Foundations and Trends® in Accounting

Financial Statement Analysis and Earnings Forecasting

Suggested Citation: Steven J. Monahan (2017), "Financial Statement Analysis and Earnings Forecasting", Foundations and Trends® in Accounting: Vol. 12, No. 2, pp 105–215. DOI: 10.1561/1400000036.

Steven J. Monahan
INSEAD, Europe Campus
Boulevard de Constance
77305 Fontainebleau, France
steven.monahan@insead.edu

now
the essence of knowledge
Boston — Delft

Contents

Financial Statement Analysis and Earnings Forecasting

Steven J. Monahan*

INSEAD, Europe Campus, Boulevard de Constance, 77305 Fontainebleau, France; steven.monahan@insead.edu

ABSTRACT

I synthesize and discuss academic research on financial statement analysis and earnings forecasting. I begin by discussing analytical and empirical evidence that shows that earnings, not dividends or free cash flows, are the payoffs that investors forecast when estimating value. This result is fundamental and it provides clear motivation for studying earnings forecasting and the role that historical accounting numbers play in the earnings-forecasting process. I then provide a detailed discussion of the research design choices that are made when developing and evaluating an earnings-forecasting approach. I describe the tradeoffs involved when making these choices and I review the extant empirical literature. An overarching theme of this discussion is that there are substantial research opportunities. For example:

- The random-walk model performs too well on a relative basis. It is inconsistent with standard economic assumptions, accounting practice and the way financial statement analysis is practiced and taught. Nonetheless, it tends to be as accurate and sometimes more accurate than other extant approaches.

*This monograph is dedicated to Larissa Ignatieva and my parents: Leo John Monahan (deceased) and Mary Eleanor Monahan.

Steven J. Monahan (2017), "Financial Statement Analysis and Earnings Forecasting", Foundations and Trends® in Accounting: Vol. 12, No. 2, pp 105–215. DOI: 10.1561/1400000036.

- Panel-data approaches that use a mix of cross-sectional and time-series data are very flexible in terms of the: (1) choice of earnings metric to predict; (2) choice of predictors; (3) choice of estimator; and (4) choice of estimation sample. At present, these approaches have not been used to their full potential.
- There is insufficient evidence regarding how to identify peers and the role that peer analysis plays in the forecasting process.
- There is insufficient evidence regarding approaches for forecasting the higher moments of future earnings, how to evaluate these forecasts and their role in determining value. Moreover, the role that accounting measurement plays in the determination of the higher moments of earnings and how accounting-measurement rules affect the usefulness of historical accounting numbers for predicting the higher moments of future earnings is not well understood.

1

Introduction

I synthesize and discuss academic research on financial statement analysis and earnings forecasting, which is the process of analyzing historical financial statement data for the purpose of developing forecasts of future earnings. This process is important because it is central to the valuation of companies and the securities they issue.

Valuation is a crucial economic activity. As discussed in Hayek (1945), security prices determine how finite resources are allocated to firms and individuals. Moreover, as discussed in Arrow (1964), when people have access to a broad set of securities, they can form diversified portfolios and share risks. Hence, to fully understand the economic role of accounting, it is imperative that we understand the role that accounting numbers play in valuation.

The above provokes an immediate question: Why earnings? Specifically, given equity values are determined by expected future dividends and enterprise value can be expressed as a function of expected future free cash flows, why study earnings? Aren't dividends and free cash flows the primitive variables that investors forecast? These are valid questions and, if they cannot be answered, the premise underlying this monograph comes into question. With this in mind, in Section 2, I delve into the question: Why earnings? I focus on dividend policy irrelevance, which implies that forecasting dividends is futile. I describe key analytical results that imply that, assuming dividend policy irrelevance, expected earnings are *the* fundamental determinant of both equity and enterprise value. I then finish the section by discussing key empirical results that imply that: (1) accrual-accounting earnings are more informative about changes in value than either dividends or cash flows and (2) accrual-accounting earnings, not dividends or free cash flows, are what investors forecast when estimating equity value.

Given the primacy of earnings, the motivation for forecasting them is clear. However, to do this, the researcher must first select the earnings metric that she will forecast. In Section 3, I discuss the issues involved in making this decision. As I explain in that section, the choice depends on the research context as well as data availability and the statistical properties of the different earnings metrics. Hence, the decision ultimately involves making a subjective tradeoff. Consequently, best practice is to clearly motivate the research question, describe the logic for selecting a particular metric or metrics, and then discuss the consequent pros and cons.

Once an earnings metric has been chosen, the natural question to ask is: How useful are historical accounting numbers for developing forecasts of that metric? In Sections 4–8, I focus on this question. In Section 4, I discuss the general role of econometric modeling. I make three points. First, when studying earnings forecasting, the goal is not to find the "best" model; rather, it is to identify models that are useful. Second, within the contexts of empirical capital markets research and practical valuation, useful models are those that are objective, replicable, generate accurate forecasts for large samples at a low cost and provide useful guidance regarding best practice. Finally, extant models are too inaccurate and, if taken at face value, extant results lead to seemingly absurd conclusions regarding best practice. Given the central role that earnings forecasting plays in valuation and the importance of valuation, these results imply that further research is necessary.

After discussing econometric modeling in general, I discuss specific types of models. In Section 5, I discuss time-series models, which were the default choice in early research studies. A key result is that, *of* *the various time-series models evaluated*, the random-walk model is the best. The superiority of the random-walk model is counterintuitive because, as discussed in Section 5, it is inconsistent with standard economic assumptions and accounting practice. However, as I argue in that section, this result is misleading because time-series models are ill-suited for developing forecasts of earnings. Hence, the fact that the random-walk model is the best time-series model does not imply that it is the best approach.

Given the limitations of time-series models, they are no longer the default choice. Rather, recent studies tend to use panel-data approaches. These approaches allow the researcher to combine cross-sectional and time-series data to arrive at a forecast of earnings. Hence, they are more flexible than time-series models and they have numerous *a priori* advantages vis-à-vis these models. In Section 6, I discuss the choices a researcher makes when using panel-data approaches and I describe the advantages of these approaches. I then discuss the extant empirical evidence. I conclude that, with regards to the usefulness of panel-data approaches, the jury is still out. Although extant results imply that panel-data approaches are not much better than the random-walk model, these studies do not exploit the full potential of panel-data approaches. Hence, in my opinion, further study of panel-data approaches is a promising research agenda.

In Section 7, I discuss the role of accounting measurement in determining the usefulness of historical accounting numbers for developing forecasts of future earnings. I begin by explaining how accounting measurement determines accruals, which, is the non-cash component of earnings. I then discuss two topics that are directly related to accounting measurement: (1) the relative persistence of cash flows and accruals and (2) accounting conservatism. The main point of the section is that, perhaps not surprisingly, accounting measurement matters. Extreme accruals imply less persistent earnings. Moreover, conservative accounting rules can lead to historical trends in profitability ratios that are misleading about future profitability ratios. Hence, colloquially speaking, either too much accounting — that is, extreme accruals — or too little accounting — that is, conservative-accounting rules — lead to historical earnings that are less useful for developing forecasts of future earnings.

In Section 8, I discuss approaches for forecasting the higher moments of future earnings. At present, this issue has not received much attention. Hence, there are ample research opportunities. I explain why higher moments are important and I briefly discuss the extant studies that develop and evaluate forecasts of them. Finally, in Section 9, I provide a summary of the monograph.

Some general caveats are warranted. First, I focus on the academic literature that relates to the pursuit of financial statement analysis for the purpose of developing forecasts of earnings. However, financial statement analysis is a broad topic and developing earnings forecasts is only one of its objectives. Nonetheless. to maintain focus and for the sake of brevity, I ignore other objectives such as the identification of mispriced securities, default prediction, employee performance evaluation, etc.

Second, I concentrate on studies in which the objective is to develop and evaluate different approaches for forecasting earnings. Hence, although I use valuation to motivate my focus on earnings forecasting, I do not attempt to survey and synthesize the large and important capital-markets literature that relates to the association between accounting numbers and security prices. Third, my primary focus is on studies that consider annual earnings. There is a large and important literature that evaluates approaches for forecasting quarterly earnings. However, a thorough treatment of that literature is beyond the scope of this monograph.

Finally, I focus on studies that develop statistical approaches for forecasting earnings. I do not discuss the literature on analysts' forecasts and their properties. My reasons are twofold. First, comprehensive reviews of this literature can be found in Ramnath *et al.* (2008) and Bradshaw *et al.* (2016). Second, my aim is to describe the literature that evaluates the role of historical accounting numbers in the forecasting process. Within this context, analysts' forecasts are a "black box." That is, the process that analysts follow when developing their forecasts and the role that accounting numbers play in that process are not observable.

2

Why Earnings?

In 1938 John Burr Williams' book *The Theory of Investment Value* was published. Williams contributed to investment theory in numerous ways. One contribution, however, clearly stands out: The dividend discount model, DDM. The DDM is seminal because it provides the core logic that underlies a fundamental result in financial economics: The value of a firm's equity equals the present value of the future dividends that the firm is expected to pay discounted by the firm's cost of equity capital. In equation form:

$$V_0 = \frac{E_0[\mathrm{DIV}_1 + V_1]}{R} = \sum_{t=1}^{\infty} \frac{E_0[\mathrm{DIV}_t]}{R^t} \tag{2.1}$$

In Equation (2.1), V_t is the equity value at date t; DIV_t is dividends paid at date t; R is one plus the cost of equity capital, r; and, $E_0[\cdot]$ is the expectation operator. Expectations are conditional on information available at date zero.

The DDM is fundamental for two reasons. First, when combined with results developed by Modigliani and Miller (1958), it leads to the net present value criterion. Consequently, the DDM continues to be a fundamental concept underlying how accounting and finance are practiced, researched, and taught. Second, the DDM is intuitive. The statement that people prefer more consumption to less is virtually self-evident. In modern exchange economies, consumption takes place via the exchange of cash for goods and services. Hence, the links between cash and consumption and between consumption and value are clear: ignoring risk and delay, higher expected future dividends imply higher future consumption and consequently higher value.

Accounting earnings, however, are not equal to dividends. Hence, a straightforward question arises: Why study earnings? That is, given that it is expected future dividends that determine value, why study how to use historical accounting numbers to forecast future earnings?

The questions posed above are the primary motivation for a large body of research that sheds light on the role that accounting numbers play in the valuation process. This research is often referred to as capital-markets research; and, I continue to use that moniker. In the remainder of this section I provide a brief overview of this literature. I do not provide a comprehensive review. Rather, I highlight the key issues that are directly related to the question that is the title of the section: Why earnings?

I begin by describing the "information perspective," which, until the mid-1990s, was the primary motivation for capital-markets research. Next, I describe some shortcomings of the information perspective. I then discuss recent analytical results that deal (either wholly or partially) with these shortcomings. These analytical results have altered how people think about the valuation role of earnings and how capital-markets research is conducted and taught. Nonetheless, they do not provide a definitive answer to the question: Why earnings? Rather, this is ultimately an empirical question. With this in mind, I discuss three key empirical studies. After doing this, I provide some concluding comments.

2.1 The information perspective

The essential logic underlying the information perspective is straightforward: Historical accounting numbers are potentially informative about expected future dividends, which, as per the DDM, determine current price.[1] The information perspective is the primary motivation underlying a vast literature that includes studies of the information content of reported earnings, earnings response coefficients, the value relevance of reported accounting numbers, the relation between historical accounting numbers and future returns, etc.

Although influential, the information perspective has two related shortcomings. First, it does not require dividend policy irrelevance, DPI. Rather, it is assumed (implicitly or explicitly) that investors' ultimate goal is to forecast future dividends. Second, although the information perspective refers to earnings, earnings are not well-defined. Rather, the word *earnings* is a label that is attached to one element of a vector of random variables that has "nice" stochastic properties.[2] I elaborate on these two limitations in the next two sections.

2.2 The importance of DPI

Per the information perspective, investors' ultimate objective is to develop expectations about future dividends. Given the DDM, this seems quite reasonable. However, it is not. The reason for this is that, per Miller and Modigliani (1961), a firm's dividend *policy* is irrelevant. That is, holding its investment opportunities constant, the firm's equity value does not depend on when future dividends are expected to be paid or the expected temporal pattern of these future payments. Hence, dividend-policy decisions are arbitrary.[3] This, in turn, implies that, although the DDM continues to hold, it cannot be implemented in

[1] Beaver (1989) contains a clear and concise description of the information perspective, which was introduced by Beaver and Demski (1979). Formal, and more-general, results regarding the implications of the information perspective for equity valuation are provided in Garman and Ohlson (1980).

[2] For example, Garman and Ohlson (1980) assume that the vector of random variables evolves via a time-independent Markovian process.

[3] This does not imply that management can choose any policy. The firm's investment opportunities must be unaffected by the choice. Conversely, it does not imply that management cannot choose a policy. They can but the choice is irrelevant. Rather, what it implies is that, *ceteris paribus*, choosing to pay relatively high (low) dividends during the early stages of the firm's existence implies that expected dividends will be relatively low (high) during the later stage of the firm's existence.

a meaningful way. This paradox is often referred to as the dividend conundrum.

First, if dividend-policy decisions are arbitrary, it is meaningless to evaluate the accuracy of dividend forecasts. This is true even if, in the obviously hypothetical situation, every element of the infinite sequence of forecast errors is observable. For example, suppose that each of the forecast errors is equal to zero. This does not imply that the underlying sequence of forecasts leads to a more accurate value estimate than another sequence of forecasts in which all of the errors are non-zero. Rather, the two sequences may lead to the same value estimate. Hence, the fact that the first sequence is more accurate is meaningless. Consequently, the DDM provides no guidance about the role of forecasting in the valuation process. In particular, it provides no insights about whether more accurate forecasts imply more accurate value estimates.

Second, DPI implies that, within the context of the DDM, assumptions about steady-state conditions are inherently *ad hoc*. This is a problem because, from Equation (2.1), implementation of the DDM requires that forecasts of all of the elements of the infinite sequence of future dividends, $\{\mathrm{DIV}_t\}_{t=1}^{\infty}$, be developed. However, it will take infinite time to do this. Hence, an assumption must be made that, after a specific future date D, dividends will evolve in a well-defined manner. This is often referred to as a steady-state condition. A common steady-state condition is to assume that $\forall t > D, E_0[\mathrm{DIV}_t] = E_0[\mathrm{DIV}_{t-1}] \times G_{\mathrm{DIV}}$ and $G_{\mathrm{DIV}} < R$. That is, expected dividends grow at the rate $G_{\mathrm{DIV}} < R$ for all dates subsequent to date D. However, this assumption is *ad hoc* because it is tantamount to assuming that at date D the firm will adopt a specific dividend policy.

2.3 The meaning of earnings within the information perspective

From the information perspective, earnings are one of a number of random variables that are informative about future dividends. Earnings are not a fundamental construct. Rather, they serve as an intermediary function. For instance, the maintained assumption might be that historical earnings and other variables are used to develop forecasts of future dividends. Alternatively, the maintained assumption might be that the goal is to develop forecasts of future earnings. However, this is only the penultimate step in the process of forecasting future dividends. Hence, earnings play an intermediate and purely statistical role in the valuation process and, in this context, the word *earnings* is just a label.

In light of the above, within the context of the information perspective, the valuation role of earnings is purely an empirical question. Moreover, this question is not well defined in the sense that the information perspective does not yield clear-cut insights regarding the nature of the empirical relations between earnings and prices, earnings and returns, etc. It is important to note that this outcome is unavoidable. As discussed in Beaver and Demski (1979), earnings is not a well-defined economic construct unless there are perfect and complete markets. However, as discussed in the next section, it is possible to go beyond the information perspective and, thus, put additional structure on the concept of earnings. Nonetheless, their valuation role remains an empirical question. That said, the question is more clearly defined.

2.4 Going beyond the information perspective: Accounting-based valuation

Accounting-based valuation involves developing a value estimate that is a function of expected future accounting numbers. It is important to note that, within this context, forecasting accounting numbers is not the penultimate step in the process of arriving at forecasts of future dividends. Rather, the forecasted accounting numbers are direct inputs into the valuation formula.

In the space below, I focus on the abnormal earnings growth valuation (AEGV), model developed by Ohlson and Juettner-Nauroth (2005) (OJ hereafter) and elaborated on in Ohlson and Gao (2006) (OG hereafter), who argue that "...the model should be accorded 'benchmark' status." The following equation describes the AEGV model.

$$\begin{aligned} V_0 &= \frac{E_0[\mathrm{EARN}_1]}{r} + \sum_{t=1}^{\infty} \frac{E_0[\Delta\mathrm{EARN}_{t+1} - r \times (\mathrm{EARN}_t - \mathrm{DIV}_t)]/r}{R^t} \\ &= \frac{E_0[\mathrm{EARN}_1]}{r} + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{AEG}_t]/r}{R^t} \qquad (2.2) \end{aligned}$$

In Equation (2.2), EARN_t is equity-level earnings for period t and AEG_t, which equals $\Delta\mathrm{EARN}_{t+1} - r \times (\mathrm{EARN}_t - \mathrm{DIV}_t)$, is abnormal earnings growth for period t. It is important to note that AEG for period t is a function of dollar growth in earnings for period $t + 1$.

An intuitive explanation of the AEGV model is as follows. To value the firm's equity, begin by capitalizing expected forward earnings. This is an incomplete value estimate, however. The reason for this is that it ignores the possibility of future earnings growth, which affects value if it differs from "normal" earnings growth arising from reinvestment. Hence, the present value of capitalized expected future *abnormal* earnings growth must be added to capitalized expected forward earnings.

Although intuitively appealing, the AEGV model raises a number of questions regarding its rigor, whether it embeds DPI, etc. As a rhetorical device, I pose and answer the questions that I think are the most likely to arise and that are the most pertinent to the subject of this section. Again, I do this as a rhetorical device, not as an attempt to provide a complete description of either the AEGV model or accounting-based valuation models in general. Readers who are interested in such a discussion can find it in OG, Gao *et al.* (2013) and the papers referenced in OG and Gao *et al.* (2013).

*Q*1: *Is the AEGV model rigorous — that is, is it consistent with the DDM?*

Yes. To see this, it is useful to take the approach described in OJ and OG. If $\frac{E_0[\mathrm{EARN}_{t+1}]/r}{R^t}$ converges to zero as t becomes large — that is, the growth rate in capitalized forward earnings eventually falls below the cost of equity capital, zero can be expressed as follows:

$$0 = \frac{E_0[\mathrm{EARN}_1]}{r} + \sum_{t=1}^{\infty} \frac{(E_0[\mathrm{EARN}_{t+1}]/r) - R \times (E_0[\mathrm{EARN}_t]/r)}{R^t}. \qquad (2.3)$$

Next, Equation (2.2) — that is, the AEGV model — is obtained by combining Equations (2.3) and (2.1). However, because this is equivalent to adding the number zero to Equation (2.1) — that is, the DDM, the AEGV model and the DDM are equivalent.

Q2: Does the AEGV model embed DPI?

The answer to Q1 above provokes a second question: Colloquially speaking, isn't the AEGV model just the DDM in disguise? Hence, doesn't it also violate DPI? Not necessarily. If some relatively innocuous assumptions are made, AEGV is consistent with DPI. This is a paradoxical result. Hence, I elaborate on it in the space below.

