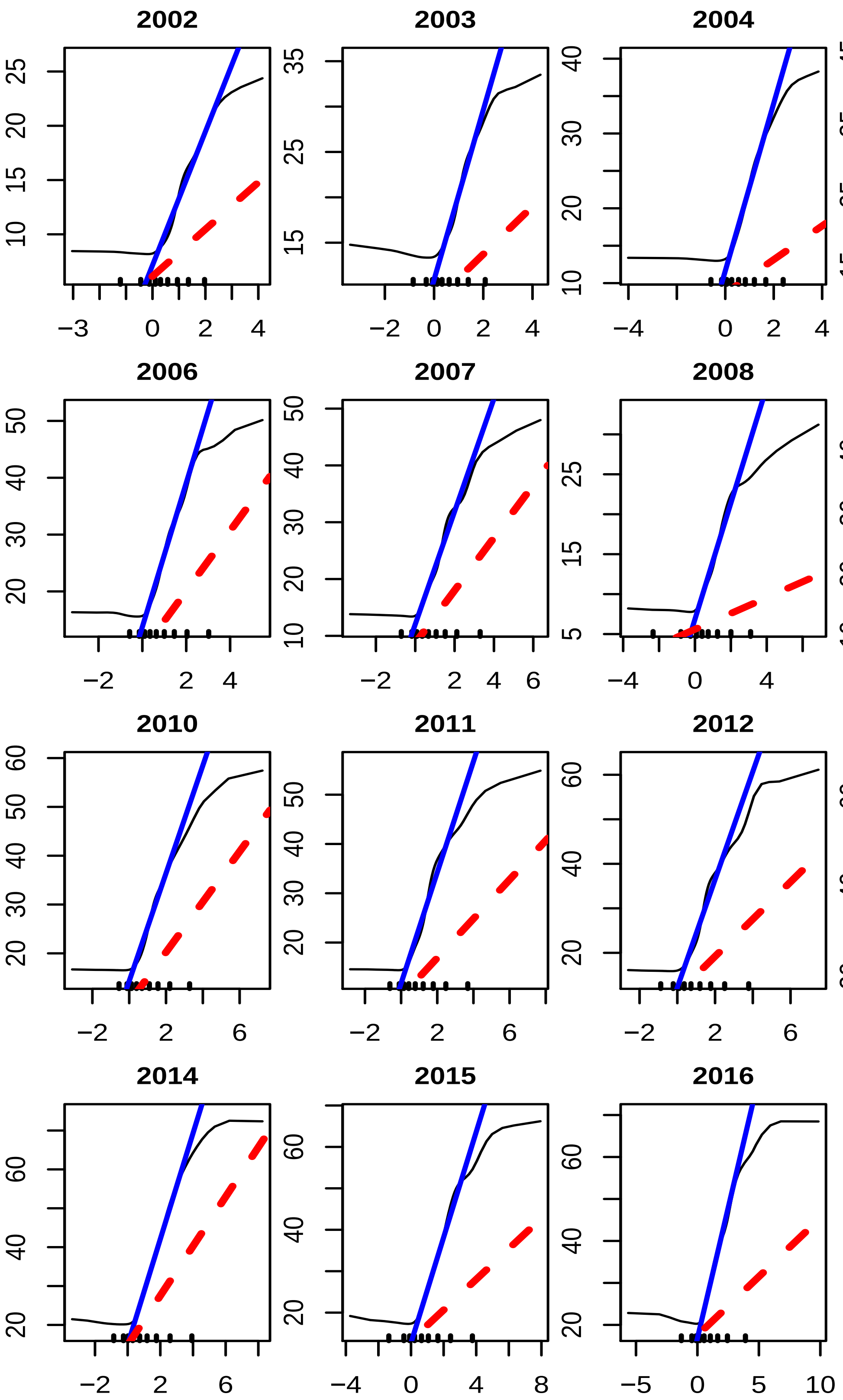


Figure 20: **Linear fit for the estimated functional relation between earnings and prices: 2002–2016.** The graphs display the estimated partial dependence function (17) for all firms in the cross-section in the title of the graph (solid, black), the linear fit for the range of earnings where the dependence is approximately linear (roughly between the third and eighth deciles) (solid, blue), and the linear fit of the regression (2) (dotted, red). We see clearly that the linear regression earnings response coefficient (ERC) is much smaller than the corresponding non-linear ERC.

