Assessing the price-earnings association in the age of machine learning

Catalin Starica*

Abstract

Most of the seminal papers mapping the relation between earnings and security prices predate the recent exponential developments in the field of machine learning. Our analysis is an example of how the new powerful non-linear estimation techniques and three dimensional visualization of data can provide the accounting researcher with new insights and/or help her document more forcefully patterns predicted by theoretical considerations.

We show that state-of-the-art linear models are problematic for hypothesis testing when fit to the non-linear relation between share prices and earnings. To bring remedy to the failure of linear modeling, we introduce a non-linear research design based on rigorous statistical considerations and accounting input and which consistently estimates prices' relation to earnings.

We validate the non-linear research design by verifying that the non-linear levels regression earnings-response coefficients (ERC) have the 'right' size and yield economically justifiable risk-premium values. Consequently, the non-linearity of the price-earnings association provides a simple explanation for the small ERCs observed by prior research based on the linear model.

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

JEL classification: G10, G30, M41

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

I thank Mary Barth, Jan Marton, Stephen Penman, Kenth Skogsvik, Jacob Thomas, and the seminar participants at Stanford University for helpful suggestions and comments that have considerably improved the paper.

1 Introduction

One of the major themes of capital markets accounting research concerns mapping the relation between accounting earnings and security prices. Most of the seminal papers addressing the estimation of this relation date from the 80's, '90s and the early 2000s. We revisit this theme availing ourselves of more recently-developed statistical and visualization methodology from machine learning and data analytics. Our analysis exemplifies how powerful non-linear estimation techniques and threedimensional visualization of data can help the accounting researcher have new insights or document more forcefully patterns predicted by theoretical considerations.

The underlying model for the price-earnings association typically states that firm value is a function of expected future dividends, which in turn are assumed to be functions of expected future earnings and current earnings. Researchers then chose commonly between two empirical specifications of the model, that is, return specification, in which returns are regressed on a scaled earnings or/and change in earnings variable and price (levels) specification, in which stock prices are regressed on earnings per share. Both specifications are overwhelmingly implemented as linear regressions estimated on cross-sections.

Motivated by Kothari and Zimmerman (1995) which states that, although return specifications are commonly preferred to levels models, the latter "are better specified in that the estimated slope coefficients from price models, but not return models, are unbiased", the present paper focuses on the price specification and investigate the impact of the non-linearity in the price-earnings relation on its econometrics.

We begin by documenting the non-linear nature of the price association to earnings. We show that commonly used linear models fit to the non-linear relation between share prices and earnings are useless for hypothesis testing. The miss-specified linear model often classifies as structurally different sub-samples from the same economic relation which differ by a small percentage of observations.

To bring remedy to the failure of the linear modeling, we introduce a non-linear research design that consistently estimates prices' relation to earnings. Using recent techniques from the fields of statistical learning and data analytics (James et al. (2014), Goldstein et al. (2015)) we uncover and visualize a complex relation between price, earnings, and proxies for risk and growth.

In particular, we bring convincing evidence that the earnings-price relationship is relatively flat for negative and large positive earnings (ranges with a large transitory earnings component) and roughly linear in the middle range. Our empirical findings are consistent with the theoretical explanation for the non-linearities in the price-earnings relation that views the equity as an option (Hayn (1995), Fischer and Verrecchia (1997)).

The non-linear research design allows for the estimation of a firm-specific earnings-response coefficient as the slope of the middle linear part. Following Kothari and Zimmerman (1995), we validate the non-linear research design by verifying that the non-linear levels regression earnings-response coefficients have the 'right' size and yield economically justifiable risk-premium values.

Since returns specifications are often motivated by first differences in the models behind price-levels regressions (see, e.g. Ohlson (1995), Easton (1999)), the econometric issues related to non-linearities we highlight in the case of levels specification will, most likely, also affect the return specification. While levels specifications yield less biased earnings response coefficient estimates in information content studies, we acknowledge that price specifications do not measure information arrival over a period. Studying the econometric specificity of the return specification is beyond the scope of this paper and it will be addressed elsewhere.

2 Related literature

Kothari and Zimmerman (1995) argue theoretically that, although return specifications are commonly preferred to levels models, the latter "are better specified in that the estimated slope coefficients from price models, but not return models, are unbiased". The empirical evaluation of the two types of specifications focuses on the extent to which the estimated slopes approximate their predicted values. The slope, commonly referred to as the *earnings response coefficient* (ERC), should be, roughly, the reciprocal of the firm's expected rate of return. Their empirical results indicate that "the slope or earnings response coefficients are substantially less biased in price models than in return models. Coefficients from the price model, but not the return model,[1] imply cost of capital estimates that are more in line with those observed in the market." The authors conclude that levels specification "gives more economically sensible earnings response coefficients".

We note that most often, the price and return specifications are assumed to be equivalent[2] and hence are expected to yield the same ERCs. The empirical evidence in Kothari and Zimmerman (1995) that price and return specifications give very different results suggests that these specifications are in fact not equivalent at all [3] (see also Christie (1987)).

[1]The median ERC from the return specification over 38 cross-sectional slope estimates, corresponding to the period 1952-1989, was 1.64. The extremely small coefficients are possibly due to the effect of prices leading earnings (Beaver et al. (1980)).

[2]Otherwise, there would be no point in comparing their performance.

[3]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

While the estimates of the earnings response coefficient from the levels specification are considerably greater than those from the return specifications, they are still smaller than the inverse of the expected rate of return. Estimating the Ohlson's linear price specification of the Residual Income model (Ohlson (1995)), Kothari and Shanken (2003) report a median (mean) estimated earnings coefficient over 34 cross-sections (1967-2000) of 3.95 (4.1) which corresponds to a median (mean) cost of equity of 25.4% (24.4%). Moreover, the third quartile of the estimates distribution was equal to 4.9, i.e., in 26 of the 34 cross-sections, the cost of equity was larger than 20.5%.

The analysis in Kothari and Zimmerman (1995) and previous literature (Freeman and Tse (1992)) suggest a possible explanation for this discrepancy: non-linearity in the price-earnings relation due to transitory earnings components. Relatively extreme earnings have a large transitory part which implies that the price-earnings relation should be relatively flat for earnings that are negative or large and positive, giving rise to a flattened *S*-shaped price-earnings relation. Consequently, fitting a linear price specification yields earnings response coefficients that are biased downward. While discussing the adverse effect of non-linearity on the estimation of the earnings response coefficients, Kothari and Zimmerman (1995) state that examining the econometric consequences of non-linearities in the price-earnings relation is beyond the scope of their paper.

In a paper addressing inference issues in price (level) specification, Kothari and Shanken (2003) offer indirect evidence of bias due to correlated-omitted variables in the inference of the slope coefficients in cross-sectional regressions of prices on financial variables (like earnings and the book value of equity). While previous research discusses this bias in detail (e.g., Christie (1987), Shevlin (1996), Easton (1998), Holthausen and Watts (2001)), their paper is the first to offer evi-

variable, i.e., the price change, becomes, a new variables.

dence of the bias.

Their results show that coefficients on earnings from levels price-earnings regressions are negatively correlated with the aggregate growth and expected rate of return proxies. They complement similar previous results on earnings response coefficients estimated in return specifications (e.g., Easton and Zmijewski (1989), Collins and Kothari (1989)). Our non-linear analysis takes into account the fact that risk and growth are economic determinants of the ERC and addresses the issues highlighted in Kothari and Shanken (2003).

Freeman and Tse (1992) use the return specification and model the non-linear relation between returns and earnings using an arc-tangent transformation. The choice of model lacks economic foundation[4]. Moreover, their analysis is time-wise extremely narrow, covering only three years, from 1984 to 1987.

In contrast to that, our non-linear design is motivated by state-of-the art valuation models as well as rigorous statistical considerations. The non-parametric feature of our design does not impose an arbitrary structure on the price-returns relation. Instead, we use modern techniques from the field of statistical learning that allows the data to "speak for itself". Moreover, the research design we propose focuses on the consistent estimation of the price-earnings relation and can be used for studying a wealth of other themes than the estimation of ER coefficients. In fact, we chose to validate the non-linear research design by verifying that the non-linear levels regression earnings-response coefficients have the 'right' size and yield economically justifiable risk-premium following Kothari and Zimmerman (1995). Other applications of the research design introduced in this paper can be found in Starica and Kang (2017) and Starica and Giosi (2019).

[4]Kothari (2001) states that "strong economic foundation for the modeling is not apparent. Therefore, researchers must exercise caution in employing ad hoc statistical refinements."

3 Variables definition and sample

The sample, obtained from Compustat (accounting information) and CRSP (prices) data bases, covers the 47-year period between 1970 and 2016 and contains all firm-year observations with non-missing values of the variables needed by the non-linear research design introduced in section 5, i.e., earnings, book value of equity, number of shares, industry. It consists of 171,347 firm-year observations and contains 16,050 distinct firms. We define earnings (*NI*) as income before extraordinary items (Compustat *IB*). The number of common shares outstanding are equal to the Compustat variable *CSHO*. The share price is equal to the monthly closing price (Compustat *PRCCM*) collected at the beginning of the fourth month after the fiscal year end. Earnings growth is mean earnings growth calculated for each firm-year using at least four values[5] of the eight most recent annual observations ($t-7$ through t). Industry is defined by the 4 digit SIC code.

Variable	Mean	SD	10%	25%	50%	75%	90%
Share price	17.56	17.00	2.12	5.00	12.23	24.80	40.62
NI	0.83	1.55	-0.71	-0.00	0.63	1.59	2.79
P/B	2.46	2.76	0.60	0.98	1.68	2.94	5.12
log(Mktv)	4.80	2.13	2.08	3.15	4.64	6.32	7.74
NI.g	-0.05	0.63	-0.88	-0.32	0.06	0.24	0.50

Table 1: **Descriptive statistics.**

Table 1 displays the summary statistics of the variables of interest. The variables were winsorized at the 1% level. Table 2 displays the values of Pearson (upper triangle) and Spearman (lower triangle) correlation between the relevant variables in the study.

[5]The sample contains firm-year with available data for at least 4 of the last 8 most recent years.

	Share price	NI	P/B	log(Mktv)	NI.g
Share price	1.00	0.60	0.24	0.68	0.17
NI	0.69	1.00	-0.03	0.34	0.18
P/B	0.37	0.06	1.00	0.25	0.07
log(Mktv)	0.73	0.37	0.41	1.00	0.11
NI.g	0.26	0.29	0.14	0.13	1.00

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

4 The economic relation between prices/returns and earnings is not linear

In this section we bring empirical evidence that the economic relation between prices/returns and earnings is not linear. We also show the consequence of this fact on testing hypothesis on their economic relation when the tests are based on an estimated (miss-specified) linear model. Such tests give unreliable results. Even small changes in the sample composition yield statistically different linear models. Although the samples are expression of the same data generating process, inference based on the miss-specified linear model often rejects the hypothesis of identical economic relation.

4.1 The econometrics of a linear relation

For the clarity of exposition and to fix the notation needed for the discussion, we begin this section by collecting well-known facts about the econometrics of modeling and estimating linear relationships. We are in no way assuming that the reader is not aware of these facts.

Consistent with modern econometric theory, two variables X and Y are said to be *linearly related* if the relation between X and Y admits a linear representation

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{1}$$

where the *orthogonality condition* of ε on X holds,

$$\mathbb{E}[\varepsilon|X] = 0. \tag{2}$$

$\mathbb{E}[\varepsilon|X]$ denotes the random variable conditional expectation of ε given X.

The *orthogonality condition* is a formal mathematical statement about all the other factors[6] (than X) that determine Y. It states that all other factors are unrelated to X, in the sense that, given a value of X, the mean of the distribution of these other factors is zero. Intuitively, it asserts that a line passes through the 'core' of the data. That is, for every value of x in the range of the first component, the mean of the error terms corresponding to pairs (X,Y) with the abscissa $X = x$ is equal to 0. For every x, the error term corresponding to pairs (X,Y) with the abscissa $X = x$ is equally often positive as it is negative.

Note that condition (2) is essential to the definition of a linear relationship. It is this condition which makes the linear representation (1) into a linear relation[7].

From an econometric point of view, the pair of equations (1) and (2) define a *linear regression*. The orthogonality condition is *necessary*[8] for its consistent statistical estimation (Stock and Watson (2012)). It ensures that no omitted variables bias the estimation of the coefficients. If it is not fulfilled, the estimated coeffi-

[6] It is their summed contribution that defines ε.

[7] The relation between the components of any random pair (X,Y) can *always* be represented in a linear form as: $Y = \beta_0 + \beta_1 X + (Y - \beta_0 + \beta_1 X) := \beta_0 + \beta_1 X + \varepsilon$. However, the relation is a linear regression if and only if $\mathbb{E}[\varepsilon|X] = \mathbb{E}[Y - \beta_0 + \beta_1 X|X] = 0$, i.e., $\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$.

[8] In particular, it implies $corr(\varepsilon,X) = 0$. The reciprocal is not true, i.e., the condition (2) is strictly stronger than being uncorrelated.

cients do not reflect the true relation inside the pair (X,Y) and cannot be used for testing hypothesis about it.

4.2 Linear representation of prices

The extant accounting literature often uses linear price specifications that are estimated cross-sectionally. Examples include:

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

$$P_{i,0} = \beta_{0,0} + \beta_{1,0} B_{i,0} + \beta_{2,0} NI_{i,0} + \varepsilon_{i,0}, \tag{4}$$

where $NI_{i,0}$ stands for firm's i earnings per share at time 0, while B is firm's per-share book value of equity. Other specifications (than the per-share ones) scale the market value and the accounting numbers with book value of equity, number of employees, sales, etc. However, the literature seems to have arrived at a certain consensus that the specification deflated by number of shares is the one that offers results more consistent with formal benchmarks backed up by extensive research on valuation (Barth and Kallapur (1996), Kothari and Shanken (2003), Barth and Clinch (2009), Aledo Martinez et al. (2019)).

Specification (4) is particularly popular and is often theoretically justified as an implementation of the Ohlson model (Ohlson (1995)) (see section 9.1 in the Appendix). This section brings empirical evidence that the price relation to earnings, as modeled by the commonly-used level specification (4), is not linear. If it were, the OLS estimated residuals of the models should approximately verify the orthogonal condition (2). While the empirical evidence we present concerns the specification theoretically motivated by the Ohlson model, the results for the other models are similar.