The first assumption is that there exists a date T such that $\forall t > T$, $E_0[\mathrm{AEG}_t] = E_0[\mathrm{AEG}_{t-1}] \times G_{\mathrm{AEG}}$ and $G_{\mathrm{AEG}} < R$. Via the law of iterated expectations, this implies the OJ model:

$$E_0[V_T] = \frac{E_0[\mathrm{EARN}_{T+1}]}{r} + \frac{E_0[\mathrm{AEG}_{T+1}]/r}{R - G_{\mathrm{AEG}}}. \tag{2.4}$$

Moreover, combining Equations (2.4) and (2.1) leads to the following expression for equity value at date zero.

$$V_0 = \sum_{t=1}^{T} \frac{\mathrm{DIV}_t}{R^t} + \frac{1}{R^T} \times \left(\frac{E_0[\mathrm{EARN}_{T+1}]}{r} + \frac{E_0[\mathrm{AEG}_{T+1}]/r}{R - G_{AEG}} \right). \tag{2.5}$$

Equation (2.5) shows that value can be separated into two components: (1) the present value of "near-term" dividends and (2) the present value of the terminal value. The terminal value is a function of expected terminal earnings, expected terminal abnormal earnings growth, the discount rate and the long-run growth rate in AEG. This is appealing because it is consistent with how valuation is often practiced and taught (e.g., McKinsey & Company Inc. *et al.*, 2015).

Second, assume that $\forall t \geq T$ expected dividends paid at date $t + 1$ are a linear combination of lagged expected earnings, lagged expected abnormal earnings growth and lagged expected dividends. Combining this assumption with the definition of AEG_{t+1} and the first assumption leads to Equation (2.6). (In Equation (2.6), I omit the expectation operator; hence, EARN, AEG, and DIV represent expectations that are based on information available at time zero.)

$$\overbrace{\begin{bmatrix}\mathrm{EARN}_{T+t+1}\\ \mathrm{AEG}_{T+t+1}\\ \mathrm{DIV}_{T+t+1}\end{bmatrix}}^{x_{T+t+1}} = \overbrace{\begin{bmatrix}R & 1 & -r\\ 0 & G_{\mathrm{AEG}} & 0\\ \omega_{31} & \omega_{32} & \omega_{33}\end{bmatrix}}^{A}\overbrace{\begin{bmatrix}\mathrm{EARN}_{T+t}\\ \mathrm{AEG}_{T+t}\\ \mathrm{DIV}_{T+t}\end{bmatrix}}^{x_{T+t}}. \tag{2.6}$$

In the above equation, the values of the top row of the matrix $\mathbf{A}$ correspond to the definition of AEG and the second row follows from the first assumption. The third row is made up of the dividend-policy parameters ω_{31}, ω_{32}, and ω_{33}. If these are "free" parameters, DPI holds.

For a given value of the vector $\boldsymbol{x_T}$, $\boldsymbol{x_{T+t}} = \mathbf{A}^t \boldsymbol{x_T}$ and the third element of $\boldsymbol{x_{T+t}}$ equals DIV_{T+t}. Hence, Equation (2.6) generates an infinite sequence of expected dividends paid after date T, $\{E_0[\mathrm{DIV}_{T+t}]\}_{t=1}^{\infty}$. This leads to the third, and final, assumption: The sequence $\{E_0[\mathrm{DIV}_{T+t}]\}_{t=1}^{\infty}$ generated by Equation (2.6) yields a finite solution to the DDM. This is equivalent to assuming that $\frac{E_0[\mathrm{DIV}_{T+t}]}{R^{-t}}$ converges to zero as t approaches infinity.

These three assumptions lead to a key result (Lemma 4.1 and Proposition 4.2 of OG): The value estimate implied by Equation (2.5) does not depend on either the dividend-policy parameters or DIV_T. The converse is also true. In particular, consider a more general version of Equation (2.6), which I refer to as Equation (2.6*). In Equation (2.6*) the matrix $\mathbf{A}$ is replaced with the matrix $\mathbf{A}^*$. The third row of $\mathbf{A}^*$ equals the third row of $\mathbf{A}$; but, the first and second rows equal $[\omega_{11} \quad \omega_{12} \quad \omega_{13}]$ and $[0 \quad \omega_{22} \quad 0]$, respectively. Now, assume that the DDM implied by Equation (2.6*) has a finite value that does not depend on the dividend-policy parameters or DIV_T. This assumption implies that $\omega_{11} = R$ and, without loss of generality, $\omega_{12} = 1, \omega_{13} = -r$, and $\omega_{22} = G_{\mathrm{AEG}}$ — that is, this assumption implies the OJ model.

The above implies that, assuming Equation (2.6), the OJ-model-based terminal value in Equation (2.5) implies DPI and *vice versa*. In colloquial terms, DPI and the OJ model are opposite sides of the same coin. This is important because it implies that the OJ model can be used to estimate value without having to make arbitrary assumptions about the dividend policy.

Q3: *Given Equation* (2.5) *requires forecasts of near-term dividends, doesn't it violate DPI?*

Not necessarily and not in the long run. Although Equation (2.5) expresses value as a function of forecasts of dividends paid in years one through T, it does not necessarily violate DPI. Moreover, if it does, the violation relates purely to the near term. First, DPI does not imply that managers cannot adopt a dividend policy. For example, management may plan that, for the *foreseeable* future, they will not pay dividends, or that they will maintain a target payout ratio, etc. The key word in the above sentence is foreseeable. If V_0 is positive, managers cannot expect to withhold dividends forever. Alternatively, a target payout ratio that is unsustainable will inevitably lead to a dividend cut. The key point is that the value created by the firm is the ultimate determinant of lifetime dividends. Hence, if the current dividend policy is too meager (generous), dividends will eventually increase (decrease).

Second, related to the first point, DPI does not imply that it is impossible to forecast near-term dividends — that is, $\{\text{DIV}_t\}_{t=1}^{T}$. For example, if management has either explicitly or implicitly adopted a policy of not paying dividends, the assumption that $\{\text{DIV}_t = 0\}_{t=1}^{T}$ may be reasonable. Alternatively, if management is credible and has stated they will maintain a constant payout ratio, it may be reasonable to apply that ratio to forecasts of near-term earnings to arrive at forecasts of near-term dividends.

Third, it is useful to keep in mind that DPI is based on the assumption that the firm's investment opportunities are unaffected by the dividend-policy choice. Hence, Equation (2.5) embeds a consistency requirement: The expected terminal value, which equals $\frac{E_0[\text{EARN}_{T+1}]}{r} + \frac{E_0[\text{AEG}_{T+1}]/r}{R-G_{\text{AEG}}}$, must be consistent with the forecasts of $\{\text{DIV}_t\}_{t=1}^{T}$. For example, suppose that, as of date zero, firm A and firm B are identical with the exception that the managers of firm A do not plan to pay dividends during the near term whereas the managers of firm B plan to maintain a constant payout ratio of $K \in (0, 1)$. Further assume that the firms are expected to be profitable, which implies that firm B is expected to pay dividends in the near term. Given that, at time zero, the firms' investment opportunities, capital structures, etc., are identical, their equity values are also identical. In order for this to be true, the expected terminal value for firm A must equal the expected terminal value for firm B plus $\sum_{t=1}^{T}(R^{T-t}-1)\times K\times E_0[\mathrm{EARN}_t^B]$. ($\mathrm{EARN}_t^B$ denotes the earnings of firm B during period t.) This amount equals the additional earnings that firm A is expected to generate because its managers are not expected to pay dividends.

The above example illustrates the essence of DPI: *Ceteris paribus*, dividend-policy choices are zero-net-present-value decisions. Firm A is expected to retain more of its earnings, which are expected to beget additional earnings. Hence, it is expected to have higher earnings than firm B. These additional expected earnings are value irrelevant, however. The reason for this is that they relate to zero-net-present-value investments — that is, the retained earnings grow at the rate $(R-1)$, which is the cost of equity capital. In fact, the example can be re-written under the assumption that there is one firm with a payout ratio that equals the random variable $K \in [0,1)$. The assumption made about K remains irrelevant as long as the forecast of the terminal value is consistent with the forecasts of near-term dividends. Hence, different assumptions about dividend policy imply the same value estimate.

Finally, although Equation (2.5) does not necessarily violate DPI, it also does not require it to hold during the near term — that is, during the period between year zero and year T. For example, during the near term, idiosyncratic phenomena such as capital-market imperfections, asymmetric information, etc. may lead to circumstances in which the dividend policy matters. However, Equations (2.5) and (2.6) embed the assumption that in the long run — that is, after date T — these idiosyncratic phenomena and their effects are irrelevant.

Q4: *Are the assumptions underlying Equations* (2.5) *and* (2.6) *overly restrictive?*

Every model is an abstraction from reality. This is unavoidable. To obtain closed-form solutions, assumptions are necessary. That said, the assumptions underlying Equations (2.5) and (2.6) are relatively innocuous. Regarding the first assumption, it is impossible to develop a detailed forecast of every element of the sequence $\{\mathrm{AEG}_t\}_{t=1}^{\infty}$. Consequently, an assumption about the long-run behavior of abnormal earnings growth is necessary. A constant-growth assumption seems natural and conforms to what is typically assumed in practice.

Regarding Equation (2.6), the first two rows of the matrix $\mathbf{A}$ follow directly from the definition of AEG and the first assumption — that is, that AEG grows at the rate $G_{\text{AEG}} < R$ in the long run. The linearity of the dividend-policy equation, however, is assumed for analytical convenience. That said, it is not a necessary condition. As discussed in OG, there exist non-linear dividend-policy equations for which DPI still holds. Finally, the assumption that $\frac{\text{DIV}_T}{R^T}$ converges to zero as T approaches infinity is standard and it reflects the fact that value created must eventually be distributed.

$Q5$: *Are earnings still undefined?*

Equation (2.3), is crucial because, when it is combined with the DDM, it leads to the AEGV model. Hence, it is awkward that Equation (2.3), and consequently Equation (2.2), holds for any variable z_t for which $\frac{z_t}{R^t}$ converges to zero as t approaches infinity. For example, z_t could be revenues, depreciation, number of employees, the shoe size of the chief executive officer, etc. This leads to the concern that the AEGV model does not really go beyond the information perspective — that is, earnings may still be just a label that has been arbitrarily assigned to the variable EARN_t.

The above concern is not completely misplaced. As discussed below, the earnings construct underlying Equation (2.2) is not fully defined. However, the label *earnings* is appropriate. The reasons for this are twofold.

First, in the AEGV model, earnings and abnormal earnings growth are the ultimate constructs of interest. They are the fundamentals that investors forecast. This is important because it leads to a testable empirical question: Are accounting-based estimates of value more accurate than value estimates obtained from the DDM? If so, earnings as defined either by generally accepted accounting principles, GAAP, or International Financial Reporting Standards, IFRS, etc. are useful. I return to this issue in a subsequent part of the section.

Second, the earnings construct underlying the OJ model — that is, Equation (2.5) — has two desirable properties, which I refer to as P1 and P2. These two properties were discovered and developed by OG (see Section 5 of OG). In order to understand P1 and P2, it is useful to define the construct aggregate earnings for the period t through $t+h$, $\text{AGG\_EARN}_{t,H}$:

$$\text{AGG\_EARN}_{t,H} = \sum_{h=1}^{H} \text{EARN}_{t+h} + \sum_{h=1}^{H} (R^{H-h} - 1) \times \text{DIV}_{t+h} \tag{2.7}$$

$\text{AGG\_EARN}_{t,H}$ is the sum of two components that accumulate from the beginning of period $t+1$ through the end of period $t+H$: (1) earnings reported by the firm and (2) additional, or foregone, earnings that the firm would have reported if it had not paid dividends.

With the definition of aggregate earnings in place, I state P1 and P2.

$$\partial E_0[\text{AGG\_EARN}_{T+t,H}]/\partial E_0[\text{DIV}_{T+t}] = -(R^H - 1) \quad \forall H \in [1, \infty] \tag{P1}$$

$$\partial E_0[\text{AGG\_EARN}_{T+t,H}]/\partial E_0[\text{EARN}_{T+t}] = \sum_{h=1}^{H} R^H \quad \forall H \in [1, \infty] \tag{P2}$$

P1 and P2 imply that, on the margin, increases in dividends paid on date $T+t$ reduce subsequent earnings at the rate r and increases in earnings for period $T+t$ beget future earnings at the rate R. P1 is a desirable property because it aligns with DPI. An increase (decrease) in dividends implies a contemporaneous decrease (increase) in investment and lower (higher) future earnings. However, on the margin, the reduction (increase) in investment relates to zero-net-present-value projects. P2 is also desirable because it coincides with the notion that managers invest in all positive-net-present-value projects, which implies that the marginal project has a net present value of zero. Consequently, on the margin, changes in investment attributable to contemporaneous changes in earnings have no effect on value.

In addition to being desirable properties, P1 and P2 are properties of an *earnings* metric — that is, a summary measure that equals the difference between revenues and expenses. There is no reason to expect that higher dividends today imply lower future revenues, or higher future depreciation, etc. Rather, higher dividends today imply less investment today and lower future earnings. Whether the subsequent decrease in earnings relates to lower revenues, higher expenses or some combination of the two is unclear. It is the net amount — that is, earnings — that is affected. Similarly, higher current values of a particular component of earnings do not necessarily imply that same component will be higher in the future. However, *ceteris paribus*, increases in earnings beget future earnings.

Not only are P1 and P2 desirable properties of earnings, they are part and parcel with the OJ model that underlies Equations (2.5) and (2.6). Specifically, P1 and P2 are obtained by taking the partial derivative of Equation (2.6) with respect to $E_0[\mathrm{DIV}_{T+t}]$ and $E_0[\mathrm{EARN}_{T+t}]$, respectively. Hence, the OJ model implies P1 and P2. The converse is also true. In particular, consider the general 3×3 matrix $[\omega_{\mathbf{i,j}}]$ for which the element in row i and column j equals $\omega_{\mathbf{i,j}}$. Next, assume that: (1) the DDM implied by the general matrix has a finite value; (2) $\omega_{22} < R$; and, (3) both P1 and P2 hold. Then, as shown in Proposition 5.1 of OG, $[\omega_{\mathbf{i,j}}] = \mathbf{A}$, which is the OJ model.

In light of the above, within the context of the OJ model, earnings have desirable properties and these properties are *earnings* properties — that is, it is nonsensical to assume that they would describe the temporal behavior of particular line item on the income statement or a non-financial metric such as the shoe size of the chief executive officer, etc. That said, P1 and P2 do not provide specificity about the underlying accounting-measurement rules. For example, either unbiased or conservative accounting-measurement rules are permissible. Hence, the word *earnings* is appropriate and not just a label. However, whether, for the purpose of estimating value, there is a "best" earnings metric, remains an empirical question.

Q6: *Does the residual income valuation, RIV, model have similar properties as the OJ model?*

Yes, but only if two additional assumptions are made. Before elaborating on this answer, it is useful to state the residual income valuation, RIV, model:

$$\begin{aligned} V_0 &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{EARN}_t - r \times B_{t-1}]}{R^t} \\ &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[(\mathrm{ROE}_t - r) \times B_{t-1}]}{R^t} = B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{RI}_t]}{R^t}. \end{aligned} \tag{2.8}$$

In the above equation, B_t is equity book value at date t, $\mathrm{ROE}_t = \frac{\mathrm{EARN}_t}{B_{t-1}}$ is return on equity for period t, and $\mathrm{RI}_t = \mathrm{EARN}_t - r \times B_{t-1} = (\mathrm{ROE}_t - r) \times B_{t-1}$ is residual income for period t.

The intuition behind the RIV model is as follows. Equity book value is a measure of the net resources controlled by the entity and, *ceteris paribus*, higher net resources imply higher expected future income. This implies that current equity book value is informative about expected future earnings and, thus, a useful starting point for determining contemporaneous equity value. However, it is only a starting point. The reason for this is that expected future rates of return on equity — that is, expected future ROE — may differ from the required rate of return on equity, r. Hence, to arrive at equity value, the present value of expected future *residual* income must be added to current equity book value.

Not only is the RIV model intuitive, when two additional assumptions hold, it is equivalent to the OJ model (i.e., Equation (2.5)) and shares all of that model's properties — that is, DPI, P1, and P2. The first assumption is that there is clean-surplus accounting in expectation. This occurs when $E_0[\mathrm{EARN}_t] = E_0[\Delta B_t + \mathrm{DIV}_t]$, which, in turn, implies that $E_0[\Delta\mathrm{RI}_{t+1}] = E_0[\mathrm{AEG}_t]$. The second assumption is that there exists a date Y such that $\forall t > Y, E_0[\mathrm{RI}_t] = E_0[\mathrm{RI}_{t-1}] \times G_{\mathrm{RI}}$ and $G_{\mathrm{RI}} < R$.

The two assumptions described earlier imply that $E_0[\mathrm{AEG}_t] = E_0[\mathrm{AEG}_{t-1}] \times G_{\mathrm{RI}} \forall t > Y$ but the converse is not true. That is, clean-surplus accounting and constant growth in RI imply constant growth in AEG but constant growth in AEG does not imply either clean surplus accounting or constant growth in RI. Hence, these two assumptions are restrictive. As discussed in Penman (2005), whether the implications of these restrictions are of first-order importance is an empirical question.

Q7: Why not use the discounted cash flow, DCF, model?

The primary objection to the DDM is that it requires forecasts of dividends, which violates DPI. However, as discussed in Gao *et al.* (2013), analogizing DPI to the enterprise level is nonsensical. In particular, there is no such thing as free cash flow *policy* irrelevance. In fact, there is no such thing as a free cash flow policy. Management can choose strategic policies that affect expected free cash flows. However, when there is uncertainty, managers do not choose free cash flows. Moreover, it is absurd to argue that the strategic policies chosen by management are, like dividend-policy choices, zero net present value. Hence, if we want to estimate value, why not use the DCF model? After all, it is intuitive, well known and considered the benchmark model in many valuation textbooks.