9.10 Market price–earnings ratio

ERC_0 improves over the market price–earnings ratio

$$\frac{(\text{average price})_0}{(\text{average } NI)_0} \tag{32}$$

that is commonly used in the literature to gauge the size of the ERCs. Using the expression in (19), the denominator becomes:

$$\frac{1}{n}\sum_i L_{i,0} + \frac{1}{n}\sum_{i\,:\,NI_{i,0}\in[L_{i,0},U_{i,0}]} k_{i,0}\,NI_{i,0} + \frac{1}{n}\sum_{i\,:\,NI_{i,0}>U_{i,0}} U_{i,0}$$

(note that $1/n\sum_i \varepsilon_{i,0} \approx \mathbb{E}[\varepsilon_0] = 0$), which yields

$$\begin{aligned}
\frac{(\text{average price})_0}{(\text{average } NI)_0} &= \frac{\frac{1}{n}\sum_i L_{i,0}}{(\text{average } NI)_0} + \frac{\frac{1}{n}\sum_{i\,:\,NI_{i,0}\in[L_{i,0},U_{i,0}]} k_{i,0}\,NI_{i,0}}{(\text{average } NI)_0} + \frac{\frac{1}{n}\sum_{i\,:\,NI_{i,0}>U_{i,0}} U_{i,0}}{(\text{average } NI)_0} \\
&= A_0 + \frac{\frac{1}{n}\sum_{i\,:\,NI_{i,0}\in[L_{i,0},U_{i,0}]} k_{i,0}\,NI_{i,0}}{\frac{1}{n}\sum_{i\,:\,NI_{i,0}\in[L_{i,0},U_{i,0}]} NI_{i,0}} \times \frac{\frac{1}{n}\sum_{i\,:\,NI_{i,0}\in[L_{i,0},U_{i,0}]} NI_{i,0}}{(\text{average } NI)_0} + B_0 \\
&= A_0 + \frac{\sum_j k_{j,0} a_{j,0}}{\sum_j k_{j,0}} \times D_0 + B_0.
\end{aligned}$$

While the term $\sum_j k_{j,0} a_{j,0}/\sum_j k_{j,0}$ could be thought of as a summary of the variability of the firm-specific ERC in the cross-section, the constants A_0, B_0, and D_0 are arbitrary and only related to the interaction between the distribution of earnings in the cross-section and the three ranges in the locally linear representation (19).

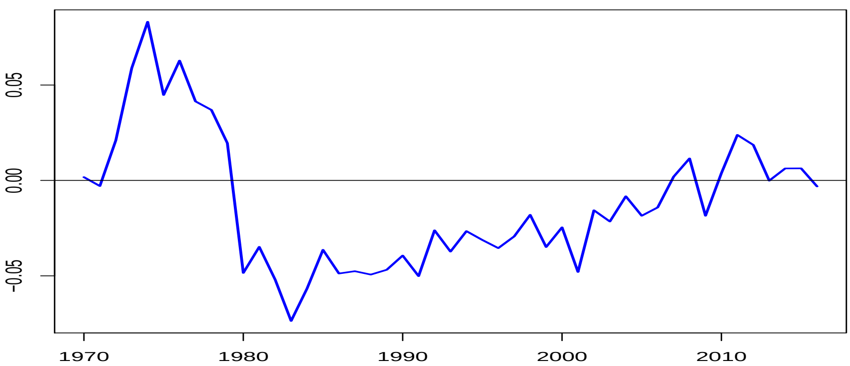


Figure 21: **The risk premium implied by the ratio of average price to average earnings per share in cross-sections.** The graphs display the time series of risk premium (33) implied by ratio of average price to average earnings per share in cross-sections. The risk premium is mostly negative for the period from 1980 to 2010.