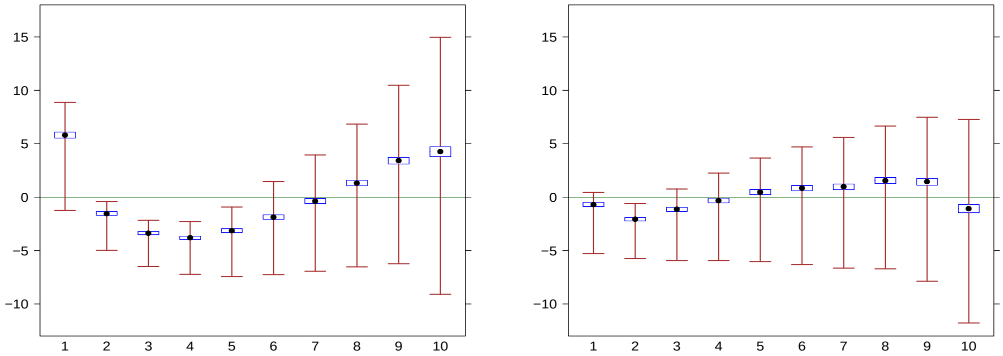

Figure 1: **Distribution of residual of model (4) conditional on the level of** NI **and** B**.** The graph displays the 25%- and the 75%-percentile (the lower end, upper end, respectively, of whiskers) together with the mean (dot) and its 95% confidence interval (box) of the cross-sectional estimated residuals of the linear model conditional on the deciles of the level of NI (left) and B (right). The linear model (4) is estimated cross-sectionally and the firms are divided in 10 equal groups (deciles) according to the level of the explanatory variable. For a given decile, all the cross-section residuals corresponding to it are used to construct the three statistics (25%- and the 75%-percentile and mean) displayed above the decile number. If the relation between the dependent and the independent variables in model (4) is linear, the twenty conditional means should be approximately 0.

Figure 1 gives strong empirical evidence of the fact that the relation between the independent and dependent variables in model (4) is not linear. The orthogonal condition for the estimated residuals, essential for the consistent estimation of a liner relation, seems strongly violated. Moreover, section 9.2 in Appendix gives a formal argument of the fact that the Ohlson model is not necessarily a linear regression.

4.3 Econometric consequences of lack of linearity in the price relation to earnings

A linear model fit to a non-linear relation is, econometrically speaking, useless. Even small changes in the sample, without any in the economic relation, causes the coefficients of the model to move significantly. Taking the linear inference face value, one would decide, falsely, against the hypothesis of equal economic relation.

In the first section, we give examples of the econometric impact of non-linearity on testing using miss-specified linear models. The second section discusses how non-linearity can interact with omitted determinants to wreak havoc hypothesis testing based on miss-specified linear models.

4.4 How non-linearity most surely ruins the day of the linear accounting researcher

This section shows how small changes in the sample induce significant changes in the coefficients of a miss-specified linear model[9]. Taking these coefficients face value, the accounting researcher would conclude erroneously that two samples issued from the same data generating process are expressions of different economic relations.

Figure 2 presents an example of such a situation. The graphs plot prices against *NI* (on the *x*-axis) for two sub-samples from the year 2015.[10] The two sub-samples were 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 squares while the circles mark the pairs in the sub-samples. The squares mark hence the pairs with first, fifth, ninth, thirteenth, etc. (up/down to the 10% threshold) largest (smallest) *NI* in the 2015 cross-section. The marked sets represent each 2.5% of the sample. The two sub-samples are expressions of the same economic relation. A statisti-

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

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

cal model that correctly reflects this fact should not reject the hypothesis of no structural difference between the two sub-samples.

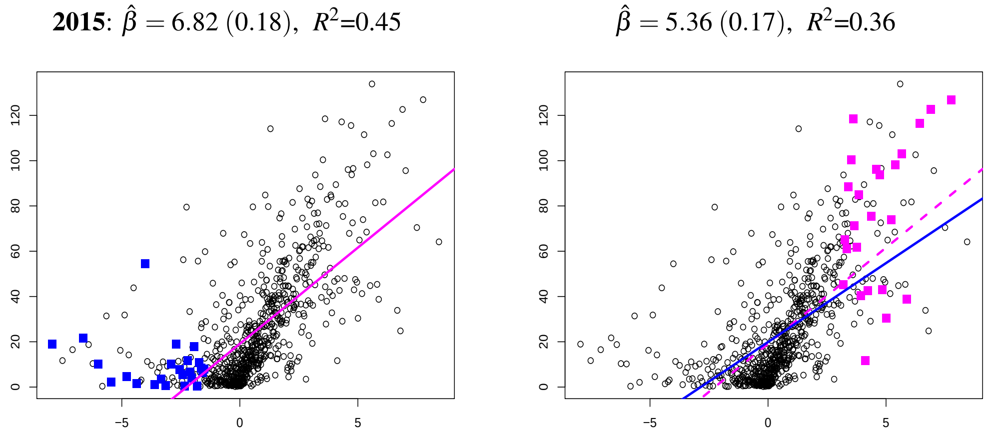

Figure 2: **Non-linearity biases the level regressions.** The graphs display prices against *NI* (on the *x*-axis) for two sub-samples from the year 2015 constructed as follows. First, we sorted the pairs in the cross-section in increasing order of their *NI* variable. Then, the first (second) sub-sample was obtained by removing every fourth observation among the pairs with 10% lowest (highest) *NI*. The removed pairs are marked in the graphs with squares. The circles mark the pairs in each of the two sub-samples. The marked sets represent each 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 (4) is displayed on each of the sub-samples. The dotted line in the second graph is the regression line estimated on the first sampler and differs significantly (see the estimates reported in the titles of the graphs) from the relation estimated on the second sub-sample.

We estimated the linear levels price earnings regression (4) on each of the sub-samples and display the two regression lines (the second graphs displays both of them). Although the two sub-samples are expression 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, while the t-statistic testing the hypothesis of equal R^2 is -3.93. 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 base. The differences we induced in the example in Figure 2, while innocuous when the economic relation between price and earning is linear, interact with the strong non-linear pattern in the data and misleads the miss-specified linear model into rejecting the true hypothesis.

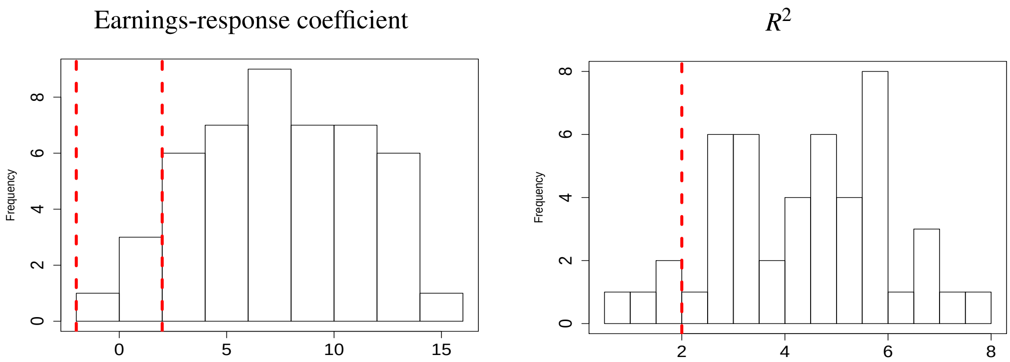

Figure 3: **Non-linearity biases the levels regressions - all cross-sections.** The graphs display the histograms of the t-statistics that test the equality of the levels regression earnings-response coefficients, as well as that of R^2, from the two regressions (4) performed on pairs of yearly cross-sectional sub-samples (from 1970 to 2016) constructed according to the description in caption of figure 2. 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 extended the analysis in Figure 2 to all the cross-sections in the sample and display the results in Figure 3. It presents the histograms of the t-statistics that test the equality of ERC, as well as that of R^2, for two linear levels price earnings regressions (4) 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 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. The graphs in Figure 3 clearly indicate that even small changes in the sample cause significant instability in the statistics that are usually used for testing hypothesis about the economic relation between prices and earnings.

4.5 How non-linearity interacts with omitted determinants to ruin the day of the linear accounting researcher

Non-linearity combines with omitted determinants to produce even more problems. Previous research (e.g., Easton and Zmijewski (1989), Collins and Kothari (1989)) documents a cross-sectional relation between earnings response coefficients from return specifications and discount rates and growth. Kothari and Shanken (2003) extends the findings to the levels regression earnings-response specification.

Table 3 presents evidence based on our sample of the relation between ERC from levels price earnings linear regression and two proxies for risk and growth, that is, size and price-to-book. Given the non-linear nature of the underlying

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 3: **Estimated earnings 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 10 regressions (4) on each of the 10 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.

relationships, we do not attach much confidence to these estimates. However, we note that the size of the differences between extreme deciles is extremely large.

The type of econometric issues related to the interaction of non-linearity and

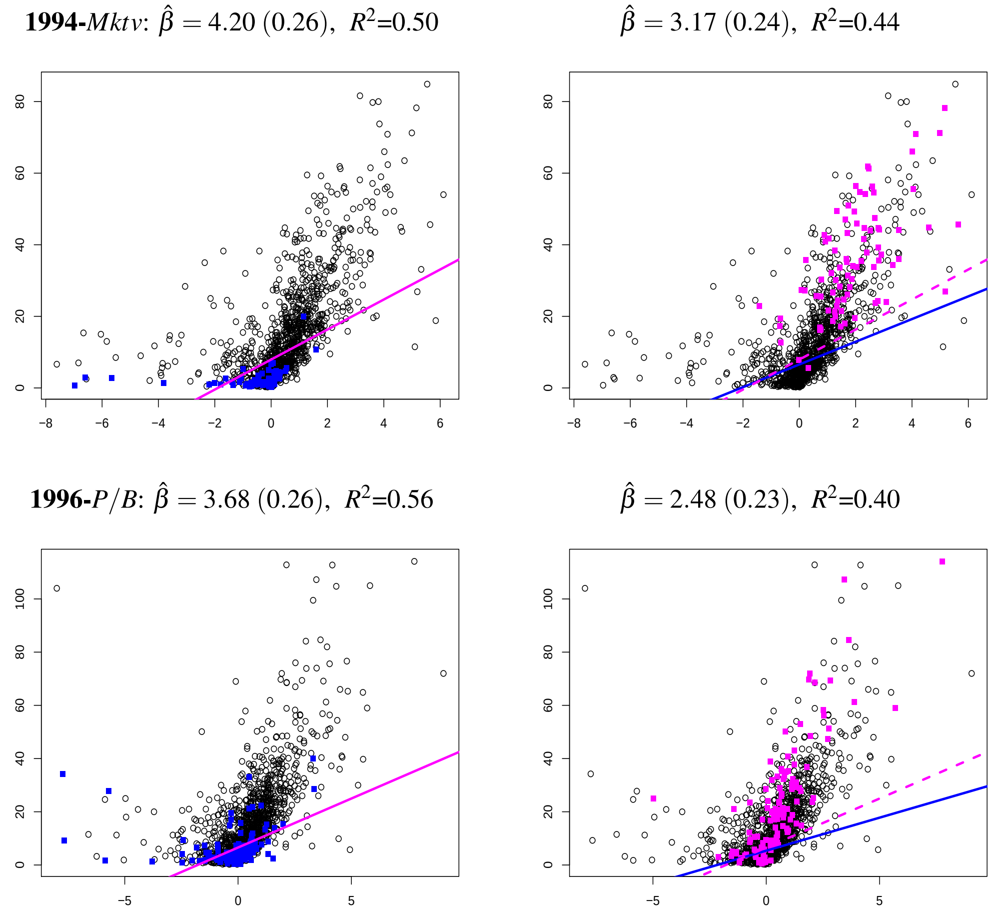

Figure 4: **Non-linearity and omitted determinants bias the level regressions.** The graphs display prices against NI (on the x-axis) for two sub-samples from the same cross-section constructed as follows. The pairs of sub-samples express the same economic relation between prices and earnings but differ in the range of values of each of two proxies. For the top (bottom) graphs, the proxy is $Mktv$ (P/B, respectively). The sub-samples in the left/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 full line represents the regression line of the sub sample in the graph while the dotted line is the regression line corresponding to the pair sub sample. We see that lower (higher) values of the proxies are associated to lower (higher) earnings and prices.

omitted determinants are exemplified in Figure 4. The graphs display pairs of samples expressing the same economic relation between prices and earnings but differ in the values of each of two proxies. The samples on the top (bottom) graphs (proxy: $Mktv$ - top, P/B - bottom) are sub-samples of the 1994/1996 cross-sections. The sub-samples in the left/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 little squares (blue on the left, magenta on the right). The full line represents the regression line of the sample in the graph while the dotted line is the regression line corresponding to the pair sample. We see that lower (higher) values of the proxies are associated to lower (higher) earnings and prices. Since the type of relation between earnings and price 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 difference between the two slopes are statistically significant: in the case of the top (lower) graphs, the t-statistic of the test of equal slopes is of 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 convincingly 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 the risk/growth profile of the firms. The interaction between non-linearities and omitted determinants exemplified in Figure 4 would then trick the miss-specified linear model into rejecting the correct hypothesis.