Before answering this question, it is useful to elaborate on the DCF model, which takes the following form:

$$V_0 = \mathrm{NFA}_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{FCF}_t]}{(1+\mathrm{wacc})^t} \tag{2.9}$$

In Equation (2.9), NFA_t is net financial assets at date t, which equal the difference between financial assets that are not used in the operations and financial obligations incurred in order to fund the operations. wacc is the weighted-average cost of capital; and, FCF_t is free cash flow for period t. It is the cash flow generated by the operations during period t that is free for distribution to investors. If there is clean surplus accounting, FCF_t is computed via the following formula:

$$\mathrm{FCF}_t = \mathrm{EARN}_t^{\mathrm{ENT}} - \Delta \mathrm{IC}_t \tag{2.10}$$

In the above equation, $\mathrm{EARN}_t^{\mathrm{ENT}}$ is enterprise-level earnings for period t, which equals equity-level earnings less tax-adjusted net financial income, NFI_t.[4] NFI_t is the difference between income earned on financial assets and expenses incurred on financial obligations. IC_t is invested capital

[4] Terminology varies. For example, McKinsey & Company Inc. *et al.* (2015) refer to $\mathrm{EARN}_t^{\mathrm{ENT}}$ as NOPAT_t, which is an acronym for net operating profit after tax for period t. Nissim and Penman (2001), on the other hand, refer to $\mathrm{EARN}_t^{\mathrm{ENT}}$ as OI_t, which is an acronym for operating income for period t.

at date t, which equals the difference between equity book value and net financial assets.[5] That is:

$$\mathrm{IC}_t = B_t - \mathrm{NFA}_t \tag{2.11}$$

A quick comparison of Equation (2.9) to Equation (2.8) reveals that the DCF and RIV models are similar. Specifically, each model expresses value as the sum of a contemporaneous stock variable from the balance sheet and the present value of expected future flows. This similarity is not a coincidence. Rather, it follows from the fact that the DCF model is a less-general version of the RIV model. To see this, first suppose that residual income is defined at the enterprise-level — that is, $\mathrm{RI}_t^{\mathrm{ENT}} = \mathrm{EARN}_t^{\mathrm{ENT}} - \mathrm{wacc} \times \mathrm{IC}_{t-1}$. This implies that, as discussed in Feltham and Ohlson (1995), Equation (2.8) can be re-written as follows[6]:

$$\begin{aligned} V_0 &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{EARN}_t^{\mathrm{ENT}} - \mathrm{wacc} \times \mathrm{IC}_{t-1}]}{(1+\mathrm{wacc})^t} \\ &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[(\mathrm{ROIC}_t - \mathrm{wacc}) \times \mathrm{IC}_{t-1}]}{(1+\mathrm{wacc})^t}. \end{aligned} \tag{2.12}$$

In Equation (2.12), $\mathrm{ROIC}_t = \frac{\mathrm{EARN}_t^{\mathrm{ENT}}}{\mathrm{IC}_{t-1}}$ is the return on invested capital for period t.

Second, suppose that accounting is done on a cash basis. This implies that equity book value equals net financial assets (i.e., $B_t = \mathrm{NFA}_t$). Hence, from Equation (2.11), invested capital equals zero (i.e., $\mathrm{IC}_t = 0$) and, from Equation (2.10), enterprise-level earnings equal free cash flow (i.e., $\mathrm{EARN}_t^{\mathrm{ENT}} = \mathrm{FCF}_t$). Combining these facts leads to the conclusion that, *when there is clean-surplus accounting and that accounting is done on a cash basis*, the DCF model is a special case of the RIV model. This,

[5] As shown in Nissim and Penman (2001), invested capital at date t is equivalent to net operating assets at date t, NOA_t. NOA_t equals the difference between operating assets and operating liabilities, which are assets and liabilities that result from transactions and events that relate to the entity's value proposition.

[6] As shown in OG, it is also possible to express the AEGV model using enterprise-level amounts in which case equity value takes the following form: $V_0 = \mathrm{NFA}_0 + \frac{E_0[\mathrm{EARN}_1^{\mathrm{ENT}}]}{\mathrm{wacc}} + \sum_{t=1}^{\infty} \frac{E_0[\Delta\mathrm{EARN}_{t+1}^{\mathrm{ENT}} - \mathrm{wacc}\times(\mathrm{EARN}_t^{\mathrm{ENT}} - \mathrm{FCF}_t)]/\mathrm{wacc}}{(1+\mathrm{wacc})^t}$.

in turn, implies that, similar to the RIV model, if there exists a date Z such that $\forall t > Z, E_0[\mathrm{FCF}_t] = E_0[\mathrm{FCF}_{t-1}] \times G_{\mathrm{FCF}}$ and $G_{\mathrm{FCF}} < R$, the DCF model is equivalent to the OJ model and has its properties — that is, DPI, P1, and P2.

In light of the above, "Why not use the DCF model?" is an empirical question. With this question in mind, I turn to the empirical evidence.

2.5 Empirical evidence

As discussed earlier, if some relatively innocuous assumptions are made, EARN is a fundamental economic variable. In particular, forecasts of EARN enter directly into the valuation formula; and, developing these forecasts is not the penultimate step in arriving at forecasts of dividends. Rather, it is the ultimate step. Hence, EARN can be used to develop a rigorous estimate of value while avoiding the dividend conundrum. Moreover, EARN has the properties of an earnings measure — that is, the word *earnings* is not just a label.

The above analytical results are fundamental and profound. They establish that the OJ model is the benchmark valuation model. That said, an issue remains: Is accrual-accounting earnings a better measure than free cash flow? As discussed in the previous section, if there is clean-surplus accounting and there exists a date Z such that $\forall t > Z, E_0[\mathrm{FCF}_t] = E_0[\mathrm{FCF}_{t-1}] \times G_{\mathrm{FCF}}$ and $G_{\mathrm{FCF}} < R$, the DCF model is a special case of the OJ model. Hence, when these assumptions hold, valuations based on either free cash flows or accrual-accounting earnings are equally valid. If this is the case, the motivation for studying accrual accounting and approaches to forecasting accrual-accounting earnings is less obvious. Colloquially speaking, if cash flows are just as good as accrual-accounting earnings, why not just focus on cash flows? After all, cash flow can be objectively calculated whereas accrual accounting requires the use of subjective estimates that are made by imperfect managers who have incentives to manipulate the reported results.

The above is an empirical question and it is the primary motivation underlying many empirical studies. In this section, I discuss three key empirical studies that compare earnings and cash flows within a valuation context. Before doing this, however, I take a step back and

I provide *a priori* arguments for why, within the context of valuation, accrual accounting is more useful than cash-basis accounting. After describing my arguments, I turn to the empirical evidence.

2.5.1 A priori arguments for accrual accounting

A defining feature of accrual accounting is that, if an investment is expected to generate future economic benefits, expenditures on that investment are initially recognized as assets on the balance sheet, and then these assets are subsequently expensed. For example, when an investment in equipment is made, the cost of the equipment is recognized as an asset. Subsequently, the carrying amount of the asset is reduced by depreciation, which is an expense on the income statement. If the depreciation expense is recognized in the same reporting periods in which revenues from the investment are earned and recognized, there is matching. Note that for matching to be achieved it must be that: (1) related revenues and expenses are matched together — that is, they are recognized in the same period — *and* (2) this period matches the period when revenue is earned.

The above implies that, accrual accounting leads to: (1) a balance sheet that is a good indicator of the net resources controlled at the end of the period and (2) a periodic earnings number that is a good indicator of economic performance for the period. This is not the case when cash-basis accounting is used. The reasons for this are twofold. First, under cash-basis accounting, all expenditures are immediately expensed; and, given there is typically a lag between the period when an investment is made and the dates when it generates revenues, matching does not occur. Second, under cash-basis accounting, revenues are recognized when cash is collected, which need not be, and often is not, the same period in which the revenue is earned.

Within the context of forecasting and valuation, the above has two implications. First, investors can analyze accrual-accounting numbers and link them to the "fundamentals." This can be done both historically and prospectively. For example, financial ratios reflect margins, cost structure, working capital management, productivity, sales growth, etc. Hence, these ratios can be linked to specific past actions, events, and circumstances — that is, the fundamentals. More importantly, they can also be linked to expected future fundamentals. Consequently, investors can link their expectations about future fundamentals to expected future ratios. These expected future ratios can then be converted into forecasts of future accounting numbers, which, in turn, can be directly input into either the AEGV model or the RIV model.[7] This process is at the heart of financial statement analysis.

Second, if there is clean-surplus accounting, $\Delta B_t = (\text{EARN}_t - \text{DIV}_t)$ and $\text{AEG}_t = \Delta\text{EARN}_{t+1} - r \times \Delta B_t = (\text{RONEC}_{t+1} - r) \times \Delta B_t$; and, $\text{RONEC}_{t+1} = \frac{\Delta\text{EARN}_{t+1}}{\Delta B_t}$ is the return on new equity capital for period $t + 1$.[8] Consequently, if there is matching, AEG_t reflects the amount of value created (or destroyed) by new investments made in period t; and, it is reasonable to assume that there exists a date T such that $\forall t > T, E_0[\text{AEG}_t] = E_0[\text{AEG}_{t-1}] \times G_{\text{AEG}}$ and $G_{\text{AEG}} < R$. In fact, if there is matching, it is often reasonable to assume that $G_{\text{AEG}} < 1$. The reason for this is that, given the information available at date zero, it is often reasonable to expect that diminishing marginal returns and product-market competition will cause value-creation potential to dissipate over the long run.

The above is crucial because it implies that, *to the extent there is matching*, the terminal value correction that is the centerpiece of the OJ model shown in Equation (2.5) is reasonable.[9] On the other

[7]If there is clean-surplus accounting, from Equation (2.10), forecasts of future financial ratios can also be used to arrive at forecasts of future free cash flows, which can be directly input into the DCF model. Hence, an argument can be made that there is nothing special about accrual accounting, the AEG model or the RIV model. This argument is self-defeating, however. The reason for this is that it would be quite unusual to argue that free cash flows are forecasted directly. Rather, forecasts of free cash flows are typically imputed from forecasted accrual-accounting balance sheets and income statements. Hence, forecasts of free cash flows require forecasts of accrual-accounting numbers but the converse is not true.

[8]Similarly, if there is clean-surplus accounting, $\Delta\text{IC}_t = (\text{EARN}_t^{\text{ENT}} - \text{FCF}_t)$ and $\text{AEG}_t^{\text{ENT}} = \Delta\text{EARN}_{t+1}^{\text{ENT}} - \text{wacc} \times \Delta\text{IC}_t = (\text{RONIC}_{t+1} - \text{wacc}) \times \Delta\text{IC}_t$; and, $\text{AEG}_t^{\text{ENT}}$ is enterprise-level abnormal earnings growth and $\text{RONIC}_{t+1} = \frac{\Delta\text{EARN}_{t+1}^{\text{ENT}}}{\Delta\text{IC}_t}$ is the return on new invested for period $t + 1$.

[9]Note that the phrase "to the extent there is matching" is important. Accounting measurement rules clearly matter; and, as discussed in Section 7, biased accounting rules imply that there is no matching.

hand, there is no obvious reason to expect that FCF will eventually grow at a rate $G_{\mathrm{FCF}} < R$. If there is clean-surplus accounting, from Equation (2.10), FCF_t grows at a constant rate whenever $\mathrm{EARN}_t^{\mathrm{ENT}}$ and IC_t both grow at the same rate. However, neither diminishing marginal returns nor product-market competition imply that this will happen. These phenomena impact the *value* of investments but not necessarily the *amount* of investment. Economic theory provides no guidance regarding the amount of zero-net-present-value investments that will be made. Rather, these investments are value neutral, which implies that they are irrelevant and that forecasts of them are arbitrary.

2.5.2 The empirical evidence

I focus on three emperical studies that establish fundamental results. My discussion of each of these studies is terse. The reason I adopt such a narrow focus is that there are a number of excellent literature reviews available. I choose to be terse partially for the sake of brevity and partially out of deference to the quality of the original papers.

The first two studies I discuss are Ball and Brown (1968) and the extension of it by Nichols and Wahlen (2004). Ball and Brown (1968) is a seminal paper. It establishes that firms that report increases (decreases) in earnings have positive (negative) contemporaneous stock returns. In their extension of Ball and Brown (1968), Nichols and Wahlen (2004) show that portfolios that consist of long (short) positions in firms that report increases (decreases) in earnings generate higher contemporaneous stock returns than portfolios that consist of long (short) positions in firms that report increases (decreases) in cash flows.

Assuming market efficiency, positive (negative) stock returns imply value creation (destruction). Hence, the results described above establish that accrual-accounting earnings are informative about contemporaneous changes in value and that they are more informative than contemporaneous cash flows. These results are fundamental but they are not the last word. The reason for this is that they only establish that accrual-accounting earnings are informative about changes in value. They do not shed light on whether accrual-accounting earnings are the payoff that investors forecast when estimating value. The study that does this is Penman and Sougiannis (1998).

The question asked by Penman and Sougiannis (1998) can be paraphrased as follows. Suppose investors are given a choice: They will be told in advance what a particular payoff (i.e., dividends, residual income or free cash flows) will be for the next F years. However, they have to pick the payoff. For example, if they choose to be told what future dividends will be, they will not be told what either future residual incomes or future free cash flows will be. Hence, the choice boils down to whether they prefer to use a finite-horizon version of the DDM, RIV model or DCF model.

The above question is both fundamental and practical. The reason it is fundamental is that, when forecasts can be developed over an infinite horizon, the DDM (i.e., Equation (2.1), AEGV model (i.e., Equation (2.2)), RIV model (i.e., Equation (2.8)) and DCF model (i.e., Equation (2.9)) yield identical value estimates. Hence, the fundamental question is: Which model is the most accurate when the forecast horizon is truncated? This is also a very practical question because it is impossible to develop explicit forecasts over an infinite horizon. Hence, truncating the forecast horizon is an unavoidable and necessary part of estimating value.

Penman and Sougiannis (1998) evaluate the above question by using average *ex post* realizations of dividends, free cash flows and GAAP earnings to develop portfolio-level value estimates per the DDM, DCF model and RIV model, respectively. They then compare these value estimates to *ex ante* equity market value and they show that the estimates based on the RIV model are more accurate than estimates based on either the DDM or DCF model. This is an important result because it is consistent with the conclusion that accrual-accounting earnings is the payoff that investors forecast when estimating value.

2.6 Summary

The main point of this section is straightforward: Accrual-accounting earnings are a fundamental economic variable and forecasting them is a central part of valuation. This conclusion is based on both analytical and empirical evidence. First, regarding the analytical evidence, models based on relatively innocuous assumptions lead to the conclusion that an earnings metric is the payoff that investors forecast when estimating value. Given dividend policy irrelevance, forecasting dividends is futile. Nonetheless, investors must forecast *something* and that something has earnings properties — that is, it is an earnings metric.

Second, although free cash flow qualifies as an earnings metric under certain conditions, *a priori* reasoning and the empirical evidence implies that it is inferior to accrual-accounting earnings. When compared to cash flow, accrual-accounting earnings is a better indicator of the contemporaneous change in value. Moreover, and more important, extant empirical evidence supports the conclusion that accrual-accounting earnings, not dividends or free cash flows, is the payoff that investors forecast when estimating value.

3

Selecting an Earnings Metric

The key point of the previous section is that, within the context of valuation, expected future earnings is the fundamental payoff of interest. Consequently, to understand value and valuation, we must understand earnings forecasting. This leads to a central question: Are historical accounting numbers useful for developing forecasts of future earnings? In the next five sections, I focus on this question. However, before doing that I first discuss how to empirically measure earnings. Hence, in this section, I discuss the choices a researcher makes when choosing an earnings metric. I first enumerate the key choices, and then I discuss each in detail. From a bird's-eye perspective, the four choices are whether to:

1. Evaluate an enterprise- or equity-level metric.
2. Evaluate comprehensive income or income per the income statement.
3. Evaluate earnings, abnormal earnings growth or residual income.
4. Evaluate profitability ratios or unscaled earnings.

3.1 Enterprise- versus equity-level

Enterprise-level earnings for period t, $\mathrm{EARN}_t^{\mathrm{ENT}}$ are the earnings that the firm would report if: (1) it held no financial assets and (2) issuances of common shares were its only source of external funding. Equity-level earnings, EARN_t, equal reported net income less preferred dividends for the period. Hence, enterprise-level earnings equal reported net income less tax-adjusted net financial income. Net financial income is the difference between income arising from financial assets (e.g., capital gains earned from selling government bonds) and expenses arising from financial obligations (e.g., interest expense on a loan from a bank). Tax-adjusted refers to the fact that each component of net financial income is reduced by the effect it has on tax expense.

The primary disadvantage of evaluating equity-level earnings is that they are affected by financing choices. A bedrock principle of finance is that, from a first-order perspective, the value of an entity's operations does not depend either on how the operations are funded or when the free cash flow generated by them is paid out to investors (Miller and Modigliani, 1961; Modigliani and Miller, 1958). However, financing choices do affect equity-level income. Hence, in order to forecast it, an explicit or implicit forecast of net financial income has to be developed. Given that financing choices are irrelevant, they are arbitrary, which makes forecasting their effect on future earnings difficult.

Enterprise-level earnings, on the other hand, are not affected by financing choices. Hence, when forecasting them, there is no need to make arbitrary assumptions about future payout and capital structure decisions. Moreover, even if the goal is to evaluate equity value, enterprise-level earnings are still useful. As discussed in the previous section, abnormal earnings growth and residual income can be defined as a function of enterprise-level earnings, and then an enterprise-level version of either the AEGV or RIV model can be used to estimate value.

In light of the above, it may seem that enterprise-level earnings are the natural choice. However, there are two issues with using them. The first issue is that whether an activity and the financial statement line items associated with it are operating or financing depends on an entity's value proposition.[1] For example, borrowing from a bank is a financing activity for a manufacturer and an operating activity for a bank. The reason for the latter is that borrowing and lending are central to the intermediation function of banks.

The above issue complicates the process of determining enterprise-level earnings. The reason for this is that firms have complicated value propositions and financial statements line items represent the aggregation of numerous activities. Hence, when using large databases to study a large sample, it is not always possible to identify and separate financing activities from operating activities. In addition, there is some subjectivity involved in determining each firm's value proposition and identifying the activities that relate to it. For instance, financial income is income earned on financial assets, which are assets that are not used in the operations. Hence, these assets can be liquidated and the proceeds distributed to investors without affecting the value of the enterprise. This is the reason why financial assets are often referred to as excess cash. Determining the amount of excess cash and the income generated by it for a single company, let alone for all the companies in a large database, is clearly a complicated, subjective process.

The second issue is that enterprise market value is not directly observable. It can be estimated by adding the fair value of net financial assets to equity market value. However, doing this involves identifying the entity's financing activities, which, as discussed earlier, is complicated. Moreover, even if financing activities can be identified, many of them do not have observable fair values. For example, many financial obligations are reported at their amortized cost, which does not reflect changes in interest rates or credit risk that occur after the date on which the liability was initially recognized on the balance sheet.

Given the measurement issues that arise when using enterprise-level numbers, most studies choose to focus on equity-level numbers, which are observable and directly related to equity market value. Although this

[1]Value proposition refers to business activities that have the *potential* to either create or destroy value. These activities do not include obtaining funding for the purpose of executing the entity's business model or investing excess cash in financial assets. Excess cash is cash that could be distributed to the entity's investors without affecting the entity's business model.

is reasonable choice, as discussed in Easton (2016), there are substantial research opportunities regarding enterprise-level accounting numbers and their role in determining value.

3.2 Comprehensive income versus income per the income statement

Comprehensive income equals net income plus changes in equity book value that: (1) do not arise from transactions with shareholders and (2) are not reflected on the income statement. For example, at present, from both GAAP and IFRS, net income does not reflect the effects of translating a foreign subsidiary's assets and liabilities from the foreign currency to the presentation currency. Rather, amounts related to these effects are included in other comprehensive income. Amounts included in other comprehensive income are often referred to as dirty-surplus items because they nullify the clean–surplus relation — that is, the change in equity book value no longer equals the difference between net income and dividends.

