

Essays on the relationship between stock prices and accounting variables

Thèse de doctorat présentée à la Faculté des sciences économiques

Institut du management de l'information

Université de Neuchâtel

Pour l'obtention du grade de docteur en Statistiques appliquées par

Pierluigi GIOSI

Évaluée par le jury de thèse :

Prof. Adrian HOLZER, Université de Neuchâtel, président du jury

Prof. Catalin STARICA, Université de Neuchâtel, superviseur

Dr. MER Paul COTOFREI, Université de Neuchâtel

Prof. Jan MARTON, University of Gothenburg, Sweden

Prof. Kenth SKOGSVIK, Stockholm School of Economics, Sweden

Soutenue le 03.12.2021

IMPRIMATUR POUR LA THÈSE

Essays on the relationship between stock prices and
accounting variables

Pierluigi GIOSI

UNIVERSITÉ DE NEUCHÂTEL
FACULTÉ DES SCIENCES ÉCONOMIQUES

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

Prof. Catalin Starica (directeur de thèse, Université de Neuchâtel)
Prof. Adrian Holzer (président du jury, Université de Neuchâtel)
Dr. MER Paul Cotofrei (Université de Neuchâtel)
Prof. Kenth Skogsvik (Stockholm School of Economics, Sweden)
Prof. Jan Marton (School of Economics, Business, and Law, University of Gothenburg, Sweden)

Autorise l'impression de la présente thèse.

Neuchâtel, le 14 décembre 2021



Le doyen
Valéry Bezençon

Preface

This thesis is the fruit of five years of Ph.D. under the supervision of Catalin Starica. I am grateful to him for the guidance, and to the members of the Ph.D. defense committee for their dedication and remarks.

Parts of this thesis—section 4.1 results—were presented at the Stockholm School of Economics, the 2018 Quantitative Finance Workshop at Roma Tre University, the 2018 International Accounting and Finance Doctoral Symposium at Bangor University, and the 2021 European Accounting Association annual congress, held online because of the Covid-19 pandemic. The presentations were based on the unpublished manuscript [Starica and Giosi \(2017\)](#), with the title “The ebbing of accrual accounting.” I thank the discussants and the participants for their comments and suggestions.

Abstract

Starica and Marton (2019) argue that a nonlinear estimation method based on simple, flexible Machine Learning algorithms can solve several econometric modeling issues in order to assess the relationship between stock prices and accounting earnings. In this thesis, we use their methodological advancements to address three problems in financial accounting.

First, we use the estimation technique they propose to measure the association that stock prices have with accounting earnings and operating cash flows.

Second, we suggest an explanation behind the stylized fact that the association between stock prices and accounting earnings has decreased in the last two decades. We argue that a change in the correlation between accruals and cash flows is the most likely reason for this phenomenon, and, consequently, our results suggest that cash flows have become at least as relevant as earnings in these last twenty years.

Third, we investigate whether the degree of association between stock prices and earnings depends on the quality of accruals. Our findings suggest that, on average, higher-quality earnings are more pertinent to valuation among firms that have comparable prices, operating uncertainty, and size of timing errors.

Keywords (in alphabetical order): Accounting, accounting quality, accruals, cash flows, earnings, Machine Learning, nonlinear regression, nonparametric statistics, Random Forest, stochastic dominance, stock prices, valuation.

Résumé

[Starica and Marton \(2019\)](#) soutiennent qu’une méthode d’estimation non linéaire basée sur des algorithmes de Machine Learning simples et flexibles peut résoudre plusieurs problèmes de modélisation économétrique afin d’évaluer la relation entre les cours des actions et les bénéfices comptables. Dans cette thèse, nous utilisons leurs avancées méthodologiques pour aborder trois problèmes de comptabilité financière.

Premièrement, nous utilisons la technique d’estimation qu’ils proposent pour mesurer l’association que les cours des actions ont avec les bénéfices comptables et les flux de trésorerie, respectivement.

Deuxièmement, nous suggérons une explication selon laquelle l’association entre les cours des actions et les bénéfices comptables semble avoir diminué au cours des deux dernières décennies. Dans cette thèse, nous utiliserons le terme anglais “accruals” pour indiquer la différence entre les bénéfices comptables (traduits comme “accounting earnings”) et les flux de trésorerie (traduits comme “operating cash flows”). Nos résultats suggèrent qu’un changement de corrélation entre accruals et flux de trésorerie est la raison la plus probable du phénomène mentionné ci-dessus, et, par conséquent, les flux de trésorerie sont devenus au moins aussi importants que les bénéfices comptables au cours de ces vingt dernières années.

Troisièmement, nous cherchons à savoir si le degré d’association entre les cours des actions et les bénéfices comptables dépend de la qualité de la comptabilité d’exercice.

Nous observons qu'en moyenne, les bénéfices comptables obtenus à partir d'une comptabilité d'exercice de meilleure qualité sont plus pertinents pour l'évaluation, parmi les entreprises dont les cours des actions sont comparables, ainsi qu'une incertitude opérationnelle et des "timing-errors" similaires.

Mots-clés (en ordre alphabétique) : Accruals, bénéfices comptables, comptabilité, comptabilité d'exercice, cours des actions, évaluation, flux de trésorerie, Machine Learning, qualité de la comptabilité d'exercice, Random Forest, régressions non linéaires, statistiques non paramétriques, stochastic dominance.

Remerciements

Ce que Catalin m'a appris, transmis ou simplement fait comprendre ces dernières années, qu'il s'agisse de statistiques ou non, sera un bagage qui m'accompagnera la vie durant. Merci, pas seulement pour le bon déroulement de mon doctorat.

Je tiens également à remercier les membres du jury pour leur dévouement et leur attention aux détails, pour leurs évaluations et pour les discussions pendant ma soutenance, sans aucun doute enrichissants.

J'aimerais dire merci à beaucoup d'autres personnes.

Jian Kang a joué un rôle important dans le bon déroulement de cette thèse. Je le remercie infiniment pour ses conseils et son implication durant ma première année de doctorat, mais aussi pour son aide vers une meilleure compréhension de notre base de données, et de son développement.

Je remercie mes collègues doctorant-e-s, peu importe de quel département et université, pour leur soutien et leur partage durant ces dernières années. Je souhaite, en particulier, bonne chance à Michael pour reprendre là où je me suis arrêté et à tous mes amis de l'IMI qui, tôt ou tard, auront l'occasion de terminer leurs projets (Eliane, je compte sur toi !).

Je suis arrivé à Neuchâtel en 2016 sans parler un mot de français. Je tiens à remercier Bahareh, Jihane, Franck, Bruno, Paula et Arnaud, Elody, pour leur accueil et le meilleur

cours de langue de ma vie. S'agissant de mon parcours en tant que doctorant, je tiens à remercier également Aronne, Bastien, Emilie, Fabio, Joseph, Juliette, Livia, Loïc, Lorenzo, Maria, Olivier, Salvatore et Sarah pour leurs conseils et les petits coups de pouce dont j'ai eu besoin. Merci encore à Lisa, parce qu'elle est spéciale.

Je tiens également à remercier mes collègues de l'École hôtelière de Lausanne, avec une mention particulière à Aurélien, Charlotte, Hedia et Petar, qui m'ont permis de mieux comprendre et apprécier les nombreux défis de l'enseignement. Je tiens également à les remercier pour la transition en douceur depuis l'Université de Neuchâtel.

Toujours en parlant de transition, je remercie mes nouveaux collègues des services de statistique et des formations postobligatoires et de l'orientation de la République et Canton de Neuchâtel pour leur chaleureux bienvenu à bord, et, tout particulièrement, à Eric pour son enthousiasme et les nombreux défis qu'il me propose au quotidien.

Pour conclure, je ne veux pas oublier de remercier ma famille pour la possibilité qu'ils m'ont offert et leur soutien tout au long de mes études et bien au-delà.

À vous tous, merci !

Contents

Preface	v
Abstract	vii
Résumé	ix
Remerciements	xi
1 Introduction and literature review	1
1.1 Why choose an alternative research design?	2
1.2 The relationship between stock prices and accounting variables	6
1.3 The relationship between stock prices and accounting quality	11
2 Research design	15
2.1 Nonparametric techniques used to model the relationship between stock prices and accounting variables	15
2.1.1 A nonlinear valuation	16
2.1.2 Estimating the expectation-formation pertinence of accounting variables	19
2.1.3 Relative and incremental expectation-formation pertinence	22
2.1.4 Stochastic dominance	25
2.1.5 Testing for a relative expectation-formation pertinence	27
2.1.6 Explaining changes in the expectation-formation pertinence	30

2.1.7	Changes in earnings' pertinence: the role of accruals smoothing	36
2.2	Econometric techniques used to estimate accounting quality and model its relationship with stock prices	39
2.2.1	The concept of accounting quality	39
2.2.2	Estimating accounting quality using a Gaussian state-space model	45
2.2.3	Estimating the relationship between earnings' expectation-formation pertinence and accounting quality	48
3	Variables and sample	51
3.1	Variables definition	51
3.1.1	Earnings, cash flows, and accruals	51
3.1.2	Regressand and regressors in the nonlinear regression model . .	52
3.1.3	Accounting constructs to explain variation in pertinence	53
3.2	Sample	60
4	Results	63
4.1	The relationship between stock prices and accounting variables	64
4.1.1	An illustration of the concept of stochastic dominance	65
4.1.2	Changes in earnings' pertinence compared to that of cash flows	68
4.1.3	A possible explanation for the decline in the earnings' pertinence	69
4.1.4	A possible explanation for the increase in the earnings volatility	78
4.1.5	Static analysis of the idiosyncrasy of earnings' relative pertinence to valuation	80

4.1.6	Dynamic analysis of the idiosyncrasy of earnings' relative pertinence to valuation	89
4.2	The relationship between earnings' expectation-formation pertinence and accounting quality	96
4.3	Robustness checks	97
5	Conclusions	103
	Appendix	105
A.1	An approximation of the distribution of the two Kolmogorov-Smirnov test statistics under the null hypothesis	105
A.2	Random Forests, a short R example	110
	References	120

List of Figures

1	Valuation errors' cumulative distribution function; c.d.f. conditional on the fiscal year	66
2	Valuation errors' cumulative distribution function; c.d.f. conditional on the correlation between accruals and cash flows	67
3	Kolmogorov-Smirnov test statistics	69
4	Earnings and cash flows volatility	71
5	Volatility of annual changes in earnings and cash flows	72
6	Earnings' pertinence relative to that of cash flows	77
7	Percentage of firms for which highly volatile earnings translate into a low earnings' pertinence relative to that of cash flows	78
8	Cross-sectional median intangible intensity, accruals volatility, and correlation between accruals and cash flows	81
9	Earnings' pertinence relative to that of cash flows; coefficients of the univariate regression models with binary variables	85
10	Earnings' pertinence relative to that of cash flows; the multivariate regression coefficients corresponding to the intangible intensity and accruals volatility	88
11	Earnings' pertinence relative to that of cash flows; the multivariate regression coefficients corresponding to the correlation between accruals and cash flows	89
12	Cross-sectional median earnings' pertinence relative to that of cash flows; unconditional median	91
13	Cross-sectional median earnings' pertinence relative to that of cash flows; median conditional on the correlation between accruals and cash flows	94

14	Sample size depending on the correlation between accruals and cash flows	95
15	Accounting quality and earnings' pertinence to valuation; cross-sectional regression coefficients corresponding to the proxies for economic performance uncertainty and the size of timing errors	98
16	Accounting quality and earnings' pertinence to valuation; cross-sectional regression coefficients corresponding to the proxy for quality	99

List of Tables

1	Stochastic dominance and expectation-formation pertinence	28
2	Kenneth R. French's twelve-industry classification	56
3	Variables definition	59
4	Sample quantiles	61

1 Introduction and literature review

In this thesis, we investigate the relationship between stock prices and accounting measures of economic performance, such as earnings and operating cash flows. [Starica and Marton \(2019\)](#) argue that a nonlinear estimation technique based on nonparametric Machine Learning algorithms can solve a number of econometric modeling issues in order to assess the relationship between stock prices and accounting earnings. We use their methodological advancements to address a series of topical problems in financial accounting.

First, we discuss the concept of expectation-formation pertinence relating to earnings and operating cash flows, explaining how we can express stock prices as nonlinear functions of firms' accounting profitability, risk, and degree of conservatism. We introduce the estimation technique proposed by [Starica and Marton \(2019\)](#) to measure earnings and cash flows' pertinence to the valuation process; that is, how we infer the association that stock prices have with earnings and cash flows. Subsections [2.1.1](#) and [2.1.2](#) consider this matter in detail.

We compare earnings and cash flows' pertinence to the valuation process, noting that the former accounting measure of economic performance has become less pertinent than the latter over the last two decades. We then investigate the potential reasons for this phenomenon; our results suggest that the most likely explanation for this is a reduction in the smoothing role played by accruals. Specifically, accruals smoothing can be measured by the correlation between accruals and cash flows, which has increased

over time. We explain all of these concepts in Subsections 2.1.3 to 2.1.6, with empirical findings given in Section 4.1. All these empirical investigations have been reported and discussed in [Starica and Giosi \(2017\)](#).

In our last investigation, we assess the relationship that accounting quality might have with the earnings' pertinence to the valuation process. To measure the accounting quality, we adopt a novel framework introduced by [Nikolaev \(2018\)](#). We examine an alternative estimation technique that is based on Gaussian state-space models, explaining its advantages when used in this empirical application. We find that, on average, higher accounting quality is associated with a greater earnings' pertinence and that the association is positive, statistically significant, and persistent over time. Section 2.2 describes this concept, before we discuss our empirical findings in Section 4.2.

1.1 Why choose an alternative research design?

We provide new insights into the open and pivotal debate on the information content of earnings and cash flows; this has been a fundamental topic for a long time, appearing in early market-based accounting literature ([Ball and Brown, 1968](#); [Lev and Ohlson, 1982](#); [Biddle et al., 1995](#)).

It is common practice to measure the association that stock prices have with a set of variables by estimating various specifications of a linear regression model. When the goal is to compare the information content in two different sets of accounting variables,

researchers usually compare regression slopes or coefficients of determination (R^2) to determine which set of variables is more relevant; see, for example, [Dechow \(1994\)](#) and [Hayn \(1995\)](#). The setup is often a simple univariate ([Dechow, 1994](#); [Hayn, 1995](#); [Guay and Sidhu, 2001](#)) or multivariate regression ([Easton and Harris, 1991](#); [Biddle et al., 1995](#); [Biddle et al., 1997](#); [Francis et al., 2003](#)). Inferences from these procedures (that is, the act of testing the statistical significance of coefficients or differences between regressions' R^2) are, of course, conditional on the degree to which the regression models are correctly specified.

This is where a research design that accounts for a possible nonlinear relationship between stock prices and their determinants can play an important role. In a seminal paper, [Holthausen and Watts \(2001\)](#) point out the lack of effective statistical modeling techniques that can be used to evaluate the complex relationship between stock prices and earnings. In the extant accounting literature, simple linear regression techniques are the most commonly used means of estimating a valuation model, despite substantial evidence suggesting that equity prices bear a highly nonlinear relationship with their determining variables, such as earnings, book value, and cash flows ([Freeman and Tse, 1992](#); [Ali, 1994](#); [Das and Lev, 1994](#); [Subramanyam, 1996](#); [Burgstahler and Dichev, 1997](#); [Lipe et al., 1998](#)).

Twenty years have passed since [Holthausen and Watts \(2001\)](#), and linear regression models are still the most common method used in the market-based accounting literature, even though the consequences of using estimating linear models to describe a nonlinear relationship are well known in econometrics ([Ramsey, 1969](#); [Greene, 1990](#)). The

misspecification of a regression model’s functional form is equivalent to the omission of relevant variables; these omitted variables represent nonlinear terms in the relationship between the dependent and the independent variables (Kmenta, 1971).¹ Ignoring these terms can bias regression coefficients if the data exhibit nonlinear patterns; therefore, estimated coefficients from a misspecified linear regression model might be meaningless, while taking these coefficients at face value can lead to the generation of erroneous inferences.²

Researchers who have studied earnings response coefficients (ERCs) have investigated

¹For example, we could face a misspecification error if we omit regressors’ higher-order monomials or possible interactions among variables.

²Achen (2005) proposes an example that shows to what extent inferences from a misspecified linear regression model can be misleading. We adapt this example to the context of this thesis to aid the understanding of the results.

Let y be a dependent variable that we explain as a function of two independent variables: a variable of interest x_1 , and a control variable x_2 . Achen (2005) simulates data to impose a precise relationship between the dependent and the independent variables, assuming the following data-generating process:

$$y = 0.10x_1 + f(x_2), \tag{1.1}$$

without even including an error term, and where f is a “slightly” nonlinear function.

Here, we say that the relationship between y and x_1 is linear because the marginal effect of x_1 on y is constant and equal to 0.10 for every observation. In contrast, the marginal effect of x_2 on y is $f'(x_2)$, the first derivative of f with respect to x_2 . That value is a function of x_2 ; therefore, we say that the relationship between x_2 and y is not linear. In this case, we could also say that f is “slightly” nonlinear because the value of $f'(x_2)$ is close to being the same for every observation. There is no error term in (1.1), the purpose of which would be to “avoid outliers and eliminate any possibility of stochastic accidents” (Achen, 2005, p. 331).

We know the true regression equation because the data have come out of a simulation; however, Achen (2005) estimates a misspecified linear regression model, as follows:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \varepsilon.$$

This model ignores the slightly nonlinear relationship between x_2 and y , resulting in a remarkably incorrect estimate of the coefficient of the variable of interest x_1 . Achen (2005) highlights a highly significant and negative coefficient $\hat{\beta}_1 = -2.78$ when estimating the misspecified linear regression model, which is 28 times larger in magnitude than the correct value of 0.10, with the wrong sign.

the consequences of using linear regression models to describe nonlinear relationships. [Freeman and Tse \(1992\)](#) show that forcing a linear specification onto an abnormal return-earnings model biases the earnings slope coefficient toward zero. [Cheng et al. \(1992\)](#) comprehensively and formally evaluate the specification of the cross-sectional, ordinary least squares (OLS) model that relates earnings to risk-adjusted security returns. They test for the existence of nonlinear effects, possible omitted variables, and systematic and random inter-firm coefficient variation, concluding that “the specification error is large enough to affect conclusions regarding economic relationships” ([Cheng et al., 1992](#), p. 580). When replicating a previously published study, the authors also demonstrate how specification problems can lead to inferences’ instability.

Basing estimation models on cross-sections ignores firm-specific effects, which can be another source of bias. Results from [Lipe et al. \(1998\)](#) suggest that including firm-specific fixed effects and considering possible nonlinear patterns have substantially improved approaches to inferring the relationship between stock returns and accounting earnings. [Kothari and Shanken \(2003\)](#) analyze how slope coefficients vary in cross-sectional regressions of prices on earnings and book values and bring indirect evidence of bias that has occurred due to correlated omitted variables. They also find a significantly negative correlation between earnings coefficients and proxies for growth and expected return. This result is consistent with [Collins and Kothari \(1989\)](#), which examines the cross-sectional determinants of the ERCs and finds that the regression coefficient is positively related to earnings persistence and negatively related to firms’ systematic risk. [Teets and Wasley \(1996\)](#) also argue that estimating a cross-sectionally constant ERC

can produce downward-biased estimates, thereby also understating the explanatory power of a linear regression model.

The novel research design used in [Starica and Marton \(2019\)](#) tackles these modeling issues and offers a solution. We discuss this approach in the next section and formalize the problem in the methodology chapter, Chapter 2, explaining how we can estimate nonlinear accounting valuation models using nonparametric regression algorithms.

1.2 The relationship between stock prices and accounting variables

In this section, we introduce our first two empirical investigations—discussed in [Starica and Giosi \(2017\)](#)—and detail the concept of the expectation-formation pertinence of accounting measures relating to economic performance. This concept is, together with the adoption of modern nonlinear regression techniques in an accounting context, the cornerstone of this thesis.

When we talk about the pertinence to the expectation-formation process of a given accounting variable of interest (such as earnings or cash flows), we refer to the strength of the association between that variable and stock prices. We refer to the same concept when writing that a given accounting variable of interest is “pertinent to the valuation process” or just “pertinent to valuation.” During this thesis, we use these expressions interchangeably when discussing whether there is an association between the accounting

variable we are studying and stock prices.

We use recent nonlinear, nonparametric regression-modeling techniques from the last two decades to consistently measure the association between stock prices and an accounting predictor of value. [Starica and Marton \(2019\)](#) explain that residuals from a nonlinear regression model can quantify a variable's pertinence to the valuation process. When a nonlinear model relates stock prices to a series of accounting variables, the residuals measure the contribution to the expectation-formation process made by information that is provided by sources other than the accounting variables we are using as regressors. [Starica and Marton \(2019\)](#) interpret these errors as investors' adjustments as a function of that alternative source of information. High absolute residuals, which are in fact valuation errors, suggest that the association between stock prices and the accounting variables we use to explain them is weak; therefore, the set of accounting variables we use should have a low pertinence to valuation. In comparison, small valuation errors should suggest that the accounting variables we are using to explain prices are helpful and, hence, that the set of regressors is pertinent to valuation.

In the first two empirical investigations, we focus on the expectation-formation pertinence of earnings and cash flows. We first compare the strength of their association with stock prices, then proceed to investigate whether specific accrual characteristics can explain the differences between earnings and cash flows' pertinence to valuation. Accruals can mitigate the timing problems inherent in cash flows if the correlation between accruals and cash flows is negative ([Dechow, 1994](#)); therefore, earnings (i.e., cash flows plus accruals) should be better able to summarize a firm's operating performance

([Dechow et al., 1998](#); [Nikolaev, 2018](#)).

These investigations revisit the issue of the superiority of earnings over cash flows (in terms of their association with stock prices) in light of new evidence relating to the changing landscape of accrual accounting over the last two decades ([Bushman et al., 2016](#)). We document a decline in earnings' association with stock prices, compared to that of cash flows. Before 1990, earnings showed a significantly stronger association with stock prices than cash flows did. From 2000 onwards, cash flows have become at least as relevant as earnings to the valuation process.

To compare earnings' pertinence to that of cash flows, we derive a relative measure of association between stock prices and these two measures of economic performance, namely, the logarithm of the ratio between earnings and cash flows' absolute relative nonlinear regression residuals, as we explain in Subsection [2.1.6](#).

We study the changes in this firm- and year-specific measure of earnings' pertinence to the valuation process compared to that of cash flows, relating these changes to changes in earnings volatility compared to changes in cash flows volatility. We find that accounting earnings are significantly more relevant than cash flows when earnings variability is lower than that of cash flows and vice versa.

This finding is consistent with previous evidence concerning the relationship between smoothness and predictability. [Dichev and Tang \(2009\)](#) observe that volatile earnings (cash flows) are less relevant in predicting future fundamentals. Concerning this aspect, we know that higher volatility means forecasts of future performance will be less reliable

(Watts and Leftwich, 1977; Gerakos and Gramacy, 2013; Nallareddy et al., 2020). An increase in earnings volatility relative to that of cash flows should reduce an investor's ability to project a firm's future performance using its current net income (Barton et al., 2010); consequently, making earnings less pertinent to the valuation process than cash flows.

We note that cash flows are more pertinent to valuation than earnings when cash flows fluctuate less than earnings. Furthermore, when the volatility of cash flows does not exceed 115% of earnings volatility, cash flows are at least as pertinent to valuation as earnings. In 1970, the volatility of cash flows did not exceed 115% of earnings volatility for roughly 10% of firms, one out of ten. However, after a remarkable increase of earnings volatility over the last half-century, in 2016, the percentage of firms for which cash flows did not exceed 115% of earnings volatility was about 65%. We argue that cash flows are at least as pertinent as earnings for roughly two firms out of three in the most recent cross-sections. For around half of the firms in the most recent cross-sections, we find that cash flows are strictly more pertinent than earnings.

In our second empirical analysis, we explore why earnings have become (surprisingly) more volatile than cash flows. Following the extant literature, we identify three possible change mechanisms. The first possible justification is the growing prominence of firms with high intangible intensity (Srivastava, 2014); that is because intangible assets may significantly alter the characteristics of earnings and their components (Kothari et al., 2002). A second mechanism relates to the increasing accruals volatility due to the rising incidence of one-time items and non-operating accruals (Givoly and Hayn, 2000;

Bradshaw and Sloan, 2002). A third competing explanation relies on an accruals' weaker ability to solve timing issues in the operating cash flow (Bushman et al., 2016). If the correlation between accruals and cash flows is significantly negative, accruals can reduce earnings variability (while having no effect on cash flows variability) (Dechow, 1994). Also, Bushman et al. (2016) show that the magnitude of the correlation between accruals and cash flows has been increasing steadily over time, moving progressively from values closer to -1.00 to less-negative ones. As a result, the difference between cash flows and earnings volatility should diminish; eventually, earnings could become even more volatile than cash flows.

We measure the intensity of firms' intangible assets using the total amount of investments in selling, general and administrative expense (SG&A) as a fraction of total expenses, historical accruals volatility to measure accrual volatility, and the smoothing role played by accruals by the correlation between accruals and cash flows. Using these proxies, we estimate a multivariate regression model to investigate whether any of these sources of earnings volatility can explain why earnings' pertinence to valuation (compared to that of cash flows) decreases.

In Subsections 4.1.5 and 4.1.6, we observe that both the intangible intensity and accruals volatility do not relate to this phenomenon, since we can find no significant association when controlling for the proxy for accruals smoothing. In contrast, we observe a statistically significant association between the smoothing role played by accruals and earnings' pertinence relative to that of cash flows. Our results indicate that the growth of non-timing-related accruals over the last five decades has severely

limited the information content of earnings, relatively speaking. Consequently, it has also significantly weakened the relevance of accrual accounting to valuation. This finding is also consistent with [Bushman et al. \(2016\)](#), who observe a dramatic decline in the relative prominence of the timing role played by accruals.