We performed the analysis in Figure 4 on all the cross-sections in the sample. The results are presented in Figure 5 which displays the histograms of the t-statistics that test the equality of slopes, as well as that of R^2, 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 top (bottom) graphs, the proxy is $Mktv$ (P/B, respectively). 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

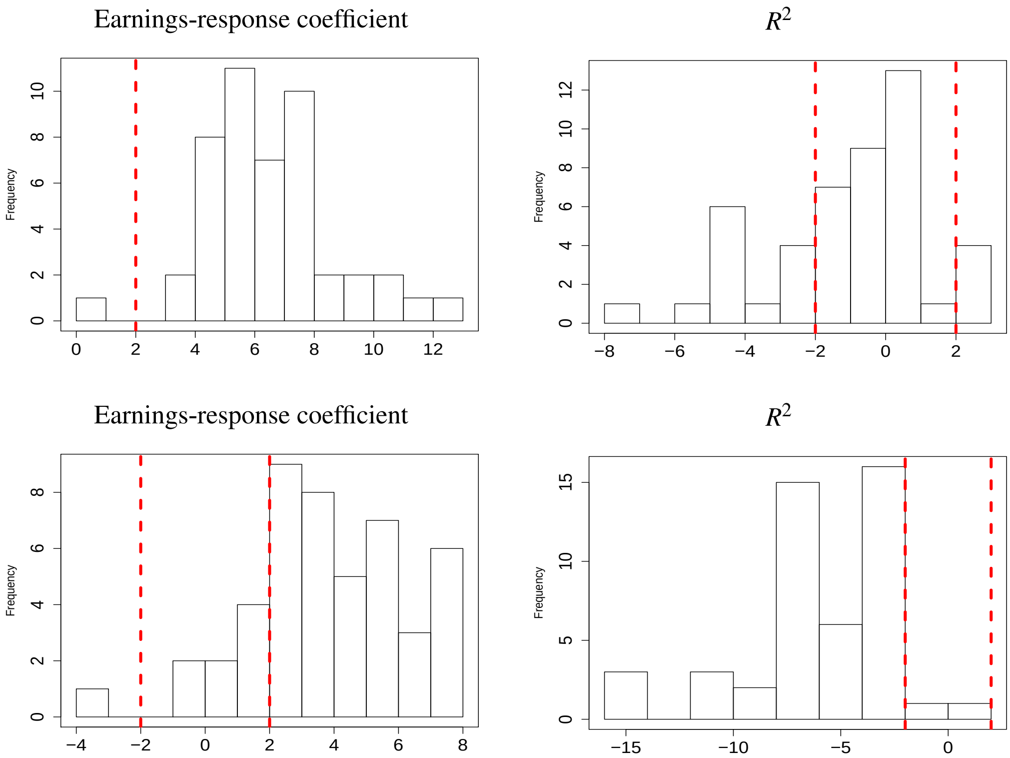

Figure 5: **Non-linearity and omitted variables biases the level regressions.** The graphs display the histograms of the t-statistics that test the equality of the levels regression earnings-response coefficients, as well as that of R^2, from the two regressions (4) performed on pairs of yearly cross-sectional sub-samples (from 1970 to 2016) constructed according to the description in caption of figure 4. 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 top (bottom) graphs, the proxy is $Mktv$ (P/B, respectively). 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 model would wrongly reject the null.

on the linear representation of the Ohlson model would wrongly reject the null. It indicates that even changes not related to earnings or prices that affect the sample structure but do not touch on the economic link between earnings and prices can cause significant instability in the statistics that are commonly used for testing hypothesis about the economic relation between prices and earnings.

5 A non-linear research design alternative

For the clarity of exposition and in the interest of making it self-contained, we begin this section by collecting a few well-known facts about the econometrics of modeling and estimating non-linear relationships.

5.1 The econometrics of a non-linear relation

While a pair (X,Y) might not be linearly related in the sense of the definition (2), a non-linear representation of the relation between X and Y for which the orthogonality condition holds always exists. More concretely, for any pair of random variables (X,Y), we can write:

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

the expectation of Y given that the variable X takes a given value x, and

$$\mathbb{E}[\varepsilon|X] = 0 \quad \text{(the orthogonality condition)}^{11}. \tag{6}$$

The function $x \to \mathbb{E}[Y|X = x]$ is, in general, a non-linear function of x. It is a good summary of the relationship between Y and X for a number of reasons.

First, we are used to thinking of averages 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 the first component X takes values close to x.

Second, it can be shown that the expected value of Y conditional on X is the

[11] Condition (2) implies $\mathbb{E}(h(X)\varepsilon) = 0$ for any function h (which explains its name). In particular, for $h(x) = x$ it implies that $corr(X,\varepsilon) = 0$ since $\mathbb{E}[\varepsilon|X] = 0$ also implies that $\mathbb{E}(\varepsilon) = 0$.

best predictor of Y given X in a sense that is made precise by the prediction property 1 in section 9.3 in the Appendix (for more details on the notion of conditional expectation and the related results see, for example, Billingsley (1995)). Consequently, the regression relation in (5) can be restated as a decomposition of Y into the best prediction of Y knowing the value of the predictor X, that is $\mathbb{E}[Y|X]$, and an orthogonal component.

The orthogonality condition is essential for statistical inference. It is the *necessary and sufficient condition* for consistent statistical estimation of the conditional expectation of Y given X in (5) (Györfi et al. (2002), James et al. (2014)).

5.2 Statistics meets accounting

In this section we explain how accounting considerations informing statistical inference methodology can provide an answer to the issue of consistent estimation of the price-earnings relationship.

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

$$P_{i,0} = f_{i,0}(NI_{i,0}) + \varepsilon_{i,0}, \quad \text{where} \quad f_{i,0}(x) := \mathbb{E}[P_{i,0}|NI_{i,0} = x], \tag{7}$$

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

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

could be inferred without bias (using proved methods from the non-linear regression estimation literature) provided we had many similar copies 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 association price-earnings realized in the pair $(NI_{i,0}, P_{i,0})$. The fact that the function f in the regression (7) bears the index $(i,0)$ formalizes the accounting fact that the association of earnings to prices depends of specific firm characteristics (ex., risk and growth). Hence, we cannot estimate the regression (7), as it is, on the entire cross-section.

To summarize, we have seen that an unbiased estimation of the price-return relation for the firm i is possible conditional on identifying a subset of firms j in the cross-section displaying a similar association of prices to earnings, that is, the subset

$$S = \left\{\text{all firms } j \text{ in cross-section 0 for which } f_{j,0} \approx f_{i,0}\right\} \tag{8}$$

of all firms j for which the regression function $f_{j,0}$ in (7) describing the association of prices to earnings, is similar to the regression function of the entity i. Once we constructed the set S, we would estimate the non-linear regression in (7) only on the sub-sample of pairs $(NI_{j,0}, P_{j,0})$ corresponding to the firms j in the set. All these pairs are an expression of the same price-earnings relation that can then be unbiasedly estimated using techniques from the field of statistical learning and data analytics. 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 miss-specifying the model.

The following accounting considerations will help us identify this set.

5.3 From the accounting 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, other accounting variables, and the equity risk price:

$$P_{i,0} = \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[f_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\ ;\ r_{i,0})]}{(1+r_{i,0})^t} \tag{9}$$

where $\mathbf{NI}_{i,t} := (NI_{i,1}, NI_{i,2}, \ldots, NI_{i,t})$, $\mathbf{O}_{i,t}$ stands for other variables than earnings (like book values or dividends), $r_{i,0}$ denotes the price of equity risk for firm i at time 0 while $\mathbb{E}_0$ stands for market's expectation conditional on all information available at time 0. Concrete examples include the Residual Income (RI) model or the Ohlson and Juettner-Nauroth (OJ) model (see section 9.1 in Appendix for details).

The modeling assumption in (9) allows us to make specific the general non-linear function in (7). More precisely, one can show (see the details in section of the Appendix) that:

$$f_{i.0}(x) = \sum_{t=1}^{\infty} \frac{\mathbb{E}[f_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_{i,0}) \,|\, NI_{i,0} = x]}{(1+r_{i,0})^t}. \tag{10}$$

Let us compare this expression to the model assumption (9). The two expression are structurally the same and differ only through the information available to the investor when forming expectations about the future income stream. In the model (9), 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 (10), 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 $\mathbb{E}[\cdot \,|\, NI_i = x]$).

This means that the regression function in (7) is also a valuation. More concretely, $f_{i,0}$ is a valuation incorporating expectations shaped only by the current level of firm's earnings (the price of disclosed earnings).

With this understanding, we can restate the relation (7) as follows. For any firm i, the price at time 0 can be thought as the sum between a valuation based on expectations of future income informed only by the current level of earnings of the firm (the value of earnings) and an orthogonal investor's adjustment that reflects all other information (then the level of earnings) available to investors:

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

where $\mathbb{E}[\varepsilon|NI] = 0$.

5.3.1 The determinants of the functions f_t

In line with the discussion in the beginning of the section, we note that the decomposition in (11) is firm-specific. The functions

$$x \to \mathbb{E}[f_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t};\ r_i) \,|\, x] \tag{12}$$

appearing in the denominators of the summands in (11) describe how current level of earnings project the future income stream and, hence, are firm-specific.

This section argues that economic considerations suggest that the persistence of earnings is determined by a number of factors, of which most are industry-specific. Consequently, one could assume that the regression function (11) of firms within the same industry and with similar levels of the determining factors are similar. This relative homogeneity of the regression functions enables the estimation on well-defined subsets in each cross-section.

Extant valuation literature lists among the determinants of the evolution of future income streams and, hence, of the shape of the functions in (12), the firm's cost of equity $r_{i,0}$ and firm's growth $g_{i,0}$.

More broadly, the economics and strategic management literature regularly identify four observable traits with an impact on profits persistence which in turn affect the evolution of future earnings: firm size, product-type, barriers-to-entry, and capital intensity (Lev (1983), Baginski et al. (1999)). For a brief survey of the literature that support this statement, see section 9.5 in the Appendix.

Accounting practices have also an effect on the persistence of abnormal earnings (Feltham and Ohlson (1995), Feltham and Ohlson (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, accounting measures depart from economic measures and a firm's abnormal ROE can be different from zero even if 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 a firm-specific obvious driver.

To summarize, beside size and, to a certain extent, accounting conservatism, the main economic and accounting factors that determine the persistence of earnings (and hence, determine the functions (12)), that is, product-type, barriers-to-entry, and capital intensity, are industry-specific.

5.3.2 How to incorporation the constraint S in statistical estimation

In line with the discussion in the previous section, we will assume that firms belonging to the same industry with similar risk, growth, size and level of conservatism have similar regression functions (7). Consequently, a possible specification of the set S in (8) for a firm i belonging to the industry I is given by:

$$S = \{\text{all firms } j \in I \text{ s.t. } r_{j,0} \approx r_{i,0},\ g_{j,0} \approx g_{i,0},\ size_{j,0} \approx size_{i,0},\ C_{j,0} \approx C_{i,0}\}.$$

In practice, to constrain the estimation of the regression (7) to the set S of firms in the same industry with similar determinants as above, we capitalize on knowledge about the functioning of an important class of non-linear estimation approaches: the tree-based methods.

How does the tree-based non-parametric estimation work. In a nut-shell, the tree-based methods partitions the space of the explanatory variables into a set of rectangles, and fit a simple model (for example, a constant) in each one of them. They are conceptually simple but powerful. To fix the ideas, let us consider the regression in (5) and let us suppose that we have more than one explanatory variable, i.e., X is a vector. For a given x, we want to estimate $f(x)$, the regression function realized in the pairs (X_i, Y_i), $i \in 1, 2, \ldots, n$.

The class of methods under discussion would then select all m pairs (X_j, Y_j) in the sample such that $X_j \approx x$ and estimate $f(x)$ as the average of the corresponding Y_js:

$$\hat{f}(x) := \frac{1}{m} \sum_{\{j\,:\,X_j \approx x\}} Y_j.$$

In other word, one approximates the regression function f at the point x with a local average of the Ys of the pairs with the explanatory variable X close in value x (X belongs to a rectangle around x). The methods belonging to this class

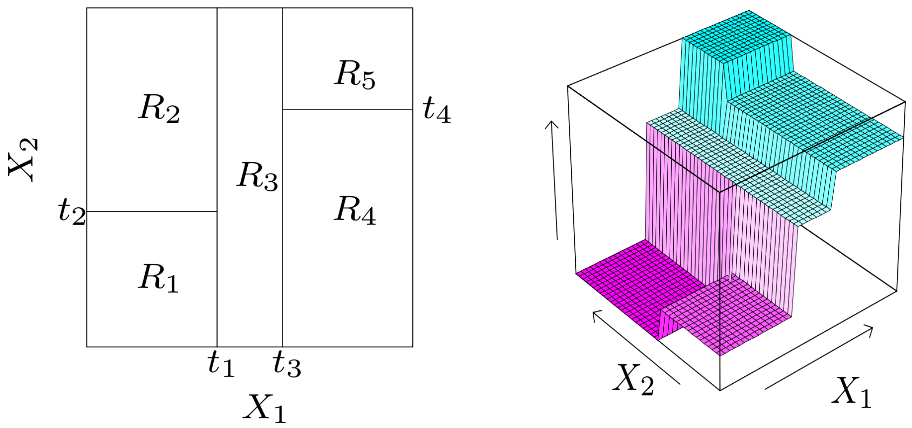

Figure 6: **Tree-based non-parametric estimation at work.** Left panel shows a partition of a two-dimensional explanatory variable space by recursive binary splitting, as used in CART. Right panal shws a plot of the estimated regression function which is locally constant on each of the five regions $R_1, R_2, \ldots, R_5$ of the explanatory variable space (reproduced from Hastie et al. (2009)).

differ only through the way they construct the set of pairs (X_j, Y_j) for which the approximation $X_j \approx x$ holds.

Figure 6 illustrates the functioning of one of the popular tree-based regression methods (Classification and Regression Tree - CART) (reproduced from Hastie et al. (2009)). To partition the space into areas where the regression function f_0 is almost constant, the method does binary recursive splits of the explanatory variable space. The steps to be repetead (hence the 'recursive' term) are as follows. Choose a variable to split and a split value (say, X_1 and t_1). Split the space into two regions (hence the 'binary' term): $X_1 \leq t_1$ and $X_1 > t_1$. Model the regression function by the mean of Y in each region. Hold on to the variable and the split point that achieve the best fit, i.e., hold on to the variable and the split that decrease the most the mean square error. Then, repeat until some stopping rule applies.