The statistical unreliability of the price–earnings ratio can be seen when analyzing the average expected rate of return implied by the sample's average price–earnings ratio. Estimating the cross-sectional risk premium as:

$$\frac{(\text{average NI})_t}{(\text{average price})_t} - r_{f,t} \tag{33}$$

yields negative risk premiums for most years t in the period from 1980 to 2010.

References

ALEDO MARTINEZ, J., J. M. GARCIA LARA, M. T. GONZALEZ PEREZ, and C. A. GRAMBOVAS. 'An empirical assessment of proposed solutions for resolving scale problems in value relevance accounting research.' Accounting & Finance (2019).

ALFORD, A. W., and P. G. BERGER. 'A simultaneous equations analysis of forecast accuracy, analyst following, and trading volume.' Journal of Accounting, Auditing & Finance 14 (1999): 219–240.

ALMEIDA, H., and M. CAMPELLO. 'Financial constraints, asset tangibility, and corporate investment.' The Review of Financial Studies 20 (2007): 1429–1460.

BAGINSKI, S. P., K. S. LOREK, G. L. WILLINGER, and B. C. BRANSON. 'The relationship between economic characteristics and alternative annual earnings persistence measures.' The Accounting Review 74 (1999): 105–120.

BAIN, J. S. Barriers to New Competition. Harvard University Press (1956).

BALACHANDRAN, S., and P. MOHANRAM. 'Is the decline in the value relevance of accounting driven by increased conservatism?' Review of Accounting Studies 16 (2011): 272–301.

BALL, R., and P. BROWN. 'An empirical evaluation of accounting income numbers.' Journal of Accounting Research 6 (1968): 159–178.

BALL, R., J. GERAKOS, J. T. LINNAINMAA, and V. NIKOLAEV. 'Accruals, cash flows, and operating profitability in the cross section of stock returns.' Journal of Financial Economics 121 (2016): 28 – 45.

BARTH, M., and G. CLINCH. 'Scale effects in capital markets-based accounting research.' Journal of Business Finance and Accounting 36 (2009): 253–288.

BARTH, M. E., W. H. BEAVER, and W. R. LANDSMAN. 'Relative valuation roles of equity book value and net income as a function of financial health.' Journal of Accounting and Economics 25 (1998): 1 – 34.

BARTH, M. E., and S. KALLAPUR. 'The effects of cross-sectional scale differences on regression results in empirical accounting research.' Contemporary Accounting Research 13 (1996): 527–567.

BARTH, M. E., W. R. LANDSMAN, M. LANG, and C. WILLIAMS. 'Are IFRS-based and US GAAP-based accounting amounts comparable?' Journal of Accounting and Economics 54 (2012): 68–93.

BEAVER, W., R. LAMBERT, and D. MORSE. 'The information content of security prices.' Journal of Accounting and Economics (1980).

BEAVER, W. H., and S. G. RYAN. 'Conditional and unconditional conservatism:concepts and modeling.' Review of Accounting Studies 10 (2005): 269–309.

BERNARD, V. L. 'The feltham-ohlson framework: Implications for empiricists.' Contemporary Accounting Research 11 (1995).

BIDDLE, G., P. CHEN, and G. ZHANG. 'When capital follows profitability: Non-linear residual income dynamics.' Review of Accounting Studies 6 (2001): 229–265.

BILLINGSLEY, P. Probability and measure. Wiley series in probability and mathematical statistics (1995).

BREIMAN, L. 'Random forests.' Machine Learning 45 (2001a): 5–32.

—. 'Statistical modeling: The two cultures (with comments and a rejoinder by the author).' Statist. Sci. 16 (2001b): 199–231.

BURGSTAHLER, D., J. JIAMBALVO, and T. SHEVLIN. 'Do stock prices fully reflect the implications of special items for future earnings?' Journal of Accounting Research 40 (2002): 585–612.

CHEN, L., R. NOVY-MARX, and L. ZHANG. 'An alternative three-factor model.' Technical report, Working Paper (Washington University in St. Louis) (2011).

CHENG, Q. ‘What determines residual income?’ The Accounting Review 80 (2005): 85–112.

CHRISTIE, A. A. ‘On cross-sectional analysis in accounting research.’ Journal of Accounting and Economics 9 (1987): 231 – 258.