1.3 The relationship between stock prices and accounting quality

In the last investigation, we focused on the relationship between accounting quality and earnings' pertinence to the valuation process. We adopt the framework introduced by [Nikolaev \(2018\)](#) to define and measure accounting quality. That is a novel methodology that can study accruals and measure their ability to fulfill their primary objective: to provide a more accurate picture of a firm's economic performance than alternative cash-based accounting measures. This approach solves a relevant methodological issue in the extant literature, opening many possible directions for future research.

[Nikolaev \(2018\)](#) explains that how to define and measure accounting quality has always been an open debate in the accounting literature. While common means of doing so include measures of the variability of earnings and their components, alternative measures consider the special items reporting frequency, measures of firms' profitability, and proxies for non-timing-related accruals. An interested reader can refer to [Schipper and Vincent \(2003\)](#) and [Dechow et al. \(2010\)](#) for a comprehensive review.

A well-known and established family of accounting-quality proxies relies on the measurement of abnormal accruals ([Jones, 1991](#); [Dechow and Dichev, 2002](#)). Usually, an underlying regression model decomposes the observed amount of accruals into both an expected component and a residual (abnormal) component. The literature considers the magnitude of the abnormal accruals, treating it as a measure of accounting quality: a higher magnitude of abnormal accruals implies lower-quality accounting, keeping everything else constant.

Measuring accounting quality using such an approach has its limitations. [Nikolaev \(2018\)](#) argues that, due to earnings management, discretionary accruals might be subject to measurement errors and distortions. Although computing accruals, earnings, or cash flow variance is straightforward, studying the relationship between these variabilities and accounting quality might also lead to incorrect inferences. We might expect that volatile earnings (or cash flows) can translate into a lower accounting quality. However, as he explains, accounting earnings (or cash flows) give us noisy measures of the underlying firm's economic performance. It is worth noting that it is not clear why the accounting numbers we observe are volatile. The volatility of accounting numbers might depend on either an unstable operating outcome, timing errors, or imperfections in the accounting measurement process.

To measure accounting quality, we need to correctly identify the firm's unobservable operating performance. [Nikolaev \(2018\)](#) offers an innovative, simple, yet effective framework that decomposes observable profitability measures into two components. The first of these is a concrete economic outcome (a signal component), while the

second is a disturbance caused by either timing problems or an imperfect accruals measurement process (a noise component). Timing problems affect the operating cash flows, while errors in the accounting measurement process affect the accounting earnings. He explains that, because of timing errors, cash flows offer a noisy measure of performance. Here, the underlying operating performance is the signal, while the timing errors are the noise. Additionally, because of accounting errors that can occur when defining accruals, earnings can also be seen as a noisy measure of the same economic performance. [Nikolaev \(2018\)](#) argues that these estimation errors can be due to ineffective assumptions, models, estimates, or to the application of subjective judgments. He also explains that local accounting standards may also affect reporting choices and earnings management, citing [Beyer et al. \(2019\)](#). For earnings, the signal is the same underlying operating performance, while the noise is the measurement errors that occur when computing accruals.

The benefits of accruals accounting depend on the relative amplitude of the two noise processes present in earnings and cash flows. Accrual accounting is more complex, and therefore more expensive than cash-based standards. Intuitively, we might want to avoid adopting accrual accounting standards when cash flow timing problems are negligible. However, accruals can play a relevant role in providing a picture of economic performance when these timing problems are substantial. We should also consider that the cost of accrual accounting increases when the accruals measurement process is imperfect and subject to higher and non-negligible errors. [Nikolaev \(2018\)](#) suggests the following relationship between accounting quality and cash flow/earnings noise

components; that is, when the timing problems exceed any problems associated with the accounting measurement process, accruals can provide investors with more relevant operating performance information.

Subsection 2.2.1 describes the rationale held by Nikolaev (2018). Specifically, we discuss the system of equations that characterizes earnings, cash flows, and accruals in light of the novel signal-noise decomposition described above. In Subsection 2.2.2, we describe how we estimate an unobservable economic performance using a state-space model. We also explain how such an approach might mitigate technical issues using the generalized method of moments estimation seen in Nikolaev (2018). In Subsection 2.2.3, we examine a multivariate regression model used to estimate the benefit of accrual accounting. Holding constant the economic performance risk and the variance of timing errors, we analyze the association between earnings' pertinence to the valuation process and the ability of accruals to provide investors with more relevant information about operating performance. Finally, we present our results in Section 4.2.

2 Research design

This chapter is divided into two sections. In Section 2.1, we discuss how we assess the relationship between stock prices, earnings, and operating cash flows—as reported in Starica and Giosi (2017). In Section 2.2, we detail how we investigate the relationship between earnings’ pertinence to the valuation process and accounting quality.

2.1 Nonparametric techniques used to model the relationship between stock prices and accounting variables

In this section, we explain how we use nonlinear econometric modeling techniques to study the association between stock prices and accounting measures of performance. In Subsection 2.1.1, we explain how simple nonlinear regression algorithms can describe an accounting valuation relationship. We also discuss the consequences of estimating a simple linear regression model without questioning whether the assumption of linearity might be valid. In Subsections 2.1.2 and 2.1.3, we formally explain the concept of expectation-formation pertinence, as well as discussing relative and incremental pertinence to the valuation process. In Subsections 2.1.4 and 2.1.5, we examine how to test for the existence of these effects using the concept of stochastic dominance. Finally, in Subsection 2.1.6, we discuss how we can use alternative regression models to consistently measure the association between our measure of pertinence and a firm’s various accounting characteristics.

2.1.1 A nonlinear valuation

Let $P_{i,t}$ be the stock price of firm i at time t , and let $\mathbf{X}_{i,t}$ be a set of accounting variables relevant to the price-formation process, such as earnings and cash flows. The alternative research design we use relies on the fact that a nonlinear regression relationship between $P_{i,t}$ and $\mathbf{X}_{i,t}$ always exists. However, as [Starica and Marton \(2019\)](#) explain, it is likely that there will not be a relationship described by a linear regression model between these two variables.

Assuming a nonlinear relationship between $P_{i,t}$ and $\mathbf{X}_{i,t}$, and following the approach taken by [Starica and Marton \(2019\)](#), we can decompose the stock price as:

$$P_{i,t} = \mathbf{m}_{i,t}^{\mathbf{X}}(\mathbf{X}_{i,t}) + \varepsilon_{i,t}, \quad (2.1)$$

where the term $\mathbf{m}_{i,t}^{\mathbf{X}}(\mathbf{X}_{i,t}) = \mathbb{E}(P_{i,t} | \mathbf{X}_{i,t})$ stands for the best nonlinear prediction of the stock price based on a set of accounting information; that is, the conditional expected value of the stock price $P_{i,t}$ given the accounting information in $\mathbf{X}_{i,t}$. Writing $\mathbf{m}_{i,t}^{\mathbf{X}}$ means that the regression function \mathbf{m} is set-specific: if the set of accounting variables changes, the regression function changes. Furthermore, the two indexes i and t suggest that the nonlinear regression function \mathbf{m} can be both firm- and year-specific. The term $\varepsilon_{i,t}$ is an adjustment that measures all the other information available to investors at a given time t . From a statistical point of view, the adjustment term $\varepsilon_{i,t}$ in (2.1) is a

regression error that is orthogonal to the predictors; that is, it satisfies:

$$\mathbb{E}(\varepsilon_{i,t} | \mathbf{X}_{i,t}) = 0. \quad (2.2)$$

A consistent estimation of the regression function \mathbf{m} in (2.1) requires the orthogonality condition found in (2.2). At this point, [Starica and Marton \(2019\)](#) recall that this condition is most likely to be violated when a linear regression model is estimated instead. We can face a so-called misspecification error when we estimate an underlying nonlinear relationship using a linear regression model. Specifically, estimating a linear regression model between the stock price and the accounting variables that are pertinent to the price-formation process when the real relationship is nonlinear can yield coefficients that do not reflect the economic relationship in (2.1). Consequently, we cannot use these estimated coefficients to test our hypotheses.³

To estimate the nonlinear regression function \mathbf{m} consistently, we can employ proven inference techniques from the field of nonparametric statistics; on this topic, [Starica and Marton \(2019\)](#) argue that the corresponding estimate $\widehat{\mathbf{m}}$ is theoretically guaranteed to be unbiased.

Following a common rationale in accounting valuation, the next step is to assume that we can express the stock price $P_{i,t}$ as a sum of discounted expectations of a number of functions of future earnings, additional variables (which may or may not represent

³[Starica and Marton \(2019\)](#) argue that the existence of a misspecification error can also represent a possible explanation of the well-known bias affecting the price levels or the returns-earnings regressions ([Lo, 2005](#); [Barth and Clinch, 2009](#)).

accounting numbers), and the price of equity risk. In this framework, let:

$$P_{i,t} = \sum_{\tau=1}^{\infty} \frac{\mathbb{E}_t[f_{i,t+\tau}(\mathbf{X}_{i,t+\tau}; \mathbf{O}_{i,t+\tau}, r_{i,t})]}{(1 + r_{i,t})^\tau}, \quad (2.3)$$

where the term $\mathbf{X}_{i,t+\tau}$ contains information about a generic accounting performance measure (such as earnings or cash flows) of firm i up to time $t + \tau$. Similarly, $\mathbf{O}_{i,t+\tau}$ includes information about variables other than this measure of performance (e.g., dividends and the book value of equity) while $r_{i,t}$ denotes the price of the equity risk of the firm i at time t . The operator $\mathbb{E}_t[\cdot]$ stands for a market expectation, conditional on all the information available at time t . Examples of the previous equation (2.3) include, but are not limited to, the residual income model and the Ohlson and Juettner-Nauroth model (Ohlson, 1995; Ohlson and Juettner-Nauroth, 2005). The equation (2.3) can represent alternative valuation models for a particular choice of functions $f_{i,t+\tau}$ and variables in the two predictor sets $\mathbf{X}_{i,t+\tau}$ and $\mathbf{O}_{i,t+\tau}$.

Following Starica and Marton (2019), the assumption in (2.3) makes it possible to express the previous nonlinear regression equation (2.1) as a valuation model. They show that:

$$\mathbf{m}_{i,t}^{\mathbf{X}}(\mathbf{X}_{i,t}) = \sum_{\tau=1}^{\infty} \frac{\mathbb{E}[f_{i,t+\tau}(\mathbf{X}_{i,t+\tau}; \mathbf{O}_{i,t+\tau}, r_{i,t}) | \mathbf{X}_{i,t}]}{(1 + r_{i,t})^\tau}, \quad (2.4)$$

comparing that with the previous model assumption (2.3). They explain that, while the structure of the two equations (2.3) and (2.4) is the same, the information available to the investors when they are forming expectations about a firm's future accounting

performance is different. While the conditional expected value $\mathbb{E}_t[\cdot]$ in (2.3) accounts for complete information available at time t , the expected value $\mathbb{E}[\cdot | \mathbf{X}_{i,t}]$ in (2.4) only accounts for the value of $\mathbf{X}_{i,t}$, which is the actual performance measure we consider. While, in the first case, investors have access to complete information, in the second case, the information that defines the conditional expected value is only partially complete. This implies that we can see the nonlinear regression function $\mathbf{m}_{i,t}^{\mathbf{X}}$ in (2.1) as a valuation that is conditional on the current level of the predictors in $\mathbf{X}_{i,t}$.

Starica and Marton (2019) interpret the previous equation (2.1) as follows. For a given firm i at time t , we can decompose the stock prices into two distinct and uncorrelated components, forming a valuation based on expectations of future income and which is conditional only on the current level of the predictors in $\mathbf{X}_{i,t}$, plus an adjustment that represents information other than $\mathbf{X}_{i,t}$. That means:

$$P_{i,t} = \sum_{\tau=1}^{\infty} \frac{\mathbb{E}[f_{i,t+\tau}(\mathbf{X}_{i,t+\tau}; \mathbf{O}_{i,t+\tau}, r_{i,t}) | \mathbf{X}_{i,t}]}{(1 + r_{i,t})^{\tau}} + \varepsilon_{i,t}, \quad (2.5)$$

where the orthogonality condition in (2.2) holds.

2.1.2 Estimating the expectation-formation pertinence of accounting variables

We have seen that a valuation model relates stock prices to a series of accounting determinants of value. Once we estimate such an economic relationship, we can infer the

association between stock prices and these accounting fundamentals. The expectation-formation pertinence (EFP) is a measure of this association. We explain the meaning of this term in this subsection and discuss how a nonlinear regression model can measure an accounting variable’s pertinence to the valuation process.

Following [Starica and Marton \(2019\)](#), the first step in this process is to obtain a consistent estimate of a nonlinear valuation model, such as the residual income in the previous equation (2.3). In doing so, we need to identify a set of determinants of the future evolution of abnormal earnings, such as the firm’s cost of equity, its growth profile, and, possibly, its level of accounting conservatism.

[Starica and Marton \(2019\)](#) investigate more than 40 financial and accounting characteristics. In line with the extant literature, they find the price-to-book ratio and a firm’s size, leverage, and industry are, on average, the most relevant accounting valuation determinants ([Fama and French, 1992](#); [Liu and Thomas, 2000](#); [Biddle et al., 2001](#); [Holthausen and Watts, 2001](#); [Kothari and Shanken, 2003](#); [Roychowdhury and Watts, 2007](#); [Penman et al., 2018](#)).

All of these results impact on how we calibrate the nonlinear regression model we use in our empirical investigations. Specifically, they indicate which regressors might better explain the accounting valuation relationship we are trying to infer.

For our analyses, we model the economic relationship in (2.3) as:

$$P_{i,t} = \mathbf{m}_{i,t}^{\mathbf{X}}(\mathbf{X}_{i,t}; \mathcal{I}_{i,t}, r_{i,t}, g_{i,t}, c_{i,t}) + \varepsilon_{i,t}, \quad (2.6)$$

where $\mathcal{I}_{i,t}$ denotes the industry to which a given firm i belongs at time t , $r_{i,t}$ stands for the firm cost of equity capital, $g_{i,t}$ denotes the income growth, and $c_{i,t}$ the level of accounting conservatism. We assume the regression error term $\varepsilon_{i,t}$ in (2.6) satisfies standard orthogonality conditions. We estimate the nonlinear regression model in (2.6) cross-sectionally, under the assumption that firms belonging to the same industry in a given year have similar nonlinear valuation functions (\mathbf{m}), once we also control for risk, growth, and level of accounting conservatism.

We estimate the regression model in (2.6) using a Random Forest, a nonlinear, non-parametric algorithm to estimate the relationship between the dependent variable and the regressors. That is a relatively recent method introduced by Breiman (2001), yet among the most versatile and frequently used algorithms in Machine Learning.

At least three arguments justify this methodological choice. First of all, Random Forests do not require any strict assumption about the relationship between the dependent and the independent variables. That means we do not have to specify any equation describing that relationship, and it is a very convenient feature because it heavily reduces the risk of regression model misspecification. Second, a Random Forest estimates the relationship between the dependent variable and the regressors on relatively small subsets of observations. That means the degree of association between stock prices and accounting earnings may vary from a group of firms to the other. That is, in general, a feature we should prefer because this kind of local modeling technique allows flexibility. Third, this algorithm is robust against the negative effect of outliers in the independent variables, and it does not suffer from overfitting issues when the underlying model

complexity increases. Both of these issues might dramatically reduce the interpretability of the estimated regression model. Accordingly, we should therefore prefer techniques that underweight outliers and do not overfit. An interested reader can find additional information about these algorithms in [Breiman \(2001\)](#), [Hastie et al. \(2009\)](#), and [Biau and Scornet \(2016\)](#).

Since we can interpret the regression model in (2.6) as a valuation, the relative magnitude of the estimated regression error term:

$$\hat{\epsilon}_{i,t}^{\mathbf{X}} = \left| \frac{P_{i,t} - \widehat{\mathbf{m}}_{i,t}^{\mathbf{X}}(\mathbf{X}_{i,t}; \mathcal{I}_{i,t}, r_{i,t}, g_{i,t}, c_{i,t})}{P_{i,t}} \right|, \quad (2.7)$$

quantifies how much information other than the level of the accounting variables in $\mathbf{X}_{i,t}$ shapes future earnings' expectations. The size of this relative error is a firm- and year-specific measure of association between the stock price and the variables in $\mathbf{X}_{i,t}$. We define the strength of that association as expectation-formation pertinence.

Holding everything else constant, higher valuation errors suggest a lower level of association between the stock price and the accounting variables in $\mathbf{X}_{i,t}$, indicating that the accounting predictors in $\mathbf{X}_{i,t}$ have a lower pertinence to the valuation process.

2.1.3 Relative and incremental expectation-formation pertinence

We can compute different absolute valuation errors using different accounting measures of performance. Therefore, comparing two valuation errors allows us to compare the

relevance to the valuation process of two sets of predictor variables: \mathbf{X} versus \mathbf{Y} . This is equivalent to measuring the strength of the association that these two sets of variables have with the stock prices; therefore, it is also equivalent to comparing their pertinence to the valuation process.

The comparison between two collections of explicatory variables can be either relative or incremental. The type of comparison depends on the predictor sets: if we want to measure the relative pertinence to valuation, the two sets (\mathbf{X} and \mathbf{Y}) are disjoint. However, if we investigate an incremental expectation-formation pertinence, one set of variables should include the other.

One of our empirical analyses focuses on a relative comparison between the expectation-formation pertinence of earnings and cash flows, with $\mathbf{X} = \{\text{EARN}\}$ being equal to accounting earnings, while $\mathbf{Y} = \{\text{CFO}\}$ is equal to operating cash flows. If we want to measure the incremental pertinence of accounting earnings compared to cash flows, the choice of predictor sets should be $\mathbf{X} = \{\text{EARN}, \text{CFO}\}$, which is equal to accounting earnings and operating cash flows, while $\mathbf{Y} = \{\text{CFO}\}$ represents operating cash flows only. This second kind of incremental comparison is beyond the scope of this thesis.

To investigate the existence of a relative pertinence to valuation, we estimate two distinct specifications of (2.6), as follows:

$$\hat{\mathbf{P}}_{i,t}^{\text{EARN}} = \widehat{\mathbf{m}}_{i,t}^{\text{EARN}}(\text{EARN}_{i,t}; \mathcal{I}_{i,t}, r_{i,t}, g_{i,t}, c_{i,t}), \quad (2.8)$$

and:

$$\hat{P}_{i,t}^{\text{CFO}} = \hat{\mathbf{m}}_{i,t}^{\text{CFO}}(\text{CFO}_{i,t}; \mathcal{I}_{i,t}, r_{i,t}, g_{i,t}, c_{i,t}). \quad (2.9)$$

Both can be seen as two possible valuations. We define the corresponding two absolute relative valuation errors as:

$$\hat{\epsilon}_{i,t}^{\text{EARN}} = \left| \frac{P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}}{P_{i,t}} \right|, \quad (2.10)$$

and:

$$\hat{\epsilon}_{i,t}^{\text{CFO}} = \left| \frac{P_{i,t} - \hat{P}_{i,t}^{\text{CFO}}}{P_{i,t}} \right|, \quad (2.11)$$

where, intuitively, smaller absolute errors denote higher pertinence to the price-formation process.

We compare the pertinence of the two sets of accounting variables in two ways. In one of our investigations, we observe that the association between stock prices and accounting earnings decreases over time, until it becomes weaker than the association we find between stock prices and operating cash flows. We arrive at this conclusion by comparing the cross-sectional cumulative distribution functions of the two valuation error terms $\hat{\epsilon}^{\text{EARN}}$ and $\hat{\epsilon}^{\text{CFO}}$ in (2.10) and (2.11), respectively, as we explain in Subsections 2.1.4 and 2.1.5, with results given in Subsections 4.1.1 to 4.1.3.

In a second empirical analysis, we explore potential reasons for the decreasing association

between accounting earnings and stock prices. We define a relative measure of pertinence to the valuation process as the logarithm of the ratio between the two absolute relative regression residuals $\hat{\epsilon}^{\mathbf{EARN}}$ and $\hat{\epsilon}^{\mathbf{CFO}}$, and test whether a firm's particular accounting characteristics can explain any changes in this measure of relative pertinence to valuation. We discuss this investigation in more detail in Subections [2.1.6](#) and [2.1.7](#), presenting our results in Subsections [4.1.5](#) and [4.1.6](#).

2.1.4 Stochastic dominance

Stochastic dominance is a comprehensive criterion that can be used to compare the pertinence of two different sets of accounting variables. In financial research, it has been used to compare the performance of different portfolios, while [Kang \(2016\)](#) applies this method to the analysis of valuation accuracy given in his assessment of financial ratios' ability to summarize value.

Following [Kang \(2016\)](#), let e be an arbitrary constant that indicates a possible valuation error. At time t , let:

$$\begin{aligned}\hat{F}_t^{\mathbf{EARN}}(e) &= \Pr\left(\left|\frac{P_{\cdot,t} - \hat{P}_{\cdot,t}^{\mathbf{EARN}}}{P_{\cdot,t}}\right| \leq e\right), \\ \hat{F}_t^{\mathbf{EARN}}(e) &= \Pr(\hat{\epsilon}_{\cdot,t}^{\mathbf{EARN}} \leq e),\end{aligned}\tag{2.12}$$

be the percentage of firms that are subject to a valuation error lower or equal to e

when the stock price is predicted using accounting earnings. Similarly, let:

$$\begin{aligned}\widehat{F}_t^{\text{CFO}}(e) &= \Pr\left(\left|\frac{P_{\cdot,t} - \widehat{P}_{\cdot,t}^{\text{CFO}}}{P_{\cdot,t}}\right| \leq e\right), \\ \widehat{F}_t^{\text{CFO}}(e) &= \Pr(\widehat{\epsilon}_{\cdot,t}^{\text{CFO}} \leq e),\end{aligned}\tag{2.13}$$

serve the same purpose for the operating cash flows. The two equations, (2.12) and (2.13), denote the empirical cumulative distribution function of the two absolute relative valuation errors $\widehat{\epsilon}_t^{\text{EARN}}$ and $\widehat{\epsilon}_t^{\text{CFO}}$, respectively.

In a given cross-section t , we say that the valuation using earnings dominates the valuation using cash flows if the percentage of firms in (2.12) is greater or equal to that in (2.13), for every possible level of e , while for some e the inequality strictly holds. If we have that $\widehat{F}_t^{\text{EARN}}(e) \geq \widehat{F}_t^{\text{CFO}}(e)$ for any e , while $\widehat{F}_t^{\text{EARN}}(e) > \widehat{F}_t^{\text{CFO}}(e)$ applies for some e , this suggests the valuation using accounting earnings is more precise than that using operating cash flows. We can rewrite the previous inequality as:

$$\Pr\left(\left|\frac{P_{\cdot,t} - \widehat{P}_{\cdot,t}^{\text{EARN}}}{P_{\cdot,t}}\right| \leq e\right) \geq \Pr\left(\left|\frac{P_{\cdot,t} - \widehat{P}_{\cdot,t}^{\text{CFO}}}{P_{\cdot,t}}\right| \leq e\right),\tag{2.14}$$

which suggest, equivalently, that accounting earnings are more pertinent than operating cash flows when the inequality in (2.14) strictly holds for some e .⁴

⁴For example, let $e = 20\%$ denote a valuation error of about 20% of the stock price. If:

$$\widehat{F}_t^{\mathbf{X}}(20\%) = \Pr\left(\left|\frac{P_{\cdot,t} - \widehat{P}_{\cdot,t}^{\mathbf{X}}}{P_{\cdot,t}}\right| \leq 20\%\right) > \widehat{F}_t^{\mathbf{Y}}(20\%) = \Pr\left(\left|\frac{P_{\cdot,t} - \widehat{P}_{\cdot,t}^{\mathbf{Y}}}{P_{\cdot,t}}\right| \leq 20\%\right),$$

the inequality suggests that estimating stock prices using the information in \mathbf{X} rather than \mathbf{Y} gives us a higher percentage of firms whose estimated stock price lies within 20% of the actual price; therefore,

Kang (2016) emphasizes that the previous inequality (2.14) also implies a clear relationship between the quantiles of the two distributions of absolute relative residuals. We can find quantiles by inverting the cumulative distribution function; therefore, we should prefer the valuation with the lowest quantiles of the distribution of absolute relative errors. For example, we can say that earnings are more pertinent than cash flows when the median of $\hat{\epsilon}^{\text{EARN}}$ is lower than that of $\hat{\epsilon}^{\text{CFO}}$; however, this comparison is not as comprehensive as a direct comparison between cumulative distribution functions.

Table 1, as in Kang (2016), summarizes the possible outcomes when evaluating a stochastic dominance between two empirical distributions.

2.1.5 Testing for a relative expectation-formation pertinence

Cross-section by cross-section, we test the null hypothesis, H_0 , concerning the equal pertinence of earnings and cash flows. We test this hypothesis against the alternative, H_1 , that one accounting measure of profitability is more expectation-formation pertinent than the other. In doing so, we assess the existence of a relative pertinence to valuation.

We test for the existence of stochastic dominance between earnings and cash flows using two comprehensive Kolmogorov-Smirnov (KS) test statistics. First, at time t , we estimate the empirical cumulative distribution function of the valuation errors that

\mathbf{X} should be better than \mathbf{Y} for this level of error. If \mathbf{X} dominates \mathbf{Y} , then, whatever error level e we may choose, there is always at least the same level of precision offered by \mathbf{X} as by \mathbf{Y} , while, in some cases (for some levels of error), the performance offered by the set \mathbf{X} is strictly better than that offered by \mathbf{Y} .