For example, in the left panel of figure 6, we first split at $X_1 = t_1$. Then the region $X_1 \leq t_1$ is split at $X_2 = t_2$ and the region $X_1 > t_1$ is split at $X_1 = t_3$, Finally, the region $X_1 > t_3$ is split at $X_2 = t_4$. As a result, the (X_1, X_2)-space is partitioned in five regions $R_1, \ldots, R_5$ shown in the figure. On each of these regions, the model approximates the regression function f_0 by a constant $c_1, \ldots, c_5$ which yields:

$$\hat{f}_0(X_1, X_2) = \begin{cases} c_1 = -5, & (X_1, X_2) \in R_1, \\ c_2 = -7, & (X_1, X_2) \in R_2, \\ c_3 = 0, & (X_1, X_2) \in R_2, \\ c_4 = 2, & (X_1, X_2) \in R_2, \\ c_5 = 4, & (X_1, X_2) \in R_5. \end{cases}$$

For illustration, the estimated $\hat{f}_0$ is represented in right panel of figure 6.

Putting together the above explanation with the specification of the set S above, we will estimate the price-earnings association by adding to the explanatory variable $NI_{i,0}$ in (7) the following five other variables: firm's cost of equity $r_{i,0}$, its growth $g_{i,0}$, its size, its level of accounting conservatism, $C_{i,0}$, and its SIC industry classification, $I_{i,0}$. In other words, we will estimate the non-linear regression:

$$P_{i,0} = f_0(NI_{i,0},\ r_{i,0},\ g_{i,0},\ size_{i,0},\ C_{i,0},\ I_{i,0}) + \varepsilon_{i,0}, \tag{13}$$

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

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

Once pertinent proxies for risk, growth and level of conservatism are specified, we will estimate the non-linear regression in (13) cross-sectionally.

5.3.3 Another interpretation of the regression (13)

The functioning of the non-parametric regression explained above yields another intuitive interpretation of the regression function in (13), familiar to the reader used to the linear regression framework. Non-linearly regressing prices $P_{i,0}$ on the vector $(NI_{i,0},\ r_{i,0},\ g_{i,0},\ size_{i,0},\ C_{i,0},\ I_{i,0})$ amounts to estimating the association between prices and earnings while holding (almost) constant the level of risk and growth, the size, the level of conservatism, and the industry, i.e., the determinants of the projections of current earnings into the future income stream put forth in section 5.3.1. In other words, the non-parametric estimation infers the price-earnings association controlling for the level of the factors that determine how current earnings of a firm project into future income.

5.4 Empirical evidence of the fit of the non-linear estimation

The analysis in the paper uses P/B and size, two of the factors in the Fama-French three factor model, as proxies for risk/growth. We measure size through the market value, or alternatively, as the amount of total assets. Smith and Watts (1992) point out that the difference between book value of equity and market value represents an approximation of the value of investment opportunities facing the firm. The market-to-book ratio is a function of gap between the firm's return on current assets and expected future investments and the required rate of return on equity. Since future earnings are affected by the growth opportunities, at the same level of disclosed earnings and risk, the higher the price-to-book ratio, the higher the expected earnings growth (see also Collins and Kothari (1989)). P/B serves also as a proxy of the level of accounting conservatism. Firms with lower P/B ratios exhibit substantially greater earnings conservatism (Pae et al. (2005)).

More recently, the accounting literature has emphasized the importance of investment and profitability proxies to valuation and market performance (**?**, **?**, **?**. **?** shows that profitability predicts economic growth, even after controlling for valuations. His measure of gross profits-to-assets predicts gross profit growth, earning's growth, as well as free cashflow growth. We include the gross profit-to-assets[12] as one of the proxies for growth.

? argue that the investment factor plays a similar role as **?**'s 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., **?**). As such, controlling for profitability, investment should be negatively correlated with expected returns. **?**We include

We use past earnings growth[13] ($NI.g$) (calculated as the median of the percent earnings change on at least four of the last seven pairs of consecutive years) as a second proxy for growth. We use the 4 digit SIC code as proxy for the industry. We hence regress non-linearly prices $P_{i,0}$ on the vector ($NI_{i,0}$, $(P/B)_{i,0}$, $\log(Mktv_{i,0})$, $NI.g_{i,0}$, $SIC_{i,0}$) and estimated the regression:

$$P_{i,0} = f_0(NI_{i,0},\ (P/B)_{i,0},\ \log(Mktv_{i,0}),\ NI.g_{i,0},\ SIC_{i,0}) + \varepsilon_{i,0}. \qquad (14)$$

The choice of proxies to use in the empirical estimation of the non-linear specification in (13) is neither unique nor the only ones that can be used. Other variables can extend or amend the set of proxies. In robustness tests, we have experimented with firm's beta as proxy for risk as well as with the conservatism index in Penman and Zhang (2002) as proxy for conservatism. The results are qualitatively

[12]Gross profits, defined as total revenue (REVT) minus cost of goods sold (COGS) is the cleanest accounting measure of true economic profitability (**?**).

[13]Growth in sales yields qualitative equal albeit weaker results.

equal to the ones we present in the sequel.

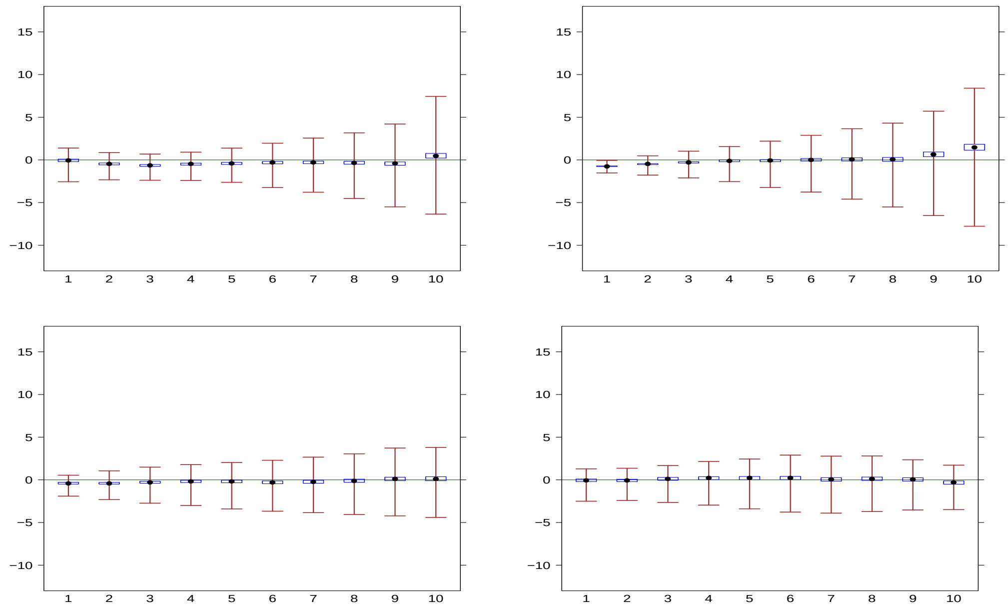

Figure 7: **Distribution of residual of specification (14) conditional on the explanatory variables.** The graph displays the 25%- and the 75%-percentile (the lower end, upper end, respectively, of whiskers) together with the mean (dot) and its 95% confidence interval (box) of the cross-sectional estimated residuals of the non-linear model (14) conditional on the deciles of the level of NI (top-left), $Mktv$ (top-right), P/B (bottom-left), and $NI.g$ (bottom-right). The non-linear model (14) is estimated cross-sectionally and the firms are divided in ten equal groups (deciles) according to the level of each explanatory variable. For a given decile of firms, all the cross-section residuals corresponding to it are used to construct the three statistics (25%- and the 75%-percentile and mean) displayed above the decile number. If the relation between the dependent and the independent variables in model (14) is correctly estimated, the ten conditional means should be approximately 0.

The results we report are obtained using the Random Forest (RF) algorithm, a popular approach in machine learning and Big Data applications of non-parametric, non-linear regression. The RF is a boosted version of CART method detailed in section 5.3.2. In a nutshell, for a firm i, the RF method approximates the function to estimate, f_0, with a local average of the prices of entities with explanatory variables close in value to those of the firm i. As such, the valuation where the expectations of future income are informed only by the level of earnings of the firm i is a local average of prices of firms from the same (or related) industry(es) that have similar level of earnings, risk, conservatism, as well as earnings growth.

In contrast to figure 1, the graphs in figure 7 confirm the non-linear estimation's fit. All four conditional distributions of the residual of the specification (14) satisfy, approximately, the orthogonality condition (6).

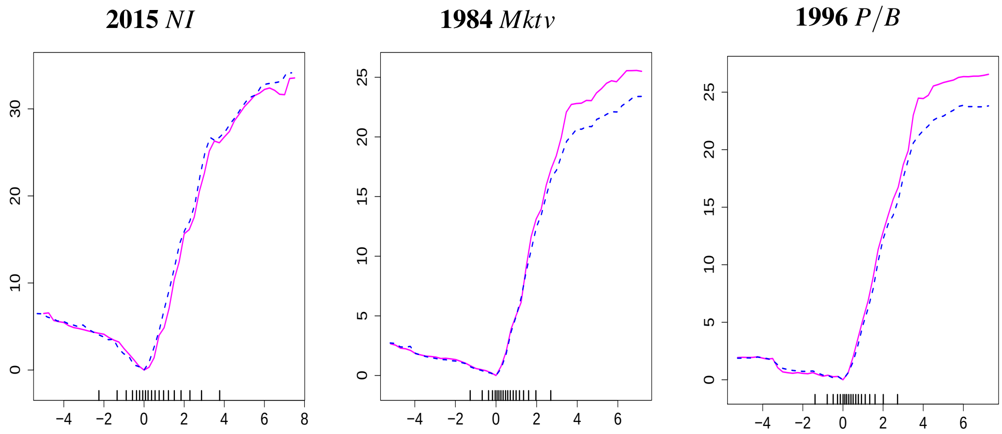

Figure 8: **The non-linear approach addresses the bias in the level regressions.** The graphs display the dependence of prices on *NI* estimated by the approach described in sections 5.2 and 5.3 for the pairs of sub-samples from the same cross-section analyzed using the linear model in figures 2 and 4. The pairs of sub-samples express the same economic relation between prices and earnings but differ in the range of values of *NI*, *Mktv*, or P/B (see the caption of the two figures for more details). The full (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., 5%-, 10%-, ..., 95%-percentiles, of the cross-section's distribution of *NI*. We see that the estimated dependency resembles very closely up to at least the 95%-percentile in all three examples.

Figure 8 displays the dependence of prices on *NI* estimated by the non-linear approach for the pairs of sub-samples from the same cross-sections analyzed using the linear level regressions in sections 4.4 and 4.5 (see figures 2 and 4). By construction, the pairs of sub-samples express the same economic relation between prices and earnings but differ in the range of values of *NI*, *Mktv*, or P/B (see the two sections for more details). We see that the estimated curves corresponding to each of the three pairs are almost identical over most of the range of earnings, i.e., at least until the 95% percentile). Moreover, the discrepancy noticeable over the extreme larger earnings is to be expected, as firms with higher values of *NI* or proxies have higher prices. As such, the samples obtained by removing firms with higher values of *NI* or proxies will have a 'thinner' upper part of the price distribution and hence a dependency function that tappers off earlier.

6 The complex relation between price, earnings, and the proxies for risk and growth

In this section, we show how the non-linear research design can be used to explore in depth the complex non-linear structure of the relation between prices, earnings and the proxies for risk and growth. Before we do that, we explain our treatment of yet another determinant of the relation between prices and earnings: their persistence.

Previous literature (Collins and Kothari (1989)) established that cross-sectional variation in ERCs to be positively related to earnings persistence,[14] i.e., the propensity of current period's earnings shocks to persist in the future and affect future earnings expectations. Earnings persistence is typically measured by estimating a simple time series earnings process (e.g., Kormendi and Lipe (1987)). We take into consideration the cross-sectional heterogeneity of the individual firm's earnings persistence by estimating an *AR* model[15] on firm's time series of past earnings and use the one-step ahead forecast[16] of the model as a proxy for the level of persistent earnings.

[14]Moreover, the extant literature also agrees that time series persistence estimates do not fully and accurately capture economic growth opportunities. However, the non-linear research design under discussion includes a proxy for the latter as a determinant of the ERC.

[15]The order of the model is chosen by the AIC criteria and varies between 0 and 3.

[16]We obtain qualitatively comparable results under the assumption of a random walk process for the earnings which motivates the use of current earnings as the best prediction of future values. The econometric evidence from our sample points, however, towards a mean-reverting (and not a random walk) dynamics of earnings. This finding might indicate that the persistence has in fact a second order impact on the price-return relationship.

The non-linear research design yields an estimate of the multidimensional economic relation between prices, on one side, and earnings and the other determinants of future income (risk and growth proxies) in model (9). This relation is expressed by the estimated regression function:

$$(NI_i,\ (P/B)_i,\ \log(Mktv_i),\ NI.g_i,\ I_i) \rightarrow\ \widehat{f}_0(NI_i,\ (P/B)_i,\ \log(Mktv_i),\ NI.g_i,\ I_i).$$

In other words, the non-linear research design estimation yields the functional form of the relation between prices and their determinants.

Although we cannot visualize the full economic relationship described by the previous expression (since this is a function of four variables), we can graphically display the relationship between prices and smaller subsets of explanatory variables. We chose to look at two aspects of the estimated economic relation. In the next section, we use use the *partial dependency plot* to visualize the association between prices and pairs of determinants, that is, $(NI, P/B)$, $(NI, \log(Mktv))$, $(NI, NI.g)$, and $(P/B, \log(Mktv))$. In section 6.2 we employ the *individual conditional expectation* (ICE) curves to visualize the firm-specific functional form of the relation between prices and earnings.