It is important to note that dirty-surplus items can relate to either operating or financing activities. Hence, even if enterprise-level earnings are evaluated, the issue of how to deal with dirty-surplus items remains. At present, most studies ignore them. The argument for this choice is that they have an expected value of zero and can exhibit high volatility. For example, if there are no arbitrage opportunities in trading foreign currencies, foreign exchange rates and the translation effects arising from them have an expected value of zero. Nonetheless, *ex post* exchange rate fluctuations will affect comprehensive income and that effect can be nontrivial. Hence, forecasts based on and of comprehensive income will be unnecessarily noisy.

As discussed in Chapter 10 of OG, if dirty-surplus items have an expected value of zero, ignoring them is a valid research-design choice. However, a related choice is not valid: The choice to ignore "special" items such as asset impairments, restructuring charges, etc. Although these items are difficult to predict, they are not inherently unpredictable. This issue is important within the context of comparing earnings forecasts obtained from econometric models to analysts' forecasts. As discussed in Bradshaw and Sloan (2002), security analysts typically forecast pro-forma numbers that exclude special items. Hence, analysts' forecasts are not directly comparable to forecasts of equity-level earnings generated by econometric models. Rather, *ceteris paribus*, analysts' forecast errors, which equal the difference between the forecast and the realized pro-forma amount, will be more accurate than the forecast errors generated by the econometric model, which equal the difference between the forecast and the realized equity-level earnings that include special items.

In light of the above, in order to compare equity-level earnings generated by econometric models to analysts' forecasts an adjustment must be made. There are two ways to do this. The first way is to develop an econometric model that generates forecasts of the pro-forma earnings number that analysts forecast, and then compute forecast errors by comparing all of the forecasts to the realized pro-forma number. The second way is to continue to use an econometric model that generates forecasts of equity-level earnings and compute forecast errors by comparing all of the forecasts, including those made by analysts, to realized equity-level earnings.

The advantage of the first way is that all the forecasts and forecast errors are comparable. Hence, unbiased evidence about relative forecast accuracy is provided. The advantage of the second way is that it provides evidence about the amount of predictable, value relevant information that is omitted from analysts' forecasts. That is, because analysts ignore special items, *ceteris paribus*, their forecasts of equity-level earnings are less accurate and this loss of accuracy will be reflected in the relative forecast errors. Given that each approach provides valuable evidence, a reasonable alternative is to use both.

3.3 Earnings versus abnormal earnings growth versus residual income

As discussed in the previous section, value can be expressed as a direct function of either abnormal earnings growth, AEG, or residual income, RI. Hence, the motivation for focusing on either AEG or RI is clear: these payoffs, not earnings *per se*, are the payoffs that investors forecast.

That said, most studies focus on forecasting earnings. The reason for this is that, unlike either AEG or RI, earnings are not a function of the discount rate, which is unobservable and difficult to estimate. Hence, earnings forecasts are not affected by measurement errors embedded in the empirical estimates of the discount rate.

The above choice is reasonable. However, there is an alternative that may make sense in certain circumstances: Use the risk-free rate as the discount rate for determining either AEG or RI. This may seem *ad hoc* but, as shown in Feltham and Ohlson (1999) and Christensen and Feltham (2009), the opposite is the case. In particular, the models discussed in Section 2 are *ad hoc* because they are based on the assumption that the discount rate is neither stochastic nor time varying. If these assumptions are relaxed and a general approach is taken, the adjustment for risk is not embedded in the denominator via the discount rate. Rather, the adjustment is added to the numerator. For example, Feltham and Ohlson (1999) show that, within the context of the residual income valuation model, equity value at time zero equals[2]:

$$\begin{aligned} V_0 &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{EARN}_t - r^f_{t-1,t} \times B_{t-1}]}{R^f_{0,t}} \\ &\quad + \sum_{t=1}^{\infty} \mathrm{COV}_0(m_{0,t}, (\mathrm{EARN}_t - r^f_{t-1,t} \times B_{t-1})). \end{aligned} \tag{3.1}$$

In Equation (3.1), $r^{\mathrm{f}}_{t-1,t}$ is the *uncertain* risk-free return from date $t-1$ to date t, $R^{\mathrm{f}}_{0,t} = (1 + r^{\mathrm{f}}_{0,t})$ and $m_{0,t}$ is the stochastic discount factor for period t.[3] The remaining variables are defined in Section 2.

[2]This result is not unique to the residual income valuation model and it also obtains at the enterprise level. In particular, Christensen and Feltham (2009) show that the abnormal earnings growth valuation model can be expressed as a function of: (1) expected year-ahead earnings divided by $r^{\mathrm{f}}_{0,1}$; (2) the present value of subsequent expected abnormal earnings growth discounted at the risk-free rate; and, (3) an additive risk adjustment that reflects the covariances between future abnormal earnings growth and future stochastic discount factors. Moreover, abnormal earnings growth is defined as a function of the risk-free rate. Christensen and Feltham (2009) also show that the approach is valid at both the equity and enterprise level.

[3]Note that for values of $t > 1$, $r^{\mathrm{f}}_{t-1,t}$ is a future amount that is not known at date zero so there is uncertainty about what it will be. Consequently, $r^{\mathrm{f}}_{t-1,t}$ is the *uncertain* risk-free return. $R^{\mathrm{f}}_{0,t}$, on the other hand, is known at date zero.

Equation (3.1) demonstrates that value is a function of: (1) the present value of expected residual income discounted at the risk-free rate and (2) an additive risk adjustment that reflects the covariances between future residual incomes and future stochastic discount factors. Moreover, residual income is a function of the *uncertain* risk-free rate, which is observable *ex post*. Consequently, historical values of residual income are observable and can be used to develop forecasts of future residual income.

The above implies that, when measured using the risk-free rate, historical AEG and RI are value-relevant variables that can be observed without error. Hence, in certain research contexts, focusing on raw earnings may be too restrictive. Rather, it may be more appropriate to focus on forecasting either AEG or RI. Examples of studies that do the latter include Baginski and Wahlen (2003), Nekrasov and Shroff (2009), and Bach and Christensen (2016).

3.4 Profitability ratios versus unscaled earnings

The two most commonly used profitability ratios are: (1) return on equity, ROE, and (2) return on invested capital, ROIC. ROE_t is the ratio of equity-level earnings for period t to either equity book value at date $t-1$ or the average of equity book value at date t and date $t-1$. ROIC_t equals the ratio of enterprise-level earnings for period t to either invested capital at date $t-1$ or the average of invested capital at date t and date $t-1$. Invested capital equals equity book value minus net financial assets.

There are two related advantages to evaluating profitability ratios vis-à-vis unscaled earnings. First, ratios are intuitive and reflect how financial statement analysis is practiced and taught. For example, as discussed further in Section 6, ROE can be decomposed into ROIC and a component that reflects the effects of financial leverage. ROIC can then be decomposed into ratios that reflect operating performance such as gross profit margin, expense ratios, working capital ratios, invested capital turnover, etc.

Second, ratios are not affected by scale, which implies that they are more comparable across firms and time. This is useful both from a fundamental analysis perspective and a statistical perspective. In particular, the moments — that is, the mean, variance, skewness, etc. — of unscaled earnings will vary with firm size whereas the moments of profitability ratios such as ROE and ROIC may be similar across firms and time. This implies that statistical models may generate more accurate predictions when they are used to predict ratios vis-à-vis when they are used to predict unscaled earnings. In addition, forecasts of ratios will generate errors that are more comparable than the errors obtained from forecasts of unscaled earnings.

The second advantage — that is, eliminating scale effects — is also the main disadvantage of using ratios. The reason for this is that value is a function of scale. Hence, if the objective of developing earnings forecasts is to use them to estimate value, forecasts of ratios are insufficient. To understand this, it is useful to consider the following version of the residual income valuation model.

$$\begin{aligned} V_0 &= B_0 + \sum_{t=1}^{\infty} R^{-t} \times E_0[(\mathrm{ROE}_t - r) \times B_{t-1}] \\ &= B_0 + \sum_{t=1}^{\infty} R^{-t} \times E_0[\mathrm{ROE}_t - r] \times E_0[B_{t-1}] \\ &\quad + \sum_{t=1}^{\infty} R^{-t} \times \mathrm{COV}_0((\mathrm{ROE}_t - r), B_{t-1}). \end{aligned} \tag{3.2}$$

From Equation (3.2), equity value is a function of: (1) expected future residual ROE, which is the difference between ROE and the discount rate; (2) expected future equity book value; and, (3) the covariances between future residual ROE and future equity book value. Moreover, evaluating the value-to-book multiple (i.e., $\frac{V_0}{B_0}$) does not solve the problem because, in addition to expected future residual ROE, $\frac{V_0}{B_0}$ is also a function of: (1) expected future compound growth rates in equity book value and (2) the covariances between these growth rates and future excess ROE.

The above does not imply that studying how to forecast profitability ratios is either inappropriate or uninteresting. Rather, it implies that the choice of earnings metric should be motivated by both the research context and econometric issues. It also implies that studies of how to forecast growth in equity book value and its covariance with residual ROE may provide useful insights that complement studies that focus only on forecasting ROE.

3.5 Summary

I summarize by making some general comments. First, when selecting an earnings metric, the primary concern is choosing the metric that best fits the research question. Second, notwithstanding the previous comment, choosing an earnings metric involves making tradeoffs. For example, the metric that best fits the research question may be unobservable or have undesirable econometric properties. To see this, consider enterprise-level residual income. To calculate it, assumptions about what constitutes an operating activity have to be made and the discount rate must be estimated. Moreover, enterprise-level residual income is affected by scale. Consequently, in certain circumstances enterprise-level residual income may be the variable that best matches the research question but another metric (e.g., return on equity) may be the best choice.

Finally, the tradeoffs described above are subjective. Consequently, best practice is to clearly motivate the research question, describe the logic for selecting a particular metric or metrics, and then discuss the consequent pros and cons.

4

The Role of Econometric Modeling

"*All models are wrong, but some are useful.*"

— George E. P. Box,
Robustness in the strategy of scientific model building.

Once an earnings metric has been selected, an obvious question arises: What is the best approach for forecasting it? In the next two sections, I discuss the use of time-series models (Section 5) and panel-data approaches (Section 6), both of which involve econometric modeling. Before doing this, I take a step back and address a broader question: Why use econometric models? Isn't earnings forecasting a highly-contextual process that involves the use of judgment? Isn't the search for a best approach to forecasting earnings a fool's errand? Doesn't econometric modeling oversimplify the forecasting process? Finally, when using econometric models don't we run the risk of over-emphasizing them and, thus, erroneously concluding that we have found the best approach?

The answer to the last four questions at the end of the previous paragraph is an unqualified *yes*. *Yes*, earnings forecasting is contextual and requires judgment. *Yes*, there either is no best approach or, if one exists, it cannot be perfectly replicated by an econometric model. *Yes*, econometric models are simplifications. Finally, *yes*, we do run the risk of over-emphasizing econometric models and over-generalizing results obtained from them.

In light of the above, some discussion of the role of econometric modeling is warranted. In this short section, I briefly address that issue. I am intentionally brief because providing a comprehensive discussion of econometrics and its general role in the knowledge-accumulation process is well beyond the scope of this monograph. Rather, I make the following four points:

1. Econometric modeling helps to bring order out of confusion.

2. Econometric models are objective and replicable.

3. Forecasts from econometric models serve as useful benchmarks.

4. The goal is to find a useful model not the best model.

4.1 Bringing order out of confusion

The primary objective of academic research is to create knowledge. This involves bringing order out of confusion. To do that successfully, the default assumption must be that there is some underlying order and that it is possible to invent a technology that will allow us to objectively observe and describe its essential characteristics. This assumption may be incorrect but before drawing that conclusion we must first convince ourselves that we have considered all reasonable approaches and that the issue is too unstructured — that is, too contextual or too subjective.

With the above in mind, I make two comments. First, although there is a substantial amount of evidence about earnings forecasting, the extant evidence is not exhaustive. More importantly, we are not at the point where we can conclude that the earnings-forecasting process is so contextual and subjective that it cannot be systematically characterized. We may never be able to fully characterize it. Nonetheless, given the importance of value and the central role that earnings forecasting plays in the valuation process, the present gap between what we do know and what we don't know is too large to be ignored.

Second, a basic premise underlying financial statement analysis is that historical accounting numbers are informative about future earnings. However, at present, the empirical evidence regarding the link between historical accounting numbers and future earnings is not well understood. Hence, further research is necessary; and, econometric modeling is the state-of-the-art technology for conducting this research. One reason for this is that econometric modeling involves making simplifying assumptions. Hence, identifying a useful econometric model involves identifying and formally modeling the essential phenomena — that is, when done well, econometric modelling brings order out of confusion.

4.2 Objectivity and replicability

Within the context of forecasting, a key part of econometric modeling is writing mathematical formulas that describe: (1) the link between historical numbers and future realizations and (2) how to evaluate the quality of the forecasts. This implies that the analytical properties of and empirical evidence generated by a particular model can be objectively evaluated. Consequently, the limitations of the model can be discovered, and then improvements can be made. This process of discovery, improvement, discovery, improvement, *ad infinitum* is a crucial part of knowledge accumulation.

4.3 Benchmarking

As discussed earlier, econometric modeling involves making simplifying assumptions. Hence, if context and subjectivity are of first-order importance, forecasts from econometric models will be significantly less accurate than forecasts developed by rational, properly incentivized agents with adequate training and experience. However, if the forecasts from the econometric model are more accurate or, if the difference in accuracy is small, something is amiss. For example, the agents making the forecasts may have behavioral biases that cloud their judgment, may not be incentivized to forecast accurately or may lack sufficient training and experience. Note that it may seem that benchmarking against econometric models is akin to "attacking a straw man." However, extant evidence provided by Bradshaw *et al.* (2012) suggests otherwise. They show that, when long horizons of a year or more are considered, forecasts obtained from the random-walk model are as accurate as analysts' forecasts. Moreover, although they find that analysts' forecasts are more accurate than forecasts obtained from the random-walk model over horizons of several months or less, the economic magnitude of the difference in accuracy is small. These are provocative results.

4.4 Usefulness is the goal

No model can every fully describe the optimal approach for forecasting earnings. Hence, identifying the best model is not the goal of econometric modeling. Rather, the goal is to develop models that are useful. Usefulness is contextual so it is infeasible to provide a general description of it. That said, there are two contexts that immediately come to mind: (1) empirical capital markets research and (2) the practice of valuation.

A substantial amount of empirical capital markets research focuses on either learning about the role of accounting numbers in the valuation process (e.g., studies of earnings response coefficients) or using these numbers to estimate key valuation parameters (e.g., the implied cost of capital, ICC). Hence, earnings forecasts are often a key variable of interest; and, given the desire for generalizability, large samples are studied. Within this context, a useful earnings forecasting model is one that is objective, replicable and can be used to generate accurate forecasts for a large sample at a low cost.

Within the context of the practice of valuation, usefulness comes in two forms. First, as discussed above, benchmarking experts' forecasts against model-based forecasts is a way of identifying experts who are either biased, poorly trained, lack proper incentives or are too inexperienced. Hence, within the context of benchmarking, usefulness is defined similarly as it is within the context of empirical capital markets research. That is, a useful model is one that is objective, replicable and generates accurate forecasts for a large sample at a low cost.

Second, a useful model is one that guides best practice. That is, it sheds light on the key issues that need to be considered and provides guidance regarding the best way to deal with them. For example, consider peer analysis. Conventional wisdom is that it is a key part of the forecasting process. However, important questions remain unanswered including: Does it lead to more accurate forecasts? If so, what are the key characteristics — for example, industry, age, lifecycle, etc. — that analysts should consider when forming peer groups? What is the best way to use the financial results of the peer companies? For instance, is the the unconditional mean of their historical earnings a sufficiently accurate forecast or is it better to take temporal trends and cross-sectional relations into account, and then use the conditional mean?

Within the contexts described earlier, econometric modeling has tremendous potential. Econometric models are objective and replicable; and, given that powerful computers and sophisticated statistical software are inexpensive, they can be used to generate forecasts for a large sample at a low cost. Furthermore, because bringing order out of confusion is a key part of econometric modeling, econometric models have a role to play to guiding best practice.

Whether econometric models generate accurate forecasts and provide useful guidance about best practice remain open questions, however. As discussed in the next two sections, extant econometric models often perform worse than or, at best, not much better than the random-walk model. Taken at face value, this result implies that a firm's current earnings is the best predictor of its future earnings. This is a provocative result because it leads to the seemingly absurd conclusion that, within the context of forecasting earnings, there is no value to peer analysis, trend analysis and using conditioning information.

Finally, the fact that extant econometric models do not outperform the random-walk model does not imply that we abandon econometric modeling. To the contrary, it implies that we need to do more research with the goal of discovering models that are useful.

4.5 Summary

Earnings forecasting lies at the heart of valuation, which is a central economic activity. Hence, studying earnings forecasting is important and econometric modeling is an indispensable, albeit imperfect, research tool. It is an objective, replicable process that helps bring order out of confusion. Hence, it has the potential to be useful. However, as discussed in the next two sections, extant models appear to be too inaccurate and provide seemingly absurd guidance about best practice. This implies that, in the spirit of the epigraph at the beginning of the section: Not only are all the models wrong, none of the present models are useful — at least, not useful enough.

5

Time-series Models

In this section, I focus on the relation between historical earnings and future earnings — that is, the time-series properties of earnings. I begin by describing autoregressive, integrated, moving-average (i.e., ARIMA) models. I then discuss a key extant result: Of the ARIMA models evaluated, the random-walk model generates the most accurate out-of-sample forecasts of either unscaled or scaled earnings. I then argue that, although well-accepted, this result is misleading. There are a number of reasons for this argument but one reason is overarching: ARIMA models are ill-suited for developing forecasts of earnings.

Given that the main conclusion of this section is that ARIMA models are ill-suited for developing forecasts of earnings, it may be unclear why it is necessary to discuss them. My reasons are threefold. First, using historical earnings to forecast future earnings is a natural thing to do. Second, there is a well-developed literature regarding the properties of earnings forecasts obtained from ARIMA models. Moreover, this literature remains influential. For example, as discussed above, the conclusion that earnings follow a random walk remains well accepted. Finally, forecasts from ARIMA models provide a useful benchmark for comparison to forecasts obtained from other sources and statistical approaches. For example, a useful question to ask when evaluating a forecast is whether it is more accurate than a forecast obtained from the random-walk model. If it is not, the validity of the forecast and the approach used to develop it come into question.

5.1 An overview of ARIMA models

Within the context of earnings, an ARIMA model is a set of assumptions about the temporal relation between earnings and random *shocks* to earnings, which are sometimes referred to as earnings innovations. I denote the shock to earnings at time t as ε_t. A standard assumption is that the sequence $\{\varepsilon_t\}_{t=-\infty}^{+\infty}$ follows a white noise process. That is, $\forall t$, the mean and variance of ε_t are zero and σ^2, respectively, and ε_t is uncorrelated with $\varepsilon_\tau \forall \tau \neq t$. Hence, earnings shocks relate to unpredictable fluctuations that are attributable to unforeseen changes in customers' tastes and preferences, the level of competition, the costs of the factors of production, etc.