Stochastic dominance	Expectation-formation pertinence
$\mathbf{X} \succ \mathbf{Y}$, or: $\hat{F}^{\mathbf{X}}(e) \geq \hat{F}^{\mathbf{Y}}(e), \forall e,$!e s.t. $\hat{F}^{\mathbf{X}}(e) > \hat{F}^{\mathbf{Y}}(e);$	Conclusion: The predictor set \mathbf{X} is more pertinent than \mathbf{Y} ;
$\mathbf{Y} \succ \mathbf{X}$, or: $\hat{F}^{\mathbf{X}}(e) \leq \hat{F}^{\mathbf{Y}}(e), \forall e,$!e s.t. $\hat{F}^{\mathbf{X}}(e) < \hat{F}^{\mathbf{Y}}(e);$	Conclusion: The predictor set \mathbf{Y} is more pertinent than \mathbf{X} ;
$\mathbf{X} \sim \mathbf{Y}$, or: $\hat{F}^{\mathbf{X}}(e) = \hat{F}^{\mathbf{Y}}(e), \forall e;$	Conclusion: The predictor set \mathbf{X} is as pertinent as \mathbf{Y} . No set stochastically dominates the other;
None of the above.	Conclusion: No set stochastically dominates the other.

Table 1: **Stochastic dominance and expectation-formation pertinence.** This table explains how the concepts of stochastic dominance and the expectation-formation pertinence of an accounting measure of economic performance are related. First, we estimate the relationship between stock prices and a set of accounting variables using a nonlinear regression model. We fit two distinct models in order to compare the pertinence of two sets of accounting variables to valuation, as in (2.8) and (2.9). We predict the value of stock prices using two sets of accounting variables: \mathbf{X} and \mathbf{Y} , before defining an absolute relative valuation error as the difference between the actual price and the estimated value, scaled by the stock price. These are error $\hat{\epsilon}^{\mathbf{X}}$, using \mathbf{X} , and error $\hat{\epsilon}^{\mathbf{Y}}$, using \mathbf{Y} , as in (2.10) and (2.11). In our case, \mathbf{X} and \mathbf{Y} denote earnings and cash flows, respectively. A cumulative distribution function (c.d.f.) gives us a comprehensive measure of the valuation error size. Therefore, testing the difference between two c.d.f., $\hat{F}^{\mathbf{X}}$ and $\hat{F}^{\mathbf{Y}}$, serves as a criterion that can decide which set of accounting variables is more pertinent to valuation. The two c.d.f. correspond to (2.12) and (2.13). The table describes the four possible outcomes that emerge from such a test; we can say that a variable is more pertinent to valuation when the size of the corresponding valuation error is smaller. See Kang (2016).

correspond to these two measures of profitability, as in (2.10) and (2.11). We record these functions as \hat{F}_t^{EARN} and \hat{F}_t^{CFO} , as in (2.12) and (2.13), respectively. Second, we calculate the most positive and negative difference between the two estimated

cumulative distribution functions; that is:

$$\overline{M}_t = \max_e \{\widehat{F}_t^{\text{EARN}}(e) - \widehat{F}_t^{\text{CFO}}(e)\}, \quad (2.15)$$

and:

$$\underline{M}_t = \min_e \{\widehat{F}_t^{\text{EARN}}(e) - \widehat{F}_t^{\text{CFO}}(e)\}, \quad (2.16)$$

before scaling them to obtain the corresponding two Kolmogorov-Smirnov test statistics:

$$\overline{KS}_t = \frac{\overline{M}_t}{\sqrt{\frac{2}{n_t}}}, \quad (2.17)$$

and:

$$\underline{KS}_t = \frac{\underline{M}_t}{\sqrt{\frac{2}{n_t}}}, \quad (2.18)$$

respectively. Finally, we compare these two statistics to a critical value that no longer depends on n_t , the sample size of the cross-section at time t .⁵

Intuitively, the two Kolmogorov-Smirnov test statistics measure the distance between

⁵We can expect a strong, positive correlation between the two absolute relative valuation errors in (2.10) and (2.11); this is because the two valuation errors have been estimated from comparable nonlinear regression models. In this case, the distribution of the two Kolmogorov-Smirnov test statistics under the null hypothesis depends on the joint distribution of the valuation errors. To find the proper critical values, we need to approximate the distribution of the two statistics via bootstrap. When comparing the pertinence of two competing predictor sets, we also have to account for the statistical error when estimating a cumulative distribution function. We explain how to address all of these issues in Appendix A.2.

the two cumulative distribution functions. In a given cross-section t , a value of \overline{KS}_t (\underline{KS}_t) above (below) the corresponding critical value suggests that earnings (cash flows) are significantly more relevant for the purpose of price setting. Otherwise, a \overline{KS}_t (\underline{KS}_t) statistic that equates to between zero and the critical value suggests that earnings (cash flows) do not stochastically dominate cash flows (earnings); therefore, we do not reject the hypothesis that earnings (cash flows) are not significantly more pertinent to the valuation process. That is the framework we use to show that the association between stock prices and accounting earnings has decreased over time.

2.1.6 Explaining changes in the expectation-formation pertinence

In this subsection, we explain how we investigate the potential reasons why a change in earnings' pertinence relative to that of cash flows might occur. For this purpose, we quantify the earnings' pertinence to the valuation process relative to cash flows as the logarithm of the ratio between two absolute relative valuation errors. At the numerator (the error from a nonlinear regression model that uses cash flows), $\hat{\epsilon}^{\text{CFO}}$ in (2.11). At the denominator (the error that uses earnings), $\hat{\epsilon}^{\text{EARN}}$ in (2.10). This measure of relative pertinence reduces to:

$$\text{Rel-EFP}_{i,t} = \ln |P_{i,t} - \hat{P}_{i,t}^{\text{CFO}}| - \ln |P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}|, \quad (2.19)$$

where, intuitively, a positive value of Rel-EFP in (2.19) suggests that accounting earnings are more expectation-formation pertinent than operating cash flows are, and

vice versa.⁶ This is because the size of the absolute valuation error we obtain when using earnings, $\widehat{\epsilon}_{i,t}^{\mathbf{EARN}}$, is less than that of $\widehat{\epsilon}_{i,t}^{\mathbf{CFO}}$.

To investigate the possible reasons for a change in the relative measure of accounting earnings' pertinence to valuation, we estimate a regression model that includes proxies for explanations for an increase in earnings volatility compared to cash flows. These are the intangible intensity, accrual volatility, and accruals smoothing, as proxied by the ratio between investments in SG&A and total expenses, the historical volatility of accruals, and the correlation between accruals and cash flows, respectively.

The increase in the intangibles significantly alters the characteristics of earnings and their components. For example, the future benefits of intangible investments are more uncertain (Kothari et al., 2002). As a result, the information content of current earnings should be less than that of cash flows. An immediate expense of investments in (internally generated) intangibles that impact the outflow of cash—such as patents, trade names, human capital, and customer relations—should also increase the ability of cash flows to predict future cash flows. Note that these investments do not generate accruals (Bushman et al., 2016). More important investments in intangible assets should increase the expenses volatility, as well as reducing the matching with concurrent revenues. As a result, earnings volatility should increase, meaning that earnings could

⁶Using the properties of logarithms:

$$\text{Rel-EFP}_{i,t} = \ln \frac{\widehat{\epsilon}_{i,t}^{\mathbf{CFO}}}{\widehat{\epsilon}_{i,t}^{\mathbf{EARN}}} = \ln \left| \frac{P_{i,t} - \widehat{P}_{i,t}^{\mathbf{CFO}}}{P_{i,t}} \right| - \ln \left| \frac{P_{i,t} - \widehat{P}_{i,t}^{\mathbf{EARN}}}{P_{i,t}} \right| = \ln |P_{i,t} - \widehat{P}_{i,t}^{\mathbf{CFO}}| - \ln |P_{i,t} - \widehat{P}_{i,t}^{\mathbf{EARN}}|,$$

as in (2.19).

become less relevant than cash flows to the valuation process.

Givoly and Hayn (2000) find that a higher accrual variability is due primarily to the rising incidence in one-time items and non-operating accruals; this is also documented in Elliott and Shaw (1988), Collins et al. (1997), Givoly and Hayn (2000), and Bradshaw and Sloan (2002). The magnitudes of both of these categories have increased drastically over the period we are examining. More importantly, the cross-sectional volatility of special and non-operating items has also increased (Bushman et al., 2016). Under these circumstances, accrual variability should increase earnings variability, decreasing its pertinence to the valuation process.

Accruals can also mitigate the variability of earnings; however, if they are to perform their smoothing role, accruals need to be negatively correlated to cash flows (Dechow, 1994). Changes in the correlation coefficient between accruals and cash flows might explain variations in the relative measure of earnings' pertinence; that is the third and last explanation we will consider in our empirical investigation.

Using a multivariate regression model, we aim to disentangle the individual association that each of these characteristics has with the measure of earnings' pertinence to valuation relative to that of cash flows. The difference in (2.19) is the dependent variable, and the proxies for the three accounting characteristics mentioned above are the regressors.⁷ The research design we use can give us an unbiased estimation of the

⁷We can estimate a multivariate regression model because the Rel-EFP difference in (2.19) is both year- and firm-specific, not cross-section specific as the coefficient of determination (R^2), which is a commonly used metric in the literature on value relevance.

individual association of various characteristics of a firm with the logarithm of the ratio of the valuation errors. This methodology has also been discussed in [Kang and Starica \(2017\)](#) when investigating how earnings attributes affect investors' expectations about future earnings reflected in market prices. Following their arguments, we justify the use of this kind of model because we can expect that the functional form of the relationship between earnings' pertinence relative to that of cash flows and the accounting characteristics mentioned above depends on their levels.

To mitigate this methodological issue, and as in [Kang and Starica \(2017\)](#), we define a set of binary variables to identify the levels of the accounting characteristics we consider. Then, we regress the Rel-EFP difference in (2.19) on these binary variables to relax the assumption of linearity.

We divide the empirical distribution of a given characteristic, Ch , into ten intervals with equal probability. For a given interval j , we define a binary variable D_j^{Ch} . That binary variable takes the value of one when the value of Ch is within the interval j ; it is equal to zero otherwise. For a given characteristic, these binary variables are equal to:

$$\begin{aligned}
 D_1^{Ch} &= I(Ch < \phi_1^{Ch}), \\
 D_2^{Ch} &= I(\phi_1^{Ch} \leq Ch < \phi_2^{Ch}), \\
 &\dots \\
 D_9^{Ch} &= I(\phi_8^{Ch} \leq Ch < \phi_9^{Ch}), \\
 D_{10}^{Ch} &= I(\phi_9^{Ch} \leq Ch),
 \end{aligned}$$

where ϕ_j^{Ch} is the j th decile of the empirical distribution of the characteristic Ch .⁸ The operator $I(\cdot)$ is an indicator function that is equal to one if the condition in square brackets holds; it is equal to zero otherwise. For example, let the accounting characteristic Ch represent the smoothing role of accruals, as measured by the correlation between accruals and cash flows (ρ). In this case, the binary variable D_1^ρ identifies those firms with the most negative correlation (i.e., closest to -1.00) between accruals and cash flows. Similarly, D_{10}^ρ is equal to one for those firms for which ρ is closest to 1.00.

In order to capture the differences in the average earnings' pertinence (relative to that of cash flows) between groups of firms with different values of a given characteristic, we regress our relative measure of pertinence in (2.19) on the binary variables we defined above. An univariate regression model appears as follows:

$$\text{Rel-EFP} = \gamma_0 + \sum_{j=2}^{10} \beta_j^{Ch} D_j^{Ch} + \varepsilon, \quad (2.20)$$

where the vector ε is an error term. Alternatively, we define a multivariate regression model as:

$$\text{Rel-EFP} = \gamma_0 + \sum_{k=1}^K \sum_{j=2}^{10} \beta_j^{Ch_k} D_j^{Ch_k} + \varepsilon, \quad (2.21)$$

where we include a total of K distinct accounting characteristics Ch_1, \dots, Ch_K .⁹

⁸This also means that $\Pr(Ch \leq \phi_j^{Ch}) = j/10$. As an example, ϕ_5^{Ch} is the median of Ch because the probability $\Pr(C \leq \phi_5^{Ch}) = 0.50$. The sum $\sum_{j=1}^{10} D_j^{Ch}$ is equal to a vector of ones by definition.

⁹The functional form of the regression models in (2.20) and (2.21) belongs to a more comprehensive

We include a total of nine binary regressors for each of these characteristics; that is because we omit the first binary regressor $D_1^{Ch_k}$ for each characteristic in order to avoid perfect multicollinearity.

The decile group we thus eliminate serves as a reference group, implying that testing the null hypothesis of the zero regression coefficient $\beta_j^{Ch_k}$ corresponding to the binary regressor $D_j^{Ch_k}$ is equivalent to testing the null hypothesis of an equal average relative expectation-formation pertinence between two groups of firms. While the first group is made up of firms for which the characteristic Ch_k takes a value within the j th interval, the second group is made of firms for which that characteristic falls within the omitted interval.¹⁰ In the multivariate regression model, we also keep the levels of all the remaining variables constant.

We adopt the following shorthand notation to simplify the previous equations (2.20) and (2.21). Let:

$$\sum_{j=2}^{10} \beta_j^{Ch_k} D_j^{Ch_k} = \mathbf{D}^{Ch_k} \gamma^{Ch_k},$$

family of step functions. We can express these functions as a finite linear combination of indicator variables. In this specific case, the indicator variables corresponding to a given quality construct define ten distinct and non-overlapping intervals.

¹⁰Stock and Watson (2012) explain that testing the difference between two means is equivalent to testing the significance of a particular regression coefficient, corresponding to a binary variable defining two distinct groups of observations. The first group serves as a reference group, and regression coefficients measure differences in the mean of the remaining groups and the reference. Instead of plugging the numerical values of firms' characteristics directly into a linear multivariate regression model, we convert them into level indicators and test whether firms with values under/above the base level have lower/higher associated expectation-formation pertinence measures. An interested reader can refer to Hardy (1993) for further details.

where:

$$\mathbf{D}^{Ch_k} = \{D_2^{Ch_k}, \dots, D_{10}^{Ch_k}\},$$

is a matrix of binary variables, while:

$$\boldsymbol{\gamma}^{Ch_k} = \{\beta_2^{Ch_k}, \dots, \beta_{10}^{Ch_k}\},$$

is a vector of the regression coefficients. In this way, we can rewrite the previous univariate regression model in (2.20), as follows:

$$\text{Rel-EFP} = \gamma_0 + \mathbf{D}^{Ch} \boldsymbol{\gamma}^{Ch} + \varepsilon, \quad (2.22)$$

while the previous univariate regression model in (2.21) becomes:

$$\text{Rel-EFP} = \gamma_0 + \sum_{k=1}^K \mathbf{D}^{Ch_k} \boldsymbol{\gamma}^{Ch_k} + \varepsilon. \quad (2.23)$$

2.1.7 Changes in earnings' pertinence: the role of accruals smoothing

The methodology we discussed above allows us to evaluate competing explanations for the decreasing earnings' pertinence to valuation relative to that of cash flows. Subsection 4.1.5 details the results that emerge in this analysis, nominating accruals smoothing as the only relevant candidate. Knowing this information, we will now study whether the persistent reduction in accruals smoothing affects the strength of the association stock prices have with earnings and cash flows using dynamic, cross-sectional

analysis.

We look at the evolution of a series of cross-sectional medians over time, estimating an unconditional, cross-sectional median earnings' pertinence to valuation relative to that of cash flows; that is, the yearly median of Rel-EFP in (2.19). We then compare this value with a cross-sectional median which is conditional on the level of accruals smoothing (smoothing measured as the correlation between accruals and cash flows).

In a given year, we consider three distinct sets of firms. The first is a comprehensive set that includes all the firms in that year. The other two are subsets of firms for which the correlation between accruals and cash flows (ρ) takes a value in a given range. In one, $\rho < -0.92$; therefore, the subset includes all the firms for which accruals smoothing is likely to be substantial. In the last, $\rho > -0.58$; therefore, the subset contains firms for which smoothing should be minor. Here, -0.92 and -0.58 denote the first and second tertiles of the distribution of ρ , respectively. While for one-third of the sample, the correlation between accruals and cash flows is lower than -0.92, ρ is higher than -0.58 for another third. We denote the smoothing role played by accruals as “average” for those observations, the remaining third, for which ρ is between -0.92 and -0.58.

For each of these sets, we estimate the following quantile regression model:

$$\text{Rel-EFP}_{i,t} = \gamma_{0,t} + \varepsilon_{i,t}, \tag{2.24}$$

where $\gamma_{0,t}$ is a time-varying intercept, while $\varepsilon_{i,t}$ is an error term. In this particular

specification, finding the intercept by minimizing the absolute instead of the squared regression errors makes the estimated coefficient $\hat{\gamma}_{0,t}$ equal to the median relative pertinence Rel-EFP in the observations used to estimate the model.¹¹

When we estimate the intercept in (2.24) using all the firms in the comprehensive set, we obtain an unconditional median Rel-EFP. However, when estimating the quantile regression model using a subset of firms, we obtain an intercept that represents a conditional median Rel-EFP as a function of the correlation between accruals and cash flows.

Let $\hat{\gamma}_{0,t}^{(-0.92)}$ and $\hat{\gamma}_{0,t}^{(-0.58)}$ denote the median for firms for which $\rho < -0.92$ and the median for firms for which $\rho > -0.58$, respectively. Here, $\hat{\gamma}_{0,t}^{(-0.92)}$ is the median Rel-EFP when accruals smoothing is likely to be substantial, and $\hat{\gamma}_{0,t}^{(-0.58)}$ is that which indicates when smoothing should be minor. When the difference between these two conditional medians is significant, this suggests that accruals smoothing relates to earnings' pertinence to valuation relative to that of cash flows. When controlling for accruals smoothing, a statistically significant value of $\hat{\gamma}_{0,t}^{(-0.92)}$, which is also positive and constant over time, supports the hypothesis that, for those firms characterized by substantial accruals smoothing, earnings are persistently more pertinent than cash flows to valuation.

¹¹Quantile regression is a model for the conditional median of the dependent variable, rather than a conditional mean like the OLS regression model. Furthermore, this technique mitigates the negative effect of extreme values. We estimate a quantile regression model so that the algorithm (we use the R function `rq()`) can easily approximate the standard error of a sample median via bootstrap or alternative techniques; in this way, we also effectively obtain a confidence interval for the median. Results do not qualitatively change when using a classic linear regression model.

2.2 Econometric techniques used to estimate accounting quality and model its relationship with stock prices

[Nikolaev \(2018\)](#) introduces a novel methodology used to define accounting quality. In Subsection [2.2.1](#), we discuss the general setting and the rationale behind this approach. We also explain how we can define and distinguish between unobservable operating performance uncertainty, operating risk, and accounting measurement errors. In Subsection [2.2.2](#), we explain why a Gaussian state-space model is an appropriate tool with which to infer these unobservable quantities (that is, the signal given by the unobservable economic performance of a firm, and the variability of the noise in terms of earnings and cash flows). In Subsection [2.2.3](#), we investigate the association between earnings' expectation-formation pertinence and accounting quality.

2.2.1 The concept of accounting quality

Let π_t denote the unobservable economic performance of a firm over a given accounting measurement period. [Nikolaev \(2018\)](#) explains that this is the sum of the economic outcomes generated by the operations a firm undertakes during that period. He argues that, while part of these operations generates profits or losses in the short term and within the accounting measurement period itself, others will instead generate profits or losses at a later date. This means that, while the economic consequences of the first type of operation are observable before the end of the measurement period, the outcome of the second type is a random variable; this is why π_t is an unobservable

quantity.

Although, at the current time, we cannot say whether the second type of operation will lead to future profits or losses, we can still have expectations regarding what the future outcome will be. Operating cash flows do not consider any expectations concerning the future economic cash flow consequences of operations taking place in the current accounting measurement period—consequences that will only be realized after the period ends. Furthermore, a fraction of a firm’s outcomes realized during the current measurement period is the consequence of operations that were undertaken before that measurement started. This also means that we should subtract the expectation about the economic outcome realized in the current period we had in the previous measurement period if we are to define a firm’s actual performance.

For these reasons, we can see that operating cash flows are fundamentally different from the unobservable performance, π_t . At time t , we can express π_t as:

$$\pi_t = \text{CFO}_{t,t} + \mathbb{E}_t(\text{CFO}_{t,t+1}) - \mathbb{E}_{t-1}(\text{CFO}_{t-1,t}), \quad (2.25)$$

where $\text{CFO}_{t,t}$ is the cash flows during the current measurement period that are the result of operations within the current measurement period (in other words, the actual number from the cash flows statement). The term $\text{CFO}_{t-1,t}$ denotes the cash flows during the current measurement period that are the result of operations that took place within the past measurement period, and $\text{CFO}_{t,t+1}$ represents the cash flows in the future measurement period that will be the result of operations that are taking

place within the current period. The operator \mathbb{E}_t denotes an expected value that is conditional to the information available at time t . The two conditional expected values in (2.25) represent timing errors at time t and $t - 1$.

Let:

$$W_t = \mathbb{E}_{t-1}(\text{CFO}_{t-1,t}) - \mathbb{E}_t(\text{CFO}_{t,t+1}),$$

denote a noise due to timing problems. Nikolaev (2018) rewrites the previous equation (2.25) as:

$$\text{CFO}_t = \pi_t + W_t, \tag{2.26}$$

showing that, because of these timing errors, operating cash flows are a noisy estimate of an unobservable economic performance. We can see the previous equation (2.26) as a signal-noise decomposition: the underlying operating performance π_t is the signal, while the timing errors in W_t are the noise.

The purpose of accrual accounting is to provide a more precise measure of a firm's underlying performance. To measure the underlying economic performance with perfect accuracy, Nikolaev (2018) points out that we should add the total amount of accruals (equal to $-W_t$) to operating cash flows to offset the existing noise. In this case, the total amount of earnings, i.e., cash flows in (2.26) plus $-W_t$, would be equal to the economic performance that cash flows should measure. However, he also explains that accountants face estimation errors due to ineffective assumptions, models, or estimates,

as well as when applying judgments in general. When accounting errors occur because of an imperfect accounting measurement process, we can see earnings as another (possibly less) noisy proxy for the same economic performance. Accordingly, Nikolaev (2018) expresses the accounting error at time t as:

$$v_t = \widehat{\mathbb{E}}_t(\text{CFO}_{t,t+1}) - \mathbb{E}_t(\text{CFO}_{t,t+1}),$$

where $\widehat{\mathbb{E}}_t(\text{CFO}_{t,t+1})$ denotes the actual accountants' correction applied to solve timing problems in the operating cash flows. Here, $\widehat{\mathbb{E}}_t(\text{CFO}_{t,t+1})$ is an estimate of the unknown (unobservable) timing error $\mathbb{E}_t(\text{CFO}_{t,t+1})$. When accountants attain a low estimation error when booking the total amount of accruals, then $\widehat{\mathbb{E}}_t$ is close to \mathbb{E}_t ; this, in turn, should suggest that accounting earnings can serve as a reliable proxy for performance. Therefore, the two terms v_t and v_{t-1} denote the difference between the timing problem at a given time and the accounting estimate of this conditional expectation, respectively.

We can then express earnings as:

$$\text{EARN}_t = \pi_t + V_t, \tag{2.27}$$

where $V_t = (v_t - v_{t-1})$ denotes a noise that occurs due to accounting measurement errors. As in the previous case, the equation (2.27) represents a signal-noise decomposition: the signal is the same underlying operating performance (π_t), while the noise V_t is given by the measurement errors when booking the total amount of accruals.

Nikolaev (2018) introduces a novel way of thinking about accruals. The two equations, (2.26) and (2.27), identify the total amount of accruals as the sum of two unobservable quantities: a performance measurement component W_t and an accounting error component V_t . That is:

$$\text{ACC}_t = W_t + V_t, \tag{2.28}$$

where W_t captures cash flow shocks arising during the operating cycle, while the variability of V_t relates to accruals' ability to measure economic performance.¹²

The three equations above define the following system:

$$\text{CFO}_t = \pi_t + W_t, \tag{2.26}$$

$$\text{EARN}_t = \pi_t + V_t, \tag{2.27}$$

$$\text{ACC}_t = V_t - W_t, \tag{2.28}$$

in which all of the variables on the right side are unobservable. Everything else constant, a higher variance of π_t denotes incertitude regarding future economic performances. Similarly, higher variability of V_t suggests higher accounting errors in reporting accruals, while a higher variability of W_t suggests that operating cash flows are being subjected to larger timing errors.

The difference between the variability of W_t and V_t captures the benefit of accruals

¹²Therefore, the variability of W_t relates to the firm's operating risk. See Nikolaev (2018), and, as cited in his paper, Hribar and Craig Nichols (2007).

accounting. Nikolaev (2018) explains that accruals meet their primary objective when the variability of W_t is greater than that of V_t ; in other words, making earnings a more precise approximation of the underlying economic performance than cash flows are.

These variance proxies are specific sources of uncertainty and underlying risk. Specifically, we cannot distinguish between these sources when measuring quality through the variability of earnings and cash flows. By looking at equations (2.26) and (2.27), we can see that cash flows volatility is a mixture of economic risk and the size of timing errors. In the same way, earnings volatility is a mixture of economic risk and the size of accounting errors. If earnings and cash flows are volatile, this does not necessarily mean that accounting quality is low.

We need to estimate economic performance volatility and timing and accounting measurement error size if we are to evaluate accounting quality, disentangling the individual contributions made by the economic performance volatility and the size of timing and accounting measurement errors. We can define accounting quality as high (low) when accruals do (do not) generate a more accurate performance measure relative to cash flows, and this is when the variance of W_t is greater (lower) than that of V_t (Nikolaev, 2018).

We can also define the quality of earnings and cash flows by looking at the variability of V_t and W_t . Keeping everything else constant, while cash flows quality should decrease when timing errors increase, earnings quality should decrease when accounting errors

increase.¹³

2.2.2 Estimating accounting quality using a Gaussian state-space model

In this subsection, we explain how we estimate accounting quality using a Gaussian state-space model, solving for the variability of operating performance and the size of timing errors and accounting process imperfections using the system of equations seen in Nikolaev (2018) in (2.26) to (2.28).