6.1 Joint dependence of price on pairs of determinants

In this section we display and discuss the *partial dependency (PD) plots* for a typical cross-section (the year 2011). The plots show the functional relationship between prices and two of the four determinants controlling for the other two. The partial dependence plots can be interpreted similarly to the coefficients in linear or logistic regression models. However, PD plots can capture more complex patterns from data, and they can be used with any model. For a formal definition as well

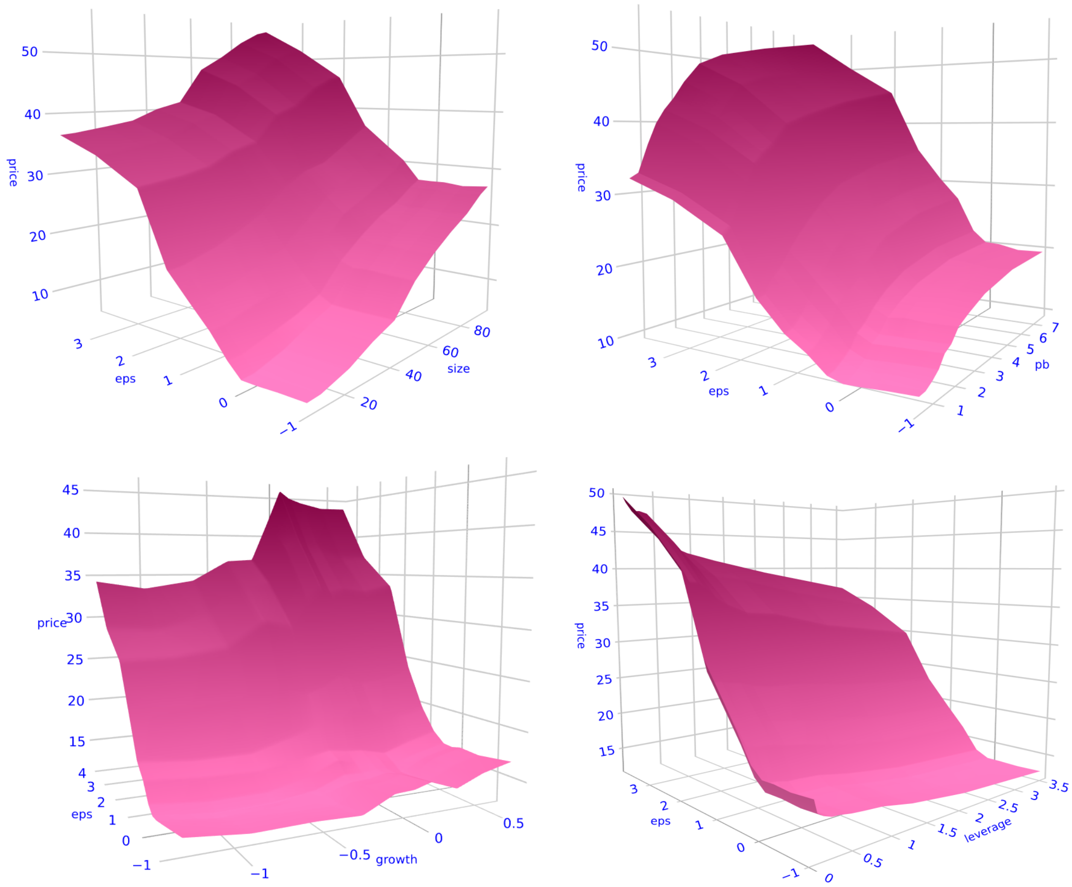

Figure 9: **Prices as a non-linear function of earnings, risk and past earnings growth.** The graphs display the estimated non-linear relation between prices and the pairs of determinants: (NI, $\log(Mktv)$) (top, left), (NI, P/B) (top, right), (NI, median NI growth) (bottom, left) and ($\log(Mktv)$, P/B) (bottom, right), keeping the third determinant constant, for a typical year in the sample (2011). The top-left graph shows a slope of the functional relation $NI \rightarrow price$ that does not vary much with size when holding constant the P/B and previous growth. The top-right graph shows that the slope of the functional relation is augmenting significantly as P/B increases. The bottom-left graph shows that previous growth has an impact on the slope of the earnings-to-price relation: positive previous growth corresponds to higher slope, all other determinants equal. Finally, the bottom-right graph shows that for the same earnings level and past growth, prices are higher for larger firms and firms with larger P/B. This confirms the fact that larger firms and firms with higher P/B have lower risk/higher growth.

Figure 10 displays the estimated non-linear functional relation between prices and the pairs: earnings and size (top, left), earnings and price-to-book ratio (top, right), earnings and past earnings growth (bottom, left), and size and price-to-book ratio (bottom, right). The graphs display significant non-linearities. We are interested, in particular, in the functional relationship *NI* $\rightarrow$ *prices* which can be followed by 'slicing' the first three graphs at different levels of the second variable in the pair, that is size, price-to-book, and *NI* growth, respectively. An important feature of the relation is its local linear nature: it is constant for negative values of earnings, it increases roughly linearly, in some cases, flattening out for large values of income. The main changing parameter of interest in this piece-wise linear relationship is the slope of the (relatively) linear middle part.

We interpret the graphs in figure 10 in the light of this changing parameter. If we 'slice' the left-top graphs at different values of size, we note that the functional relation *NI* $\rightarrow$ *prices* (which is flat for negative values of earnings) flattens also for large *NI*, in the case of small size firms, while the roughly linear increase of prices as a function of earnings continues for the whole range for larger firms. In the case of the top-right graph, the linear relation in the middle range of earnings flattens for all values of the price-to-book ratio. The top two graphs differ also in the way the slope of the earnings-to-price relation varies with the determinant. In the left graph, the slope of the earnings-to-price relation does not vary much with size when holding constant the P/B and previous growth, while the right graph shows that the slope of the functional relation is augmenting significantly as P/B increases. On the right, the slope augments rapidly up to a ratio of $P/B = 4$ and then more slowly.

The two-top graphs also show that, for the same level of earnings, large firms and firms with larger price-to-book ratio are priced higher in line with the intuition that risk and growth have a negative, respectively positive, marginal effect on share price.

The bottom-left graph shows that previous growth has an impact on the slope of the relation between earnings and price as well as on the level of the price. Positive previous growth corresponds to higher slope and higher price per unit of earnings, all other determinants equal. Finally, the bottom-right graph shows that for the same earnings level and past growth, prices are are higher for larger firms and firms with larger P/B. This confirms the fact that larger firms and firms with higher P/B have lower risk/higher growth.

6.2 Firm-specific dependence of price on NI

In this section we take a look at the *individual conditional expectation* (ICE) curves displaying the firm-specific association of earnings to prices. An ICE plot visualizes the dependence of price for each firm in a cross-section separately, resulting in one line per firm. One can think of each ICE curve as a simulation that shows what would happen to a given firm's stock price if only the level of earnings of the firm varies while the other determinants are constant (at the values they take for the firm under investigation). The simulation is based on the estimated model (14). For more details about how the ICE curves are constructed, see section 9.7 and Goldstein et al. (2015).

Figure 11 displays the *ICE* curves for four different cross-sections: 1974, 1983, 1995, 2013. Each curve depicts the association of prices with earnings for a specific firm in the given cross-section. The graphs in the figure unveil a locally linear structure of the relation between earnings and prices. They give clear evidence that the earnings-price relationship is flat in the beginning and concave in the end of the range 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 evolution of the curves is typically linear,

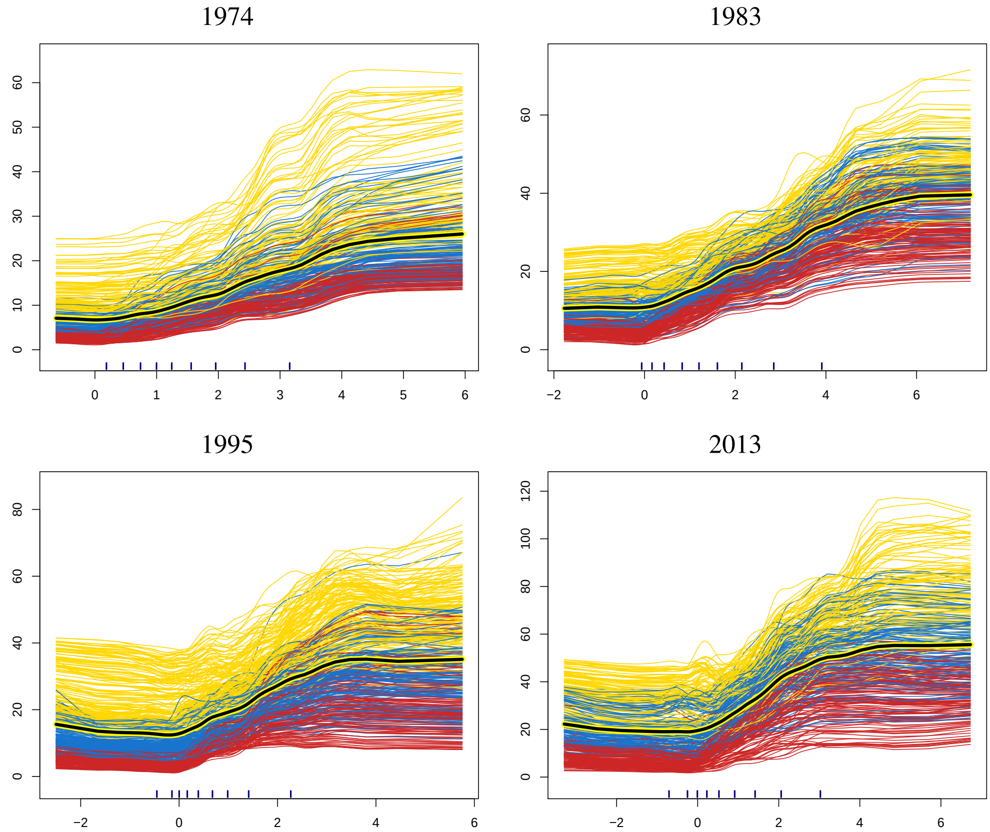

Figure 10: **Individual Conditional Expectation curves (ICE curves).** The graphs display the estimated functional relationship between earnings and prices for all the firms in the cross-section specified above each graph. More precisely, each ICE curve shows what would happen to prices if one varied only the level of earnings of a particular firm in the cross-section holding the other determinants 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 33%(66%)-percentile (above) of the size variable in the cross-section are colored in red (yellow). The firms with the size in the middle range of the cross-section are depicted in blue. The graphs give clear evidence of the common structure of the functional form of the earnings-price relationship for all firms: flat in the beginning and in the end of the range and roughly linear in the middle. It is also shows a clear positive dependence between the price level and size (for the same value of disclosed earnings).

with curves in the bottom (lower prices, i.e., higher/lower risk/growth) displaying lower slopes than the curves on 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 these ranges an increase of one unit of earnings is associated with a small (or no) increase in the share price. The thick black line (bordered by yellow) is the mean of all curves (firms) in a given year.

The firm-specific price-earnings relation visible in Figure 11 is consistent with the transitory nature of extreme earnings. Share prices associated with extreme earnings are not proportionately as small/large as those associated with the 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 sustained extremely high level of profitability unlikely. At the other end of the earnings range, one can argue that losses are likely to be temporary as shareholders can always liquidate the firm rather than suffer indefinite losses (Hayn (1995)).

The type of price-earnings non-linearity we document is consistent with the assumption of limited liability (Fischer and Verrecchia (1997)). When liability is limited, equity holders have a call option on the firm. This option captures all the upside potential of a positive movement in price while the down-side risk is limited. This implies a non-linear effect of disclosure on prices: their response to news that exceeds expectations is stronger than to news that falls short of expectations.

At the extreme positive end of earnings, the concave inflection in the price-earning relationship fits the theoretical prescription of the case of convertible debt. When debt is convertible, equity price is increasing in earnings, bounded below by zero, strictly convex for low levels of earnings, and strictly concave for high levels of earnings (Proposition 4 in Fischer and Verrecchia (1997)). The response of debt prices to information differs from that of equity because debt holders have a different type of limited liability. While they have no liability on the downside (much like equity owners), they have no potential for gain on the upside, in contrast to equity holders.

At the other end of the earnings range, the shape of the non-linearity put forth 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." Consequently, only firms expecting 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, hence, negative earnings will not be associated with a decline in the 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, and a linear function of X, the expected earnings per period in perpetuity, $k * X$ above that level. The slope[17] of the linear function k is the ER coefficient.

Consistent with these considerations, the graphs in figure 11 suggest a straightforward extension of Hayn (1995)'s abandonment option model which allows for a concave inflection at the positive extreme end of earnings range, as implied by the assumption of convertible debt (Fischer and Verrecchia (1997)). The functional

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

form of the depicted dependency can be well-approximated by a local linear function with three components (instead of two as in the cited model). Two of them, the first and the third[18], are flat, while the second component has a slope which is firm-specific. Formally, the extension of the abandonment option model looks like:

$$\hat{P}_{i,0}(X) = \begin{cases} L_{i,0}, & X \leq 0, \\ L_{i,0} + k_{i,0} \times X, & 0 \leq X \leq U_{i,0}, \\ L_{i,0} + k_{i,0} \times U_{i,0} + \tilde{k}_{i,0} \times (X - U_{i,0}), & U_{i,0} \leq X \end{cases} \tag{15}$$

where $X = \mathbb{E}_0[NI_{i,1}]$, the expected earnings next period and $\tilde{k}_{i,0} \leq k_{i,0}$. The constants $L_{i,0}$, $U_{i,0}$, and the slopes $k_{i,0}$ and $\tilde{k}_{i,0}$ are function of growth and risk and are firm-specific. The slope $k_{i,0}$ is the non-linear firm-specific levels earnings-response coefficient. The slope $\tilde{k}_{i,0}$, which models the price-earnings relation for extreme earnings is often equal to 0. The main differences with the abandonment option model is that the expression in (15) takes into account also the transitory nature of extreme positive earnings and that it acknowledges that the parameters of the model are firm-specific.

In the sequel, we summarize the cross-sectional values of the firm-specific levels earnings-response coefficient $k_{i,0}$ by the mean slope over the cross-section 0:

$$ERC_0 := \underset{i \in \text{cross-section } 0}{\text{mean}} k_{i,0} \tag{16}$$

that we interpret as a cross-section levels earnings-response coefficient. ERC_0

[18] The third component might be a linear relation with a slope that is smaller than that of the second component

improves over the market price-earnings ratio

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

which is commonly used in the literature to gauge the size of the earnings-response coefficients (see section 9.8) and section 7.2).

Visually, ERC_0 is the slope of the linear part of the thick black line (bordered by yellow) in the graphs in figure 11. It gives the synthesis, over all firms in a cross-section, of the local linear relation which ties prices to earnings in the central range of earnings.

The firms (curves) in Figure 11 are color-coded. The firms with size below (above) the 33%(66%)-percentile (above) of the size variable in the cross-section are colored in red (yellow). The firms with size in the middle range of the cross-section's size values are depicted 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 on top). To the extent to which size is a proxy for risk, the positive association between the valuation based on expectations informed by the level of earnings and firm size is consistent with risk negatively affecting value.