A general ARIMA model is a combination of two processes: (1) a pth-order autoregressive (i.e., AR(p)) model and (2) a qth-order moving-average (i.e., MA(q)) model. The letters p and q denote the number of lagged values included in the model. For example, if the *level* of earnings follows an AR(p) process, Equation (5.1) applies. On the other hand, if the *change* in earnings follows an MA(q) process, Equation (5.2) applies.

$$E_t[\text{EARN}_{t+1}] = c + \sum_{\tau=1}^{p} \phi_\tau \times \text{EARN}_{t+1-\tau} \tag{5.1}$$

$$E_t[\Delta\text{EARN}_{t+1}] = \mu - \sum_{\tau=1}^{q} \theta_\tau \times \varepsilon_{t+1-\tau} \tag{5.2}$$

In Equations (5.1) and (5.2), c and μ are constants and ϕ_τ and θ_τ are the τth-order autoregressive and moving-average coefficients, respectively. Because Equation (5.1) relates to the level of earnings, the order of integration is zero and earnings follows an ARIMA($p, 0, 0$) process. However, Equation (5.2) relates to the first difference of earnings; hence, the order of integration is 1 and earnings follows an ARIMA(0,1,q)

process. More generally, if the variable obtained after differencing earnings d times follows a process that is a combination of an AR(p) process and an MA(q) process, the order of integration is d and earnings follows an ARIMA(p, d, q) process.

Although ARIMA(p, d, q) models are complex in general, the models that are used in practice tend to be fairly simple. For example, when modeling annual earnings, it is often the case that when p is positive q is not and *vice versa*. Moreover, when $p(q)$ is non-zero, it is rare for it to be greater than 2. With regards to the order of integration, d, it is typically either zero or one. Consequently, as discussed below, a fairly complete understanding of ARIMA models within the context of forecasting annual earnings can be obtained by considering two simple models: (1) the ARIMA(0,1,1) model and (2) the ARIMA(1,0,0) model.

5.2 The ARIMA$(0, 1, 1)$ model

Suppose there is a history of length h — that is, earnings are observable for period $t - h$ through period t — and $E_{t-h-1}[\mathrm{EARN}_{t-h}] = \mu$, from Equation (5.2), the ARIMA(0,1,1) implies:

$$\begin{aligned} E_t[\mathrm{EARN}_{t+1}] &= \mu + \mathrm{EARN}_t - \theta \times \varepsilon_t \\ &= \mu + (1-\theta) \times \mathrm{EARN}_t + \theta \times E_{t-1}[\mathrm{EARN}_t] \\ &= \mu \times \eta + (1-\theta) \times \left\{ \mathrm{EARN}_t + \sum_{k=1}^{h} \theta^k \mathrm{EARN}_{t-k} \right\}. \end{aligned} \tag{5.3}$$

The second equality follows from the fact that $\varepsilon_t = \mathrm{EARN}_t - E_{t-1}[\mathrm{EARN}_t]$. The third equality is obtained by recursively substituting for $E_{t-\tau-1}[\mathrm{EARN}_{t-\tau}]$, assuming $\theta < 1$, and then setting the ratio of $[1 - \theta^{h+1}]/(1-\theta)$ equal to η.

If $\theta \in [0, 1]$ and $\mu = 0$, Equation (5.3) implies that there are three equivalent ways of arriving at the optimal date t forecast of earnings for period $t + 1$.

1. Assume that a fraction θ of the current shock is transitory. Hence, if current earnings are higher (lower) than expected, deduct (add) from (to) them θ of the absolute value of their unexpected

component — that is, $|\varepsilon_t|$. This is a form of an error correction model — that is, the realization is "corrected" by removing the transitory portion of the "error" from it.

2. Compute a weighted average of current earnings and the most-recent forecast of them. Hence, the forecast is a *moving average* of the lagged forecasts and the lagged realizations. This explains why ARIMA$(0, d, q)$ models are referred to as moving-average models.

3. When $\theta < 1$, assume that more recent values of earnings are more informative about future earnings, which implies that the forecast of earnings is an exponentially weighted moving average of the historical values of earnings. This is another reason why the ARIMA$(0, d, q)$ model is referred to as a moving-average model. It also illustrates that the ARIMA$(0, 1, 1)$ model is equivalent to: (1) the simple exponential smoothing (i.e., SES) model and (2) an ARIMA$(h, 0, 0)$ model in which the gth-order autoregressive coefficient equals $(1 - \theta) \times \theta^{g-1}$.

Equation (5.3) leads to the conclusion that ARIMA(0,1,1) models lie on a continuum between the "constant" model and the random-walk model. In particular, if $\theta = 0$, we obtain the random-walk model and the optimal forecast of EARN_{t+1} equals $\mu + \mathrm{EARN}_t$. Hence, the random-walk model is equivalent to assuming that: (1) none of the current shock to earnings is transitory and (2) only current earnings are informative about future earnings — that is, values of earnings realized prior to period t contain no information. On the other hand, if $\theta = 1$, we obtain the constant model in which we ignore all past values of earnings and our optimal forecast is $\mu \times (h + 2)$.

As shown in Figure 5.1, for values of θ between zero and one, ARIMA$(0, 1, 1)$ processes imply that the local trend in earnings exhibits a cyclical pattern. For unscaled earnings, this may be a reasonable approximation. For example, the use of historical cost accounting implies that positive economic news is only recognized when uncertainty is resolved. Hence, positive economic news will affect both current and subsequent earnings. Moreover, although extreme negative economic news tends to be recognized immediately, moderately negative economic news is often recognized over time. For example, decreases in demand

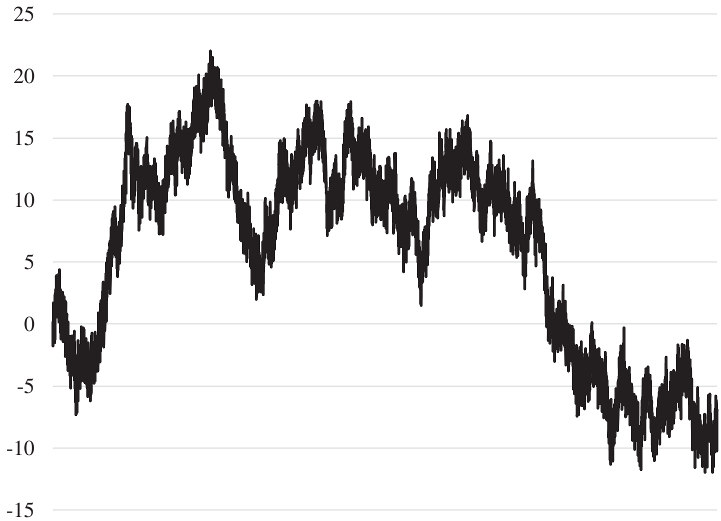

Figure 5.1: Simulation of an ARIMA$(0,1,1)$ process. A graph of EARN_{t+1} on $t+1$. $E_t[\Delta\text{EARN}_{t+1}] = \mu - \sum_{\tau=1}^{q} \theta_\tau \times \varepsilon_{t+1-\tau}$. $\mu = 0, q = 1, \theta_1 = 0.75$ and $\{\varepsilon_t\}_{t=-\infty}^{+\infty}$ follows a white-noise process and $\sigma(\varepsilon_t) = 1 \forall t$.

do not necessarily trigger asset impairments. Rather, the reduction in expected revenues leads to a reduction in investment and future earnings.

5.3 The ARIMA$(1,0,0)$ model

Suppose there is a history of length h — that is, earnings for period $t-h$ through period t are observable — and $E_{t-h-1}[\text{EARN}_{t-h}] = c$, from Equation (5.1), the ARIMA$(1,0,0)$ implies:

$$\begin{aligned} E_t[\text{EARN}_{t+1}] &= c + \phi \times \text{EARN}_t \\ &= c \times \eta + \sum_{k=1}^{h+1} \phi^k \times \varepsilon_{t+1-k} \\ &= c \times \eta + \phi \times \left(\varepsilon_t + \sum_{k=1}^{h} \phi^k \varepsilon_{t-k} \right). \end{aligned} \tag{5.4}$$

The second equality is obtained by replacing EARN_t with $\varepsilon_t + E_{t-1}[\text{EARN}_t]$, recursively substituting for $E_{t-\tau-1}[\text{EARN}_{t-\tau}]$, assuming $|\phi| < 1$, and then setting the ratio of $[1 - \phi^{h+1}]/(1 - \phi)$ equal to η.

If $\phi \in [0, 1]$ and $c = 0$, Equation (5.4) implies that there are three equivalent ways of arriving at the optimal date t forecast of earnings for period $t + 1$.[1]

1. Assume that earnings follow a mean-reverting process and that a fraction $(1 - \phi)$ of current earnings is purely transitory. This is different from the ARIMA$(0, 1, 1)$ model in which the fraction θ of the *shock* to current earnings is purely transitory.

2. When $|\phi| < 1$, assume that earnings follow an ARIMA$(0, 0, h + 1)$ process and that the gth-order moving average coefficient, θ_g equals ϕ^g.[2]

3. When $|\phi| < 1$, assume that more recent values of the *shocks* to earnings are more informative about future earnings, which implies that the forecast of earnings is an exponentially weighted moving average of the historical shocks.

Similar to ARIMA$(0, 1, 1)$ models, ARIMA$(1, 0, 0)$ models lie on a continuum between the constant model and the random-walk model. In particular, if $\phi = 1$, we obtain the random-walk model and the optimal forecast of EARN_{t+1} equals $c + \text{EARN}_t$, which, in turn, equals $c \times (h + 2)$ plus the sum of all the historical shocks. Hence, the random-walk model is equivalent to assuming that earnings shocks accumulate over time. On the other hand, if $\phi = 0$, we obtain the mean model in which we ignore all past values of earnings and our optimal forecast is c.

As shown in Figure 5.2, ARIMA$(1, 0, 0)$ processes imply that earnings follow a mean-reverting process. If the model includes a time

[1] I ignore cases in which $-1 \leq \phi < 0$. In these cases, earnings is expected to oscillate in a mean-reverting fashion around the mean (or a time trend). That is, for a firm with positive (negative) current earnings the optimal forecast is to assume that the firm will generate a loss (profit) in the subsequent year.

[2] Note that Equation (5.4) is a linear combination of the shocks whereas Equation (5.2) is obtained by subtracting a linear combination of the shocks. Hence, when interpreting the ARIMA$(1, 0, 0)$ as an ARIMA$(0, 0, h + 1)$, $(1 - \theta_g)$, not θ_g, is the fraction of the shock occurring in period $t - g$ that is transitory.

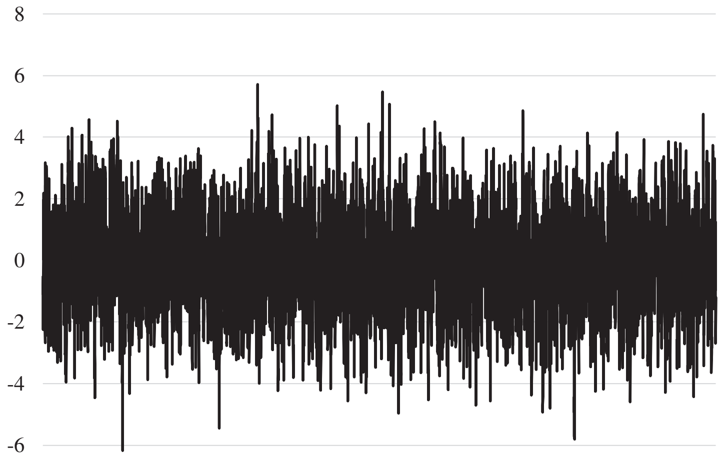

Figure 5.2: Simulation of an ARIMA$(1, 0, 0)$ process. A graph of EARN_{t+1} on $t+1$. $E_t[\text{EARN}_{t+1}] = c + \sum_{\tau=1}^{p} \phi_\tau \times \text{EARN}_{t+1-\tau}$. $c = 0, p = 1, \phi_1 = 0.75$ and $\{\varepsilon_t\}_{t=-\infty}^{+\infty}$ follows a white-noise process and $\sigma(\varepsilon_t) = 1 \forall t$.

trend, this is equivalent to assuming that earnings mean revert to the long-run trend. Although this assumption may be reasonable as a first approximation, it rules out the possibility of earnings exhibiting cyclical behavior. Hence, an ARIMA$(0, 1, 1)$ model may generate superior forecasts of dollar earnings. On the other hand, if above-average (below-average) values of profitability ratios are attributable to good (poor) economic performance, diminishing marginal returns and product-market competition (the market for corporate control and competition in the labor market) will cause the ratio to revert to the mean. This, in turn, suggests that that the ARIMA$(1, 0, 0)$ model may provide relatively accurate forecasts of profitability ratios such as ROE, ROIC and their components.

5.4 Model selection and estimation

Model selection begins by determining the order of integration — that is, the order of differencing. The objective of differencing is to generate a time series that is covariance-stationary — that is, the *unconditional* mean, variance, and autocovariances are not time varying. For example, a series that exhibits a time trend is not stationary.[3] Whether the series is stationary is typically determined by both visual inspection — that is, graphing the series on a time scale — and formal statistical testing — for example, testing for the presence of a unit root.

Once the order of integration has been determined, the model type is chosen. This is typically done via the visual inspection of graphs of autocorrelations and partial autocorrelations. For example, if the autocorrelations of the series fade to zero as the lag length increases but there is a discontinuity in the partial autocorrelation function at lag k, the series has an AR(k) signature. Alternatively, if the partial autocorrelations of the series fade to zero as the lag length increases but there is a discontinuity in the autocorrelation function at lag k, the series has an MA(k) signature.

When choosing a model, it is often the case that the best model is not immediately obvious. For example, a situation in which the first partial autocorrelation is close to one (negative one) but the remaining partial autocorrelations are close to zero is indicative of both an AR(1) model as well as a situation in which the series is under-differenced (over-differenced). Hence, when choosing a model, it is common to initially evaluate a set of candidate models. This is done by estimating the parameters for each model, and then evaluating the accuracy of the in- and out-of-sample forecasts. There are various measures of forecast accuracy such as root mean squared error, RMSE, mean absolute error, MAE, mean absolute percentage error, MAPE, etc. However, RMSE is typically used because: (1) it places greater weight on errors that are large in absolute value and (2) it is consistent with the objective function underlying the optimization program that generates the parameter estimates.

In light of the above, three points about model selection and estimation warrant mentioning. First, both objective statistical testing and the use of judgment are important parts of the process. This implies

[3] If the series includes a linear time trend — for example, $\mathrm{EARN}_{t+1} = (t+1) \times \mu + \varepsilon_{t+1}$, the unconditional mean and variance of EARN_{t+1} are $(t+1) \times \mu$ and σ^2, respectively; whereas, the unconditional mean and variance of $\Delta\mathrm{EARN}_{t+1}$ are μ and $2 \times \sigma^2$. Hence, differencing generates a stationary series.

that reasonable people may disagree about the "best" model, which, in turn, has implications for replicability. It also implies that, when evaluating a sample that contains many firms, the researcher must either: (1) expend substantial amounts of time and effort, which is costly, or (2) forego the use of judgment, which reduces the quality of the forecasts.

Second, out-of-sample forecasts play a crucial role in the model-validation process. (A common heuristic is that at least 20% of the observations in the observable time series should be used as a hold-out sample.) The reason out-of-sample forecasts are so important is that, given the emphasis on judgment in the model-selection process, the risk of overfitting is substantial. Finally, the estimated parameters of an ARIMA model are chosen to minimize the mean of the one-period-ahead squared forecast errors. Hence, there is no guarantee that the model will generate the most accurate forecasts of earnings for periods subsequent to period $t+1$.

5.5 Extant evidence

The key extant result is that the random-walk model generates out-of-sample forecasts that are either as accurate as or more accurate than out-of-sample forecasts generated by other ARIMA models (e.g., Ball and Watts, 1972; Albrecht *et al.*, 1977; Watts and Leftwich, 1977). This result was established in the early 1970s and remains well accepted. For example, when discussing the superiority of the random-walk model, Brown (1993, p. 295) states "*... the issue was pretty much resolved in the late 1970s.*" Similarly, Kothari (2001, p. 145) states "*A large body of evidence suggests a random walk or random walk with drift is a reasonable description of the time-series properties of annual earnings.*" Finally, Bradshaw *et al.* (2012, p. 948) state that "*Numerous studies examine the time-series properties of annual earnings and generally find that annual earnings approximate a simple [random walk] process.*"

5.6 Interpreting the success of the random-walk model

Although the superiority of the random-walk model is well accepted, I argue that this result is misleading. I have several reasons. First, the random-walk model is inconsistent with standard assumptions about economics and accounting. Second, the superiority of the random-walk model is inconsistent with other well-accepted empirical results. Third, the random-walk model is at odds with how financial statement analysis is practiced and taught. Finally, ARIMA models are ill-suited for developing forecasts of earnings. I elaborate on each of these reasons in the space below.

5.6.1 Inconsistent with standard assumptions about economics and accounting

Three boilerplate assumptions in economics are that: (1) people maximize their expected utility: (2) expected utility is increasing in expected wealth; and, (3) there are finite resources. Consequently, people compete for resources. Although agency costs and other frictions may lead to some misalignment between the objectives of a firm's managers and those of its shareholders, this is a second-order issue. The reason for this is that the market for corporate control, labor-market competition, compensation contracts, reputation, etc. imply that a manager's utility is an increasing function of the value of the firm that employs her. Hence, the primary way for her to compete is by attempting to increase the value of the firm that she manages. This, in turn, implies that managers have strong incentives to innovate. These innovations will take the form of improvements in the production process, disruption of existing markets, creation of new markets and sources of demand, etc.

The above implies that successful innovations are mimicked, improved upon, and eventually become obsolete. Hence, positive (negative) changes in a firm's expected future earnings that accompany successful innovations by its managers (competitors) are not permanent. This, in turn, implies that earnings will not follow a random walk. There are also three accounting principles that augur against the random-walk model.

First, the use of historical-cost accounting implies that there is delayed recognition of economic news. For example, revenues are recognized when they are earned, measurable, and collectible. Hence, economic news affects both current and subsequent earnings, which, in turn, implies non-zero autocorrelation in earnings levels and changes. Second, the amount of delay in the recognition of economic news depends on the sign and extremity of the news. In particular, extreme negative economic news tends to be recognized in the reporting period in which the news arrives whereas positive economic news is only recognized when the underlying uncertainty is resolved via the occurrence of an arms-length exchange. This implies that negative earnings changes are less persistent than positive earnings changes. In fact, extreme negative changes may exhibit negative autocorrelation.

Finally, accounting tends to be conservative in the sense that investments in certain long-lived economic assets are immediately expensed. For example, research and development spending, marketing expenditures, etc. are typically not capitalized. As I discuss further in Section 7, this implies that when there is high growth in investments that are accounted for conservatively, earnings and accounting rates of return are lower than what they would be in the absence of conservatism. However, when the growth rate becomes low, earnings and accounting rates of return are higher than they would be if the investments were capitalized. This implies a temporal pattern in earnings and accounting rates of return that is inconsistent with the random-walk model.