The first two equations are as follows:

$$\text{CFO}_t = \pi_t + W_t, \tag{2.26}$$

$$\text{EARN}_t = \pi_t + V_t. \tag{2.27}$$

These describe the dynamics of accounting earnings and operating cash flows, respectively. In this setting, the economic performance π_t represents a common and unobservable state, which we assume is a Gaussian random walk with a constant drift parameter. This means that period-to-period changes in the state are white noise following a normal distribution; that is, $\Delta\pi_t \sim N(0, \sigma_{\pi,t}^2)$ where $\mathbb{E}(\Delta\pi_t \Delta\pi_j) = 0$ when $t \neq j$, while $\mathbb{E}(\Delta\pi_t \Delta\pi_t) = \sigma_{\pi,t}^2$ measures the variability of the underlying economic performance. The two quantities W_t and V_t in (2.26) and (2.27) measure cash flows and earnings departures from the unobservable state π_t , respectively, and (as we have

¹³In this thesis, we only focus on accrual quality and its relationship with earnings' pertinence to the valuation process; the analysis of the quality of earnings and cash flows is beyond the scope of this work.

explained in the previous subsection) we can see these quantities as two noises.

Let:

$$(V_t, W_t)' \sim N(0, \Sigma_t), \quad (2.29)$$

be the joint distribution of the two error terms V_t and W_t . As in [Nikolaev \(2018\)](#), the covariance matrix Σ_t is assumed to be diagonal. This means that timing errors occur regardless of the measurement errors in the accruals. In principle, the higher the variances of these departures are, the more ineffective accounting earnings and operating cash flows should be at summarizing the firm's unobservable economic performance. In what follows, $\sigma_{\pi,t}^2$ denotes uncertainty behind the operating performance, $\sigma_{W,t}^2$ measures the size of the timing issues, and $\sigma_{V,t}^2$ quantifies the size of the accounting measurement errors.

The three assumptions—(2.26), (2.27), and (2.29)—together with $\Delta\pi_t$ as white noise, correspond with the strand of literature that suggests that a random walk can effectively describe earnings and cash flows' temporal dynamics ([Brown, 1993](#); [Finger, 1994](#); [Gerakos and Gramacy, 2013](#)). These assumptions define a dynamic linear model, from which we can estimate the variance of the three unobservable quantities π_t , V_t , and W_t via maximum likelihood estimation. These estimates are conditional on the information about accounting earnings and operating cash flows.

The state-space model described above gave us a more stable solution than the gen-

eralized method of moments proposed by [Nikolaev \(2018\)](#). When we estimated the variance of π_t , V_t , and W_t using a generalized method of moments, the algorithm found a different optimal solution that depends on an arbitrary starting point. That is equivalent to saying that estimates of these variances are conditional on random inputs of the algorithm rather than purely on the value of earnings and cash flows. Therefore, changing the starting value of the algorithm may let the generalized method of moments converge toward many different local optima (alternative and imperfect solutions to the same optimization problem), thus making the method unstable. Estimating a Gaussian state-space model has mitigated this issue; additionally, it is also more convenient in terms of the computational burden.¹⁴ An interested reader can find details about state-space models in [Lütkepohl \(2005\)](#) and [Tsay \(2005\)](#).

We use the following three constructs to study the association between earnings' expectation-formation pertinence and accounting quality:

$$\lambda_t^\pi = \ln \hat{\sigma}_{\pi,t}^2, \tag{2.30}$$

$$\lambda_t^{\text{CFO}} = \ln \hat{\sigma}_{W,t}^2, \tag{2.31}$$

$$\lambda_t^{\text{ACC}} = \ln \hat{\sigma}_{W,t}^2 - \ln \hat{\sigma}_{V,t}^2. \tag{2.32}$$

The first construct is an estimate of economic performance variability; keeping everything else constant, this means that higher values of λ_t^π should imply that higher levels of

¹⁴We can also consider alternative specifications of the state-space model mentioned above. For example, results are robust when estimating a dynamic linear model by considering changes in earnings and cash flows instead of the same variables in levels.

uncertainty lie behind a firm’s profitability. The second construct is a proxy for cash flow’s ability to capture that underlying performance. Keeping everything else constant, higher values of λ_t^{CFO} should imply greater timing issues; therefore, we can view this second construct as a measure of operating risk. The last construct is a measure of the accrual accounting process’s effectiveness. The variance of V_t should capture the magnitude of accounting errors when defining accruals; therefore, keeping everything else constant, the difference between the two variabilities in (2.32) should capture the benefit of accrual accounting.

Consistent with Nikolaev (2018), when W_t fluctuates more than V_t , accruals meet their objective, making accounting earnings a more precise measure of the underlying economic performance than operating cash flows are. It is in this interpretation that we refer to the third construct as accounting quality.

2.2.3 Estimating the relationship between earnings’ expectation-formation pertinence and accounting quality

In this subsection, we explain how we investigate if the proxy for accounting quality in (2.32) has a significant association with our measure of earnings’ pertinence to the price-formation process. To answer the research question, we consider the following multivariate panel regression model:

$$\ln |P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}| = \text{FE}_{i,t} + b_t^{\text{P}} \ln P_{i,t} + b_t^{\pi} \lambda_{i,t}^{\pi} + b_t^{\text{CFO}} \lambda_{i,t}^{\text{CFO}} + b_t^{\text{ACC}} \lambda_{i,t}^{\text{ACC}} + \varepsilon_{i,t}, \quad (2.33)$$

estimated on a rolling window of multiple years. The dependent variable is the logarithm of the absolute error from the valuation model in (2.6). We consider four independent variables: the logarithm of the stock price and three constructs from the system of equations of Nikolaev (2018). Of these, λ_t^π measures the economic performance variability, λ_t^{CFO} measures operating risk, λ^{ACC} proxies accounting quality. The term FE denotes both firm- and time-fixed effects. ε is a regression error term.

A statistically significant value of the regression coefficient b_t^{ACC} in (2.33) offers evidence in favor of an association between accounting quality and earnings' pertinence to the valuation process. A negative value suggests that lower-quality earnings are less pertinent to the valuation process for firms that have comparable levels of economic uncertainty and operating risk, and similar stock prices.¹⁵ We can assume that this association is also persistent through time if the coefficient b_t^{ACC} is significant in every cross-section.

¹⁵Due to natural differences in scale between valuation errors, the stock price is on the right side of the equation in order to avoid bias. When the stock price is high, a valuation model is likely to provide a higher absolute error when keeping everything else constant. Alternatively, we can see the conditional variance of signed (i.e., non-absolute value) valuation errors as an increasing function of the stock price. We need to hold the stock price constant if we are to investigate the association between absolute errors and proxies for accounting quality. Under these circumstances, and when prices correlate with accounting quality, omitting the information about stock prices would bias the coefficient of interest b_t^{ACC} in (2.33).

3 Variables and sample

3.1 Variables definition

Table 3 contains a summary of all of the following definitions.

3.1.1 Earnings, cash flows, and accruals

We define earnings (EARN) as income before extraordinary items (Compustat IB). Firms did not have to disclose the statement of cash flows before the Statement of Financial Accounting Standards (SFAS) 95 in 1987 (Bushman et al., 2016). When the Compustat item OANCF is available, we define cash flows (CFO) as operating activities (Compustat OANCF). In this case, we derive the corresponding amount of accruals as the difference $ACC = EARN - CFO$. When Compustat OANCF is missing, we define operating cash flows as $CFO = EARN - ACC$, where we estimate the total amount of accruals using a balance-sheet approach; that is, $ACC = (\Delta ACT - \Delta LCT) - (\Delta CHE - \Delta DLC) - DP$, where the terms are annual changes in current assets (Compustat item ACT), current liabilities (Compustat LCT), cash and short-term investments (Compustat CHE), total debt in current liabilities (Compustat DLC), net to depreciation and amortization (Compustat DP).

3.1.2 Regressand and regressors in the nonlinear regression model

The following variables allow us to compute a firm- and year-specific measure of earnings and cash flows' pertinence to the valuation process, as we explain in Section 2.1. The dependent variable is the stock price (P), being the monthly closing price (CRSP PRCCM) at the beginning of the fourth month after the fiscal year-end. We estimate the nonlinear regression models using earnings and cash flows per share; i.e., EARN and CFO scaled by common shares outstanding (Compustat CSHO). Beside of earnings and cash flows, we include the following independent variables. A categorical variable identifies the industry to which a firm belongs, using either the twelve-industry classification available on Kenneth French's website (see **Table 2**) or the SIC code, from two to three digits. We also include the total market value (Mkt), i.e., the stock price times the total number of common shares outstanding. We use the market value of equity as a second proxy for risk, with industry membership being the other.¹⁶ The book value of equity is the Compustat item CEQ. The market-to-book ratio (MB) is equal to Mkt divided by CEQ from the previous fiscal year. We use the market-to-book ratio and industry membership to proxy the level of firms' conservatism. We compute an average change in sales (Δ Sales) to measure firms' growth, where sales are the Compustat SALE. We calculate that average change using time-series information that corresponds to the most recent eight years of data. We compute this time-series measure of change for each firm and year, using at least four values out of the eight most recent annual

¹⁶It is also common to use the logarithm of the market value. Another advantage of Random Forests is that they are invariant to the monotonic transformation of the regressors. For example, taking the logarithm to reduce the impact of extreme values is not necessary for this setting.

observations in a period from year $t - 7$ up to t .

Our results do not change when using alternative proxies for size (either sales or total assets, which is the Compustat AT) or the C-index of [Penman and Zhang \(2002\)](#) as an alternative proxy for accounting conservatism. [Penman and Zhang \(2002\)](#) define the total amount of estimated reserves created by the conservatism as equal to the inventory reserves (Compustat LIFO), R&D assets (Compustat XRD), and advertising expenses (Compustat XAD). As in their paper, we assume amortizations of five and two years for the R&D assets and advertising expenses, respectively, using the sum of the years' digits method. To obtain the C-score of [Penman and Zhang \(2002\)](#), we divide the estimated reserve by the previous year's total assets.

3.1.3 Accounting constructs to explain variation in pertinence

As in [Givoly and Hayn \(2000\)](#) and [Dichev and Tang \(2009\)](#), we scale accounting earnings, operating cash flows, and accruals by total assets.¹⁷ For our analyses, we consider the average total assets at the beginning and end of the fiscal year (current and past lags in Compustat AT). This is done in order to estimate the standard deviation of accounting earnings, operating cash flows, and accruals, as well as the correlation between accruals and cash flows. We compute these four time-series measures for each firm and year, using at least four values out of the eight most recent annual observations, in a period from year $t - 7$ up to t . We denote the volatility of earnings, cash flows, and accruals

¹⁷We deflate none of these variables by the number of common shares outstanding to compute the accounting constructs to explain variation in pertinence.

as $\sigma(\text{EARN})$, $\sigma(\text{CFO})$ and $\sigma(\text{ACC})$, while the correlation coefficient between accruals and cash flows is ρ . The ratio between earnings and cash flows' volatility is equal to $\sigma(\text{EARN})$ divided by $\sigma(\text{CFO})$, denoted as $\delta\sigma$.

For the remaining construct, and following [Dichev and Tang \(2008\)](#) and [Srivastava \(2014\)](#), we define total expenses (EXP) as the difference between total revenues (Compustat REVT) and earnings. The investment in SG&A (SG&A) is the Compustat XSGA, while we define the corresponding intangible intensity (SGAI) is the ratio between the investment in SG&A and total expenses.

Industry description	
1.	Consumer nondurables: food, tobacco, textiles, apparel, leather, toys SIC: 0xxx, 2000-2399, 2700-2749, 2770-2799, 31xx, 3940-3989
2.	Consumer durables: cars, TVs, furniture, household appliances SIC: 2500-2519, 259x, 3630-3659, 3710-3711, 3714, 3716, 3750-3751, 3792, 3900-3939, 399x
3.	Manufacturing: machinery, trucks, planes, off. furn., paper, com. printing SIC: 2520-2589, 26xx, 2750-2769, 30xx, 3200-3569, 3580-3629, 370x, 3712-3713, 3715, 3717-3749, 3752-3791, 3793-3799, 383x, 3860-3899
4.	Oil, gas, and coal extraction and products SIC: 1200-1399, 29xx
5.	Chemicals and allied products SIC: 2800-2829, 2840-2899
6.	Business equipment: computers, software, and electronic equipment SIC: 3570-3579, 3660-3692, 3694-3699, 3810-3829, 737x
7.	Telephone and television transmission SIC: 48xx
8.	Utilities SIC: 4900-4949
9.	Wholesale, retail, and some services (laundries, repair shops) SIC: 5xxx, 72xx, 76xx
10.	Healthcare, medical equipment, and drugs SIC: 283x, 3693, 3840-3859, 80xx

- 11. Finance
SIC: 6xxx
 - 12. Other
-

Table 2: **Kenneth R. French’s twelve-industry classification.** This table shows the twelve-industry classification available on Kenneth R. French’s website. We group firms classified as ‘Other’ into industries according to the one-digit SIC code to address heterogeneity.

Variable	Description
EARN	Accounting earnings Compustat IB.
ACC	Accruals When Compustat OANCF is not missing: $ACC = EARN - \text{Compustat OANCF},$ or, when Compustat OANCF is missing: $ACC = (\Delta ACT - \Delta LCT) - (\Delta CHE - \Delta DLC) - DP,$ where the terms are annual changes in current assets (Compustat ACT), current liabilities (Compustat LCT), cash and short-term investments (Compustat CHE), total debt in current liabilities (Compustat DLC), net to depreciation and amortization (Compustat DP).
CFO	Operating cash flows Compustat OANCF when available, while: $CFO = EARN - ACC,$ when Compustat OANCF is missing.

AvgAT	<p>Average total assets</p> <p>The average between the current and past values of total assets (Compustat AT). We use this quantity to scale earnings, cash flows, and accruals before computing any of the time-series constructs given below.</p>
$\sigma(\text{EARN})$	<p>Earnings volatility</p> <p>The square root of earnings' time-series variance. This is computed using firm-specific data from the most recent eight years; i.e., the volatility at time t takes into account data from year $t - 7$ to t. We require at least four observations in this interval, and scale earnings by the average total assets before computing their volatility.</p>
$\sigma(\text{CFO})$	<p>Cash flows volatility</p> <p>The square root of cash flows' time-series variance. This is computed using firm-specific data from the most recent eight years; i.e., the volatility at time t takes into account data from year $t - 7$ to t. We require at least four observations in this interval, and scale cash flows by the average total assets before computing their volatility.</p>
$\sigma(\text{ACC})$	<p>Accruals volatility</p> <p>The square root of accruals' time-series variance. This is computed using firm-specific data from the most recent eight years; i.e., the volatility at time t takes into account data from year $t - 7$ to t. We require at least four observations in this interval, and scale accruals by the average total assets before computing their volatility.</p>
ρ	<p>Proxy for the smoothing role played by accruals</p> <p>The time-series correlation coefficient between accruals and cash flows. This is computed using firm-specific data from the most recent eight years; i.e., the correlation at time t takes into account data from year $t - 7$ to t. We require at least four observations in this interval, and scale accruals and cash flows by the average total assets before computing that correlation coefficient.</p>
$\delta\sigma$	<p>Ratio between the volatility of earnings and cash flows</p> <p>For a given firm i and year t, this ratio is equal to:</p> $\delta\sigma_{i,t} = \frac{\sigma(\text{EARN})_{i,t}}{\sigma(\text{CFO})_{i,t}},$ <p>where a value greater than one suggests that accounting earnings fluctuate more than operating cash flows do.</p>

EXP	Total expenses Compustat REVT minus EARN.
SGAI	Intangible intensity Compustat SGA divided by EXP.
P	Stock price We take CRSP PRC from the monthly stock file as at the end of the third month after the corresponding firm's fiscal year-end. When this is missing, we replace this definition with the average between the bid and ask prices: CRSP BID and ASK, respectively. We eliminate all the observations for which the stock price is less than one USD.
Mkt	Market value of equity P times Compustat CSHO. We eliminate all the firm-year observations for which this value is negative.
B	Book value of equity Compustat CEQ. We eliminate all the firm-year observations for which this value is negative.
MB	Market-to-book ratio Mkt divided by B.
$\hat{\varepsilon}$	Valuation error This is the difference between the actual stock price and the estimated value from a nonlinear regression model; that is: $\hat{\varepsilon}_{i,t} = P_{i,t} - \hat{P}_{i,t},$ where the definitions from Subsection 2.1.2 apply.
$\hat{\varepsilon}$	Absolute relative valuation error This is the absolute value of a valuation error $\hat{\varepsilon}$ divided by the actual stock price.
\overline{KS}	Kolmogorov-Smirnov test statistic (maximum) This is the first measure of distance between the two empirical cumulative-distribution functions (c.d.f.) of absolute relative valuation errors using earnings and cash flows. We examine this further in Subsection 2.1.5.

KS	<p>Kolmogorov-Smirnov test statistic (minimum)</p> <p>This is the second measure of distance between the two c.d.f., see Subsection 2.1.5.</p>
Rel-EFP	<p>Earnings' pertinence relative to that of cash flows</p> <p>The logarithm of the ratio between two valuation errors, one using cash flows and the other using earnings; that is:</p> $\text{Rel-EFP}_{i,t} = \ln P_{i,t} - \hat{P}_{i,t}^{\text{CFO}} - \ln P_{i,t} - \hat{P}_{i,t}^{\text{EARN}} ,$ <p>where a positive value suggests that accounting earnings are more expectation-formation pertinent than operating cash flows, and vice versa. This is because the size of the absolute valuation error we obtain when considering earnings is less than the one we obtain when considering cash flows. We examine this further in Subsection 2.1.6.</p>
λ^π	<p>Operating performance volatility</p> <p>That is the logarithm of the estimated operating performance variance. We examine this further in Subsection 2.2.2.</p>
λ^{CFO}	<p>Size of timing errors</p> <p>This is the logarithm of the estimated timing error variance. We examine this further in Subsection 2.2.2.</p>
λ^{ACC}	<p>Accounting quality</p> <p>This is the difference between the logarithm of the estimated timing error variance and the logarithm of the estimated errors in the accruals booking process. A positive value suggests earnings are a more precise measure of the underlying economic performance than operating cash flows are. We examine this further in Subsection 2.2.2.</p>
\mathcal{I}	<p>Industry</p> <p>This is a categorical variable identifying the industry. See Table 2.</p>

Table 3: **Variables definition.** This table defines the main variables in our empirical investigations.

3.2 Sample

The data in this thesis come from Compustat and CRSP. The final sample is made of 188,237 firm-year observations for a total of 16,817 distinct firms over a time horizon of 56 years, between 1960 and 2016.

These observations have non-missing values for those variables that define the valuation errors (2.10) and (2.11); furthermore, we require non-missing values for the three fundamental firm characteristics we use in the multivariate analysis described in Subsections 2.1.6 and 2.1.7. When calculating these time-series constructs, we require that a given firm has at least four non-missing observations in the eight most recent years.

Following the prior literature, we also exclude financial firms (for which the two-digit SIC code is between 60 and 69). We also eliminate stock prices lower than one, negative values of total assets, book values of equity, and common shares outstanding. All the variables are winsorized at the 1% level to mitigate the effect of extreme values.

Panel A in **Table 4** shows sample quantiles for earnings, cash flows, and accruals. Panel B serves the same purpose for the two measures of earnings and cash flows' pertinence to the valuation process, plus the ratio between them (not in logarithm). Panel C shows sample quantiles of the explanatory firms' characteristics we considered to explain variation in our measure of earnings' pertinence relative to cash flows.

Variable description		Q-10%	Q-25%	Q-50%	Q-75%	Q-90%
Panel A - Earnings, cash flows, and accruals						
EARN	Earnings	-13.33	-0.54	3.02	22.55	100.40
CFO	Cash flows	-9.14	-0.34	4.88	38.38	173.00
ACC	Accruals	-87.31	-19.03	-2.37	0.56	7.02
Panel B - Pertinence to the valuation process						
$\hat{\epsilon}^{\text{EARN}}$	Earnings' pertinence	0.04	0.10	0.23	0.47	0.85
$\hat{\epsilon}^{\text{CFO}}$	Cash flows' pertinence	0.05	0.11	0.24	0.49	0.88
Rel-EFP	Relative pertinence	0.27	0.58	1.04	1.94	4.35
Panel C - Explanatory characteristics						
$\sigma(\text{EARN})$	Earnings volatility	0.01	0.02	0.04	0.09	0.17
$\sigma(\text{CFO})$	Cash flows volatility	0.02	0.04	0.06	0.10	0.21
$\sigma(\text{ACC})$	Accruals volatility	0.02	0.04	0.06	0.10	0.17
ρ	Accruals smoothing	-0.98	-0.94	-0.81	-0.52	-0.09
SGAI	Intangible intensity	0.07	0.13	0.22	0.34	0.48

Table 4: **Sample quantiles.** This table shows the sample quantiles for the main variables used in our analysis. For ease of understanding, Rel-EFP is not in logarithm in this table; therefore, the numbers refer to the distribution of the ratio between $\hat{\epsilon}^{\text{CFO}}$ and $\hat{\epsilon}^{\text{EARN}}$. We can see that the absolute residuals from a nonlinear regression model using cash flows are greater than those of earnings for more than 50% of the firm-year observations. We can also see that the earnings volatility tends to be lower than that of cash flows, while the correlation between accruals and cash flows is negative for most of the sample. In our analysis, we focus on the evolution of all of these measures over time. Following the literature cited in the previous Subsection 3.1.3, we scale earnings, cash flows, and accruals by the average total assets between $t - 1$ and t , instead of the number of shares outstanding, when computing the corresponding three volatilities, plus the correlation between accruals and cash flows. Table 3 defines all the variables above.

4 Results

In this chapter, we show the empirical findings from the analysis of the accounting earnings and operating cash flows' pertinence to the valuation process. The first section—section 4.1—is about the relationship that stock prices have with earnings and cash flows. The results in this section have been reported in [Starica and Giosi \(2017\)](#). The second section—section 4.2—is about the relationship between earnings pertinence and accounting quality.

As we have already discussed in Subsection 2.1.2, we base our conclusions on the distribution of the two absolute relative valuation errors, defined as follows:

$$\begin{aligned}\hat{\epsilon}_{i,t}^{\text{EARN}} &= \frac{|P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}|}{P_{i,t}}, \\ \hat{\epsilon}_{i,t}^{\text{EARN}} &= \frac{|\hat{\epsilon}_{i,t}^{\text{EARN}}|}{P_{i,t}},\end{aligned}\tag{4.1}$$

and:

$$\begin{aligned}\hat{\epsilon}_{i,t}^{\text{CFO}} &= \frac{|P_{i,t} - \hat{P}_{i,t}^{\text{CFO}}|}{P_{i,t}}, \\ \hat{\epsilon}_{i,t}^{\text{CFO}} &= \frac{|\hat{\epsilon}_{i,t}^{\text{CFO}}|}{P_{i,t}}.\end{aligned}\tag{4.2}$$

We obtain these by consistently estimating the corresponding two nonlinear regression models, as follows:

$$P_{i,t} = \mathbf{m}_{i,t}^{\text{EARN}}(\text{EARN}_{i,t}; \mathcal{I}_{i,t}, \Delta\text{Sales}_{i,t}, \text{Mkt}_{i,t-1}, \text{MB}_{i,t-1}) + \varepsilon_{i,t}^{\text{EARN}},\tag{4.3}$$

and:

$$P_{i,t} = \mathbf{m}_{i,t}^{\text{CFO}}(\text{CFO}_{i,t}; \mathcal{I}_{i,t}, \Delta\text{Sales}_{i,t}, \text{Mkt}_{i,t-1}, \text{MB}_{i,t-1}) + \varepsilon_{i,t}^{\text{CFO}}, \quad (4.4)$$

where all of the definitions from the earlier Subsection 2.1.2 apply. As we have already explained, the two nonlinear regression functions—denoted as \mathbf{m} in (4.3) and (4.4)—are specific to each predictor set and are estimated cross-sectionally.

The growth rate of sales (ΔSales) is used as a proxy for future income growth. We include a categorical variable indicating the industry membership (\mathcal{I}) to control for both risk and conservatism. The market value of equity (Mkt) is another proxy of risk, while the market-to-book ratio (MB) is a second measure of the level of conservative accounting. The absolute relative valuation errors $\hat{\varepsilon}_{i,t}^{\text{EARN}}$ in (4.1) ($\hat{\varepsilon}_{i,t}^{\text{CFO}}$ in (4.2)) are firm- and year-specific measures of earnings' (cash flows') pertinence to valuation.

4.1 The relationship between stock prices and accounting variables

We divide this broad section of results into six subsections, each of which refers to empirical findings concerning a specific aspect of the relationship between stock prices and accounting variables.

4.1.1 An illustration of the concept of stochastic dominance

First, we show the estimated cumulative distribution functions of the two absolute relative valuation errors: $\hat{\epsilon}_{i,t}^{\text{EARN}}$ and $\hat{\epsilon}_{i,t}^{\text{CFO}}$ in (4.1) and (4.2), respectively. These functions are conditional on the year and the level of the correlation coefficient between accruals and operating cash flows, ρ , which serves as a measure of the smoothing role of accruals.

Figure 1 shows the cumulative distribution function of the two valuation errors in 1968, 1980, 1990, and 2009 (logarithmic scale). The black (blue) line denotes the cumulative distribution function of valuation errors from the model using earnings (cash flows).

The four graphs illustrate the notion of stochastic dominance, with one cumulative distribution function being above the other. First, the ordering of the curves has become inverted over time. In 1968 and 1980, the distribution of the absolute relative errors corresponding to a valuation informed by earnings stochastically dominates that of the errors that correspond to operating cash flows. This suggests that the accounting earnings provide a more precise valuation and have a stronger association with stock prices, and, therefore, that they are more pertinent than cash flows to valuation in that period. In 1990, the two functions were almost identical, suggesting that earnings were as pertinent as cash flows. Then, in 2009, we observe that stock prices are more associated with cash flows than earnings. **Figure 3** shows the results of hypothesis testing on the difference between the two cumulative distribution functions.