Figure 12 displays the *ICE* curves for the 2006 cross-sections and sheds some light on the relationship between earnings' value and the three proxies used in the estimation of the model (13) for the risk and growth determinants. It also shows the variability of the firm-specific cost of equity implied by the firm-specific estimated slope $k_{i,2006}$. Each of the first 3 graphs color-codes the firms (curves) based on a different proxy which is mentioned in graph's title. The firms with proxy levels below (above) the cross-section proxy median are colored in red (blue). We note a certain color-homogeneity in all three graphs consistent with the role as-

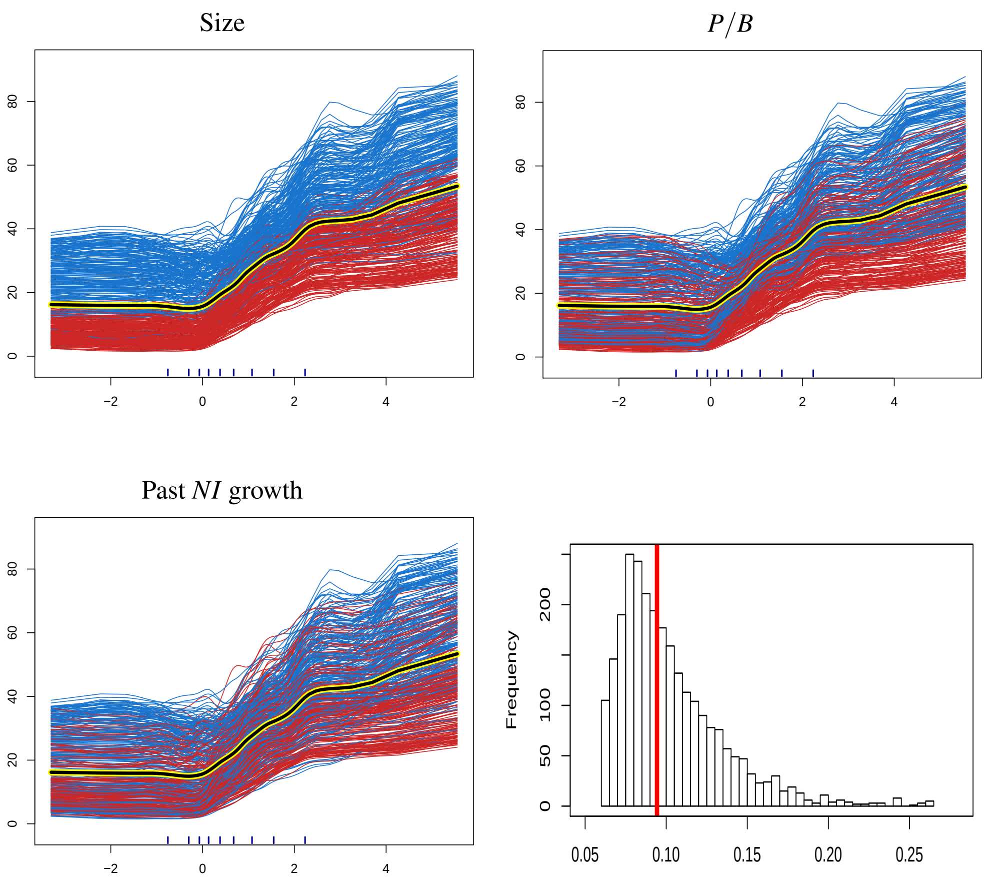

Figure 11: **Individual Conditional Expectation curves (ICE curves) for year 2006 (relation to proxies) and individual ERC.** The graphs on top and left-bottom display the estimated functional relationship between earnings and prices for all the firms in the 2006 cross-section color-coded for each of the three proxies/determinants named in graph's title. The firms with proxy values below (above) the cross-section proxy median are colored in red (blue). The thick black line (bordered by yellow) is the median of all curves (firms) in a given year. The graphs show a certain color-homogeneity consistent with the role assigned to the three proxies: larger firms should be less risky (higher values), firms with higher P/B or higher past earnings growth are expected to have stronger growth, and hence, higher prices for the same level of current NI. The degree however varies from size being most homogeneous and past earnings growth being least homogeneous. The bottom-right graph displays the histogram of the firm-specific cost of equity, i.e., $1/k_{i,2006}$. The vertical line corresponds to the median firm-specific cost of equity.

signed to the three proxies. To the extent to which P/B and $NI.g$ are proxies for growth, the second graph and the third graph are evidence of a positive marginal effect of a firm's growth on the price of earnings (the red curves tend to fall at the bottom of the graph compared to the blue ones) The individual cost of equity of the 2006 cross-section has a median of 9.5% and a mode of 7.5% which corresponds to a market risk premium of 4.6% or 2.6%, respectively.

To summarize, the relationships illustrated by the graphs are in line with the economic intuition. The graphs indicate a positive association between the firm-specific earnings' value and size, price-to-book ratio, and previous earnings growth proxies.

7 The non-linear version of 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 levels regression earnings-response coefficients it yields (16) have the 'right' size and imply economically justifiable risk-premium values.

In the previous section we defined the cross-section's non-linear earnings-response coefficient as the slope of the linear segment of the PD plot displaying the functional relationship between earnings and share prices (the thick black curves in Figure 11). Figure 13 displays a few instances[19] of the estimation of the *non-linear levels regression earnings-response coefficient*, i.e., the linear fit of a regression line on the linear segment of the 1-way PD plot. It also shows, for comparison, the price-earnings linear relation implied by the estimated linear price levels regression in (4) (dotted, red). We see clearly that the the slope of the linear fit of Ohlson's linear specification of the RI model is much smaller than the corresponding slope of the local linear fit of the non-linear research design.

[19]Due to space constraints, we abstain from showing all the 47 cross-sections.

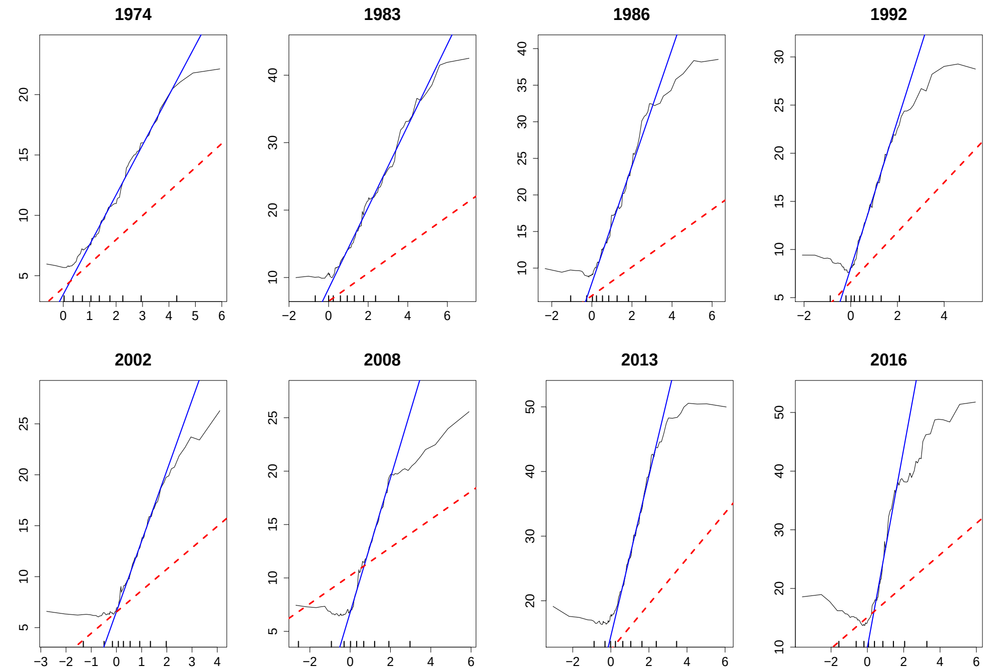

Figure 12: **Linear fit for the estimated functional relationship between earnings and prices - selected years.** The graphs display the estimated functional relationship between earnings and prices for all firms in the cross-section in the title of the graph (full, black), the linear fit for the range of earnings where the dependency is approximately linear (roughly between the third and the eight decile) (full, blue) and the linear fit of the levels regression (4) (dotted, red). We see clearly that the levels regression earnings-response coefficient is much smaller than the corresponding non-linear ERC.

7.1 The non-linear levels ERC has the ’right’ size

The size of the earnings response coefficients has been a outstanding puzzle since the beginning of the ’80s (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[20] of 6-20 (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)).

[20] The average price-earnings ratio in our sample ranges between 6.6 and 21.8.

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

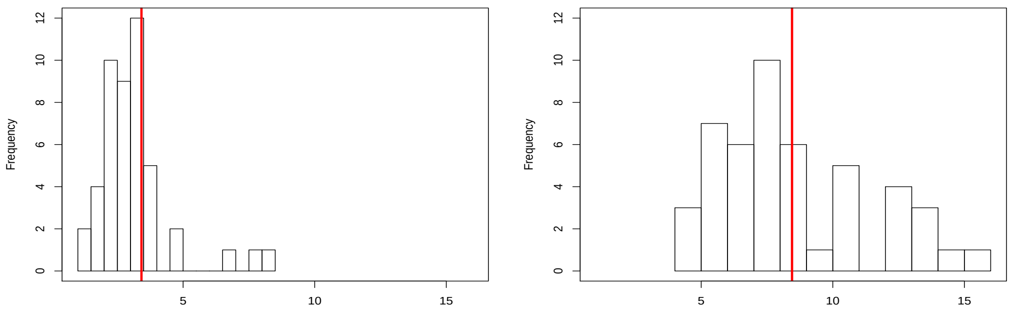

Figure 13: **The linear and the non-linear ERC.** The graphs show the histograms of the 47 cross-sectional levels linear ERC, $\beta_{2,0}$, in (4) (left) and cross-sectional mean slope of the linear component (for all firms in a given year) in the non-linear design, a_0 (right), respectively. The vertical line marks the mean ERC: 2.91 for the linear model and 8.01 for the non-linear design. The standard deviatins are of 1.39 and 2.50, respectively.

Figure 14 displays the histograms of the 47 values of the linear ERC obtained from the cross-sectional estimation of the linear levels regression in (4) (on the left-hand side) and of the *non-linear level regression earnings-response coefficient* ERC_0 defined in (16) (right-hand side), respectively. The mean ERC is equal to 2.91 for the linear model and to 8.01 for the non-linear design. While, as previously reported in the literature, the cross-sectional levels linear regression ERC are relatively small (compared to their predicted values), the non-linear level regression ERCs seem to have the right size.

7.2 Implied risk premium comparison

To better gauge the values of the ERC coefficients estimated in cross-section, we construct the cross-sectional risk premium implied by the two types of ERC coefficients. Barth and Kallapur (1996) argue that 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 earnings coefficients that yield unreasonably high cost of capital estimates as signs of the miss-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:

$$r_{e,0}^{(non-lin)} = 1/\widehat{ERC_0} - r_{f,0}, \tag{17}$$

where $r_{f,0}$ stands for the cross-sectional risk-free rate.[21] The linear approach based on Ohlson's linear specification (4) of the RI model yields another estimate:

$$r_{e,0}^{(lin)} = 1/\widehat{\beta}_{2,0} - r_{f,0}. \tag{18}$$

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 16 displays the time series of the risk-premiums implied by the linear research design (18) (top) as well as by the non-linear research design estimate (17) (bottom). The linear approach yields unreasonable large risk-premium estimates. In contrast to this, the risk-premium implied by the non-linear approach is economically justifiable and comes close to estimates produced by alternative approaches.

The criteria that we employ to validate the non-linear research design are nei-

[21] We used the 10 year US government T bond.

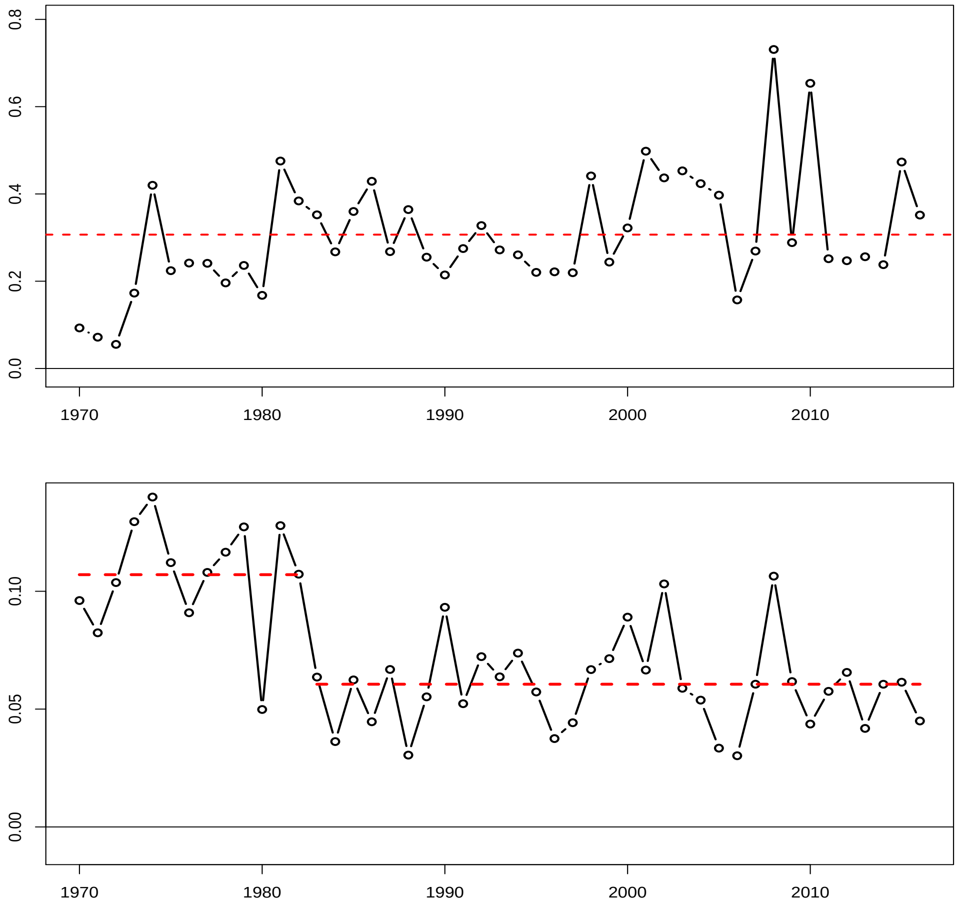

Figure 14: **The risk premium implied by the linear and the non-linear ERC.** The graphs display the time series of risk premiums implied by the levels regression earnings-response coefficient β_2 in (4), given by the expression in (17) (top) and by the non-linear ERC (16), defined in (18), respectively (bottom). 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%).

ther unique nor the only ones that can be used. Depending upon one's research design and loss function, different criteria will apply. Commonly (see Kothari and Zimmerman (1995), Kothari (2001)), the estimated earnings response coefficients are compared to the market price-earnings ratios. We avoid using this criteria for sound statistical reasons. The earnings-price 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.[22] As such, it is an extremely unreliable estimator, i.e., with an extremely large variance. As further evidence of the lack of statistical reliability, we note that estimating 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 9.8 in the Appendix).