5.6.2 Contradictory empirical evidence

In addition to being reasonable *a priori*, the economic and accounting assumptions described above are supported by several well-known empirical results. First, Beaver and Morse (1978) show that price-to-earnings ratios are mean reverting. Assuming that equity prices are formed in efficient capital markets, this result implies that earnings do not follow a random walk. Second, a number of studies provide evidence that both unscaled earnings (e.g., Brooks and Buckmaster, 1976; Ramakrishnan and Thomas, 1992; Lipe and Kormendi, 1994) and profitability ratios such as ROE and ROIC (e.g., Freeman *et al.*, 1982; Nissim and Penman, 2001) are mean reverting and that the speed of reversion is greatest when earnings are extreme relative to either the cross-sectional or historical average. Third, as shown in Basu (1997) negative earnings changes (deflated by lagged price) are less persistent than positive earnings changes. Fourth, Penman and Zhang (2002) and Monahan (2005) demonstrate that accounting conservatism has implications for the time-series properties of ROIC amd ROE that are inconsistent with the random-walk model. Finally, there is considerable evidence that the persistence of earnings varies with their composition. For example, Lipe (1986) shows that persistence varies across the line items of earnings such as gross profit, general and administrative expense, etc. and Sloan (1996) shows that accruals are less persistent than cash flows.

5.6.3 Inconsistent with teaching and practice

Although standard pedagogy and practice may reflect flawed conventional wisdom, it is still noteworthy that they are very much at odds with the random-walk model. In particular, it would be unusual for an instructor teaching financial statement analysis to not discuss how to evaluate trends in the components of ROIC and how to compare these trends across firms. It is also standard practice amongst analysts to discuss notions of earnings momentum, etc. These concepts, however, are irrelevant if the random-walk model is descriptive because, in that case, the most-recent realization of earnings is all that matters.

5.6.4 Limitations of ARIMA models

In light of the above, the superiority of the random-walk model may seem puzzling. However, I argue that this is a misinterpretation of the evidence. Rather, I argue that ARIMA models are ill-suited for generating forecasts of earnings. Hence, the fact that the random-walk model performs well in comparison to other ARIMA models is both not too surprising and a moot issue — that is, it may simply be the best of a bad lot. There are a number or reasons for this conclusion including: (1) ARIMA models ignore information embedded in the components of earnings; (2) ARIMA models ignore information embedded in other financial statement line items; (3) ARIMA models ignore information about the contemporaneous performance of the firm's rivals, suppliers, etc.; and, (4) obtaining a stationary time series by transforming earnings is nontrivial. Although all of these reasons are important, I emphasize the last reason because it is central to ARIMA modeling.

As discussed above, the first step in ARIMA modeling is to obtain a covariance-stationary time series. With respect to earnings, this is a nontrivial undertaking for two related reasons. First, annual earnings are a low-frequency variable. Hence, in order to obtain sufficient data to develop precise estimates of ARIMA-model parameters and conduct out-of-sample validation, it is necessary to have observations that are drawn from multiple decades. Second, during any reasonable length of time (e.g., one decade) the likelihood that there will be a significant disruption at the firm-level, industry-level, macro-level or all three, is high. For example, per the National Bureau of Economic Research, the United States experienced 11 recessions during the years spanning 1948 and 2009, which is an average of one recession during each five-and-a-half-year period. In addition, between 1971 and 2009 the Financial Accounting Standards Board, FASB, promulgated 168 new accounting standards, which is an average of 4.42 standards a year.

Combining these two points, I conclude that obtaining a stationary series of the necessary length for developing precise ARIMA-model parameter estimates is typically infeasible. This implies that ARIMA forecasts that are a function of estimated autoregressive or moving-average coefficients will be biased and imprecise. Hence, given that the random-walk model either requires no parameter estimates or only an estimate of the drift parameter, it is not too surprising that it performs relatively well vis-à-vis more complicated ARIMA models.

5.7 Summary

Early research in earnings forecasting treated ARIMA models as the default choice. Hence, when the random-walk model proved to be superior to other ARIMA models, it became well accepted that the random walk was the best model. This conclusion was arguably premature. The fact that the random walk is the best ARIMA model does not imply that it is the best approach in general. ARIMA models are not the only approach for forecasting earnings. Rather, panel-data approaches that develop forecasts on the basis of both time-series and cross-sectional data are also feasible. Moreover, these approaches have many *a priori* advantages vis-à-vis ARIMA models. Hence, they are the subject of the next section.

6

Panel-data Approaches

The inaccuracy of ARIMA models and their *a priori* limitations led researchers to consider alternative forecasting approaches. In this section, I discuss panel-data approaches, which use a mix of cross-sectional and time-series data — that is, panel data — to develop a forecast. I begin by discussing the different ways that panel data can be used and the choices that need to be made when using them. I then describe the *a priori* advantages of using panel-data approaches vis-à-vis ARIMA models and whether there is empirical evidence supporting these advantages. Finally, I summarize the discussion. Note that I do not attempt to provide a detailed literature review. Rather, throughout the discussion, I emphasize key issues and the studies that illustrate them.

The main conclusion of this section is that, although panel-data approaches have clear *a priori* advantages, the jury is still out on their usefulness. At present, the evidence regarding their usefulness is relatively scant. Moreover, and perhaps more important, for the purposes of predicting earnings, it is unclear whether panel-data approaches are substantially better than the random-walk model.

Finally, note that, in this section, I ignore the implications of accounting measurement for the predictability of earnings and the use of panel-data approaches for forecasting the higher moments — that is, the variance, skewness and kurtosis — of earnings. I discuss these issues in Section 7 and 8, respectively.

6.1 An overview of panel-data approaches

Panel data approaches provide the researcher with a broad set of choices regarding what she predicts, how she predicts it, etc. Given this fact, the easiest way to discuss them is to first enumerate the five key choices that must be made, and then elaborate on each choice. The five key choices are:

1. Choose what to predict.
2. Choose the predictors.
3. Choose an estimator.
4. Choose an estimation sample.
5. Choose how to evaluate the predictions.

6.2 Choose what to predict

Three issues warrant discussion. Selecting an earnings metric is the first issue; however, because I discuss this topic at length in Section 3, I only mention it here for completeness. The second issue relates to the flexibility of panel-data approaches. In addition to facilitating the prediction of continuous variables (e.g., year-ahead net income), panel-data approaches can be used to predict a binary outcome. For example, Ou (1990) uses estimates from Logit regressions to predict whether the change in year-ahead earnings per share will be greater that the average of the changes for the previous five years. Similarly, Joos and Plesko (2005) use Logit regressions to predict loss reversals — that is, whether a firm that reported negative earnings in year T will report positive earnings in year $T + 1$.

Finally, the forecast horizon is an explicit choice and horizons beyond one year are feasible. For example, Hou *et al.* (2012), Gerakos and

Gramacy (2013), and Li and Mohanram (2014) use estimates from ordinary least squares, OLS, regressions to predict one-year-ahead, two-years-ahead and three-years-ahead net income.

6.3 Choose the predictors

The predictors refer to a set of variables that are observable in the year when the prediction is made. I refer to the set of predictors for firm i as $\boldsymbol{x}_{i,T}$, which is a $1 \times k$ vector of variables that are observable in year T. Note that the firm subscript i does not imply that the predictors are specific to firm i. Although this is often the case, it is not necessary. For example, one of the predictors could be an industry-level variable. In addition, the time subscript T only implies that the predictors are observable in year T; hence, variables that were realized prior to year T can also be used as predictors.

To elaborate on the above, suppose the goal is to form a prediction in year T of firm i's earnings in year $T+1$, $\mathrm{EARN}_{i,T+1}$. Further suppose that, based on arguments in Li and Mohanram (2014), earnings persistence is assumed to vary with the sign of earnings. Hence, the predictors are: $\mathrm{EARN}_{i,T}$; the indicator variable $\mathrm{LOSS}_{i,T}$ that equals one (zero) if firm i reported negative (non-negative) earnings in year T; and, the interaction term $\mathrm{EARN}_{i,T} \times \mathrm{LOSS}_{i,T}$. (That is, $\boldsymbol{x}_{i,T}$ is the 1×3 vector that equals $[\mathrm{EARN}_{i,T} \quad \mathrm{LOSS}_{i,T} \quad \mathrm{EARN}_{i,T} \times \mathrm{LOSS}_{i,T}]$.) Predictions are then formed by combining the predictors with a set of parameters that are estimated using data observable in year T. For example, if an ordinary least squares, OLS, regression that includes a constant term is used to estimate the parameters β_0, β_1, β_2, and β_3, the prediction is:

$$\beta_0 + \beta_1 \times \mathrm{EARN}_{i,T} + \beta_2 \times \mathrm{LOSS}_{i,T} + \beta_3 \times (\mathrm{EARN}_{i,T} \times \mathrm{LOSS}_{i,T}). \tag{6.1}$$

The above example illustrates two key attributes of panel-data approaches. First, they allow the researcher to use multiple predictors and to assign a different weight to each predictor. For example, earnings components can be used to forecast net income; the well-known DuPont decomposition can be used to separate ROE into its components, and then use these components to predict future ROE; etc.

Second, the prediction does not have to be a linear function of either the predictors or the parameters. The simplest way of allowing for non-linear relations is to use one or more interaction terms. For instance, in the example described above, the variable $\text{EARN}_{i,T} \times \text{LOSS}_{i,T}$ implies that the temporal relation between future earnings and current earnings is not linear. Rather, it depends on the sign of current earnings. More complex non-linear relations can also be accommodated. For example, suppose $\boldsymbol{\beta}$ is $k \times 1$ vector of parameters estimated via a Logit regression, the predicted probability that the binary outcome of interest will occur (e.g., firm i will experience a loss in year $T+1$) is $(1 + e^{-\boldsymbol{x}_{i,T}\boldsymbol{\beta}})^{-1}$.

Given the wide range of possibilities regarding the choice of the predictors, *how* the predictors are chosen is the overarching issue. I discuss three approaches: (1) statistical learning; (2) appealing to conventional wisdom; and, (3) appealing to accounting-based valuation models. In the space below, I discuss each approach separately. However, it is important to note that they are not mutually exclusive and they can be, and often are, used in conjunction.

6.3.1 Statistical learning

Statistical selection involves evaluating different sets of predictors and functional forms, and then choosing the set of predictors and the functional form that are "best." Best is defined by the researcher. Her definition may be tied to an explicit loss function or be the result of *ad hoc* arguments.

An example of statistical learning is described in Ou (1990). As discussed above, her objective is to predict above-average changes in year-ahead earnings per share. She uses a two-step process to identify the final set of predictors from her original set of 61 candidate variables, most of which are financial ratios (e.g., ROE). In the first step, she separately evaluates each candidate variable. To do this, she uses it as the sole independent variable in a Logit regression. If the estimated coefficient on the candidate variable is statistically significant at the 10% level, she retains it for evaluation in her second stage. Thirteen candidate variables are held over to the second stage; and, in that stage, she uses these 13 variables as the independent variables in a multivariate Logit regression. She retains the eight variables with estimated coefficients that are statistically significant at the 10% level. These eight variables constitute her final set of predictors.

The primary advantage of statistical learning is that it allows the researcher to "let the data speak." There is no need for an explicit theory regarding the choice of predictors, their relations with each other and the predicted value, etc. This is also the primary disadvantage of using statistical learning. Because there is no need for an explicit theory, there are no constraints regarding candidate variables, candidate functional forms, etc. Hence, there is considerable risk of overfitting the in-sample data.

Finally, it is important to note that since the study by Ou (1990) there have been tremendous advances in statistical learning. (For example, see Hastie *et al.* (2009).) Whether these advances are useful within the context of forecasting earnings is unclear. Gerakos and Gramacy (2013), which I discuss further in the space below, provide some evidence regarding the performance of some of the recently developed approaches. However, there is much more that can be done.

6.3.2 Appealing to conventional wisdom

I use the term conventional wisdom to refer to extant practice and economic intuition that does not take the form of a formal analytical model. An example of a study that chooses the predictors by appealing to practice is Abarbanell and Bushee (1997). Abarbanell and Bushee (1997) evaluate the in-sample relations between nine fundamental signals observed in year t and: (1) $\Delta\text{EPS}_{i,t+1}/P_{i,t}$ ($\Delta\text{EPS}_{i,t+1}$ is the change in firm i's earnings per share in year $t+1$ and $P_{i,t}$ is the share price of firm i at the end of year t) and (2) the compound annual growth rate in net income between years t and $t+5$. The nine signals are chosen by reference to arguments made in Lev and Thiagarajan (1993), who identify the signals by conducting a survey of articles in the financial press (e.g., articles in the *Wall Street Journal* that contain the term "earnings quality"), written pronouncements of financial analysts (e.g., newsletters issued by major securities firms) and other practitioner-oriented publications.

Regarding the use of economic intuition, I highlight two studies. The first is by Fairfield *et al.* (1996) (FSY hereafter), who evaluate a fundamental question: Is it useful to disaggregate earnings into components, and then allow for different weights on the different components? Using line-item definitions from GAAP, FLS consider five different disaggregation schemes. The simplest is to not disaggregate whereas the most complex is to separate earnings into 10 components. For each scheme, FLS use a linear model to generate out-of-sample forecasts of ROE. They find that an intermediate scheme in which earnings is separated into four components has the lowest mean absolute forecast error.[1] This is an important result as it provides empirical support for a basic tenet of financial statement analysis: Different line items have different persistence and, thus, line items should be evaluated separately.

The second study that I highlight is by Fama and French (2000), who argue that profitability is mean reverting. I focus on the intermediate model that they use as part of testing their prediction. In particular, to test for mean reversion, Fama and French (2000) use a partial adjustment model in which the change in profitability in year $t+1$, $\mathrm{CP}_{i,t+1}$, is a function of $\mathrm{CP}_{i,t}$ and $\mathrm{DFE}_{i,t}$. $\mathrm{DFE}_{i,t}$ is the difference between realized profitability in year t and its year $t-1$ expected value. Hence, in order to implement their model, they need to estimate the year $t-1$ expected value of profitability in year t. The intermediate model that they use to do this is the inspiration for a number of studies that use earnings forecasts and/or expectations (e.g., Hou and Robinson, 2006; Hou and van Dijk, 2008; Hou *et al.*, 2012).

Referring to practice and economic intuition to choose predictors is reasonable. Accounting and fundamental analysis are inherently practical endeavors. Consequently, it is not surprising that practice often leads theory and serves as the motivation for empirical tests.

[1]The four components, each of which are deflated by beginning equity book value, are: (1) OPINC, which consists of gross margin, selling, general and administration expense, depreciation and amortization expense, interest expense and minority interest income; (2) NOPTAX, which is the sum of non-operating income and income tax expense; (3) special items; and, (4) the sum of discontinued operations and extraordinary items.

However, the risk of overfitting remains; and, there are extant results that suggest that this risk is manifest. For example, Abarbanell and Bushee (1997) find that the associations between future earnings changes and two of the fundamental signals that they evaluate have signs that are the opposite of what practitioners expect. Although this does not necessarily imply overfitting of the in-sample data, it does suggest that practitioners' understanding is incomplete. Hence, using that understanding to guide the choice of predictors may lead to spurious results.

6.3.3 Appealing to accounting-based valuation models

Given that there is a valuation context underlying many of the studies that focus on earnings forecasting, it is not surprising that a number of studies use accounting-based models to motivate both the choice of what to predict and the predictors. Typically, the RIV model is used as motivation. There are two reasons for this. First, two influential studies of the RIV model by Ohlson (1995) and Feltham and Ohlson (1995) introduce linear information dynamics, LIDs, in which earnings in year $T+1$ is a linear combination of accounting variables observable in year T and a random error. These LIDs are used in a number of studies to motivate earnings forecasting models (e.g., Dechow *et al.*, 1999; Myers, 1999; Monahan, 2005; Li and Mohanram, 2014).

Second, as discussed in Section 3, residual income can be decomposed into a combination of growth in equity book value and return on equity, ROE, which can be further decomposed into a number of well-known financial ratios such as return on invested capital, ROIC, operating profit margin, invested capital turnover, etc. These components can then be used as predictors of future ROE or other financial ratios.

Regarding the decomposition of ROE into its components, the studies by Nissim and Penman (2001) and Nissim and Penman (2003) (hereafter NP1 and NP2) are authoritative. Although these studies make a number of contributions, I focus on the two that I believe are the most pertinent to earnings forecasting. First, they develop and describe a modified version of the well-known DuPont decomposition. In this modified version, return on equity is expressed as follows:

$$\overbrace{\frac{\text{EARN}_{i,t}}{B_{i,t-1}}}^{\text{ROE}_{i,t}} = \overbrace{\frac{\text{EARN}^{\text{ENT}}_{i,t}}{\text{IC}_{i,t-1}}}^{\text{ROIC}_{i,t}} + \overbrace{\left(\overbrace{\frac{\text{EARN}^{\text{ENT}}_{i,t}}{\text{IC}_{i,t-1}}}^{\text{ROIC}_{i,t}} - \overbrace{\frac{\text{NFE}_{i,t}}{\text{NFO}_{i,t-1}}}^{\text{NFR}_{i,t}}\right)}^{\text{FINSPREAD}_{i,t}} \times \overbrace{\frac{\text{NFO}_{i,t-1}}{\text{B}_{i,t-1}}}^{\text{FLEV}_{i,t-1}}. \tag{6.2}$$

In Equation (6.2), firm i's year t return on equity, $\text{ROE}_{i,t}$, equals the ratio of its year t equity-level income, $\text{EARN}_{i,t}$, to its common stockholders' equity in year $t-1$, $B_{i,t-1}$. This ratio is then decomposed into two components. The first component is firm i's contemporaneous return on invested capital, $\text{ROIC}_{i,t}$, which equals the ratio of year t enterprise-level earnings, $\text{EARN}^{\text{ENT}}_{i,t}$, to invested capital in year $t-1$, $\text{IC}_{i,t-1}$. As discussed in NP1, $\text{EARN}^{\text{ENT}}_{i,t}$ is the difference between all operating revenues and all operating expenses: and, $\text{IC}_{i,t-1}$ is the difference between all operating assets and all operating liabilities. Hence, $\text{ROIC}_{i,t}$ is a measure of operating performance that is not affected by financing activities.

The second component of $\text{ROE}_{i,t}$ reflects the effect of firm i's capital structure choices. This effect is determined by the: (1) the degree of financial leverage, $\text{FLEV}_{i,t-1}$, and (2) the extent to which the operations generate a return in excess of the cost of borrowing, $\text{FINSPREAD}_{i,t}$. $\text{FLEV}_{i,t-1}$ equals the ratio of net financial obligations in year $t-1$, $\text{NFO}_{i,t-1}$, to $B_{i,t-1}$. $\text{NFO}_{i,t-1}$ is a comprehensive measure of debt and is net of financial assets, which are those assets that are unrelated to the business's value-generating activities. $\text{FINSPREAD}_{i,t}$ equals the difference between $\text{ROIC}_{i,t}$ and the net financial rate, $\text{NFR}_{i,t}$, which is the ratio of after-tax net financial expense, $\text{NFE}_{i,t}$, to $\text{NFO}_{i,t-1}$.[2] Similar to $\text{NFO}_{i,t-1}$, $\text{NFE}_{i,t}$ is a measure of the cost of debt and is net of any financial income earned in year t.