Figure 2 shows the cumulative distribution function of the two absolute relative valuation errors conditional on the value of the correlation between accruals and cash

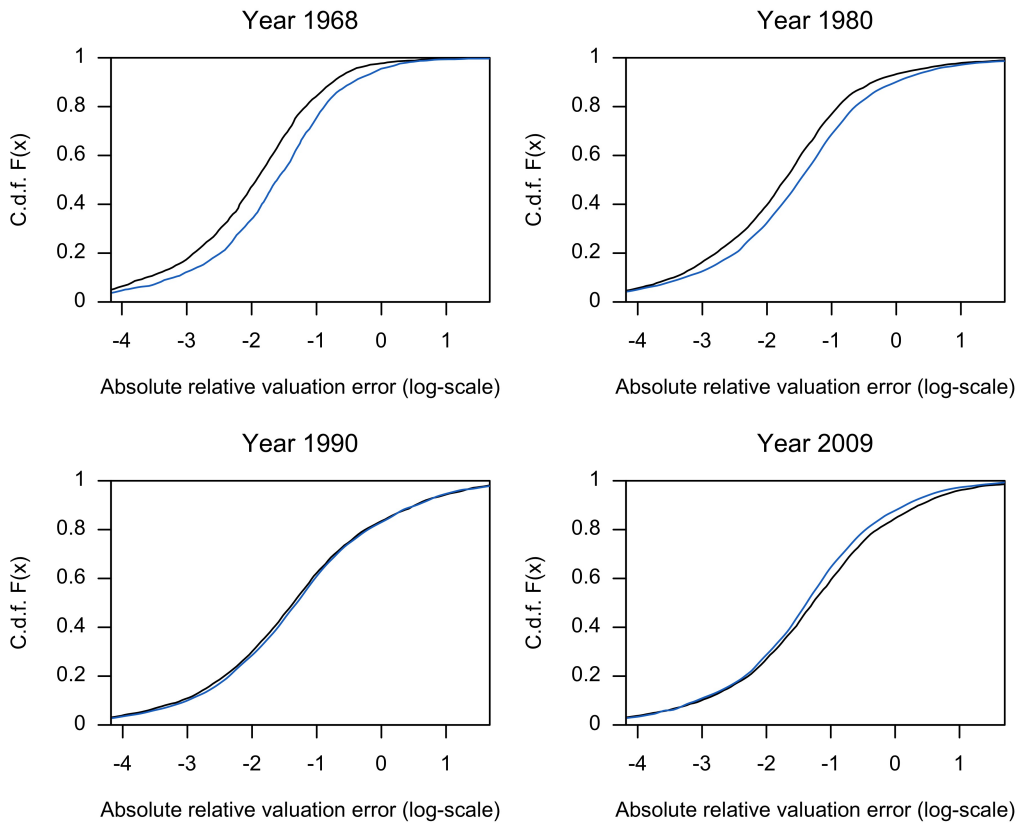


Figure 1: **The cumulative distribution function of valuation errors; c.d.f. conditional on the fiscal year.** These graphs show the cumulative distribution function of the two valuation errors $\hat{\epsilon}_{i,t}^{\text{EARN}}$ and $\hat{\epsilon}_{i,t}^{\text{CFO}}$ in (4.1) and (4.2), respectively, as in 1968, 1980, 1990, and 2009 (logarithmic scale). The black (blue) line denotes the cumulative distribution function of valuation errors from the model using earnings (cash flows). The four graphs illustrate the notion of stochastic dominance, with one cumulative distribution function being above the other. First, the ordering of the curves has become inverted over time. In 1968 and 1980, the distribution of the absolute relative errors corresponding to a valuation informed by earnings stochastically dominates that of the errors that correspond to operating cash flows. This suggests that the accounting earnings provide a more precise valuation and have a greater association with stock prices, and, therefore, that they are more pertinent than cash flows to valuation in that period. In 1990, the two functions were almost identical, suggesting that earnings were as pertinent as cash flows. Then, in 2009, we observe that stock prices are more associated with cash flows than earnings. The following Figure 3 shows the results of hypothesis testing on the difference between the two cumulative distribution functions.

flows (ρ). The black (blue) line denotes the cumulative distribution function of the valuation errors from the model that uses earnings (cash flows).

The first graph in Figure 2 refers to firms for which the correlation between accruals and cash flows is strongly negative and lower than the first decile (-0.98). In this instance, the valuation informed by earnings stochastically dominates the one that corresponds to cash flows. When the correlation between accruals and cash flows is higher than the third quartile (-0.50), the second graph suggests that cash flows become at least as pertinent as earnings. Hypothesis testing on the difference between the two cumulative distribution functions is shown in the following Figure 6.

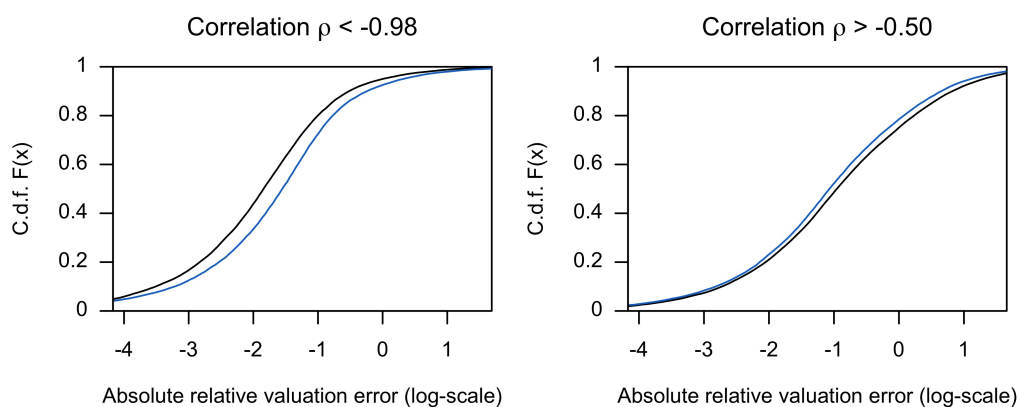


Figure 2: Valuation errors' cumulative distribution function; c.d.f. conditional on the correlation between accruals and cash flows. These graphs show the cumulative distribution function of the absolute relative valuation errors $\hat{\epsilon}_{i,t}^{\text{EARN}}$ and $\hat{\epsilon}_{i,t}^{\text{CFO}}$ in (4.1) and (4.2), respectively, conditional on the value of the correlation between accruals and cash flows (ρ). The black (blue) line denotes the cumulative distribution function of valuation errors from the model that uses earnings (cash flows). The first graph refers to firms for which the correlation between accruals and cash flows is strongly negative and lower than the first decile (-0.98). In this instance, the valuation informed by earnings stochastically dominates the one corresponding to cash flows. When the correlation between accruals and cash flows is higher than the third quartile (-0.50), the second graph suggests that cash flows become at least as pertinent as earnings. Hypothesis testing on the difference between the two cumulative distribution functions is shown in the following Figure 6.

4.1.2 Changes in earnings' pertinence compared to that of cash flows

Figure 3 shows the evolution of the two Kolmogorov-Smirnov test statistics between 1962 and 2016. The blue (red) solid line denotes the maximum \overline{KS}_t in (2.17) (minimum \underline{KS}_t in (2.18)). The shaded grey area marks the corresponding acceptance region of a one-sided, 95% test. For the test statistic \overline{KS}_t (\underline{KS}_t), the acceptance region is the interval between zero and the corresponding critical value, the blue (red) dotted line.

Figure 3 presents the results of a large number of hypothesis tests, one for each cross-section. A value of the test statistic \overline{KS}_t (\underline{KS}_t) above (below) the critical value is against the null hypothesis that earnings and operating cash flows have equal pertinence to valuation, supporting an alternative: that accounting earnings (cash flows) are more pertinent than cash flows (accounting earnings). We do not reject the null hypothesis mentioned above when the two statistics do not exceed the corresponding critical values.

Up to around the year 2000, the test statistic \overline{KS}_t lies outside the acceptance region. A statistically significant value of \overline{KS}_t suggests that, before this year, valuations informed by accounting earnings were more precise than those informed by cash flows. With only one exception, the two statistics are within the acceptance region between 2000 and 2007, suggesting that the two accounting variables have comparable associations with stock prices. Between 2008 and 2011, the \underline{KS}_t lies outside of the acceptance region. A statistically significant value of \underline{KS}_t suggests that, during and immediately after the subprime financial crisis, cash flows were more pertinent than earnings to the valuation process.

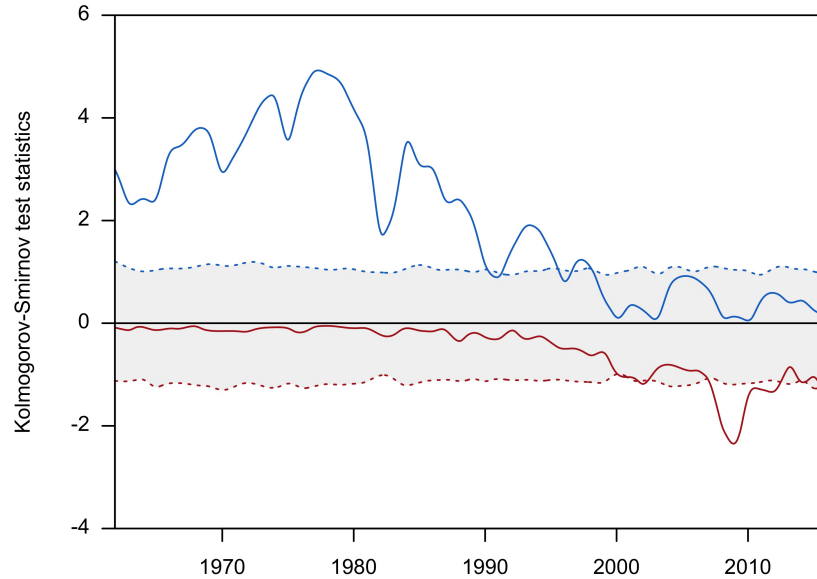


Figure 3: **Kolmogorov-Smirnov test statistics.** This graph shows the evolution of the two Kolmogorov-Smirnov test statistics between 1962 and 2016. The blue (red) solid line denotes the maximum \overline{KS}_t in (2.17) (minimum \underline{KS}_t in (2.18)). The shaded grey area marks the corresponding acceptance region of a one-sided, 95% test. For the test statistic \overline{KS}_t (\underline{KS}_t), the acceptance region is the interval between zero and the corresponding critical value, the blue (red) dotted line. The figure presents the results of a large number of hypothesis tests, one for each cross-section. A value of the test statistic \overline{KS}_t (\underline{KS}_t) above (below) the critical value is against the null hypothesis that earnings and operating cash flows have equal pertinence to valuation, supporting an alternative: that accounting earnings (cash flows) are more pertinent than cash flows (accounting earnings). We do not reject the null hypothesis mentioned above when the two statistics do not exceed the corresponding critical values. Up to around the year 2000, the test statistic \overline{KS}_t lies outside the acceptance region. A statistically significant value of \overline{KS}_t suggests that, before this year, valuations informed by accounting earnings were more precise than those informed by cash flows. With only one exception, the two statistics are within the acceptance region between 2000 and 2007, suggesting that the two accounting variables have comparable associations with stock prices. Between 2008 and 2011, the \underline{KS}_t lies outside of the acceptance region. A statistically significant value of \underline{KS}_t suggests that, during and immediately after the subprime financial crisis, cash flows were more pertinent than earnings to the valuation process.

4.1.3 A possible explanation for the decline in the earnings' pertinence

We have seen that accounting earnings are becoming less pertinent to the valuation process compared to operating cash flows. In this subsection, we document that earnings

volatility, compared to that of cash flows, is also increasing, arguing that a change in earnings' smoothness compared to that of cash flows can justify the reduction in its pertinence to valuation. In Subsection 4.1.4, we also investigate whether this increase in volatility may be due to economic and accounting developments.

The blue solid (dotted) line in **Figure 4** shows the cross-sectional, median earnings' (cash flows') volatility. We can see that the variability of the two accounting measures of profitability evolves differently. The volatility of cash flows has an upward trend until the end of the 1980s, from which point cash flows become as smooth or smoother than earnings by the end of the sample. The cross-sectional median earnings volatility increases threefold between the 1960s and the year 2000, decreasing slightly afterward. As a result of these changing dynamics, the cross-sectional median volatility of cash flows is equal to or smaller than the median volatility of accounting earnings after 2000.

The red dotted line in **Figure 4** shows the difference between the volatilities of earnings and cash flows (see the scale on the right-hand side of the graph). That difference converges to zero from around 1980 to 2000. In summary, cash flows have been less volatile than earnings over the last ten years.

In **Figure 5**, we repeat the same analysis, this time considering annual changes in earnings and cash flows. This figure shows the cross-sectional, median volatility of yearly changes in the total amount of earnings and cash flows, and the difference between these two volatilities. The colors are as in **Figure 4**; while the blue solid (dotted) line shows the cross-sectional median volatility of annual changes in earnings (cash flows), the red dotted line is the difference between the two volatilities (see the scale on the right-hand side of the graph). Conclusions regarding the yearly changes

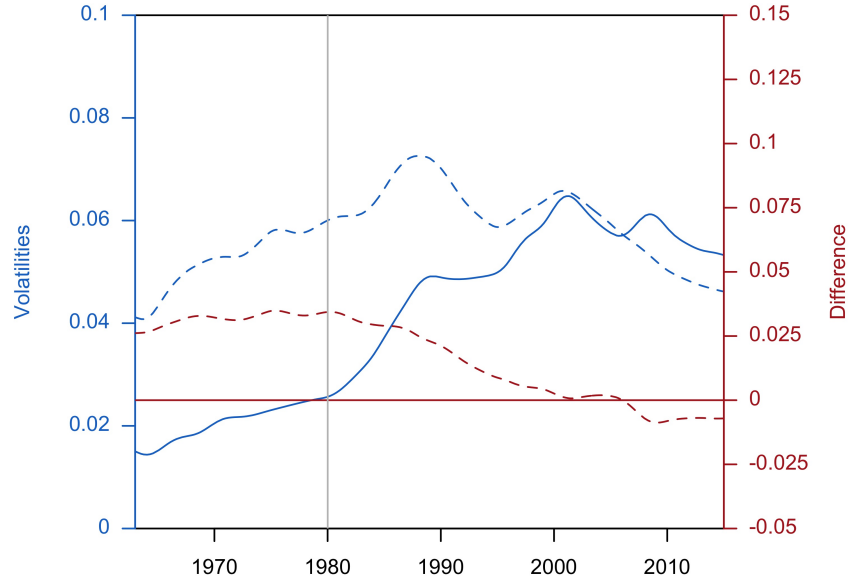


Figure 4: **Earnings and cash flows volatility.** The blue solid (dotted) line shows the cross-sectional, median earnings' (cash flows') volatility. We can see that the variability of the two accounting measures of profitability evolves differently. The volatility of cash flows has an upward trend until the end of the 1980s, from which point cash flows become as smooth or smoother than earnings by the end of the sample. The cross-sectional median earnings volatility increases threefold between the 1960s and the year 2000, decreasing slightly afterward. As a result of these changing dynamics, the cross-sectional median volatility of cash flows is equal to or smaller than the median volatility of accounting earnings after 2000. The red dotted line shows the difference between the volatilities of earnings and cash flows (see the scale on the right-hand side of the graph). That difference converges to zero from around 1980 to 2000. In summary, cash flows have been less volatile than earnings over the last ten years.

in these accounting measures of profitability are robust; the patterns in Figure 5 are comparable to those seen in Figure 4.

Next, we postulate that the reduction in earnings' pertinence to valuation documented in Subsection 4.1.1 may be associated with changes in the relationship between the smoothness of earnings and cash flows. This association is also consistent with previous evidence concerning the relationship between smoothness and predictability (Dichev and Tang, 2009); furthermore, our argument is in line with extensive evidence in the

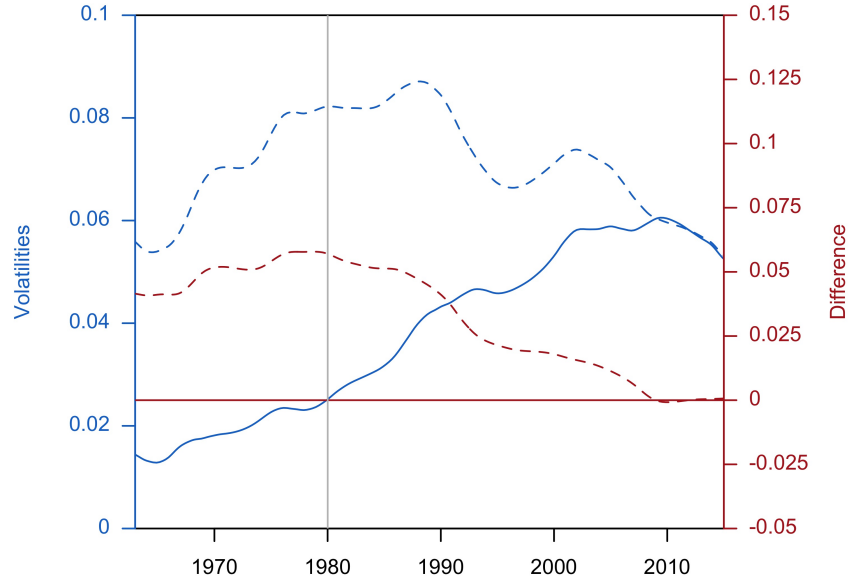


Figure 5: **Volatility of annual changes in earnings and cash flows.** The blue solid (dotted) line shows the cross-sectional, median volatility of annual changes in the total amount of earnings (cash flows). We can see that the variability of the two accounting measures of profitability evolves differently. The volatility of cash flows has an upward trend until the end of the 1980s, from which point cash flows become as smooth or smoother than earnings by the end of the sample. The cross-sectional median earnings volatility increases threefold between the 1960s and 2008, decreasing slightly afterward. As a result of these changing dynamics, the cross-sectional median volatility of cash flows is equal to or smaller than the median volatility of accounting earnings after 2008. The red dotted line shows the difference between the volatilities of earnings and cash flows (see the scale on the right-hand side of the graph). That difference converges to zero from around 1980 to 2008. These patterns are similar to those seen in Figure 4.

literature regarding earnings and cash flow forecasting.

Recent accounting research shows that cross-sectional OLS regressions provide relatively well-behaved forecasts for both earnings and cash flows (Hou et al., 2012; Gerakos and Gramacy, 2013; Nallareddy et al., 2020).

Specifically, Gerakos and Gramacy (2013) examine regression-based, out-of-sample net income forecasts using both linear and nonlinear methods. In their paper, they

note that: “Overall, we find that forecasts generated using ordinary least squares and lagged net income are broadly more accurate [for predicting net income]” (Gerakos and Gramacy, 2013, p. 1). They also argue that: “Moreover, at a one year horizon, the random walk model performs as well as modern sophisticated methods that use larger predictor sets. This finding echoes an old result that, given recent applications of forecasts in the literature, may have been forgotten” (Gerakos and Gramacy, 2013, p. 1). We can summarize their findings in their own words: “Random walk and AR(1) are hard to beat” (Gerakos and Gramacy, 2013, p. 3).

Brown (1993) states that: “Thus, while the random walk with drift, IMA(1,1) and AR(1) models are reasonable characterizations of the annual earnings time-series process, they can readily be improved on once additional information beyond past annual earnings is allowed for in the conditioning information set” (Brown, 1993, p. 299).

Finger (1994) performs an out-of-sample forecasting comparison between the random walk and predictions from an AR(2) model. She concludes that: “Out-of-sample forecasts show that random walk models outperform individually estimated earnings models for one-year but not for four- or eight-year horizons” (Finger, 1994, p. 220); for four- or eight-year horizons, the autoregressive model is slightly better. Furthermore, she points out that: “Earnings, used alone and with cash flow, are a significant predictor of cash flow for the majority of firms” (Finger, 1994, p. 220). Also: “Cash flow is a better short-term predictor of cash flow than are earnings, both in and out of sample, and the two are approximately equivalent long-term” (Finger, 1994, p. 220).

In summary, the literature seems to agree that random walks or simple linear models up to a few lags can provide reasonable forecasts for both earnings and cash flows.

When the underlying model is a random walk, the prediction accuracy is a function of period-to-period changes. The prediction accuracy of other linear time-series models is a function of the variability of the process itself; for example, for an autoregressive process, the one-step-ahead mean squared error is a multiple of the process variance.¹⁸ Therefore, higher earnings or cash flow volatility translates into less precise expectations of future income. Everything else constant, this should also imply that these accounting measures of economic performance have a lower pertinence to the valuation process.

The increase in the volatility of accounting earnings over the last two decades should reduce an investor's relative ability to project a firm's future performance using the current level of net income (Barton et al., 2010). As a consequence, earnings should become less pertinent to price formation when compared to cash flows. We can draw the same conclusions when assuming that the underlying process is a random walk.¹⁹

To investigate this argument, we define a relative measure of variability as the ratio

¹⁸For example, let Y_t satisfy a locally stationary AR(2) process, such that:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \quad (4.5)$$

where ε_t in (4.5) is white noise. This also implies that moments of Y_t do not change between $t - 2$ and t .

The corresponding one-step ahead mean squared error is equal to:

$$\mathbb{E}(Y_{t+1} - \mathbb{E}_t(Y_{t+1}))^2 = \left(1 - \left(\frac{1 + \phi_2}{1 - \phi_2}\right)\phi_1^2 + \phi_2^2\right)\sigma_Y^2, \quad (4.6)$$

where σ_Y^2 is the variance of the process Y_t . The expression in (4.6) reduces to:

$$\mathbb{E}(Y_{t+1} - \mathbb{E}_t(Y_{t+1}))^2 = (1 - \phi_1^2)\sigma_Y^2, \quad (4.7)$$

when the underlying model is an autoregressive of order one.

For an arbitrary autoregressive order p , the one-step ahead mean squared error of $Y_t \sim \text{AR}(p)$ is still proportional to the process variance σ_Y^2 .

¹⁹In this instance, the h -step-ahead mean squared error is equal to $h^2\sigma_{\Delta Y}^2$; that is, it is a multiple of the variance of the temporal changes.

between the volatilities of accounting earnings and operating cash flows, as follows:

$$\delta\sigma_{i,t} = \frac{\sigma(\text{EARN})_{i,t}}{\sigma(\text{CFO})_{i,t}}. \quad (4.8)$$

We also document a significant association between the size of this measure and Rel-EFP, the accounting earnings' pertinence to valuation relative to that of cash flows. Results from a series of Kolmogorov-Smirnov tests suggest that low (high) values of the ratio of earnings to cash flows volatility are associated with a higher (lower) earnings' pertinence to the valuation process compared to that of cash flows.

Figure 6 shows the two Kolmogorov-Smirnov test statistics \overline{KS}_t and \underline{KS}_t in (2.17) and (2.18), respectively, as functions of the ratio $\delta\sigma$ in (4.8). The values on the x-axis denote decile groups in that ratio; these range from 1 (earnings volatility is very low compared to cash flows volatility) up to 10 (earnings volatility is very high compared to cash flows volatility). The blue (red) solid line denotes the maximum \overline{KS}_t (minimum \underline{KS}_t). The shaded grey area marks the corresponding acceptance region of a one-sided test, at a 95% level. For the test statistic \overline{KS}_t (\underline{KS}_t), the acceptance region is the interval between zero and the corresponding critical value, the blue (red) dotted line.

Figure 6 presents the results of a large number of hypothesis tests, one for each decile group of the ratio $\delta\sigma$. A value of \overline{KS}_t (\underline{KS}_t) above (below) the critical value is against the null hypothesis that earnings and operating cash flows have equal pertinence to valuation, supporting an alternative: that earnings (cash flows) are more pertinent than cash flows (earnings). We do not reject the null hypothesis mentioned above when the two statistics do not exceed the corresponding critical value. The graph suggests that, for the seventh to tenth decile groups of the ratio, cash flows are as pertinent

as earnings or more. These decile groups of the variance ratio correspond to firms for which $\delta\sigma$ is higher than 0.86, i.e., those firms for which the volatility of cash flows does not exceed 115% of earnings volatility.²⁰ The graph also suggests that operating cash flows are strictly more pertinent than accounting earnings for the eighth to tenth decile groups, i.e., when earnings fluctuate more than cash flows, or, in other words, when the variance ratio $\delta\sigma$ is above 1.00.

Next, we investigate what happens to the proportion of firms for which the cash flows are at least as pertinent as earnings. **Figure 7** shows the evolution of the percentage of firms for which the ratio between the volatilities of earnings and cash flow belongs to given deciles. We investigate deciles for which cash flows are at least as pertinent (i.e., the seventh to tenth deciles of $\delta\sigma$, shown in black) or strictly more (i.e., the eighth to tenth deciles of $\delta\sigma$, shown in blue) than earnings.

The percentage of firms for which $\delta\sigma$ belongs to these decile groups increases sharply after 1980, possibly explaining the dynamics of the two Kolmogorov-Smirnov test statistics in **Figure 6**. The percentage of firms for which highly volatile earnings translate into a low earnings' pertinence compared to that of cash flows rises from 10% in the 1960s to around 65% in 2016. For two-thirds of the firms in the most recent cross-sections, cash flows are at least as pertinent as earnings to valuation. For around half of the firms in the most recent cross-sections, cash flows are strictly more pertinent than earnings.

²⁰0.86⁻¹ is rounded to 1.15.