8 Conclusions

Most of the seminal papers investigating the association of earnings to prices were written before the recent exponential developments of non-linear modeling and prediction methodology in the fields of statistical and machine learning. Since the accounting literature is highly consensual on the fact that the relation of prices to earnings is non-linear, advances in non-linear modeling offer a promising avenue for accounting research in capital markets. Our paper is a step in this direction.

We show with concrete examples how non-linearities affect a miss-specified linear model. Even small changes in a sample, that do not affect in any way the economic relation expressed, causes the coefficients of a linear model to move significantly. Taking the linear inference face value, one would decide, falsely, against the hypothesis of equal economic relation. A linear model fit to a non-linear relation is, econometrically speaking, useless.

We outline a non-linear research design in which accounting considerations inform modern statistical inference methodology to address the issue of consistent estimation of the non-linear relation between earnings and prices. We start

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

with a general accounting model and show how to deduce a non-linear regression specification of the economic relation in the model. Besides the level of earnings, the specification contains (as independent variables) proxies for the determinants of firm's future income stream. The non-parametric feature of our design does not impose an arbitrary structure on the price-returns relation. Instead, we use modern techniques from the field of statistical learning that allows the data to "speak for itself".

We estimate the non-linear regression using the Random Forest (RF) approach, a popular methodology in machine learning and Big Data applications of non-parametric, non-linear regression. The RF method approximates the price-earnings relation of a given firm, with a local average of the prices of firms with earning and determinants of future income close in value to those of the given firm.

Using recently developed visualizations techniques from the of data analytics (partial dependency plot, ICE curves), we uncover and visualize a complex relation between price, earnings, and proxies for risk and growth. The estimated price-earnings relation is consistent with limited liability and the option-view of equity (Hayn (1995), Fischer and Verrecchia (1997)): equity price is increasing in earnings, bounded below by zero, strictly convex for low levels of earnings, and strictly concave for high levels of earnings.

The non-linear research design allows for the estimation of a firm-specific earnings-response coefficient. Following Kothari and Zimmerman (1995), we validate the novel research design by verifying that the non-linear levels regression earnings-response coefficients have the 'right' size and yield economically justifiable risk-premium values.

The research design we propose focuses on the consistent estimation of the price-earnings relation and can be used for studying a wealth of other themes than the estimation of ER coefficients. Other applications of the non-linear research design can be found in Starica and Kang (2017) and Starica and Giosi (2019).

9 Appendix

9.1 Ohlson and Ohlson Juettner-Nauroth models

The Ohlson model is a linear expression of the *RI* valuation relation (Preinreich (1936) and (1938), Edwards and Bell (1961), Peasnell (1982)),

$$\begin{aligned} P_{i,0} = B_{i,0} + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[NI_{i,t} - r_{i,0} \times B_{i,t-1}]}{(1+r_{i,0})^t} = B_{i,0} + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[RI_{i,t}]}{(1+r_{i,0})^t} \\ = B_{i,0}\left(1 + \sum_{t=1}^{\infty} \frac{\mathbb{E}_0[RI_{i,t}/B_i]}{(1+r_{i,0})^t}\right), \end{aligned} \tag{19}$$

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 (19), Ohlson (1995) makes two additional assumptions on the dynamics of *RI*:

$$\begin{aligned} RI_{i,t} &= \omega_i RI_{i,t-1} + \nu_{i,t-1} + \varepsilon_{1,i,t}, \\ \nu_{i,t} &= \delta_i \nu_{i,t-1} + \varepsilon_{2,i,t}, \end{aligned} \tag{20}$$

where $0 \leq \omega_i, \delta_i < 1$ are two constants determined exogenously by firms economic environment and its accounting practices, $k_{i,0} = \omega_i r_{i,0}/(1+r_{i,0}-\omega_i)$, $\alpha_{i,0} = 1 + r_{i,0}/(1+r_{i,0}-\omega_i)(1+r_{i,0}-\delta_i)$. 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 (19) admits a linear representation that explicits prices as a linear combination of current book values and residual earnings:

$$P_{i,0} = B_{i,0} + k_{i,0} r_{i,0}^{-1} \times RI_{i,0} + \alpha_{i,0} \times \nu_{i,0}. \tag{21}$$

However, this linear representation is not necessarily a regression and hence cannot motivate 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 (aka AEG model) that allows for deviations from the clean surplus relation that are to be expected under current major accounting standards (IFRS and US GAAP). 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* :

$$P_0(OJ) = \frac{E_0[NI_{i,1}]}{r_{i,0}} + \sum_{t=1}^{\infty} \frac{E_0[NI_{i,t+1} + r_{i,0} dps_{i,t} - (1 + r_{i,0}) NI_{i,t}]}{r_{i,0}(1 + r_{i,0})^t}. \tag{22}$$

This model is more general than the 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.

9.2 Ohlson's linear representation of the RI model is not necessarily a regression

This section identifies supplementary second order condition on the dynamics of *RI* that are needed for the Ohlson's linear representation of the RI model to be a linear regression. There is no particular reason to assume that similar conditions actually hold in the data.

Proposition 1. *Assume prices verify the RI valuation model (19) and Ohlson's assumptions (20). If the sequence* $(\varepsilon_{1,t}, \varepsilon_{2,t})$ *is second-order stationary 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 (21) to be a linear regression. $\gamma_X(h) := Cov(X_t, X_{t-h})$ *denotes the auto-covariance function of the series* X_t *while* $\gamma_{\varepsilon_1,\varepsilon_2}(h) := Cov(\varepsilon_{1,t}\varepsilon_{2,t-h})$.

Proof. Note that the linear dynamics assumptions (20) imply (we suppress the index making the quantities firm-specific) that:

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

and that

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

for all t. Hence

$$cov(\nu_t, \nu_{t-h}) := \gamma_\nu(h) = \sum_{i,j=0} \delta^{i+j} \gamma_{\varepsilon_2}(h+j-i)$$

Note that

$$cov(\nu_t, RI_t) = Cov(\nu_t, \nu_{t-1} + \omega \nu_{t-2} + \omega^2 \nu_{t-3} + \ldots)$$
$$+ Cov(\nu_t, \varepsilon_{1,t} + \omega \varepsilon_{1,t-1} + \omega^2 \varepsilon_{1,t-2} + \ldots).$$

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

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

while the second term looks like:

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

Since $cov(\nu_t, RI_t) = 0$ is a necessary condition for linear expression in (21) 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 why 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 (21) 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 0.*

It 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[23] 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* could be similar in consecutive years if the economic conditions in which the firms 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 *RI* this year is likely to be followed by a large/small correction to get next year's *RI*. Unless $corr(\nu_0, \omega RI_{-1} + \varepsilon_{1,0})$ is negative and compensates exactly $corr(\nu_0, \nu_{-1})$, the 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 (21) is not a linear regression associating prices with book values and *RI*,.

9.3 Non-linear regression

Definition 1. *The decomposition:*

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

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

is called a regression *if*

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

or equivalently, if the orthogonality condition

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

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

In particular, the pair (X,Y) is related by a *linear regression*[24] if $\mathbb{E}[Y|X]$, the expected value of Y conditional on X, is a linear function of X:

$$\mathbb{E}[Y|X] = \beta_0 + \beta_1 X,$$

or equivalently, if

$$\mathbb{E}[Y - \beta_0 - \beta_1 X|X] = 0.$$

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

[24] The relation between the components of any random pair (X,Y) can *always* be represented in a linear form as: $Y = \beta_0 + \beta_1 X + (Y - \beta_0 + \beta_1 X) := \beta_0 + \beta_1 X + \varepsilon$. However, the relation is a linear regression if and only if $\mathbb{E}[\varepsilon|X] = \mathbb{E}[Y - \beta_0 + \beta_1 X|X] = 0$, i.e., $\mathbb{E}[Y|X] = \beta_0 + \beta_1 X$.

The orthogonality condition, essential to the definition of a regression, is *necessary*[25] for its consistent estimation (Stock and Watson (2012), Györfi et al. (2002)). The following proposition states the inferentiability 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*[26] *of approaches that yield consistent estimates of the regression function* $f(X)$ *(Györfi et al. (2002), James et al. (2014)).*

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

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

This section gives the theoretical motivation of the passage from the accounting model in (9) to a non-linear regression relation in (13).

Succinctly, our approach is intuitive and works as follows. We start with a set of predictors of future income. Market's 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 term complements the information in the observed data. Formally, it is orthogonal

[25] 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.

[26] 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.

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 constraint. The resulting valuation specification is hence a non-linear regression that can be consistently estimated using techniques from the field of non-parametric statistics

Denote by `PREDICT.NI` a set of predictors of future earnings available at time 0 (for example, 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 (9) that are regressions (in the sense of definition 1):`

   Suppose prices are given by equation (9) and let `PREDICT.NI` *be a set of predictors of future NI available at time* 0 *(as specified above). Then, there exists* $f_{i,0}$*, a possibly non-linear, firm- and predictor-specific function, and* ε_i*, an error term, such that:*

   $$P_{i,0} = f_{i,0}(\texttt{PREDICT.NI}_{i,0};\ r_{i,0}) + \varepsilon_{i,0} \qquad (23)$$

   where

$$\mathbb{E}[\varepsilon_0 \,|\, \texttt{PREDICT.NI}_0] = 0. \tag{24}$$

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

   Moreover,

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

   i.e., the regression function in the decomposition (23) 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 investors correction based on other information (than the levels of the predictor set) available at time 0

   The error term in the decomposition (23), ε_i, amounts to investors 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 predictors levels inform expectations of future earnings incorporated in prices. Higher absolute values indicate a lower contribution of the predictors levels to pricing.

Condition (24) guarantees that the non-linear regression specifications of the accounting valuation (9) in (23) can be consistently estimated.The issue of omitted variable 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 (9) for testing hypotheses about the relation between prices and any choice of accounting constructs associated to prices. For a given set of predictors of future income streams, the result states the existence of a specification of the general valuation relation (9) 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's setting of firm's price. The researcher can test her hypotheses on the association between prices and the set of predictors 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, not limited to):

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

if the research concerns earnings;

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

when interested in the pertinence of bottom-line items or

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

for studies of 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 split its derivation in a couple of steps which better convey the intuition behind the decomposition in (23).

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 are used to projects future financial statement data. Second, the link between those forecasts and current value must be specified: in our case, by the valuation expression in (9). While observed levels of the predictors project future earnings, other information is available to investors. They use this 'other information' to correct these projections into expectations of future RI (first link). Consequently, the prices, given by the valuation expression in (9), 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 $(\texttt{PREDICT.NI}_i,\ f_t(\mathbf{NI}_{i,t},\mathbf{O}_{i,t}\ ;\ r_{i,0}))$. For the firm i, we write:

$$\begin{aligned} f_t(\mathbf{NI}_{i,t},\mathbf{O}_{i,t}\ ;\ r_{i,0}) &= \mathbb{E}[f_t(\mathbf{NI}_{i,t},\mathbf{O}_{i,t}\ ;\ r_{i,0})\,|\,\texttt{PREDICT}_{i,0}] + \delta_{i,0}(t) \\ &:= g_{i,0}(\texttt{PREDICT.NI}_{i,0};\ t) + \delta_{i,0}(t). \end{aligned} \tag{25}$$

In words, we decompose the function of future *NI* *t* time units ahead, into its best projection[27] by the current values of the predictors, $g_{i,0}(\texttt{PREDICT.NI}_{i,0};\, t)$, and a left-over piece, orthogonal to the projection. The orthogonality condition reads:

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

The left-over piece amounts to the part of the function of future NIs 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[28] the persistence of earnings and are an expression of the factors that determine the abnormal earnings: firm size, product-type, capital intensity, barriers-to-entry, accounting practices (see also sections 5.3.1 and 9.5 in the 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 (25) for the process of expectation formation, we condition with the set of information available at time 0 to get market's expectation of the function f_t:

$$\begin{aligned}\mathbb{E}_0[f_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\,;\, r_{i,0})] &= \mathbb{E}_0[g_{i,0}(\texttt{PREDICT.NI}_i;\, t)] + \mathbb{E}_0[\delta_{i,0}(t)] \\ &= g_{i,0}(\texttt{PREDICT.NI}_i;\, t) + \mathbb{E}_0[\delta_{i,0}(t)] \\ &= g_{i,0}(\texttt{PREDICT.NI}_{i,0};\, t) + \varepsilon_{i,0}(t), \qquad (26)\end{aligned}$$

[27]Recall that, according to proposition 1, $\mathbb{E}[f_t(\mathbf{NI}_{i,t}, \mathbf{O}_{i,t}\,;\, r_{i,0})\,|\,\texttt{PREDICT.NI}_{i,0}]$ 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 $g_{i,0}(\cdot\,;\,t)$.

[28]The empirical specification of the valuation regression relation will assume that they are roughly similar for comparable firms.

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 later. Its complementarity is reflected by its orthogonality to the information in the current levels of the predictor set. To see this, note that, by the theorem of iterated expectations, one has:

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

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 a term collecting the contributions of the observed predictors to expectation formation and another assembling the corrections brought by other available information.