Equation (6.2) is important because it shows that the well-known and intuitive DuPont decomposition can be made rigorous. In particular,

[2]Note that net financial expense is the additive inverse of net financial income — that is, $\text{NFE}_t = -\text{NFI}_t$ and net financial obligations is the additive inverse of net financial assets — that is, $\text{NFO}_t = -\text{NFA}_t$.

ROE can be cleanly separated into an operating component and a financing component. Hence, Modigliani and Miller (1958) separation (i.e., MM separation), which is a bedrock principle of finance, is integrated into the analysis. To elaborate, MM separation refers to the fact that a firm's operations are the primary source of value and its financing activities are of second-order importance. This implies that, when analyzing ROE, operating and financing activities should be evaluated separately. Equation (6.2) allows for this separation while preserving the flavor of DuPont analysis. For example, $\mathrm{ROIC}_{i,t}$ can be decomposed into operating profit margin (i.e., the ratio of $\mathrm{EARN}_{i,t}^{\mathrm{ENT}}$ to contemporaneous revenues) and invested capital turnover (i.e., the ratio of revenues in year t to $\mathrm{IC}_{i,t-1}$). These two ratios can then be decomposed into ratios that can be further decomposed, etc.

The second key contribution made by NP1 and NP2 is their empirical analyses of the ratios underlying Equation (6.2). They document the cross-sectional distribution of each ratio and how median values of that ratio vary across time. They also evaluate whether the ratio exhibits mean reversion; and, if so, to what degree. Finally, they document how the different ratios are related to one another. These analyses are meticulous; and, they provide a set of valuable baselines for developing forecasts. For example, by understanding the typical value of ROIC, its distribution and how it tends to vary with growth in invested capital, investors can perform a "reality check" on the forecasts of enterprise-level residual income that they use as inputs into their value estimates.

Although important, NP1 and NP2 have one omission: They do not contain evidence regarding the usefulness of the modified DuPont framework for developing forecasts. In particular, neither NP1 nor NP2 develop in-sample forecasting models, and then evaluate the accuracy of the out-of-sample forecasts obtained from these models. Fortunately, Fairfield and Yohn (2001) fill the gap. They develop models that use predictors observable in year T to forecast the year $T+1$ change in return on invested capital, $\Delta\mathrm{ROIC}_{i,T+1}$. The main difference between the models is whether $\Delta\mathrm{ROIC}_{i,T}$ or its components — that is, the change in profit margin, $\Delta\mathrm{PM}_{i,T}$, the change in invested capital turnover, $\Delta\mathrm{ICT}_{i,T}$, and the interaction of the two, $\Delta\mathrm{PM}_{i,T} \times \Delta\mathrm{ICT}_{i,T}$ — are used as predictors. Fairfield and Yohn (2001) show that the model that uses the components of $\Delta\text{ROIC}_{i,T}$ has a higher in-sample r-squared and generates lower out-of-sample mean absolute forecast errors. These results are important because they imply that the DuPont decomposition is, in addition to being intuitive, useful.

Using accounting-based valuation models and the DuPont decomposition of ROE to motivate the choice of predictors seems natural. This approach is also arguably more rigorous than using either statistical selection or appealing to conventional wisdom. For example, evidence in Soliman (2008) shows that, when predicting changes in year-ahead ROIC, current changes in operating profit margin and invested capital turnover have incremental explanatory power vis-à-vis a set of competitor predictors that are described in Abarbanell and Bushee (1997) and Richardson *et al.* (2005). One caveat, however, is that these results are in-sample and the competitor predictors also have incremental explanatory power. Hence, at present there is neither a set of dominant predictors nor a dominant approach for selecting them.

6.4 Choose the estimator

The choice of estimator is a complex issue that extant studies tend to ignore. In particular, ignoring studies that forecast binary outcomes, OLS regressions are by far the most-popular estimator. There are two arguments in favor of using OLS regressions. First, they are easy to estimate and interpret. Second, as discussed in Kennedy (1992), when the relation between the predicted value and the predictors is linear and the forecast errors are normally distributed, OLS regressions generate in-sample forecasts that are unbiased and forecast errors that have minimum variance.

Although they are the most popular estimator, OLS regressions are not the obvious choice. Moreover, and more importantly, an obvious choice does *not* exist. The reasons for this are threefold. First, earnings exhibit negative skewness and positive kurtosis (e.g., Basu, 1997; Chang *et al.*, 2017), which, in turn, suggests that forecast errors are not normally distributed. Rather, their true distribution is unknown and varies across firms and time. Hence, even in the unlikely case when both the predictors and the functional form are known, a minimum-variance, unbiased estimator may not exist. This implies that non-parametric approaches that do not require assumptions about the distribution may be superior to OLS. Second, neither the predictors nor the functional form are known. Consequently, non-parametric estimators that use statistical learning to identify the predictors and/or the functional form (e.g., regression trees allow for non-linear relations) may outperform OLS.

Finally, the choice of estimator depends on the loss function implied by the economic motivation for the study. For example, if the objective is to minimize mean squared forecast error, an estimator that embeds an explicit bias-variance tradeoff (e.g., lasso) may be superior to OLS. On the other hand, if the forecaster wants to minimize mean absolute forecast error, forecasts of the median obtained from a quantile regression may be superior. More generally, the loss function depends on the context and the utility function implied by the research question. However, context is often ill-defined and, as discussed in Harsanyi (1990), interpersonal utility comparisons are meaningless, which implies that the choice of utility function is arbitrary.

In light of the above, choosing an estimator requires making a subjective decision about the loss function, and then making a trade-off. On the one hand, OLS regressions embed unnecessary assumptions that may lead to estimators that are too simple vis-à-vis more sophisticated approaches. On the other hand, the degree of overfitting tends to increase with complexity. The best way to make this trade-off is an empirical question; and, within the context of forecasting earnings, there is very little evidence on the matter. Gerakos and Gramacy (2013) is the only study I know of that considers alternative estimators; and, their overarching conclusion on the matter is that it depends. In particular, they find that, when forecasting unscaled net income, more sophisticated estimators tend to outperform OLS in terms of root mean squared error. On the other hand, when forecasting scaled net income, the OLS estimator tends to dominate.

The results in Gerakos and Gramacy (2013) reinforce the point made above: The choice of estimator is not obvious and it is an empirical question. This, in turn, implies that further study of the issue is needed and that the OLS estimator should not be considered the default choice.

Rather, best practice is to propose and empirically evaluate a set of estimators.

6.5 Choose the estimation sample

The estimation sample is the panel of data that are used to train the estimator. Hence, it is sometimes referred to as either the in-sample data or the training data. The estimation sample is chosen by making a tradeoff between size and homogeneity. *Ceteris paribus*, larger samples are better. First, as the sample size increases parameter estimates become more precise. Second, some estimators are biased in small samples but unbiased asymptotically; hence, as the sample size increases the bias declines. Finally, statistical learning approaches require large amounts of data to learn from.

Although larger samples are better *ceteris paribus*, it is only optimal to maximize sample size when the predictors, functional form and parameters are the same for every observation in the sample — that is, the sample is homogeneous. *A priori*, there is no reason to expect homogeneity. However, at present, there is not a lot of evidence about the degree of heterogeneity and what determines it. In particular, I know of only two studies that explicitly consider the issue within the context of forecasting earnings: (1) Fairfield *et al.* (2009) (FRLY hereafter) and (2) Gerakos and Gramacy (2013).

FRLY compare out-of-sample forecasts obtained from industry-level estimation samples to those obtained from an economy-wide estimation sample. Industry-level estimation samples are formed by grouping observations on the basis of their GICS codes. FRLY evaluate forecasts of the year $T + 1$ values of three growth measures (growth in net operating assets, growth in equity book value and growth in sales) and two profitability ratios (ROE and ROIC). They base their forecasts on estimated coefficients obtained from OLS regressions.

When estimating the regression coefficients described above, FRLY use two different regressions: (1) a regression that uses all of the data drawn from the years $T - 10$ through T — that is, the economy-wide sample — and (2) a regression that uses the subset of the economy-wide sample that consists of the firms that are in the same industry as firm i — that is, the industry-level sample. FRLY then compare the forecasts implied by the regression coefficients obtained from the economy-wide sample to those implied by the industry-level sample.

FRLY's research design allows them to evaluate an important question: Does industry-level peer analysis lead to more accurate forecasts of growth and profitability? The manner in which fundamental analysis is practiced and taught implies that the answer is yes. For example, when conducting fundamental analysis, analysts tend to compare a firm's financial ratios to the financial ratios of firms in the same industry. This reflects implicit assumptions that the factors that drive growth and profitability are industry-specific, that growth and profitability tend to revert to an industry-level mean and that the rate of mean reversion varies with industry. These assumptions are based on more primitive assumptions that the degree and form of competition, production technologies, cost structures, markups, accounting measurement, etc. tend to vary by industry.

In light of the conventional wisdom, FRLY's results are provocative. They find that, with one exception, the regressions that use the industry-level sample do not generate more accurate forecasts of either one-year-ahead growth or one-year-ahead profitability. (The exception relates to sales growth.) FRLY also evaluate forecasts of five-year compound annual growth rates and five-year-ahead profitability. Although there is some improvement in the relative accuracy of the forecasts that are based on the industry-level sample, the overarching conclusion remains: Constraining the sample to industry peers does not generate statistically significant improvements in forecast accuracy.

While FRLY consider cross-sectional heterogeneity, Gerakos and Gramacy (2013) evaluate temporal heterogeneity. That is, they evaluate the relation between forecast accuracy and the length of the panel — that is, the number of years included in the panel of data that are used to train the estimator. They consider lengths of one year and five years and they evaluate forecasts of net income scaled by equity market value and unscaled net income. They find that when forecasting scaled (unscaled) net income, shorter (longer) panels lead to more accurate forecasts. However, there is no *a priori* logic for these results. Hence, the key conclusions regarding panel length are that: (1) it matters and (2) it cannot be assumed; rather, it must be evaluated.

6.6 Choose how to evaluate the predictions

Predictions are typically evaluated on a relative basis. Specifically, the typical study involves proposing a new forecasting approach — for example, a different set of predictors, a different estimator, etc., and then providing evidence about the relative accuracy of the new approach vis-à-vis existing approaches. To do this, a choice must be made about how to evaluate and compare the out-of-sample predictions obtained from the different approaches. (In order to avoid over-fitting, it is imperative that the *out-of-sample* predictions are evaluated.) This leads to two issues: (1) evaluating forecast errors and (2) evaluating economic magnitudes. It is well beyond the scope of this monograph to provide an in-depth discussion of these issues. Hence, in the space below, I provide an overview. Readers interested in an in-depth discussion should refer to Hastie *et al.* (2009) and the references therein.

6.6.1 Evaluating forecast errors

It is important to note that the choice of how to evaluate the forecast errors of an approach and the choice of estimator used as part of that approach are linked. The reason for this is that estimators embed implicit or explicit loss functions. However, as discussed above, loss functions are subjective choices that depend on the context, which is often ill-defined. Consequently, best practice is to provide detailed descriptive statistics — for example, mean, standard deviation, key percentiles — about the distribution of the forecast errors generated by each approach and to use multiple criteria to identify the best approach.

A potential outcome of adopting the "best practice" described above is that there is no dominant approach. That is, different criteria may lead to different conclusions regarding the identity of the "best" approach. What to do if this occurs depends on the circumstances. If the approaches being evaluated share a common loss function, the natural choice is to give priority to the criterion that is consistent with the loss function (Gerakos and Gramacy, 2013). For example, the OLS estimator has a loss function that penalizes mean squared error. Hence, a study that evaluates different OLS models, should use mean squared forecast error as the main criterion. On the other hand, if there is no common loss function, there is no clear criterion. If this is the case, the research question and context should be clearly explicated, and then used to motivate the main criterion.

The issues involved in evaluating forecasts depend on whether a binary or continuous outcome is being predicted. When a binary outcome such as the sign of net income is being predicted, there are two common ways of assessing model performance: (1) discrimination and (2) calibration. Discrimination relates to how well the model discriminates between positive and negative outcomes. Calibration, on the other hand, reflects the extent to which the predicted probability of a positive outcome matches the true probability of a positive outcome. Note that, in this context, the terms positive and negative relate to the binary outcome being predicted. For example, if the objective is to predict accounting losses in year $T + 1$, a positive (negative) outcome occurs when year $T + 1$ income is negative (not negative).

Regarding discrimination, receiver operator characteristic, ROC, curves are quite useful. An ROC curve is created by plotting the true positive rate, TPR_a, against the false positive rate, FPR_a, for different thresholds, a. TPR_a is the fraction of positive outcomes that were correctly predicted by the model whereas FPR_a is the fraction of negative outcomes that were incorrectly predicted by the model. The threshold, a, is the amount that is used to impute forecasts from model scores. For example, if a Logit model is being used to forecast accounting losses, the predicted probabilities generated by the model are the scores. Hence, a threshold of 0.50 implies that, if the predicted probability for firm i in year T is greater than (not greater than) 0.50, the forecast is that firm i will (will not) report a loss in year $T+1$. Different thresholds imply different values of TPR_a and FPR_a; and, TPR_a and FPR_a are positively related. For instance, if the threshold for the Logit model is set to zero (one), TPR and FPR both equal one (zero).

ROC curves have many nice properties, two of which are particularly noteworthy. First, as discussed in Choi (1998), they are informative about likelihood ratios — that is, the ratio of the probability of obtaining a particular test result for positive outcomes to the probability of obtaining the same test result for negative outcomes. Hence, the likelihood ratio of a positive test equals $\mathrm{TPR}_a \div \mathrm{FPR}_a$, which equals the slope of the line between the origin and the point on the ROC curve that corresponds to the threshold value a. Moreover, the slope of the tangent line to a point on the ROC curve equals the likelihood ratio of obtaining a score equal to the threshold value corresponding to that point.

Second, as discussed in Hanley and McNeil (1982), the area under the ROC curve, AUC, equals the ratio of the number of concordant pairs to the total number of pairs that consist of one positive and one negative outcome. Hence, the AUC equals the probability that a randomly chosen positive outcome will receive a higher model score than a randomly chosen negative outcome. This implies that the AUC can be used to evaluate how well a model discriminates between true and false positives.

Although ROC curves are informative about discrimination, they can be noisy indicators of calibration. Rather, as discussed in Cook (2007), a well-calibrated model may discriminate poorly and *vice versa*. Hence, if the goal is to forecast the *risk* of a positive outcome — for example, the risk of reporting an accounting loss, both discrimination and calibration should be considered. This introduces two complications: (1) choosing a scoring rule (e.g., the Brier score, Brier (1950)) to evaluate calibration and (2) making a tradeoff between discrimination and calibration. Both of these are subjective choices that depend on the loss function. Hence, as discussed above, best practice is to provide detailed descriptive statistics and consider multiple criteria.

When evaluating forecasts of continuous variables such as net income, the main issues are: (1) choosing a metric for evaluating accuracy and (2) dealing with differences in scale. Regarding the choice of metric, there are many possibilities. However, the most commonly used metrics are the mean of the forecast errors (i.e., bias), the mean of the absolute values of the forecast errors (i.e., mean absolute error) and the mean of the squared forecast errors (i.e., mean squared error). As discussed above, the natural approach is to adopt the metric that is consistent with the loss function underlying the estimator. Hence, given that most extant approaches embed a loss function that penalizes mean squared error, mean squared error is often the natural choice. Moreover, mean squared error has a nice property: It is increasing in both the variance of the forecast errors and the square of the bias of the forecasts.

Differences in scale relate to the fact that forecast errors drawn from a large panel of firm-years may not be comparable because of temporal and cross-sectional variation in size. For example, if forecasts of unscaled net income are evaluated, it is likely that the variance of the forecast errors will be increasing over time and in firm size. Hence, a relatively poor forecast of a small firm's net income may lead to a much smaller forecast error than a relatively good forecast of a large firm's net income. One way to deal with scale is to forecast profitability ratios such as ROE and its components.

Forecasting ratios is not always appropriate, however. Rather, as discussed in Section 3, the research question or context may imply that unscaled earnings is the variable of interest. If this is the case, scale can be dealt with by forecasting unscaled earnings, and then deflating the forecast errors. The seemingly natural choice of deflator is the realized value of the variable being forecasted. Unfortunately, this is not a good choice when earnings is the variable being forecasted. The reason for this is that earnings are often either negative or very close to zero. Consequently, another deflator must be chosen. Unless a clear case can be made for an alternative, equity market value at the end of the period in which the forecast is *made* is the default choice. There are two reasons for this. First, equity market value is a good scale proxy. Second, the ratio of the forecast error to lagged equity market value reflects the degree to which the forecast error affects the inverse of the price-to-forward-earnings ratio, which is a natural valuation multiple to evaluate given the AEGV model described in Section 2.

6.6.2 Evaluating economic magnitude

Colloquially speaking, economic magnitude relates to the question: Does it matter? That is, given the economic issue that motivates the study, are the results material? Clearly, how to evaluate economic magnitude depends on the context. That said, given that most studies of earnings forecasts have a valuation context, I will focus on valuation. Within the context of valuation, a direct way to evaluate economic magnitude is to compare the relative valuation errors obtained using the earnings forecasts from the different approaches evaluated in the study. Although this is a natural comparison to make, there are two drawbacks. First, most forecasting approaches relate to relatively short horizons such as three years or less; hence, in order to develop a value estimate, assumptions about how earnings will evolve over the long run must be made. To the extent that the errors in these assumptions are correlated with the forecasts of near-term earnings, the test is confounded.

Second, in order to arrive at a value estimate, the discount rate must be estimated. However, there is no well-accepted approach for doing this; and, again, to the extent that the errors in estimates of the discount rate are correlated with the forecasts, the tests are confounded. One way to deal with this issue is to evaluate the accuracy of the discount rate implied by current price and the different forecasts — that is, the implied cost of capital. (This is the approach taken in Hou *et al.* (2012) and Li and Mohanram (2014).) However, this is also imperfect because assumptions about long-run earnings are still required and, as discussed in Easton and Monahan (2005) and Easton and Monahan (2016), implied cost of capital proxies often exhibit low reliability.

The above implies that the relative accuracy of value estimates and implied costs of capital are imperfect indicators about the accuracy of earnings forecasts. This does not imply that tests of these phenomena are inappropriate. Rather, it implies that, to the extent possible, these tests should be augmented by other tests. For example, Hou *et al.* (2012) evaluate earnings response coefficients, ERCs, which measure the relation between stock returns and contemporaneous forecast errors.

The motivation for evaluating ERCs is that, as discussed in Brown *et al.* (1987), if a particular forecasting approach is better, it will generate forecast errors that are closer to contemporaneous revisions in investors' expectations about future cash flows — that is, cash flow news. Consequently, its forecast errors will exhibit a relatively high association with contemporaneous stock returns. This is a reasonable assumption but there are two issues. First, another implicit assumption underlying the test is that forecast errors are uncorrelated with contemporaneous revisions in investors' expectations about future discount rates — that is, discount rate news. If this is not the case, proxies for discount rate news must be included in the regression. However, if these proxies are measured with error, the tests are confounded. Second, differences in ERCs cannot be directly translated into differences in the accuracy of value estimates or implied costs of capital. Hence, although evaluating differences in ERCs provides a nice complement to tests of the relative accuracy of value estimates and implied costs of capital, they are not a substitute for these tests.