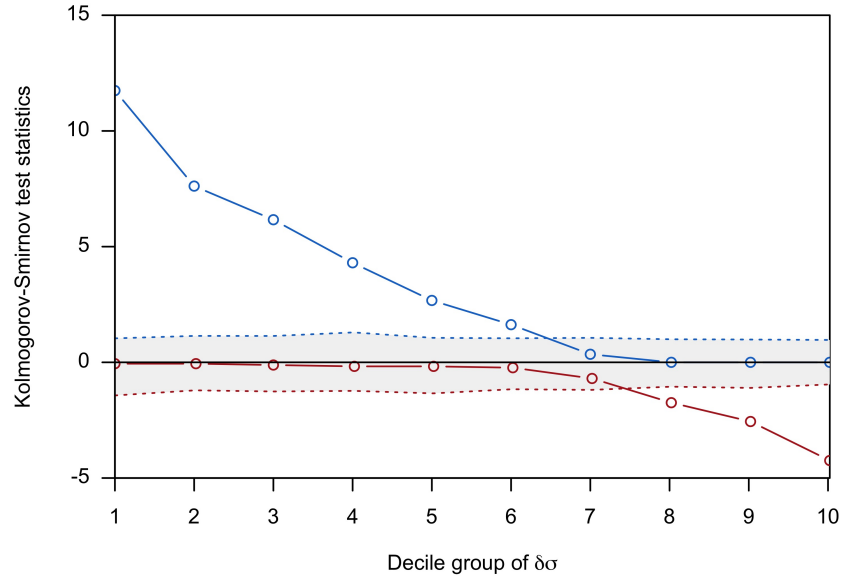


Figure 6: **Earnings’ pertinence relative to that of cash flows.** The graph shows the two Kolmogorov-Smirnov test statistics \overline{KS}_t and KS_t in (2.17) and (2.18), respectively, as functions of the ratio $\delta\sigma$ in (4.8). The values on the x-axis denote decile groups in that ratio; these range from 1 (earnings volatility is very low compared to cash flows volatility) up to 10 (earnings volatility is very high compared to cash flows volatility). The blue (red) solid line denotes the maximum \overline{KS}_t (minimum KS_t). The shaded grey area marks the corresponding acceptance region of a one-sided test, at a 95% level. For the test statistic \overline{KS}_t (KS_t), the acceptance region is the interval between zero and the corresponding critical value, the blue (red) dotted line. The figure presents the results of a large number of hypothesis tests, one for each decile group of the ratio $\delta\sigma$. A value of \overline{KS}_t (KS_t) above (below) the critical value is against the null hypothesis that earnings and operating cash flows have equal pertinence to valuation, supporting an alternative: that earnings (cash flows) are more pertinent than cash flows (earnings). We do not reject the null hypothesis mentioned above when the two statistics do not exceed the corresponding critical value. The graph suggests that, for the seventh to tenth decile groups of the ratio, cash flows are as pertinent as earnings or more. These decile groups of the variance ratio correspond to firms for which $\delta\sigma$ is higher than 0.86, i.e., those firms for which the volatility of cash flows does not exceed 115% of earnings volatility.²⁰ The graph also suggests that operating cash flows are strictly more pertinent than accounting earnings for the eighth to tenth decile groups, i.e., when earnings fluctuate more than cash flows, or, in other words, when the variance ratio $\delta\sigma$ is above 1.00. Results from a series of Kolmogorov-Smirnov tests suggest that low (high) values of the ratio of earnings to cash flows volatility are associated with a higher (lower) earnings’ pertinence to the valuation process compared to that of cash flows.

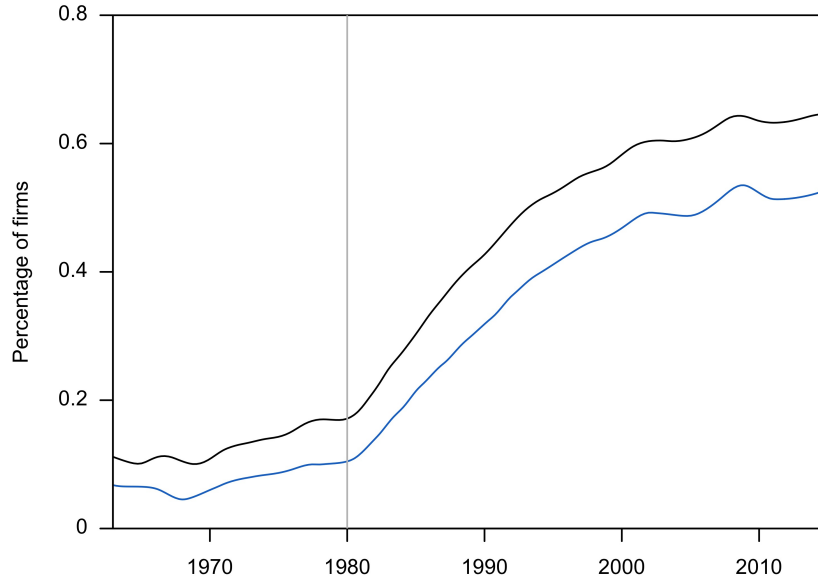


Figure 7: **Percentage of firms for which highly volatile earnings translate into a low earnings' pertinence relative to that of cash flows.** We divide the empirical distribution of the ratio $\delta\sigma$ in (4.8) into ten decile intervals. These intervals range from 1 (earnings volatility is very low compared to cash flows volatility) up to 10 (earnings volatility is very high compared to cash flows volatility). Results from Figure 6 suggest that the seventh to tenth deciles correspond to firms for which cash flows are at least as pertinent as earnings, while the eight to tenth deciles relate to firms for which cash flows are strictly more pertinent than earnings. The black (blue) line in Figure 7 corresponds to firms in the seventh to tenth (eighth to tenth) deciles. The percentage of firms for which $\delta\sigma$ belongs to these decile groups increases sharply after 1980, possibly explaining the dynamics of the two Kolmogorov-Smirnov test statistics in Figure 6. The percentage of firms for which highly volatile earnings translate into a low earnings' pertinence compared to that of cash flows rises from 10% in the 1960s to around 65% in 2016. For two-thirds of the firms in the most recent cross-sections, cash flows are at least as pertinent as earnings to valuation. For around half of the firms in the most recent cross-sections, cash flows are strictly more pertinent than earnings.

4.1.4 A possible explanation for the increase in the earnings volatility

Several economic and accounting developments could explain why earnings are becoming more volatile than cash flows. As we have discussed in Section 1.2, the existing literature leads us to three possible explanations: (i) the growing prominence of firms that exhibit a high intangible intensity; (ii) an increase in accruals volatility; and (iii) a reduction

in the smoothing role of accruals.

We then consider four different characteristics of a firm. We measure accruals smoothing using the correlation coefficient between accruals and cash flows (ρ). We proxy changes in the earnings' smoothness compared to that of cash flows using the ratio between the volatility of these accounting measures of profitability ($\delta\sigma$ in (4.8)). To capture the growing prominence of firms that exhibit a high intangible intensity, we also measure the ratio between investments in SG&A and total expenses (this ratio is denoted as SGAI). Finally, we proxy accruals volatility using the corresponding sample statistic ($\sigma(\text{ACC})$).

First, we note that intangible intensity is increasing over time. The first graph in **Figure 8** documents a 125% increase in cross-sectional median SGAI. The median value of SGAI has risen from about 0.12 at the beginning of the 1960s, to 0.27 by around the year 2000. As we have already explained, we can expect that accounting earnings will fluctuate more when investments in intangible assets are substantial.

Second, we observe that accruals now fluctuate more than they did before. The second graph in **Figure 8** shows the evolution of cross-sectional median accruals volatility. The cross-sectional median of $\sigma(\text{ACC})$ increased from 4% at the beginning of the sample to about 6% by around the year 2000. As we have anticipated, we can expect a positive association between accruals and earnings volatility.

Lastly, we investigate how the smoothing role played by accruals is changing over time. It is known that accruals can mitigate earnings variability when they are negatively correlated to cash flows. The last graph in **Figure 8** shows a steady increase in the correlation between accruals and cash flows over the previous two decades. We observe

an increase in the correlation: from around -0.93 at the beginning of the sample to -0.55 after the year 2000. As we have already discussed, we can expect that the less negative values of that correlation coefficient may justify the increase in the earnings volatility. As a result, the positive difference between the cash volatility and the earnings volatility should diminish and may become negative.

4.1.5 Static analysis of the idiosyncrasy of earnings' relative pertinence to valuation

We have argued that accounting earnings' pertinence to the valuation process is decreasing compared to operating cash flows, hypothesizing that this decrease could be associated with changes in accruals volatility, the intangible intensity, or the correlation between accruals and cash flows. In this subsection, we investigate the most likely reason behind this phenomenon.

We proceed as follows: first, we perform a static analysis by estimating a regression model using all the firm-year observations in our sample. We then perform a dynamic, cross-sectional investigation (we describe the results of this second analysis in the next subsection). In both cases, and from now on, we no longer quantify relative pertinence via the two Kolmogorov-Smirnov test statistics. Instead, we compute a firm- and year-specific measure of the association between stock prices and accounting measures of economic performance, such as earnings and cash flows, because this is the best approach to use with analyses that use a multivariate regression model.

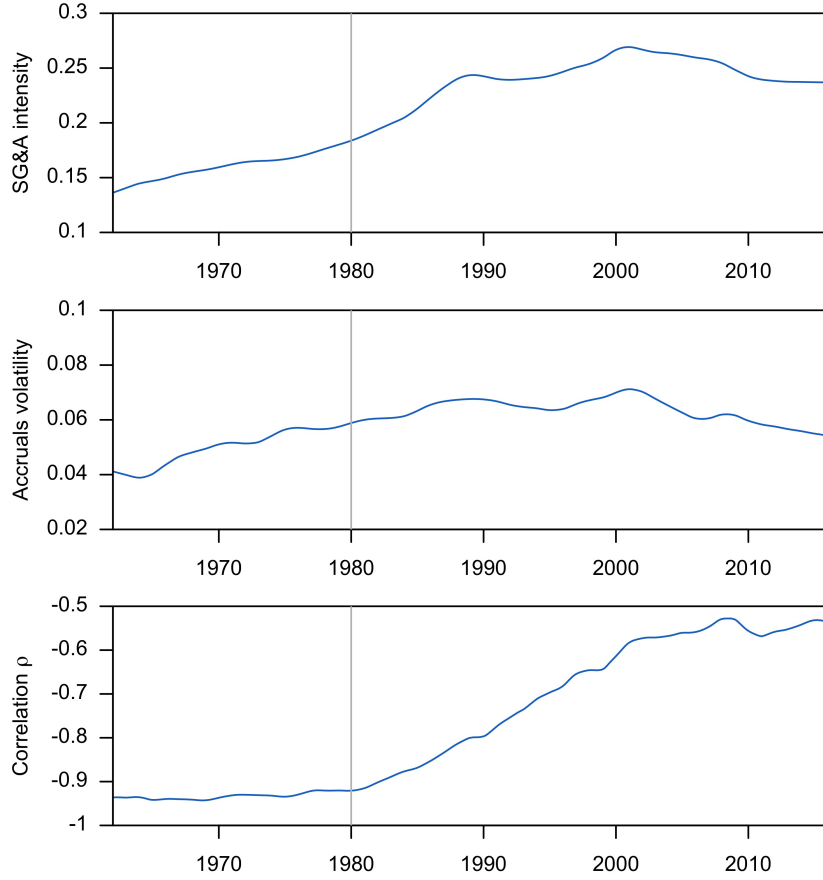


Figure 8: **Cross-sectional median intangible intensity, accruals volatility, and the correlation between accruals and cash flows.** The figure shows the time series of the cross-sectional median intangible intensity (SGAI, top graph), accruals volatility ($\sigma(\text{ACC})$, middle graph), and the correlation between accruals and cash flows (ρ , bottom graph). The top graph documents a 125% increase in cross-sectional median SGAI over the last four decades. The median value of SGAI has risen from about 0.12 at the beginning of the 1960s, to 0.27 by around the year 2000. The middle graph shows the evolution of cross-sectional median accruals volatility ($\sigma(\text{ACC})$). Median volatility increased from 4% at the beginning of the sample to around 6% by around the year 2000. The bottom graph shows the time evolution of the cross-sectional median correlation between accruals and cash flows (ρ). We document an increase from around -0.93 at the beginning of the sample to -0.55 after the year 2000.

The measure of relative pertinence we use is as given in (2.19), as follows:

$$\text{Rel-EFP}_{i,t} = \ln |P_{i,t} - \hat{P}_{i,t}^{\text{CFO}}| - \ln |P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}|, \quad (2.19)$$

where all of the previous definitions given in Subsection 2.1.6 apply.

In our first static analysis, we document a statistically significant association between earnings' pertinence to the valuation process relative to that of cash flows and the smoothing role played by accruals. However, we do not find any significant association between our relative measure of pertinence and either the intangible intensity or accruals volatility for firms with comparable accruals smoothing.

First, we estimate four univariate regression models as in (2.20) and (2.22). In each one, we regress our relative measure of pertinence on only one of the four accounting characteristics mentioned above. These four univariate regression models are as follows:

$$\text{Volatility ratio : Rel-EFP} = \gamma_0^{\delta\sigma,U} + D^{\delta\sigma}\gamma^{\delta\sigma,U} + \varepsilon^{\delta\sigma,U}, \quad (4.9)$$

$$\text{Intangible intensity : Rel-EFP} = \gamma_0^{\text{SGAI},U} + D^{\text{SGAI}}\gamma^{\text{SGAI},U} + \varepsilon^{\text{SGAI},U}, \quad (4.10)$$

$$\text{Accruals volatility : Rel-EFP} = \gamma_0^{\sigma(\text{ACC}),U} + D^{\sigma(\text{ACC})}\gamma^{\sigma(\text{ACC}),U} + \varepsilon^{\sigma(\text{ACC}),U}, \quad (4.11)$$

$$\text{Accruals smoothing : Rel-EFP} = \gamma_0^{\rho,U} + D^{\rho}\gamma^{\rho,U} + \varepsilon^{\rho,U}, \quad (4.12)$$

where all the definitions seen in Subsection 2.1.6 apply. In that subsection, we explained how we interpret the regression coefficients $\hat{\gamma}^U$. The superscript U found in (4.9) to (4.12) means these coefficients come out of a regression model in which we do not hold constant other competing characteristics (U stands for uncontrolled). Because of this, we should note that these estimated coefficients give a measure of association that might be subject to a bias due to the omission of relevant and correlated regressors.

Let us also recall that $\hat{\gamma}^U$ is a vector of nine regression coefficients. In the following pictures, we show three curves for each characteristic. A solid blue line traces the nine

coefficients that correspond to the binary regressors D in (4.9) to (4.12). The other two blue dotted lines indicate the lower and the upper bounds of the corresponding nine two-sided 95% confidence intervals.²¹

Figure 9 features four graphs that contain nine blue points each. These points represent regression coefficients for each of the four characteristics under investigation, plus a two-sided 95% confidence interval. Each point denotes a difference in the average relative earnings' pertinence of two groups of firms. While the first group is made up of firms for which the given characteristic takes a value within the decile group indicated on the x-axis, the second group is made up of firms for which the value is within a reference group. The reference in the top two graphs is the tenth decile group, while the one in the bottom two graphs is the first decile group. The top graphs correspond to the correlation between accruals and cash flows and the ratio between earnings and cash flows' volatility, while the bottom ones correspond to accruals volatility and intangible intensity.

For example, let us consider the estimated coefficients that correspond to the levels of the correlation between accruals and cash flows (the top-left graph). The first point is at about 0.25, which suggests that firms that have the most negative correlation between accruals and cash flows (that is, the first decile group of ρ) have, on average, a relative pertinence greater than 0.25 compared to those firms with the most positive correlation (that is, the tenth decile group). This first graph suggests that the difference between these two groups of firms is also significant at 95% because the corresponding confidence interval does not include zero. If we look at the remaining decile groups,

²¹To compute the confidence intervals, we use a heteroskedasticity-consistent (HC) estimator of the covariance matrix of the regression coefficients.

the graph suggests that earnings' relative pertinence to the valuation process might be associated with accruals smoothing.

Similarly, the last blue point on the bottom-right graph of Figure 9 is at about -0.10. That means that, on average, those firms that exhibit the highest intangible intensity (that is, the tenth decile group of SGAI) have an earnings' pertinence relative to that of cash flows that is about 0.10 points lower than that of those firms that have the lowest fraction of intangible assets (that is, the first decile group). This second graph suggests that the difference between these two groups of firms is also significant at 95% because the corresponding confidence interval does not include zero.

The top-right graph of Figure 9 confirms the analysis based on the Kolmogorov-Smirnov test statistics shown in the previous Figure 6. We can see a significant and monotonically decreasing association between the ratio $\delta\sigma$ and the year- and firm-specific measure of relative pertinence (Rel-EFP in (2.19)). This graph suggests that, when earnings are less volatile than cash flows, earnings yield valuations that are significantly more accurate than those informed by the firm's cash flow level.

Looking at the whole of Figure 9, we can hypothesize a significant association between our proxy for earnings' pertinence to valuation relative to that of cash flows and the three possible explanations under investigation, although the degree of association among them varies. In this preliminary and univariate analysis, lower levels of either the accrual volatility, the SG&A intensity, and the correlation between accruals and cash flows seem to correspond to earnings with more pertinence than cash flows. We also can see that the correlation between accruals and cash flows has the strongest association with the measure of relative pertinence to valuation in (2.19).

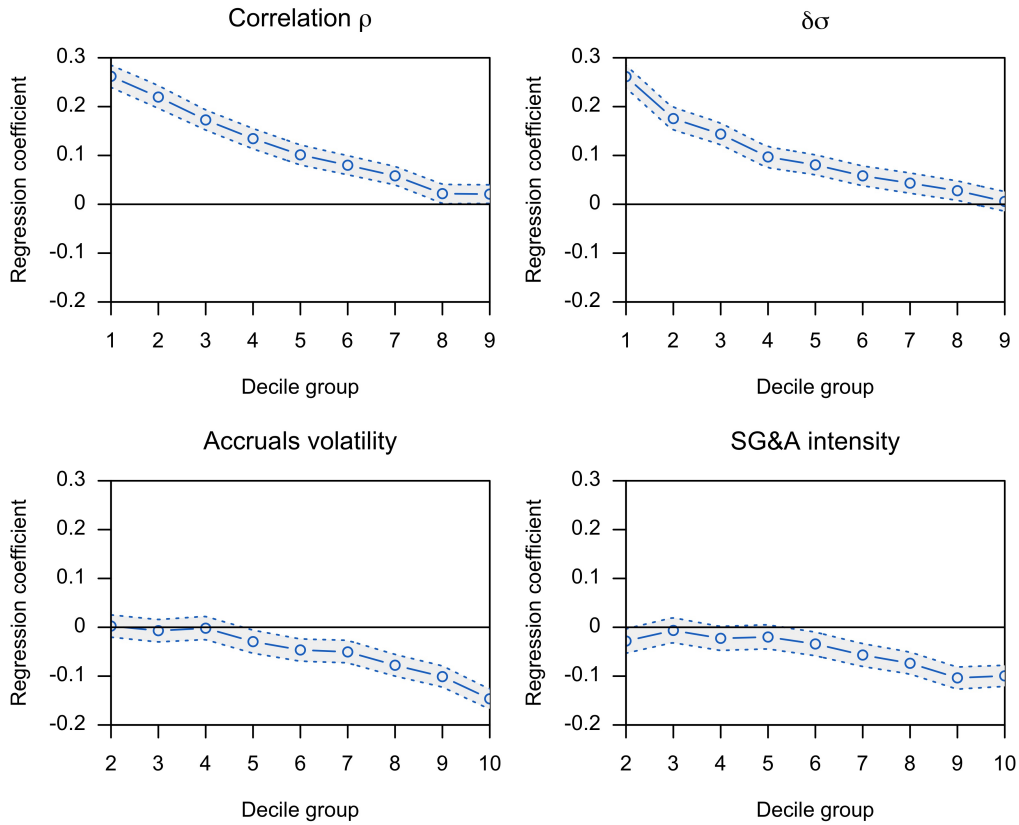


Figure 9: **Earnings' pertinence relative to that of cash flows; coefficients of the univariate regression models with binary variables.** The four graphs contain nine blue points each, representing regression coefficients for each of the four characteristics under investigation, plus a two-sided 95% confidence interval. Each point denotes a difference in the average relative earnings' pertinence of two groups of firms. While the first group is made up of firms for which the given characteristic takes a value within the decile group indicated on the x-axis, the second group is made up of firms for which the value is within a reference group. The reference in the top two graphs is the tenth decile group, while the one in the bottom two graphs is the first decile group. Looking at the whole of Figure 9, we can hypothesize a significant association between our proxy for earnings' pertinence to valuation relative to that of cash flows and the three possible explanations under investigation, although the degree of association among them varies. In this preliminary and univariate analysis, lower levels of either the accrual volatility, the SG&A intensity and the correlation between accruals and cash flows seem to correspond to earnings with more pertinence than cash flows. We also can see that the correlation between accruals and cash flows has the strongest association with the measure of relative pertinence to valuation in (2.19).

As we explained in this subsection, the univariate regression models (4.9) to (4.12) might lead to misleading inferences because of the omitted variable bias. This is as in a plain-vanilla linear regression model (that is, one in which we do not hold constant competing characteristics when estimating these univariate models).

To mitigate this issue, we also estimate a multivariate regression model, as in (2.21) and (2.23), regressing our relative measure of pertinence to the valuation process on the intangible intensity, the accruals volatility, and the correlation between accruals and cash flows. The multivariate regression model we estimate is:

$$\text{Rel-EFP} = \gamma_0^C + D^{\text{SGAI}}\gamma^{\text{SGAI},C} + D^{\sigma(\text{ACC})}\gamma^{\sigma(\text{ACC}),C} + D^\rho\gamma^{\rho,C} + \varepsilon^C, \quad (4.13)$$

where all the definitions seen in Subsection 2.1.6 apply. The superscript C in the equation above means that we estimate regression coefficients by controlling for competing explanations and interpreting them as more consistent estimates of the idiosyncratic association between each of these accounting characteristics and the relative measure of pertinence to valuation.²²

The results in **Figure 10** suggest that the association between earnings' pertinence relative to that of cash flows and both accrual volatility and SG&A intensity weakens once we control for the level of accruals smoothing, which can be measured by the correlation between accruals and cash flows.

On the left-hand side of **Figure 10** are regression coefficients that correspond to the

²²We should also explain why we cannot (directly) include the ratio $\delta\sigma$ from (4.8) in the previous multivariate regression model (4.13). When we hold the accruals volatility constant, the correlation between accruals and cash flows implies that $\delta\sigma$ is also held constant. This is why including that ratio in the equation would offer only a redundant source of variability.

intangible intensity, $\hat{\gamma}^{\text{SGAI},C}$, while coefficients corresponding to accrual volatility, $\hat{\gamma}^{\sigma(\text{ACC}),C}$, are given on the right. These regression coefficients correspond to nine decile groups; each point denotes a difference in the average relative pertinence of earnings of two groups of firms. While the first group is made up of firms for which the firms' characteristic takes a value within the decile group indicated on the x-axis, the second group is made up of firms within the first decile group. In contrast to the previous analysis, we estimate the multivariate regression model in (4.13) by holding constant the remaining characteristics (i.e., the proxies for the competing explanations behind the decline in earnings' pertinence to valuation relative to that of cash flows, including the correlation between accruals and cash flows). Figure 10 suggests that, first, there is a weak relationship between earnings' relative pertinence and accruals volatility, and second, there is no significant association between earnings' relative pertinence and intangible intensity.

In contrast, the results given in **Figure 11** suggest that the association between our measure of earnings' pertinence to the valuation process relative to that of cash flows and the correlation between accruals and cash flows remains statistically significant when controlling for the levels of the other two competing characteristics. The graph shows the regression coefficients $\hat{\gamma}^{\rho,C}$ in (4.13) for nine decile groups, with each point denoting a difference in the average earnings' relative pertinence to valuation between two groups of firms. The first group is made up of firms for which the correlation between accruals and cash flows takes a value within the decile group indicated on the x-axis, while the second group is made up of firms for which the value of that correlation (ρ) is within the tenth decile group (i.e., ρ closest to 1.00). Contrary to the previous analysis, we hold constant the remaining firms' characteristics when estimating the

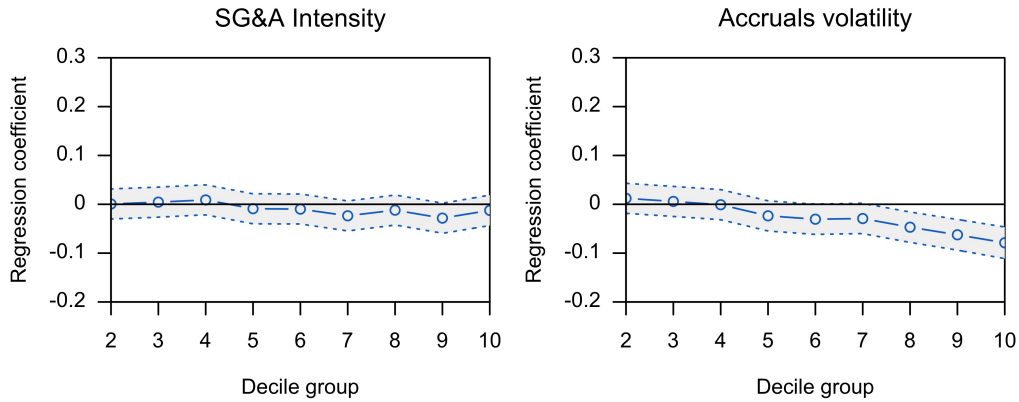


Figure 10: **Earnings’ pertinence relative to that of cash flows; the multivariate regression coefficients corresponding to the intangible intensity and accruals volatility.** On the left-hand side of the figure are regression coefficients that correspond to the intangible intensity, $\hat{\gamma}^{SG\&A,C}$, while coefficients corresponding to accrual volatility, $\hat{\gamma}^{\sigma(ACC),C}$, are given on the right. These regression coefficients correspond to nine decile groups; each point denotes a difference in the average relative pertinence of earnings of two groups of firms. While the first group is made up of firms for which the firms’ characteristic takes a value within the decile group indicated on the x-axis, the second group is made up of firms within the first decile group. In contrast to the previous analysis, we estimate the multivariate regression model in (4.13) by holding constant the remaining characteristics (i.e., the proxies for the competing explanations behind the decline in earnings’ pertinence to valuation relative to that of cash flows, including the correlation between accruals and cash flows). Figure 10 suggests that, first, there is a weak relationship between earnings’ relative pertinence and accruals volatility, and second, there is no significant association between earnings’ relative pertinence and intangible intensity.

multivariate regression model (4.13).