COLLINS, D. W., and S. KOTHARI. ‘An analysis of intertemporal and cross-sectional determinants of earnings response coefficients.’ Journal of Accounting and Economics 11 (1989): 143 – 181.

CORE, J. E., W. R. GUAY, and A. VAN BUSKIRK. ‘Market valuations in the New Economy: an investigation of what has changed.’ Journal of Accounting and Economics 34 (2003): 43–67.

CORE, J. E., and C. M. SCHRAND. ‘The effect of accounting-based debt covenants on equity valuation.’ Journal of Accounting and Economics 27 (1999): 1 – 34.

CORRADO, C., C. HULTEN, and D. SICHEL. ‘Measuring capital and technology: An expanded framework.’ In: Measuring Capital in the New Economy. The University of Chicago Press (2005).

CREADY, W. M., T. J. LOPEZ, and C. A. SISNEROS. ‘Negative special items and future earnings: Expense transfer or real improvements?’ The Accounting Review 87 (2012): 1165–1195.

DECHOW, P. M., and I. D. DICHEV. ‘The quality of accruals and earnings: The role of accrual estimation errors.’ The Accounting Review 77 (2002): 35–59.

EASTON, P. D., and M. E. ZMIJEWSKI. ‘Cross-sectional variation in the stock market response to accounting earnings announcements.’ Journal of Accounting and Economics 11 (1989): 117 – 141.

EDWARDS, E., and P. BELL. The Theory and Measurement of Business Income. University of California Press (1961).

ENACHE, L., and A. SRIVASTAVA. ‘Should intangible investments be reported separately or commingled with operating expenses? new evidence.’ Management Science 64 (2018): 3446–3468.

FAMA, E., and K. FRENCH. 'The cross-section of expected stock returns.' Journal of Finance (1992).

FAMA, E. F., and K. R. FRENCH. 'Common risk factors in the returns on stocks and bonds.' Journal of Financial Economics 33 (1993): 3–56.

—. 'Industry costs of equity.' Journal of Financial Economics 43 (1997): 153 – 193.

—. 'Profitability, investment and average returns.' Journal of Financial Economics 82 (2006): 491 – 518.

—. 'A five-factor asset pricing model.' Journal of Financial Economics 116 (2015): 1 – 22.

FELTHAM, G. A., and J. A. OHLSON. 'Valuation and clean surplus accounting for operating and financial activities.' Contemporary Accounting Research 11 (1995): 689–731.

—. 'Uncertainty resolution and the theory of depreciation measurement.' Journal of Accounting Research 34 (1996): 209–234.

FISCHER, P. E., and R. E. VERRECCHIA. 'The effect of limited liability on the market response to disclosure.' Contemporary Accounting Research 14 (1997): 515–541.

FREEMAN, R. N., and S. Y. TSE. 'A nonlinear model of security price responses to unexpected earnings.' Journal of Accounting Research 30 (1992): 185–209.

GOLDSTEIN, A., A. KAPELNER, J. BLEICH, and E. PITKIN. 'Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation.' Journal of Computational and Graphical Statistics 24 (2015): 44–65.

GOW, I., D. LARCKER, and P. REISS. 'Causal inference in accounting research.' Journal of Accounting Research 54 (2016): 477–523.

GU, S., B. KELLY, and D. XIU. 'Empirical asset pricing via machine learning.' Review of Financial Studies 33 (2020): 2223–2273.

GYÖRFI, L., M. KOHLER, A. KRZYZAK, and H. WALK. 'A distribution-free theory of nonparametric regression.' Springer Series in Statistics (2002).

HASTIE, T., R. TIBSHIRANI, and J. FRIEDMAN. The Elements of Statistical Learning. Springer (2009).

HAYN, C. 'The information content of losses.' Journal of Accounting and Economics 20 (1995): 125–153.

HOLTHAUSEN, R. W., and R. L. WATTS. 'The relevance of the value-relevance literature for financial accounting standard setting.' Journal of Accounting and Economics 31 (2001): 3–75.

JAMES, G., D. WITTEN, T. HASTIE, and R. TIBSHIRANI. An Introduction to Statistical Learning with Applications in R. Springer (2013).