Expression (9) together with the decomposition in equation (26) imply:

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

where the errors ε_0 verify:

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

This ends the proof. □

9.5 Literature on the determinants of the functions f_t

Larger firms have more financial resources to diversify and hence to stabilize growth which leads to a more persistent earnings flow (Scherer (1973), Martin (1988)). The less stable patterns of demand for durable goods firms result in less sustainable residual earnings relative to non-durable goods firms (Friedman (1957), Darby (1972), Zarnowitz (1972), Caves (1987)). As the persistence of abnormal earnings is a function of the market pressures facing the firm and its production technology, higher barriers-to-entry decrease competition by limiting entry into an industry which in turn leads to market share stability, and to sustainable earnings growth (Stigler (1963), Mueller (1977), Kamerschen (1968)). Earnings persistence is hence an increasing function of barriers-to-entry, a measure of competition (Lev (1983)). Capital-intensive firms display more volatility of earnings relative to the smoother earnings stream of low capital-intensive firms, possibly due to operating leverage reflected by the proportion of fixed cost to total cost (Scherer (1973), Lev (1983)).

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[29], managerial

[29] Following the literature, we define non-discretionary conservatism as conservative accounting resulting from the unbiased application of GAAP (Lawrence et al. (2013), Ahmed et al. (2000)). Unconditional conservatism consists of the aspects of the accounting process determined at the inception of assets and liabilities that yield an understatement of the book value of net assets

preferences are a firm-specific obvious driver[30].

In an economic-theory-motivated investigation of the firm characteristics determining abnormal earnings, Cheng (2005) documents[31] that firm differential abnormal ROE increases with market share, firm size, firm level barriers to entry, and firm conservative accounting factors. For the empirical specifications of the regression valuation model, one might plausibly assume that size is a good proxy for market share.

9.6 Partial dependency (PD) plot

A partial dependence plot displays the functional relationship between a small number of input variables and the dependent variable. They show how the Y partially depend on values of the input variables of interest. For example, a partial dependence plot can show whether the price of a share increases linearly with earnings. It can show whether small size will decrease the value of a share (all other input variables constant). The partial dependency plot can also show the type of relationship, such as a step function, curvilinear, locally linear and so on.

The simplest partial dependency plots are 1-way plots, which show how a model's dependent variable changes with a single input. Examples of 1-way plots are the thick black curves in the graphs in Figures 11 and 12.

The partial dependency plots are constructed in two steps and are easy to comprehend. In the sequel we explain the construction of the partial dependency be-

---

relative to their market value (Beaver et al. (2005)).

[30] Previous research indicates that managers voluntarily engage in conservative accounting given certain contracting incentives (e.g., Ahmed et al. (2002)), for reputational reasons (Diamond (1991)) or under threat of litigation (Qiang (2007))

[31] Cheng (2005) assumes that abnormal ROE follows a first-order auto-regressive process and allows the auto-regressive parameter (i.e., persistence) to vary with economic rent proxies and conservative accounting factors.

tween NI and prices. The reader can replace the NI variable with any other independent variable or pairs of independent variables among $(NI, P/B, Mktv, NI.g)$. The formal definition is given for a 2-way PD plot associated to the pair $(NI, P/B)$.

In the first step, we fix a firm, and create variants of it by replacing the actual NI value of the firm in the cross-section with values from a grid while keeping all other independent variables the same, i.e. at their values corresponding to the firm under investigation.For each of the newly created firms we make predictions with the Random Forest model (14). The result is a set of points with the NI value from the grid in abscissa and the respective predictions as the ordinate. We do Step 1 for all firms in the cross-section which yields n (the number of firms in cross-section) functions relating firm share prices to NI. In Step 2, we average the n functions to produce one function of NI, the partial dependency of prices on NI for a given cross-section.

Let us now give the formal definition for a 2-way PD plot. We have estimated the regression function in (14):

$$(NI,\ P/B,\ Mktv,\ NI.g\ SIC) \to \widehat{f}(NI,\ P/B,\ Mktv,\ NI.g,\ SIC),$$

and, to fix the ideas, suppose we want to visualize the relation between prices and the pair $(NI, P/B)$ implied by our estimation. For a given pair of values $(NI, P/B)$, the *partial dependency plot* is defined as:

$$\widehat{f}_{(NI,P/B)}(NI, P/B) := \frac{1}{n}\sum_{i=1}^{n} \widehat{f}(NI;\ P/B,\ Mktv_{i,0},\ NI.g_{i,0},\ SIC_{i,0}).$$

9.7 Individual Conditional Expectation (ICE) plot

Partial dependency plots provide a coarse view of a model's functioning. They are a global method as they do not focus on specific firms, but on an overall average. The equivalent to a partial dependency plot for an individual firm is called Individual Conditional Expectation (ICE) plot (Goldstein et al. (2015)). An ICE plot visualizes the dependence of price for each firm in a cross-section separately, resulting in one line per firm, compared to one line overall in partial dependence plots. A partial dependency plot is the average (median) of the lines of an ICE plot.

For a given independent variable of interest (say NI), the values for a line (that is, a firm) in the ICE plot are computed by keeping all other independent variables the same (at their values corresponding to the firm under investigation), creating variants of the firm by replacing the NI value with values from a grid and making predictions with the Random Forest estimation of the model (14) for these newly created firms. The result is a set of points for our firm with the NI value from the grid in abscissa and the respective predictions as the ordinate.

9.8 Market price-earnings ratio

ERC_0 improves over the market price-earnings ratio

$$\frac{(\text{average price})_0}{(\text{average } NI)_0} \tag{27}$$

which is commonly used in the literature to gauge the size of the earnings-response coefficients. Using the expression in (15), 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 earnings-response coefficient 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 (15).

The statistical unreliability of the earnings-price 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:

$$\frac{(\text{average NI})_t}{(\text{average price})_t} - r_{f,t} \tag{28}$$

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

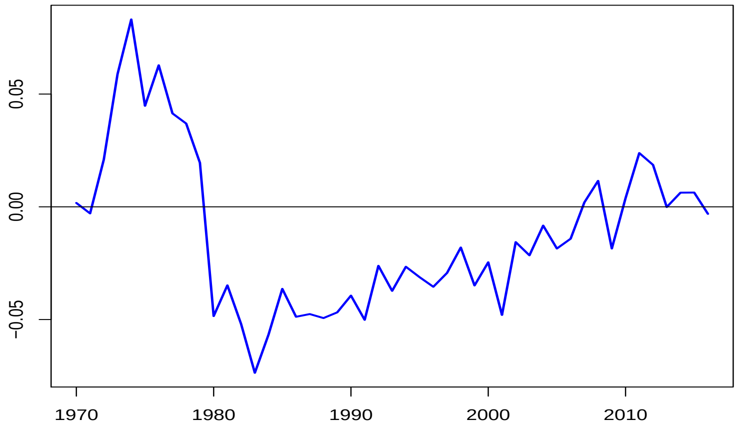

Figure 15: **The risk premium implied by the ratio of average price to average earnings per share in cross-sections.** The graphs display the time series of risk premium (28) implied by ratio of average price to average earnings per share in cross-sections. The risk premium is mostly negative for the period from 1980 to 2010.

References

Ahmed, A., Billings, B. K., Morton, R. M., and Stanford-Harris, M. (2002). The role of accounting conservatism in mitigating bondholder-shareholder conflicts over dividend policy and in reducing debt costs. *The Accounting Review*, 77(4):867–890.

Ahmed, A. S., Morton, R. M., and Schaefer, T. F. (2000). Accounting conservatism and the valuation of accounting numbers: Evidence on the Feltham-Ohlson (1996) model. *Journal of Accounting, Auditing & Finance*, 15(3):271–292.

Aledo Martinez, J., Garcia Lara, J. M., Gonzalez Perez, M. T., and Grambovas, C. A. (2019). An empirical assessment of proposed solutions for resolving scale problems in value relevance accounting research. *Accounting & Finance*.

Baginski, S. P., Lorek, K. S., Willinger, G. L., and Branson, B. C. (1999). The relationship between economic characteristics and alternative annual earnings persistence measures. *The Accounting Review*, 74(1):105–120.

Barth, M. and Clinch, G. (2009). Scale effects in capital markets-based accounting research. *Journal of Business Finance and Accounting*, 36(3-4):253–288.

Barth, M. E. and Kallapur, S. (1996). The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research*, 13(2):527–567.

Beaver, W., Lambert, R., and Morse, D. (1980). The information content of security prices. *Journal of Accounting and Economics*.

Beaver, W., McNichols, M., and Rhie, J.-W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, 10(1):93–122.

Bernard, V. L. (1995). The feltham-ohlson framework: Implications for empiricists. *Contemporary Accounting Research*, 11(2).

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

Caves, R. (1987). American industry: Structure, conduct, performance. *Englewood Cliffs, NJ: Prentice Hall*.

Cheng, Q. (2005). What determines residual income? *The Accounting Review*, 80(1):85–112.

Christie, A. A. (1987). On cross-sectional analysis in accounting research. *Journal of Accounting and Economics*, 9(3):231 – 258.

Collins, D. W. and Kothari, S. (1989). An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *Journal of Accounting and Economics*, 11(2–3):143 – 181.

Darby, M. R. (1972). The allocation of transitory income among consumers' assets. *The American Economic Review*, 62(5):928–941.

Diamond, D. (1991). Monitoring and reputation: the choice between bank loans and directly placed debt. *Journal of Political Economy*.

Easton, P. D. (1998). Discussion of revalued financial, tangible, and intangible assets: Association with share prices and non-market-based value estimates. *Journal of Accounting Research*, 36:235–247.

Easton, P. D. (1999). Security returns and the value relevance of accounting data. *Accounting Horizons*, 13(4):399–412.

Easton, P. D. and Zmijewski, M. E. (1989). Cross-sectional variation in the stock market response to accounting earnings announcements. *Journal of Accounting and Economics*, 11(2):117 – 141.

Edwards, E. and Bell, P. (1961). *The Theory and Measurement of Business Income*. University of California Press.

Feltham, G. A. and Ohlson, J. A. (1995). Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research*, 11(2):689–731.

Feltham, G. A. and Ohlson, J. A. (1996). Uncertainty resolution and the theory of depreciation measurement. *Journal of Accounting Research*, 34(2):209–234.

Fischer, P. E. and Verrecchia, R. E. (1997). The effect of limited liability on the market response to disclosure. *Contemporary Accounting Research*, 14(3):515–541.

Freeman, R. N. and Tse, S. Y. (1992). A nonlinear model of security price responses to unexpected earnings. *Journal of Accounting Research*, 30(2):185–209.

Friedman, M. (1957). A theory of the consumption function. *New York, NY: National Bureau of Economic Research, Inc*.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1):44–65.

Györfi, L., Kohler, M., Krzyzak, A., and Walk, H. (2002). A distribution-free theory of nonparametric regression. *Springer Series in Statistics*.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Hayn, C. (1995). The information content of losses. *Journal of Accounting and Economics*, 20(2):125–153.

Holthausen, R. W. and Watts, R. L. (2001). The relevance of the value-relevance literature for financial accounting standard setting. *Journal of Accounting and Economics*, 31(1–3):3–75.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning with Applications in R*. Springer.

Kamerschen, D. R. (1968). The influence of ownership and control on profit rates. *The American Economic Review*, 58(3):432–447.

Kormendi, R. and Lipe, R. (1987). Earnings innovations, earnings persistence, and stock returns. *The Journal of Business*, 60(3):323–345.

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

Kothari, S. and Zimmerman, J. L. (1995). Price and return models. *Journal of Accounting and Economics*, 20(2):155 – 192.

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

Lawrence, A., Sloan, R., and Sun, Y. (2013). Non-discretionary conservatism: Evidence and implications. *Journal of Accounting and Economics*, 56(2–3, Supplement 1):112 – 133.

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

Martin, S. (1988). Market power and/or efficiency? *The Review of Economics and Statistics*, 70(2):331–335.

Mueller, D. C. (1977). The persistence of profits above the norm. *Economica*, 44(176):369–380.

Ohlson, J. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11(2):661–687.

Ohlson, J. A. and Juettner-Nauroth, B. E. (2005). Expected eps and eps growth as determinantsof value. *Review of Accounting Studies*, 10(2):349–365.

Pae, J., Thornton, D. B., and Welker, M. (2005). The link between earnings conservatism and the price-to-book ratio*. *Contemporary Accounting Research*, 22(3):693–717.

Peasnell, K. (1982). Some formal connections between economic values and yields and accounting numbers. *Journal of Business Finance and Accounting*, 9(3):361–381.

Penman, S. H. and Zhang, X.-J. (2002). Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review*, 77(2):237–264.

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

Qiang, X. (2007). The effects of contracting, litigation, regulation, and tax costs on conditional and unconditional conservatism: Cross-sectional evidence at the firm level. *The Accounting Review*, 82(3):759–796.

Scherer, R. M. (1973). *Industrial Market Structural and Economic Performance*. Rand McNally, Chicago.

Shevlin, T. (1996). The value-relevance of nonfinancial information: A discussion. *Journal of Accounting and Economics*, 22(1):31 – 42.

Smith, C. W. and Watts, R. L. (1992). The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics*, 32(3):263 – 292.

Starica, C. and Giosi, P. (2019). The ebbing of accrual accounting (March 21, 2019). In *SSRN: https://ssrn.com/abstract=3072287 or http://dx.doi.org/10.2139/ssrn.3072287*.

Starica, C. and Kang, J. (2017). Why have measures of earnings quality changed over time? A competing narrative. *Working paper*.

Stigler, G. (1963). *Capital and Rates of Return in Manufacturing Industries.* Princeton, NJ.: Princeton University Press.

Stock, J. H. and Watson, M. W. (2012). *Introduction to Econometrics*. Addison-Wesley.

Zarnowitz, V. (1972). Forecasting economic conditions: Record and prospects. In Zarnowitz, V., editor, *The Business Cycle Today*. National Bureau of Economic Research, NY.

Zhang, G. (2000). Accounting information, capital investment decisions, and equity valuation: Theory and empirical implications. *Journal of Accounting Research*, 38(2):271–295.