Comparing Figure 11 with the top-right graph in Figure 9, we notice that the association between the relative measure of pertinence and the correlation between accruals and cash flows is consistent and statistically significant. However, a comparison between Figures 9 and 10 suggests that the association between the earnings’ relative pertinence and the accrual volatility or intangible intensity might be biased.

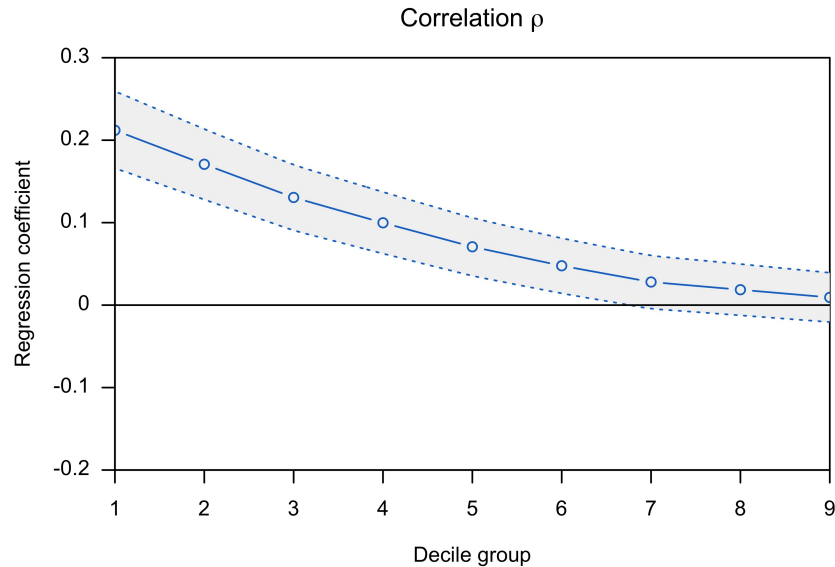


Figure 11: **Earnings’ pertinence relative to that of cash flows; the multivariate regression coefficients corresponding to the correlation between accruals and cash flows.** The graph shows the regression coefficients $\hat{\gamma}^{\rho,C}$ in (4.13) for nine decile groups, with each point denoting a difference in the average earnings’ relative pertinence to valuation between two groups of firms. The first group is made up of firms for which the correlation between accruals and cash flows takes a value within the decile group indicated on the x-axis, while the second group is made up of firms for which the value of that correlation (ρ) is within the tenth decile group (i.e., closer to 1.00). Contrary to the previous analysis, we hold constant the remaining firms’ characteristics when estimating the multivariate regression model (4.13). Comparing Figure 11 with the top-right graph in Figure 9, we notice that the association between the relative measure of pertinence and the correlation between accruals and cash flows is consistent and statistically significant. However, a comparison between Figures 9 and 10 suggests that the association between the earnings’ relative pertinence and the accrual volatility or intangible intensity might be biased.

4.1.6 Dynamic analysis of the idiosyncrasy of earnings’ relative pertinence to valuation

The results of a dynamic investigation confirm the existence of a significant relationship between the smoothing role played by accruals and earnings’ relative pertinence to the valuation process.

First, we compute a cross-sectional median value of the relative measure of expectation-formation pertinence seen in (2.19). We can then use a simple quantile regression model to find that median, plus a confidence interval. The model is as follows:

$$\text{Rel-EFP}_{i,t} = \gamma_{0,t} + \varepsilon_{i,t}, \tag{2.24}$$

as in Subsection 2.1.6, and where all the previous definitions apply. As we have already explained, when including only an intercept, and when that intercept is found by minimizing the absolute, instead of the squared, regression errors, the coefficient $\hat{\gamma}_{0,t}$ represents the median pertinence among observations used to estimate the model.

In the following two graphs, we show an unconditional median of Rel-EFP, followed by two conditional medians as a function of the correlation between accruals and cash flows. To find the unconditional median of Rel-EFP, we estimate the quantile regression above using all the firms in a given cross-section. To find conditional medians as a function of accruals smoothing, we estimate the same quantile regression considering observations for which the correlation between accruals and cash flows takes a value that lies within a specific range.

The solid black line in **Figure 12** shows the unconditional, cross-sectional median of the measure of earnings' relative pertinence (Rel-EFP in (2.19)), plus a two-sided 99% confidence interval, indicated by the two black dotted curves. When statistically significant, values of the median above (below) zero suggest that earnings are more (less) pertinent, relative to cash flows' pertinence, for half of the firms in a given cross-section.

We observe a significant downward trend that sharpens after 1980, suggesting that earnings' pertinence to valuation is lessening compared to that of cash flows. Before

1990, the median relative pertinence of earnings is significantly positive, implying that higher valuation errors occur when using cash flows instead of earnings for half of the firms in each cross-section. In the 1990s, that quantity falls to around zero, and after 2000 we observe a significantly negative median relative pertinence, suggesting that higher valuation errors occur when using earnings compared to when using cash flows.

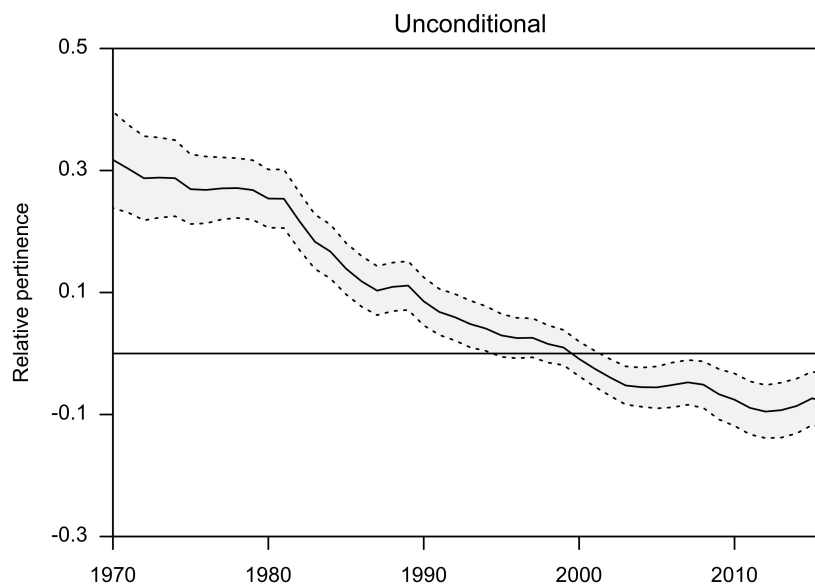


Figure 12: **Cross-sectional median earnings' pertinence relative to that of cash flows; unconditional median.** The solid black line shows the unconditional, cross-sectional median of the measure of earnings' relative pertinence (Rel-EFP in (2.19)), plus a two-sided 99% confidence interval, indicated by the two black dotted curves. When statistically significant, values of the median above (below) zero suggest that earnings are more (less) pertinent, relative to cash flows' pertinence, for half of the firms in a given cross-section. We observe a significant downward trend that sharpens after 1980, suggesting that earnings' pertinence to valuation is lessening compared to that of cash flows. Before 1990, the median relative pertinence of earnings is significantly positive, implying that higher valuation errors occur when using cash flows instead of earnings for half of the firms in each cross-section. In the 1990s, that quantity falls to around zero, and after 2000 we observe a significantly negative median relative pertinence, suggesting that higher valuation errors occur when using earnings compared to when using cash flows.

The solid blue (red) line in **Figure 13** shows the cross-sectional median of Rel-EFP, which is now conditional on the correlation between accruals and cash flows in the

range $\rho < -0.92$ (in the range $\rho > -0.58$). As we have already said in Subsection 2.1.7, -0.92 and -0.58 denote the first and second tertiles of the distribution of ρ , respectively. We have that $\rho < -0.92$ for one-third of the sample, i.e., for those firms for which the accruals smoothing can be substantial. For another third of the firms, we have that $\rho \in (-0.92, -0.58)$, denoting observations for which the accruals smoothing might be “average.” $\rho > -0.58$ for the last third of the sample, corresponding to those firms for which the accruals smoothing might be ineffective.

The two dotted, blue (red) curves in Figure 13 denote the corresponding two-sided 99% confidence interval. When statistically significant, values of the conditional median above (below) zero suggest that earnings are more (less) pertinent relative to cash flows for half of the firms in a given cross-section, firms for which the correlation ρ is also within the given range. The behavior differs depending on the aforementioned correlation coefficient. For those firms for which accruals smoothing might be weak, the red curve indicates a significant downward trend, with conclusions that are comparable to the unconditional analysis in the left-hand side graph. In comparison, the blue line indicates that earnings are persistently preferable to cash flows for those firms for which the correlation between accruals and cash flows is closest to -1.00 .

The results of this investigation reinforce the hypothesis that earnings’ relative pertinence strongly depends on accruals smoothing. This is because, after 1985, holding the smoothing role of accruals constant leads to opposite results, and the corresponding two confidence intervals are far from overlapping. Two non-overlapping, 99% confidence intervals bring strong statistical evidence in favor of a rejection of the null hypothesis, at 5%, of having the same conditional median between two groups of observations.²³

²³We are applying Bonferroni’s correction: if two 97.50% confidence intervals do not overlap, we

Furthermore, we can see that the earnings' relative pertinence is constant when we consider firms with the most negative and similar values for the correlation between accruals and cash flows.

Figure 14 suggests why the length of the two confidence intervals in **Figure 13** changes over time, showing the number of firms for which the correlation between accruals and cash flows, ρ , takes a value within a given range. For the blue line, we are counting firms for which $\rho < -0.92$, while for the red, we are counting firms for which $\rho > -0.58$.

When estimating the median of a random variable (like any given statistic of interest), the length of a confidence interval depends on the confidence level, the variability of the random variable, and the sample size. Everything else constant, a larger sample size should reduce the length of a confidence interval.

The blue line shows the number of firms for which the correlation is $\rho < -0.92$. This number is decreasing over time, which can be a reason why the length of the confidence interval in blue in **Figure 13** increases. Additionally, the number of firms for which $\rho > -0.58$ increases over time, which can be the reason why the confidence interval in red in the previous figure, **Figure 13**, shrinks.

We conclude this first broad section of results with a final remark about the definition and use of a firm- and year-specific measure of pertinence. [Granger and Newbold \(1974\)](#) note that, when we regress a non-stationary variable on another, the bias toward rejection of the hypothesis of no relationship between them is proportional to the sample size. We might observe a spurious, non-zero correlation between two non-stationary

could reject the null hypothesis, at 5%, of having the same conditional median between two groups of observations. In this case, we consider intervals at 99% to further reinforce our conclusions.

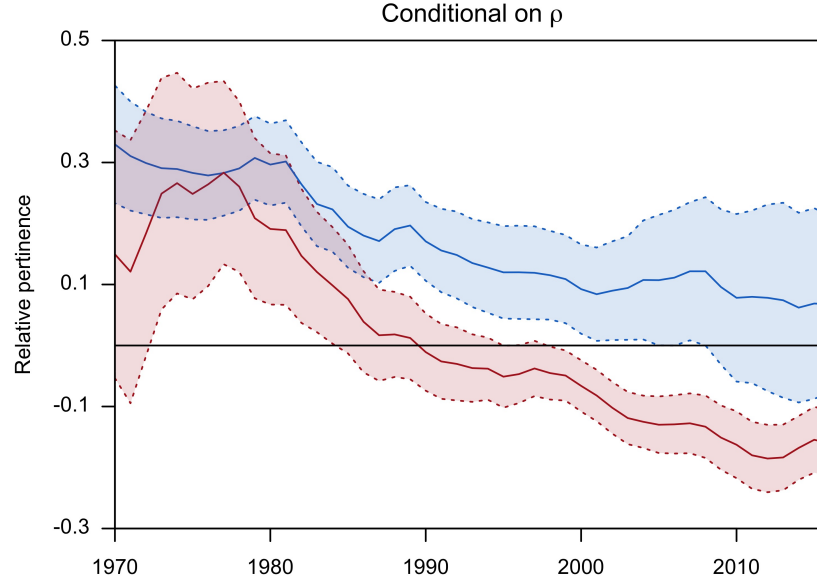


Figure 13: **Cross-sectional median earnings’ pertinence relative to that of cash flows; median conditional on the correlation between accruals and cash flows.** The solid blue (red) line in **Figure 13** shows the cross-sectional median of Rel-EFP, which is now conditional on the correlation between accruals and cash flows in the range $\rho < -0.92$ (in the range $\rho > -0.58$). As we have already said in Subsection 2.1.7, -0.92 and -0.58 denote the first and second tertiles of the distribution of ρ . The two dotted, blue (red) curves denote the corresponding two-sided 99% confidence interval. When statistically significant, values of the median above (below) zero suggest that earnings are more (less) pertinent relative to cash flows for half of the firms in a given cross-section, firms for which the correlation ρ is also within the given range. The behavior differs depending on the aforementioned correlation coefficient. For those firms for which accruals smoothing might be weak, the red curve indicates a significant downward trend, with conclusions that are comparable to the unconditional analysis in the left-hand side graph. In comparison, the blue line indicates that earnings are persistently preferable to cash flows for those firms for which the correlation between accruals and cash flows is closest to -1.00 . The results of this investigation reinforce the hypothesis that earnings’ relative pertinence strongly depends on accruals smoothing. This is because, after 1985, holding the smoothing role of accruals constant leads to opposite results, and the corresponding two confidence intervals are far from overlapping. Two non-overlapping, 99% confidence intervals bring strong statistical evidence in favor of a rejection of the null hypothesis, at 5%, of having the same conditional median between two groups of observations.²³ Furthermore, we can see that the earnings’ relative pertinence is constant when we consider firms with the most negative and similar values for the correlation between accruals and cash flows.

variables when, in fact, they are independent.

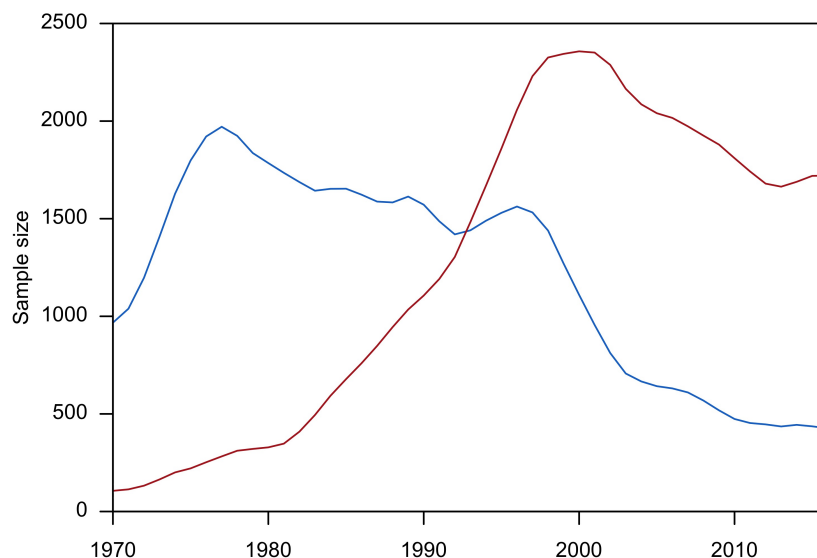


Figure 14: Sample size depending on the correlation between accruals and cash flows. The graph suggests why the length of the two confidence intervals in Figure 13 changes over time, showing the number of firms for which the correlation between accruals and cash flows, ρ , takes a value within a given range. For the blue line, we are counting firms for which $\rho < -0.92$, while for the red, we are counting firms for which $\rho > -0.58$. When estimating the median of a random variable (like any given statistic of interest), the length of a confidence interval depends on the confidence level, the variability of the random variable, and the sample size. Everything else constant, a larger sample size should reduce the length of a confidence interval. The blue line shows the number of firms for which the correlation is $\rho < -0.92$. This number is decreasing over time, which can be a reason why the length of the confidence interval in blue in Figure 13 increases. Additionally, the number of firms for which $\rho > -0.58$ increases over time, which can be the reason why the confidence interval in red in the previous figure, Figure 13, shrinks.

We could not regress a cross-sectionally constant measure of pertinence, like the two Kolmogorov-Smirnov test statistics in Figure 3, against the correlation coefficient in the bottom graph of Figure 8. In these types of time-series regression models, we should only take stationary variables into account. This is why, in our second investigation, we considered the Rel-EFP in (2.19) to study the association between earnings' pertinence to valuation and possible accounting explanations.

4.2 The relationship between earnings' expectation-formation pertinence and accounting quality

In this last section of results, we show the empirical findings concerning the relationship between prices and accounting quality. These findings are based on the yearly coefficients we have obtained from the multivariate regression model in (2.33), based on a rolling window of 15 consecutive years.

The regression model we consider is as follows:

$$\ln |P_{i,t} - \hat{P}_{i,t}^{\text{EARN}}| = \text{FE}_{i,t} + b_t^P \ln P_{i,t} + b_t^\pi \lambda_{i,t}^\pi + b_t^{\text{CFO}} \lambda_{i,t}^{\text{CFO}} + b_t^{\text{ACC}} \lambda_{i,t}^{\text{ACC}} + \varepsilon_{i,t}, \quad (2.33)$$

to which all the definitions from Subsection 2.2.1 apply. On the left-hand side of the equality, the dependent variable is the logarithm of an absolute valuation error, a measure of earnings' pertinence to the valuation process. On the right, the regressors are the logarithm of the stock price (P), plus proxies for the economic performance uncertainty (that is, $\lambda^\pi = \ln \hat{\sigma}_\pi^2$), the size of timing errors ($\lambda^{\text{CFO}} = \ln \hat{\sigma}_W^2$), and accounting quality ($\lambda^{\text{ACC}} = \ln \hat{\sigma}_W^2 - \ln \hat{\sigma}_V^2$). We have found these quantities via a maximum likelihood estimation of a Gaussian state-space model, a statistical tool we use to model the system of equations proposed by Nikolaev (2018), as discussed in Subsection 2.2.2. The term $\text{FE}_{i,t}$ denotes both firm- and time-fixed effects. Standard errors are also clustered by firm and year.

The solid blue line in the first graph in **Figure 15** shows the estimated regression coefficient that corresponds to the economic performance volatility, \hat{b}_t^π , over time, plus a 95% confidence interval, denoted by the dotted blue lines. Before 2005, we find a

significant association between this volatility and earnings' pertinence to the valuation process, at a 5% level; however, this seems to disappear in later years. The other solid blue line in the second graph shows the yearly regression coefficient \hat{b}_t^{CFO} , which corresponds to the proxy for operating risk, plus a 95% confidence interval, denoted by the dotted blue lines. Results suggest a significant association between this measure of timing errors and earnings' pertinence between 1990 and 2010, at a 5% level. On average, higher timings errors seem to be associated with a lower earnings' pertinence to the valuation process, holding the other variables constant.

The solid blue line in **Figure 16** shows \hat{b}_t^{ACC} , the yearly regression coefficient corresponding to the proxy for accounting quality, plus a 95% confidence interval, denoted by the dotted blue lines. We observe a stable and persistent association between this measure of accounting quality and earnings' pertinence over the whole period under investigation, also significant at a 5% level. These findings indicate that earnings' pertinence to the valuation process benefits from accruals accounting. On average, earnings seem to be more pertinent to the valuation process for those firms that exhibit a higher accounting quality, keeping everything else constant. That is because the estimated coefficient \hat{b}_t^{ACC} is significant and negative, suggesting that we could expect the valuation errors using earnings to decrease once the earnings quality increases, once again, on average, and holding constant the other regressors.

4.3 Robustness checks

To estimate the year-specific nonlinear regression model in (2.6), we changed the industry definition. We consider the first two and three SIC digits and the eight-

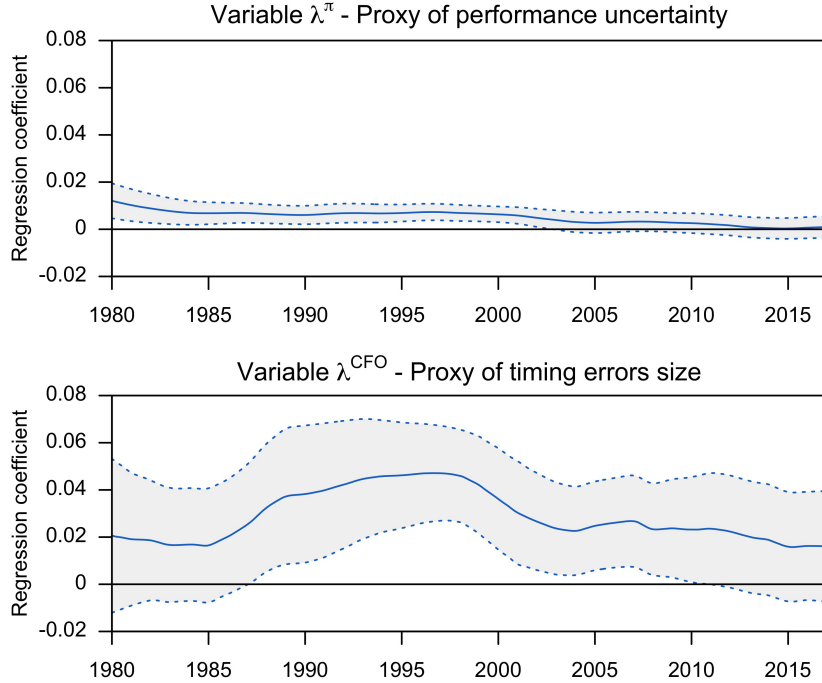


Figure 15: **Accounting quality and earnings' pertinence to valuation; cross-sectional regression coefficients corresponding to the proxies for economic performance uncertainty and the size of timing errors.** We estimate the following regression model to measure the association between prices and accounting quality. The equation is:

$$\ln |P_{i,t} - \widehat{P}_{i,t}^{\text{EARN}}| = \text{FE}_{i,t} + b_t^P \ln P_{i,t} + b_t^\pi \lambda_{i,t}^\pi + b_t^{\text{CFO}} \lambda_{i,t}^{\text{CFO}} + b_t^{\text{ACC}} \lambda_{i,t}^{\text{ACC}} + \varepsilon_{i,t}, \quad (2.33)$$

where all the definitions of Section 2.2 apply. On the left side of the equality, the dependent variable is the logarithm of an absolute valuation error, a measure of earnings' pertinence to the valuation process. On the right, the regressors are the logarithm of the stock price (P), plus proxies for the economic performance uncertainty (λ^π), size of timing errors (λ^{CFO}), and accounting quality (λ^{ACC}). The term $\text{FE}_{i,t}$ denotes both firm- and time-fixed effects. Standard errors are also clustered by firm and year. The solid blue line in the first graph shows the estimated regression coefficient corresponding to the economic performance volatility, b_t^π , over time, plus a 95% confidence interval, denoted by the dotted blue lines. Before 2005, we find a significant association between this volatility and earnings' pertinence to the valuation process, which seems to disappear in later years. The other solid blue line in the second graph shows the yearly regression coefficient b_t^{CFO} , corresponding to the proxy for operating risk, plus a 95% confidence interval, denoted by the dotted blue lines. Results suggest a significant association between this measure of timing errors and earnings' pertinence between 1990 and 2010. On average, higher timings errors seem to be associated with a lower earnings' pertinence to the valuation process, holding the other variables constant.

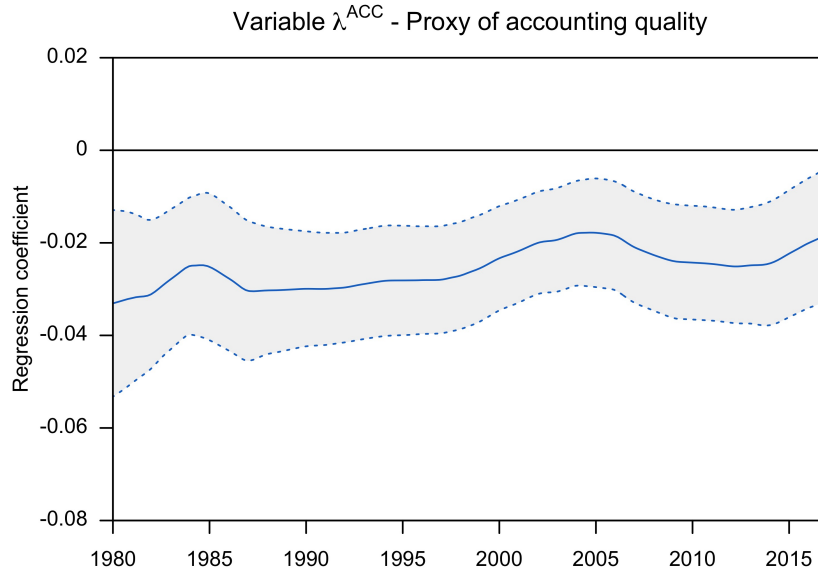


Figure 16: **Accounting quality and earnings’ pertinence to valuation; cross-sectional regression coefficients corresponding to the proxy for quality.** We estimate the following regression model to measure the association between prices and accounting quality. The equation is:

$$\ln |P_{i,t} - \widehat{P}_{i,t}^{\text{EARN}}| = \text{FE}_{i,t} + b_t^P \ln P_{i,t} + b_t^\pi \lambda_{i,t}^\pi + b_t^{\text{CFO}} \lambda_{i,t}^{\text{CFO}} + b_t^{\text{ACC}} \lambda_{i,t}^{\text{ACC}} + \varepsilon_{i,t}, \quad (2.33)$$

where all the definitions of Section 2.2 apply. On the left-hand side of the equality, the dependent variable is the logarithm of an absolute valuation error, a measure of earnings’ pertinence to the valuation process. On the right, the regressors are the logarithm of the stock price (P), plus proxies for economic performance uncertainty (λ^π), the size of timing errors (λ^{CFO}), and accounting quality (λ^{ACC}). The term $\text{FE}_{i,t}$ includes both firm- and time-fixed effects. Standard errors are also clustered by firm and year. The solid blue line in this graph shows $\widehat{b}_t^{\text{ACC}}$, the yearly regression coefficient corresponding to the proxy for accounting quality, plus a 95% confidence interval, denoted by the dotted blue lines. We observe a stable and persistent association between this measure of accounting quality and earnings’ pertinence over the whole period under investigation, also significant at 5% level. These findings indicate that earnings’ pertinence to the valuation process benefits from accruals accounting. On average, earnings seem to be more pertinent to the valuation process for those firms that exhibit a higher accounting quality, keeping everything else constant. That is because the estimated coefficient $\widehat{b}_t^{\text{ACC}}$ is significant and negative, suggesting that we could expect the valuation errors using earnings to decrease once the earnings quality increases, once again, on average, and holding constant the other regressors.

and twelve-industry definitions of Kenneth French, which are available on his website. We adopt the C-score from Penman and Zhang (2002) as an alternative measure of

conservatism and use total assets or sales as alternative proxies for firm size. We measure knowledge intensity using the expenditure in R&D.