KAHLE, K. M., and R. M. STULZ. 'Is the US Public Corporation in Trouble?' Journal of Economic Perspectives 31 (2017): 67–88.

KAROLYI, A., and S. VAN NIEUWERBURGH. 'New methods for the cross-section of returns.' Review of Financial Studies (2020).

KORMENDI, R., and R. LIPE. 'Earnings innovations, earnings persistence, and stock returns.' The Journal of Business 60 (1987): 323–345.

KOTHARI, S. 'Capital markets research in accounting.' Journal of Accounting and Economics 31 (2001): 105–231.

KOTHARI, S., and J. L. ZIMMERMAN. 'Price and return models.' Journal of Accounting and Economics 20 (1995): 155 – 192.

KOTHARI, S. P., and J. SHANKEN. 'Time-series coefficient variation in value-relevance regressions: A discussion of Core, Guay, and Van Buskirk and new evidence.' Journal of Accounting and Economics 34 (2003): 69–87.

KOTHARI, S. P., and C. WASLEY. 'Commemorating the 50-year anniversary of ball and brown (1968): The evolution of capital market research over the past 50 years.' Journal of Accounting Research 57 (2019): 1117–1159.

LEV, B. 'Some economic determinants of time-series properties of earnings.' Journal of Accounting and Economics 5 (1983): 31–48.

LIU, J., and J. THOMAS. 'Stock returns and accounting earnings.' Journal of Accounting Research 38 (2000): 71–101.

MODIGLIANI, F., and M. H. MILLER. 'The cost of capital, corporation finance and the theory of investment.' The American Economic Review 48 (1958): 261–297.

NOVY-MARX, R. 'The other side of value: The gross profitability premium.' Journal of Financial Economics 108 (2013): 1 – 28.

OHLSON, J. 'Earnings, book values, and dividends in equity valuation.' Contemporary Accounting Research 11 (1995): 661–687.

OHLSON, J. A., and B. E. JUETTNER-NAUROTH. 'Expected eps and eps growth as determinantsof value.' Review of Accounting Studies 10 (2005): 349–365.

PEASNELL, K. 'Some formal connections between economic values and yields and accounting numbers.' Journal of Business Finance and Accounting 9 (1982): 361–381.

PENMAN, S. H., F. REGGIANI, S. A. RICHARDSON, and I. TUNA. 'A framework for identifying accounting characteristics for asset pricing models, with an evaluation of book-to-price.' European Financial Management 24 (2018): 488–520.

PENMAN, S. H., and X.-J. ZHANG. 'Accounting conservatism, the quality of earnings, and stock returns.' The Accounting Review 77 (2002): 237–264.

PETERSEN, M. A. 'Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches.' The Review of Financial Studies 22 (2008): 435–480.

PREINREICH, G. 'The fair value and yield of common stock.' The Accounting Review (1936).

ROYCHOWDHURY, S., and R. L. WATTS. 'Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting.' Journal of Accounting and Economics 44 (2007): 2 – 31.

SHIN, S. H. 'Disclosures and asset returns.' Econometrica 71 (2003): 105–133.

SKOGSVIK, K., S. SKOGSVIK, and H. THORSELL. 'Disentangling the enterprise book-to-price and leverage effects in stock returns - a commentary on penman, richardson and tuna (2007).' (2012). Available at SSRN: https://ssrn.com/abstract=2046522 or http://dx.doi.org/10.2139/ssrn.2046522.

SRIVASTAVA, A. ‘Why have measures of earnings quality changed over time?’ Journal of Accounting and Economics 57 (2014): 196 – 217.

STOCK, J. H., and M. W. WATSON. Introduction to Econometrics. Addison-Wesley (2012).

SUBRAMANYAM, K. R. ‘Uncertain precision and price reactions to information.’ The Accounting Review 71 (1996): 207–219.

TAGGART, R. A. ‘Consistent valuation and cost of capital expressions with corporate and personal taxes.’ Financial Management 20 (1991): 8–20.

ZHANG, G. ‘Accounting information, capital investment decisions, and equity valuation: Theory and empirical implications.’ Journal of Accounting Research 38 (2000): 271–295.