A member of the Ph.D. committee underlined that the book value of equity, a second theoretical valuation driver together with earnings, was not directly included in the nonlinear regression models used to measure pertinence. Indeed, the book value can be seen as a driver of value, but we did not find evidence in favor of the book value influencing the relationship between prices and earnings. However, including both the market-to-book ratio and the market value of equity implies that we also indirectly capture the book value, as the Random Forest produces deep trees to capture complex interactions among regressors when estimating the conditional expected value of the stock price. As a second possible approach, we also used a Random Forest to estimate price-to-book ratios among comparable firms, then obtained a stock price estimate by multiplying the estimated multiple by the actual book value per share. Our results do not qualitatively change when adopting this methodology based on multiples.

We also estimated the nonlinear regression model in (2.6) using other nonlinear methods, such as Projection-Pursuit Regression (PPR) and Multivariate Adaptive Regression Splines (MARS). In this specific application, we have also found that residuals from the Random Forest do not significantly change when applying different tuning parameters; for example, found via cross-validation. Our results are also robust if we replace the values of the independent variables with their ranks when estimating the nonlinear regression model.

When comparing earnings and cash flows' pertinence to valuation, instead of computing the two Kolmogorov-Smirnov statistics, we also perform hypothesis testing on the

average difference between absolute valuation errors. In doing so, we approximate the distribution of the corresponding test statistics under the null hypothesis of equal pertinence using bootstrap. Results are robust to all of these changes.

In our last empirical investigation, our results are robust if we consider the total amount of earnings and cash flows in annual changes when estimating the Gaussian state-space model. Furthermore, we found that our results do not depend on differences between cohorts or differences between low- and high-tech industries ([Srivastava, 2014](#)).

5 Conclusions

When the goal is to estimate a relationship between stock prices and a set of accounting determinants of value, nonparametric regression techniques are flexible tools that ask for a minimal number of modeling assumptions. In this thesis, we leverage the flexibility of these techniques to investigate and revisit several research questions regarding this fundamental relationship in accounting and finance.

Using Random Forests, a popular and effective algorithm from Machine Learning, we have first measured, then compared, the association that stock prices have with earnings and cash flows. We have found that the association between stock prices and accounting earnings has decreased over the past two decades. After investigating several possible explanations, we have argued that this phenomenon is likely to be driven by persistent changes in the smoothing role played by accruals.

Accounting earnings are equal to cash flows plus accruals. By adjusting cash flows using accruals, accrual accounting attempts to address the timing and matching issues that can cause cash flows to be a noisy measure of firm performance. Accruals can perform their noise-dampening role when their correlation with cash flows is strongly negative. Accruals smoothing has changed in the last two decades as the correlation between accruals and cash flows has increased; this seems to be related to the weakened association between accounting earnings and stock prices.

[Bushman et al. \(2016\)](#) describe how economic and reporting developments can explain the decline in accruals' effectiveness at solving timing issues. They state: "The growth in the frequency and the magnitude of non-timing-related accruals has increasingly

obscured the expected negative accrual-cash flow relation” (Bushman et al., 2016, p. 44). In line with their conclusions, our results seem to indicate that, over the last five decades, this phenomenon has severely weakened the association between stock prices and earnings and, consequently, significantly weakened the pertinence of accrual accounting to prices.

Bushman et al. (2016) also argue that: “While we document strong evidence that the correlation between accruals and cash flows has significantly declined, many questions remain unanswered and thus are open to future research. For example, have earnings become less meaningful than before because of the changing nature of the information content of accruals?” (Bushman et al., 2016, p. 75). Hence, we can give an affirmative answer to this question. Our results suggest yes, in the last decade, cash flows are at least as relevant to valuation compared to earnings for as much as two-thirds of the firms we studied. During these years, the association between stock prices and operating cash flows is shown to be stronger than that of accounting earnings for roughly half of the firms we studied.

To conclude, in our third empirical investigation, we observe that firms with a higher accounting quality provide earnings that are significantly more associated with stock prices. Our results suggest that earnings are more pertinent to the valuation process when firms exhibit an accounting measurement error that is lower than their cash flows timing problems, holding constant the uncertainty surrounding firms’ operating performance and the size of their timing errors.

Appendix

A.1 An approximation of the distribution of the two Kolmogorov-Smirnov test statistics under the null hypothesis

In a given cross-section, let $\hat{\epsilon}^{\mathbf{X}} = (\hat{\epsilon}_1^{\mathbf{X}}, \dots, \hat{\epsilon}_n^{\mathbf{X}})$ and $\hat{\epsilon}^{\mathbf{Y}} = (\hat{\epsilon}_1^{\mathbf{Y}}, \dots, \hat{\epsilon}_n^{\mathbf{Y}})$ indicate two samples of n absolute relative valuation errors (hereafter, this term will be abbreviated as “valuation errors” or just “errors”). These errors are as in (2.10) and (2.11). The notation emphasizes that the valuation errors depend on a given set of predictor variables (\mathbf{X} and \mathbf{Y}). In our investigation, \mathbf{X} is equal to accounting earnings, while \mathbf{Y} represents operating cash flows.

For each sample of valuation errors, we can define a cumulative distribution function as follows:

$$\hat{F}_t^{\mathbf{X}}(e) = \Pr\left(\left|\frac{P_{\cdot,t} - \hat{P}_{\cdot,t}^{\mathbf{X}}}{P_{\cdot,t}}\right| \leq e\right),$$

and:

$$\hat{F}_t^{\mathbf{Y}}(e) = \Pr\left(\left|\frac{P_{\cdot,t} - \hat{P}_{\cdot,t}^{\mathbf{Y}}}{P_{\cdot,t}}\right| \leq e\right),$$

respectively, and where all of the previous definitions given in Subection 2.1.4 apply. The estimates above give the percentage of errors that are less than or equal to an arbitrary threshold e .

To test if there are significant differences between the two cumulative distribution

functions mentioned above, we will compute two Kolmogorov-Smirnov statistics as follows:

$$\overline{KS} = \sqrt{\frac{n}{2}} \max_e \{\Delta(e) = \hat{F}^{\mathbf{X}}(e) - \hat{F}^{\mathbf{Y}}(e)\},$$

and:

$$\underline{KS} = \sqrt{\frac{n}{2}} \min_e \{\Delta(e) = \hat{F}^{\mathbf{X}}(e) - \hat{F}^{\mathbf{Y}}(e)\},$$

as we have seen in Subsection [2.1.3](#).

When the two errors $\hat{\varepsilon}^{\mathbf{X}}$ and $\hat{\varepsilon}^{\mathbf{Y}}$ are independent, the asymptotic distribution of these two Kolmogorov-Smirnov test statistics is known. However, we observe a correlation between vectors of valuation errors that we have estimated from comparable nonlinear regression models in [\(2.8\)](#) and [\(2.9\)](#). In this case, the distribution of \overline{KS} and \underline{KS} under the null hypothesis of the test is complex and depends on the joint density of the valuation errors; that is why we need to approximate such a distribution using bootstrap.

For this purpose, we have to draw pairs of simulated valuation errors that satisfy two conditions: these artificial valuation errors must have the same marginal distribution; and, at the same time, they must be drawn from the same joint distribution of the original pairs of errors.

This exercise requires a few steps. First of all, we have to assign a rank to the original

pairs of valuation errors, and divide these ranks by the sample size. For example, let:

$$\hat{\epsilon}^{\mathbf{X}} = (1.50\%, 2.00\%, 0.50\%, 3.00\%, 3.50\%),$$

and:

$$\hat{\epsilon}^{\mathbf{Y}} = (1.20\%, 2.05\%, 0.30\%, 3.10\%, 3.05\%),$$

be two vectors of valuation errors of size $n = 5$. If we estimate a nonlinear regression model using the regressors in \mathbf{X} , then we obtain a valuation error of 1.50% corresponding to the first observation, 2.00% for the second, and so on. The valuation error corresponding to the first observation reduces to 1.20% if we estimate the nonlinear regression model using the regressors in \mathbf{Y} , while the error increases to 2.05% for the second observation. 1.50% is the second-lowest valuation error in $\hat{\epsilon}^{\mathbf{X}}$; therefore, the rank of 1.50% is equal to 2, and we will obtain $2/5 = 0.40$ if we divide this rank by the sample size. Similarly, 3.10% is the highest error in $\hat{\epsilon}^{\mathbf{Y}}$ corresponding to the fourth observation; therefore, we will obtain $5/5 = 1.00$ from dividing the rank by the sample size.

We will then collect these ranks into two vectors, as follows:

$$R^{\mathbf{X}} = \left(\frac{2}{5}, \frac{3}{5}, \frac{1}{5}, \frac{4}{5}, \frac{5}{5} \right) = (0.40, 0.60, 0.20, 0.80, 1.00),$$

and:

$$R^{\mathbf{Y}} = \left(\frac{2}{5}, \frac{3}{5}, \frac{1}{5}, \frac{5}{5}, \frac{4}{5} \right) = (0.40, 0.60, 0.20, 1.00, 0.80),$$

and we will further group them into a matrix, as follows:

$$R = \begin{pmatrix} R^{\mathbf{X}} \\ R^{\mathbf{Y}} \end{pmatrix} = \begin{pmatrix} 0.40, & 0.60, & 0.20, & 0.80, & 1.00 \\ 0.40, & 0.60, & 0.20, & 1.00, & 0.80 \end{pmatrix}. \quad (\text{A1})$$

The next step is to combine the two vectors of valuation errors. In this example, let:

$$\hat{\epsilon}^c = (0.30\%, 0.50\%, 1.20\%, 1.50\%, 2.00\%, 2.05\%, 3.00\%, 3.05\%, 3.10\%, 3.50\%),$$

be the result. The superscript c denotes this comprehensive set of errors.

At this point, we will choose at random and with replacement a total of $n = 5$ pairs of ranks from the columns of the matrix R . Let $(1.00, 0.80)$, $(0.80, 1.00)$, $(0.20, 0.20)$, $(0.20, 0.20)$, and $(0.80, 1.00)$ be the first five random pairs of ranks, corresponding to the fifth, fourth, third, third, and fourth columns of the matrix in (A1), respectively.

These random pairs of ranks generate the following random pairs of valuation errors:

$$\begin{aligned} (q^c(1.00), q^c(0.80)) &= (3.50\%, 3.06\%), \\ (q^c(0.80), q^c(1.00)) &= (3.06\%, 3.50\%), \\ (q^c(0.20), q^c(0.20)) &= (1.06\%, 1.06\%), \\ (q^c(0.20), q^c(0.20)) &= (1.06\%, 1.06\%), \\ (q^c(0.80), q^c(1.00)) &= (3.06\%, 3.50\%), \end{aligned}$$

where $q^c(\alpha)$ is the α -quantile of $\hat{\epsilon}^c$; for example, $q^c(0.50) = 2.025\%$ is the median of $\hat{\epsilon}^c$.

In this way, we have obtained two vectors of simulated valuation errors, as follows:

$$\hat{\epsilon}_b^{\mathbf{X}} = (3.50\%, 3.06\%, 1.06\%, 1.06\%, 3.06\%), \quad (\text{A2})$$

and:

$$\hat{\epsilon}_b^{\mathbf{Y}} = (3.06\%, 3.50\%, 1.06\%, 1.06\%, 3.50\%), \quad (\text{A3})$$

where the subscript b means that the two terms in (A2) and (A3) are bootstrap replications of the original vectors of valuation errors $\hat{\epsilon}^{\mathbf{X}}$ and $\hat{\epsilon}^{\mathbf{Y}}$.

This process ensures that the pairs of errors we have generated have the same marginal distribution. At the same time, these pairs of artificial errors are drawn from the same joint density as the original pairs. As we have explained above, these are the requirements of the null hypothesis of the test we want to implement. To conclude, we estimated two cumulative distribution functions using these simulated errors, evaluating the difference between these two functions to compute the corresponding Kolmogorov-Smirnov test statistics for a given bootstrapped sample of valuation errors. To approximate the distribution of the Kolmogorov-Smirnov statistics under the null hypothesis, we need to repeat the whole process a large number of times; see, e.g., [Efron and Tibshirani \(1994\)](#).

A value of a Kolmogorov-Smirnov test statistic that is far from the mean of the distribution under the null hypothesis allows us to reject the null hypothesis with a certain degree of confidence. While, for \overline{KS} , the 95%-quantile of the distribution under the null hypothesis will bound the acceptance region of a one-sided test at 5%, the

5%-quantile will bound that of KS . As usual, we reject a null hypothesis when a test statistic exceeds the corresponding critical value.

A.2 Random Forests, a short R example

A member of the committee asked for a technical appendix to better clarify the codes used to estimate a nonlinear regression model using a Random Forest algorithm.

For the estimation of this kind of regression model we used the software R. Let y be an R object indicating a dependent variable, and x_1 and x_2 two R objects indicating two regressors. Let `dataset` be an R object containing x_1 , x_2 , and y . The first line of code below is used to estimate a classical, OLS linear regression model, whereas the other lines are used to estimate a nonlinear regression model using a Random Forest algorithm:

```
# OLS regression:
lm(y ~ x1 + x2, data = dataset)

# Random Forest (it requires a package):
install.packages("randomForest")
library(randomForest)
randomForest(y ~ x1 + x2, data = dataset)
```

Random forests have been introduced by [Breiman \(2001\)](#). The R package `randomForest` has been developed by Andy Liaw and Matthew Wiener.

It is worth noticing that, even if the syntax of the function is the same, the standard OLS method asks for the specific assumption of linearity between the dependent and independent variables. In contrast, a standard Random Forest does not ask to assume a particular type of relationship a priori.

Furthermore, the Random Forest may be tuned considering additional parameters. For technical details about those tuning parameters (e.g., the number of regression trees, trees depth, number of candidate variables for a split), please refer to [Liaw et al. \(2002\)](#), or the technical R vignette (R code: `?randomForest::randomForest`). For our analysis, different choices of these tuning parameters were leading to highly correlated regression residuals; therefore, we kept the default tuning parameters to ease the code and its replicability.

The R code used in this thesis is conceptually similar to the following one:

```
# A R object representing the dataset:
view(data)

# A list of quantitative variables for the regression:
regression_variables <- c('price', 'earnings', 'size')

year <- data$year # The year corresponding to each observation
year_unique <- sort(unique(year)) # The time span covered

# An empty vector (it will contain the estimated regression residuals):
res_hat <- rep(NA, n)
```

```

# A loop:
for (t in 1:length(year_unique)){

  # At time t, we select the observations and the variables involved:
  df_year <- data %>%
    filter(year == year_unique[t]) %>%
    select(regression_variables)

  # Random Forest and regression residuals:
  rf_reg <- randomForest::randomForest(price ~ ., data = df_year)
  rf_residuals <- df_year$price - rf_reg$predicted

  # Final allocation:
  res_hat[which(year == year_unique[t])] <- rf_residuals

}

```

Once we collect data, we select the period (in years) for estimating the nonlinear regression and we specify the variables to include. A loop is used to iterate the estimation across time. At each step, new regression residuals are saved.

The code above has been simplified, and it is indicative only. In this thesis, we include different variables, and we have generalized the algorithm, for example, to include categorical regressors (e.g., the industry code), to convert quantitative regressors into ranks, and to train the regression model using different tuning parameters.

References

- C. H. Achen. Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, 22(4):327–339, 2005.
- A. Ali. The incremental information content of earnings, working capital from operations, and cash flows. *Journal of Accounting Research*, 32(1):61–74, 1994.
- R. Ball and P. Brown. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2):159–178, 1968.
- M. Barth and G. Clinch. Scale effects in capital markets-based accounting research. *Journal of Business Finance and Accounting*, 36(3-4):253–288, 2009.
- J. Barton, T. B. Hansen, and G. Pownall. Which performance measures do investors around the world value the most—and why? *The Accounting Review*, 85(3):753–789, 2010.
- A. Beyer, I. Guttman, and I. Marinovic. Earnings management and earnings quality: Theory and evidence. *The Accounting Review*, 94(4):77–101, 2019.
- G. Biau and E. Scornet. A Random Forest guided tour. *Test*, 25(2):197–227, 2016.
- G. Biddle, P. Chen, and G. Zhang. When capital follows profitability: Non-linear residual income dynamics. *Review of Accounting Studies*, 6(2):229–265, 2001.
- G. C. Biddle, G. S. Seow, and A. F. Siegel. Relative versus incremental information content. *Contemporary Accounting Research*, 12(1):1–23, 1995.

- G. C. Biddle, R. M. Bowen, and J. S. Wallace. Does EVA beat earnings? Evidence on associations with stock returns and firm values. *Journal of Accounting and Economics*, 24(3):301–336, 1997.
- M. T. Bradshaw and R. G. Sloan. GAAP versus The Street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research*, 40(1): 41–66, 2002.
- L. Breiman. Random Forests. *Machine learning*, 45(1):5–32, 2001.
- L. D. Brown. Earnings forecasting research: Its implications for capital markets research. *International Journal of Forecasting*, 9(3):295–320, 1993.
- D. C. Burgstahler and I. D. Dichev. Earnings, adaptation and equity value. *The Accounting Review*, 72(2):187–215, 1997.
- R. M. Bushman, A. Lerman, and X. F. Zhang. The changing landscape of accrual accounting. *Journal of Accounting Research*, 54(1):41–78, 2016.
- C. S. A. Cheng, W. S. Hopwood, and J. C. McKeown. Non-linearity and specification problems in unexpected earnings response regression model. *The Accounting Review*, 67(3):579–598, 1992.
- D. Collins, E. Maydew, and I. Weiss. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics*, 24(1): 39–67, 1997.
- D. W. Collins and S. Kothari. An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *Journal of Accounting and Economics*, 11 (2-3):143–181, 1989.

- S. Das and B. Lev. Nonlinearity in the returns-earnings relation: Tests of alternative specifications and explanations. *Contemporary Accounting Research*, 11(1):353–379, 1994.
- P. Dechow, W. Ge, and C. Schrand. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3):344–401, 2010.
- P. M. Dechow. Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics*, 18(1):3–42, 1994.
- P. M. Dechow and I. D. Dichev. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review*, 77(1):35–59, 2002.
- P. M. Dechow, S. Kothari, and R. L. Watts. The relation between earnings and cash flows. *Journal of Accounting and Economics*, 25(2):133–168, 1998.
- I. D. Dichev and V. W. Tang. Matching and the changing properties of accounting earnings over the last 40 years. *The Accounting Review*, 83(6):1425–1460, 2008.
- I. D. Dichev and V. W. Tang. Earnings volatility and earnings predictability. *Journal of Accounting and Economics*, 47(1):160–181, 2009.
- P. D. Easton and T. S. Harris. Earnings as an explanatory variable for returns. *Journal of Accounting Research*, 29(1):19–36, 1991.
- B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.
- J. A. Elliott and W. H. Shaw. Write-offs as accounting procedures to manage perceptions. *Journal of Accounting Research*, 26:91–119, 1988.

- E. F. Fama and K. R. French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2):427–465, 1992.
- C. A. Finger. The ability of earnings to predict future earnings and cash flow. *Journal of Accounting Research*, 32(2):210–223, 1994.
- J. Francis, K. Schipper, and L. Vincent. The relative and incremental explanatory power of earnings and alternative (to earnings) performance measures for returns. *Contemporary Accounting Research*, 20(1):121–164, 2003.
- R. N. Freeman and S. Y. Tse. A nonlinear model of security price responses to unexpected earnings. *Journal of Accounting Research*, 30(2):185–209, 1992.
- J. Gerakos and R. Gramacy. *Regression-based earnings forecasts*. Chicago Booth Research Paper No. 12-26, 2013. URL <https://ssrn.com/abstract=2112137>.
- D. Givoly and C. Hayn. The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become more conservative? *Journal of Accounting and Economics*, 29(3):287–320, 2000.
- C. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of Econometrics*, 2(2):111–120, 1974.
- W. Greene. *Econometric analysis*. Pearson Prentice Hall, 1990.
- W. R. Guay and B. K. Sidhu. The usefulness of long-term accruals. *Abacus*, 37(1): 110–131, 2001.
- M. A. Hardy. *Regression with dummy variables*. Sage, 1993.

- T. Hastie, R. Tibshirani, and J. Friedman. *The elements of Statistical Learning: Data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- C. Hayn. The information content of losses. *Journal of Accounting and Economics*, 20(2):125–153, 1995.
- R. W. Holthausen and R. L. Watts. The relevance of the value-relevance literature for financial accounting standard setting. *Journal of Accounting and Economics*, 31(1-3):3–75, 2001.
- K. Hou, M. A. van Dijk, and Y. Zhang. The implied cost of capital: A new approach. *Journal of Accounting and Economics*, 53(3):504–526, 2012.
- P. Hribar and D. Craig Nichols. The use of unsigned earnings quality measures in tests of earnings management. *Journal of Accounting Research*, 45(5):1017–1053, 2007.
- J. Jones. Earnings management during import relief investigations. *Journal of Accounting Research*, 29(2):193–228, 1991.
- J. Kang. *New insights into equity valuation using multiples*. Ph.D. thesis, Université de Neuchâtel, 2016.
- J. Kang and C. Starica. *What kind of earnings shape more market expectations?* Working paper, 2017. URL <http://ssrn.com/abstract=2958777>.
- J. Kmenta. *Elements of econometrics*. Macmillan New York, 1971.
- S. Kothari and J. Shanken. Time-series coefficient variation in value-relevance regressions: A discussion of Core, Guay, and Van Buskirk and new evidence. *Journal of Accounting and Economics*, 34(1-3):69–87, 2003.

- S. P. Kothari, T. E. Laguerre, and A. J. Leone. Capitalization versus expensing: Evidence on the uncertainty of future earnings from capital expenditures versus R&D outlays. *Review of Accounting Studies*, 7(4):355–382, 2002.
- B. Lev and J. A. Ohlson. Market-based empirical research in accounting: A review, interpretation, and extension. *Journal of Accounting Research*, 20:249–322, 1982.
- A. Liaw, M. Wiener, et al. Classification and regression by randomForest. *R news*, 2(3):18–22, 2002.
- R. C. Lipe, L. Bryant, and S. K. Widener. Do nonlinearity, firm-specific coefficients, and losses represent distinct factors in the relation between stock returns and accounting earnings? *Journal of Accounting and Economics*, 25(2):195–214, 1998.
- J. Liu and J. Thomas. Stock returns and accounting earnings. *Journal of Accounting Research*, 38(1):71–101, 2000.
- K. Lo. *The effects of scale differences on inferences in accounting research: Coefficient estimates, tests of incremental association, and relative value relevance*. MIT Sloan Working Paper No. 4490-04, Sauder School of Business Working Paper, 2005. URL <http://ssrn.com/abstract=555684>.
- H. Lütkepohl. *New introduction to multiple time series analysis*. Springer Science & Business Media, 2005.
- S. Nallareddy, M. Sethuraman, and M. Venkatachalam. Changes in accrual properties and operating environment: Implications for cash flow predictability. *Journal of Accounting and Economics*, 69(2), 2020. URL <https://doi.org/10.1016/j.jacceco.2020.101313>.

- V. V. Nikolaev. *Identifying accounting quality*. Chicago Booth Research Paper No. 14-28, 2018. URL <https://ssrn.com/abstract=2484958>.
- J. Ohlson. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*, 11(2):661–687, 1995.
- J. A. Ohlson and B. E. Juettner-Nauroth. Expected EPS and EPS growth as determinants of value. *Review of Accounting Studies*, 10(2):349–365, 2005.
- S. H. Penman and X.-J. Zhang. Accounting conservatism, the quality of earnings, and stock returns. *The Accounting Review*, 77(2):237–264, 2002.
- S. H. Penman, F. Reggiani, S. A. Richardson, and I. Tuna. A framework for identifying accounting characteristics for asset pricing models, with an evaluation of book-to-price. *European Financial Management*, 24(4):488–520, 2018.
- J. B. Ramsey. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(2):350–371, 1969.
- S. Roychowdhury and R. L. Watts. Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting. *Journal of Accounting and Economics*, 44(1-2):2–31, 2007.
- K. Schipper and L. Vincent. Earnings quality. *Accounting Horizons*, 17:97–110, 2003.
- A. Srivastava. Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*, 57(2):196–217, 2014.
- C. Starica and P. Giosi. *The ebbing of accrual accounting*. Working paper, 2017. URL <http://ssrn.com/abstract=3072287>.

- C. Starica and J. Marton. *Assessing the price-earnings relation via machine learning*. Working paper, 2019. URL <http://ssrn.com/abstract=3813048>.
- J. H. Stock and M. W. Watson. *Introduction to econometrics*. Addison-Wesley, 2012.
- K. R. Subramanyam. Uncertain precision and price reactions to information. *The Accounting Review*, 71(2):207–219, 1996.
- W. R. Teets and C. E. Wasley. Estimating earnings response coefficients: Pooled versus firm-specific models. *Journal of Accounting and Economics*, 21(3):279–295, 1996.
- R. S. Tsay. *Analysis of financial time series*. John Wiley & Sons, 2005.
- R. L. Watts and R. W. Leftwich. The time series of annual accounting earnings. *Journal of Accounting Research*, 15(2):253–271, 1977.