Finally, it is important to note that each of the above tests embeds the implicit assumption that security markets are efficient — that is, the expectations underlying observed equity prices are unbiased and reflect all contemporaneously available public information. Although market efficiency is a reasonable hypothesis, it is not an axiom. Hence, an alternative approach for evaluating economic magnitude is to test whether the earnings forecasting approach being studied can be used as part of a trading strategy that generates abnormal returns (e.g., Ou, 1990). I refer to tests of this type as "trading-strategy" tests.

There are two issues with trading-strategy tests. First, there is the well-known joint-hypothesis problem, which is described in Fama (1991). That is, when testing a trading strategy, the researcher is testing both: (1) market efficiency and (2) whether she has fully controlled for risk. Given that the second hypothesis cannot be separately evaluated, abnormal returns generated by the strategy may be attributable to either mispricing or risk. Hence, results of trading-strategy tests are difficult to interpret.

Second, although market efficiency is not an axiom, it is a reasonable first-order approximation of how security prices are formed. Relative to other markets — for example, the market for cars, houses, etc. — securities markets are liquid and have low transactions costs. Hence, arbitrage opportunities are likely to be rare and fleeting. This implies that trading-strategy tests set a very high bar. In particular, a forecasting approach that cannot be used as part of a successful trading strategy may still be a good approach. If the marginal investor is rational and the limits to arbitrage are low, a good forecast of year t earnings will equal the sum of: (1) the expectation of year t earnings that is embedded in price and (2) a low-variance, zero-mean random variable. Hence, even if it is a good approach. the approach that generated this forecast cannot be used to develop a profitable trading strategy.

6.7 Advantages of panel-data approaches

Vis-à-vis ARIMA models, panel-data approaches have five *a priori* advantages. First, they allow the researcher to consider a broad set of choices of what to predict. For example, they can be used to predict probabilities, which is not possible with ARIMA models.

Second, panel-data approaches can accommodate multiple predictors and different weights can be assigned to different predictors. This is crucial given extant empirical results and the manner in which fundamental analysis is practiced and taught. For example, earnings components (e.g., revenues, gross margin, special items, etc.) can now be used to forecast net income; operating profit margin and invested capital turnover can be used to forecast return on invested capital; negative earnings can be given a different weight than positive earnings; etc.

Third, parameter estimates obtained from panel-data models are based on information taken from both the cross section and the time series. Moreover, there is flexibility regarding the choice of the information set. For example, observations can be grouped by a particular attribute (e.g., industry), and then separate parameter vectors can be estimated for each group. In addition, different predictor sets can be used for different groups. Hence, panel-data approaches allow the researcher to evaluate the role that peer-based benchmarking plays in the forecasting process.

Fourth, panel-data approaches are less subject to concerns about non-stationarity. Because information from the cross section is used to estimate the parameters, a short time series can be chosen. Hence, only recent data, which are more likely to come from the same data-generating process, are used. This is a clear advantage vis-à-vis ARIMA models, which will only yield precise forecasts if there is a long, covariance-stationary time series available, which, as discussed in the previous section, is typically not the case.

Finally, and related to the previous point, panel-data approaches yield a larger number of predictions than ARIMA models. The reason for this is that, even if a firm does not contribute observations to the estimation sample, a forecast can still be generated as long as that firm reports non-missing values of the predictors in year T.

In light of their *a priori* advantages, it is not surprising that researchers looked to panel-data approaches as a way of dealing with the deficiencies of ARIMA models. However, extant empirical evidence regarding the relative advantages of panel-data approaches is mixed. Two sets of results are particularly noteworthy. First, the results in FRLY, which are discussed above, suggest that peer analysis does not lead to more accurate forecasts. Second, extant evidence in Gerakos and Gramacy (2013) and Li and Mohanram (2014) suggest that forecasts obtained from panel-data approaches are not substantially more accurate than forecasts obtained from the random-walk model.[3] For example, as shown in Table 2 of Li and Mohanram (2014), the random-walk model is ranked as the most accurate approach more frequently than any of the three alternative approaches evaluated; and, it performs well in terms of mean and median absolute forecast error. Its mean (median) absolute forecast errors are always lower than the mean (median) absolute forecast errors generated by the model developed in Hou *et al.* (2012). The remaining two models considered each generate forecasts of year-ahead earnings that have lower mean absolute errors than forecasts obtained from the random-walk model. However, there is no difference in the median error. Moreover, when longer horizons are considered, the random-walk model is either as accurate or more accurate than either of these two approaches.

The two sets of results described above are noteworthy because, without evidence to the contrary, they lead to the conclusion that two fundamental tenets of financial statement analysis are incorrect. These tenets are: (1) how well a company is performing relative to its peers is informative about its future earnings and (2) vis-à-vis

[3]Studies that pre-date Gerakos and Gramacy (2013) and Li and Mohanram (2014) lead to a similar conclusion. Although the more elaborate approaches developed in these studies generate more accurate forecasts than the random-walk model, the differences in accuracy are typically small.

bottom-line earnings, financial ratios and financial statement line items are incrementally informative about future earnings. Are these fundamental tenets incorrect? Is it truly the case that the best advice to give students and investors is that they should simply assume that future earnings will equal current earnings? The process of providing and interpreting financial reports that are comparable across companies and time and that are disaggregated into details about sources of revenues, expense, assets, liabilities, etc. is expensive. However, if comparability and disaggregation do not lead to reports that are more useful in terms of estimating value, are we as a society overspending on preparation and analysis?

The above questions are central to accounting research. Hence, they imply that there are tremendous research opportunities regarding the use of historical financial data to predict future earnings. Moreover, given their *a priori* advantages, panel-data approaches are the state of the art at present. Consequently, further study of them is, in my opinion, a promising research agenda.

6.8 Summary

Relative to ARIMA models, panel-data approaches have clear *a priori* advantages. However, at present, the empirical evidence regarding their superiority vis-à-vis ARIMA models is inconclusive. This fact can be seen as either a reason to become discouraged or as an opportunity. In my opinion, it is the latter. Panel-data approaches are remarkably flexible in terms of choice of: (1) what to predict; (2) potential predictors; (3) type of estimator; and, (4) the composition of the estimation sample. Although the extant literature generates a number of interesting insights, it does not fully explore the possibilities. Moreover, the question of whether historical accounting numbers are useful for forecasting earnings is central to accounting research. Hence, in my opinion, additional study of panel-data approaches is necessary.

7

Accounting Measurement and Earnings Predictability

Accounting measurement refers to the rules and practices that determine whether a transaction or event generates a corresponding asset or liability on the balance sheet; and, if so, what amount is recognized, when is that amount recognized and how is the asset or liability accounted for during the post-recognition period. For example, GAAP and IFRS have standards regarding whether an expenditure is recognized as an asset on the balance sheet or immediately expensed. These standards also state how the asset is measured both at the time of recognition and subsequently. For example, per GAAP, an expenditure to acquire a piece of equipment is recognized on the balance sheet as a long-lived tangible asset. The asset is initially recognized at its cost, and then it is depreciated. Hence, during the post-recognition period, the asset's value declines and depreciation expense is recognized on the income statement.

Accounting measurement is an important, broad and complex subject that continues to receive tremendous attention from practitioners, standard setters and academics. In this section, I focus on a relatively-narrow yet important question: How does accounting measurement affect earnings predictability? I begin by describing how accounting measurement choices affect the properties of reported earnings. Next, I discuss the evidence regarding the relative persistence of accruals and cash flows. I then discuss the issue of conservatism. Finally, I summarize this section and use that summary to motivate the next section.

7.1 Accounting measurement and the properties of reported earnings

As discussed in Section 2, clean-surplus accounting implies the following:

$$\mathrm{EARN}_t = \mathrm{DIV}_t + \Delta B_t, \tag{7.1}$$

$$\mathrm{EARN}_t^{\mathrm{ENT}} = \mathrm{FCF}_t + \Delta \mathrm{IC}_t. \tag{7.2}$$

That is, equity-level earnings, EARN_t, (enterprise-level earnings, $\mathrm{EARN}_t^{\mathrm{ENT}}$) for period t equals the sum of: (1) contemporaneous dividends, DIV_t, (free cash flow, FCF_t) and (2) the contemporaneous change in equity book value, ΔB_t, (invested capital, $\Delta \mathrm{IC}_t$). Hence, earnings has both a cash flow component (i.e., either DIV_t or FCF_t) and an accrual component (i.e., either ΔB_t or $\Delta \mathrm{IC}_t$).

Equations (7.1) and (7.2) establish that the balance sheet and income statement are inextricably linked and that accounting-measurement rules affect earnings via their effects on accruals. For example, if an expenditure made on date t to obtain a piece of equipment is recognized as an asset, both ΔB_t and $\Delta \mathrm{IC}_t$ are higher than they would have been had the expenditure been immediately expensed. Hence, equity- and enterprise-level earnings for period t are also higher. However, subsequent earnings are lower. The reason for this is that, assuming the asset has a depreciable life of $L, \forall \tau \leq L, \Delta B_{t+\tau}$ and $\Delta \mathrm{IC}_{t+\tau}$ are reduced by depreciation expense.

The above leads to three conclusions. First, accruals are the reason that current earnings are more informative about future dividends than current cash flows (e.g., Ball and Brown, 1968; Nichols and Wahlen, 2004). Moreover, this outcome is not surprising. To the extent that accounting measurement is done in a manner such that recognized assets and liabilities reflect resources and claims on those resources, $\Delta B_{t+\tau}$ reflects the change in net resources. If this change is positive (negative), the firm has more (less) resources and, thus, expected dividends are higher (lower).

Second, the term earnings quality is somewhat of a misnomer; rather, it is the quality of the balance sheet that ultimately matters.[1] This is true even if there is dirty-surplus accounting and, thus, Equations (7.1) and (7.2) do not hold. The reason for this is that even if accruals do not reflect the entire change in the balance sheet, every accrual reflects a change in a balance-sheet item.

Finally, although the quality of the income statement is ultimately a function of the quality of the balance sheet, a low-quality balance sheet does not necessarily imply a low-quality income statement. The reason for this is that earnings are a function of the *change* in the balance sheet. For example, if inventory is consistently over-valued by the same amount, cost of goods sold, gross profit and net income are unaffected. This is often referred to the cancelling-error property of earnings; and, as discussed below, it implies that growth is an important determinant of how accounting measurement affects the relation between historical earnings and future earnings.

7.2 The relative persistence of cash flows and accruals

Accruals are based on estimates, which are made by managers who have imperfect information, behavioral biases and incentives to manipulate reported financial statements. Moreover, as discussed in Beaver and Demski (1979), in a world of imperfect and incomplete markets, there is no best accounting-measurement scheme. Cash flow, on the other hand, can be objectively measured and verified.

In light of the above, a reasonable hypothesis is that accruals are less persistent than cash flows and that when historical earnings contain extreme accruals (either positive or negative), they are less useful for predicting future earnings. This hypothesis is the primary motivation

[1] This is not always the case. For example, choices about how to assign revenues and expenses to the various line items of the income statement typically do not affect the balance sheet but can affect the information content of earnings. For example, if ongoing operating expenses are combined with transitory impairment charges, investors' perceptions about the persistence of earnings may be affected.

for the study by Sloan (1996). He estimates the following regression:

$$\mathrm{ROA}_{t+1} = \gamma_0 + \gamma_1 \times \mathrm{ACC}_t + \gamma_2 \times \mathrm{CF}_t + \varepsilon_{t+1}. \quad (7.3)$$

In Equation (7.3), ROA_{t+1} is GAAP income from continuing operations in year $t+1$ divided by average total assets in year t. ACC_t and CF_t are the accrual and cash flow components of ROA_t, respectively; and, ε_{t+1} is an error term. The numerator of ACC_t equals the change in non-cash current assets less the change in current liabilities (exclusive of short-term debt and taxes payable) less depreciation. The numerator of CF_t is the difference between the numerator of ROA_t and the numerator of ACC_t. Average total assets in year t is the denominator of both ACC_t and CF_t.

Sloan (1996) shows that $\gamma_1 < \gamma_2$ and that the difference between the two parameters is statistically significant. This result is robust. For example, he prepares time-series plots of $\mathrm{ROA}_{t-5}, \ldots, \mathrm{ROA}_{t+5}$ for firm-years in the extreme deciles of $\mathrm{ROA}_t, \mathrm{ACC}_t$, and CF_t, respectively. These plots illustrate that firms with extreme values of ACC_t experience rapid mean reversion in return on assets whereas firms with extreme values of either ROA_t or CF_t do not. Taken together, these results imply that the accrual component of earnings is less persistent than the cash flow component.

In addition to showing that extreme accruals imply less persistent earnings, Sloan (1996) also provides evidence that suggests that, during his sample period, investors did not understand the implications of current accruals for future earnings. He shows that an investment portfolio that is long (short) firms with values of ACC_t in the bottom (top) decile earns positive size-adjusted returns in years $t+1$ and $t+2$. These results are consistent with investors functionally fixating on bottom-line earnings and ignoring the information about mean reversion implied by the relative size of the accrual component.

Within the context of earnings forecasting, the results in Sloan (1996) are important because they suggest that accounting-measurement rules and the manner in which managers implement these rules affects the usefulness of historical earnings for developing forecasts of future earnings. Extensions of Sloan's study support this conclusion. For example, Richardson *et al.* (2005) and Richardson *et al.* (2006) provide evidence of a positive association between: (1) the reliability of accruals and earnings persistence and (2) extreme accruals and allegations of earnings manipulation.

7.3 Accounting conservatism and earnings predictability

The study by Sloan (1996) and extensions of it lead to the conclusion that, colloquially speaking, too much accounting is a bad thing. That is, when accounting accruals are extreme, historical earnings are less useful for predicting future earnings. An immediate question arises: Is the opposite also true? That is, colloquially speaking, is too little accounting also a bad thing?

To put structure on the above question, I express equity value at date t, V_t, as follows:

$$V_t = B_t + \mathrm{DIFF}_t \tag{7.4}$$

In the above equation, DIFF_t is the date t difference between equity value and equity book value. DIFF_t has two, non-mutually exclusive sources: (1) non-zero-net-present-value projects and (2) accounting bias. Regarding the first source, under most accounting-measurement rules, $\mathrm{DIFF}_t \neq 0$ when current or future investments are expected to generate returns that are unequal to the cost of equity capital. Regarding the second source, accounting-measurement rules are biased. Consequently, even if all investment projects have zero net present value, differences between equity market value and equity book value arise and $\lim_{t\to\infty} E_0[\mathrm{DIFF}_t] \neq 0$.

Conceptually, accounting bias can arise from either aggressive or conservative accounting-measurement rules.[2] However, in practice, accounting-measurement rules tend to be conservative. For example, consider the treatment of research and development, R&D, expenditures. With some exceptions, present GAAP and IFRS do not allow capitalization of R&D expenditures. Rather, these expenditures are immediately expensed. Hence, unless all of these expenditures are

[2] Accounting bias leads to a lack of matching, which is described in section 2. Conservative (aggressive) accounting-measurement rules imply that: (1) expenses are recognized before (after) revenues or (2) revenues are recognized after (before) they are earned.

made on projects that are worthless, which is unlikely, $\mathrm{DIFF}_t > 0$. Moreover, if the firm is expected to continue to invest in R&D, $\lim_{t\to\infty} E_0[\mathrm{DIFF}_t] > 0$ even if the expected net present value of future investments converges to zero as $t \to \infty$.

The accounting treatment of R&D is not the only example of conservative accounting-measurement rules. Other examples include: (1) immediate expensing of expenditures on internally generated intangibles — e.g., branding and marketing, supplier relations, etc.; (2) impairment rules that require assets to be written down if their fair value is less than their carrying value but that do not allow these write downs to be reversed if the fair value subsequently increases; (3) when inventory quantities and costs are increasing, the last-in-first-out (LIFO), cost-flow assumption for valuing inventory; etc. Moreover, accountants tend to follow informal guidelines such as "recognize losses when expected but gains when realized," which imply that $\lim_{t\to\infty} E_0[\mathrm{DIFF}_t] > 0$.

In light of the above, in the remainder of the discussion, I mainly focus on accounting-measurement rules that are conservative — that is, $\lim_{t\to\infty} E_0[\mathrm{DIFF}_t] > 0$. In addition, for ease of exposition, I assume that the net present value of every project is equal to zero. As discussed in Rajan *et al.* (2007), this assumption can be relaxed.

By combining Equation (7.4) with either Equation (7.1) or (7.2), I obtain the following expressions for equity- and enterprise-level earnings. When doing this, I assume that net financial assets are reported on the balance sheet at their fair value; hence, financing activities do not affect DIFF_t.

$$\mathrm{EARN}_t = (\mathrm{DIV}_t + \Delta V_t) - \Delta\mathrm{DIFF}_t \tag{7.5}$$

$$\mathrm{EARN}_t^{\mathrm{ENT}} = (\mathrm{FCF}_t + \Delta V_t^{\mathrm{ENT}}) - \Delta\mathrm{DIFF}_t. \tag{7.6}$$

Equations (7.5) and (7.6) lead to an important conclusion: Reported earnings equal the amount of value created or destroyed during the period unless: (1) the accounting-measurement rules are biased *and* (2) the bias changed during the period. Hence, conservative accounting only affects the relation between earnings and value creation when there is growth in the bias attributable to conservatism. Empirical evidence in Monahan (2005) is consistent with this argument.

Equations (7.5) and (7.6) also show that conservatism embodies the notion of "too little" accounting. In particular, if $\frac{\mathrm{DIFF}_t}{V_t^{\mathrm{ENT}}}$ is large $\forall t$, $\frac{\Delta V_t^{\mathrm{ENT}} - \Delta\mathrm{DIFF}_t}{V_t^{\mathrm{ENT}}} \cong 0 \forall t$ — that is, there are no extreme accruals. For example, if an expenditure made at time t is immediately expensed, there are no contemporaneous positive accruals or future negative accruals. Moreover, cash-basis accounting can be thought of as an example of extreme conservatism and much too little accounting. The reason for this is that, under cash-basis accounting, $\frac{\mathrm{DIFF}_t}{V_t^{\mathrm{ENT}}} = 1 \forall t$ and $\frac{\Delta V_t^{\mathrm{ENT}} - \Delta\mathrm{DIFF}_t}{V_t^{\mathrm{ENT}}} = 0 \forall t$.

Next, I evaluate the effects of biased accounting on return on equity, ROE_t, and return on invested capital, ROIC_t, which are shown below:

$$\mathrm{ROE}_t = \frac{\mathrm{EARN}_t}{B_{t-1}} = \frac{(\mathrm{DIV}_t + \Delta V_t) - \Delta\mathrm{DIFF}_t}{V_{t-1} - \mathrm{DIFF}_{t-1}} \tag{7.7}$$

$$\mathrm{ROIC}_t = \frac{\mathrm{EARN}_t^{\mathrm{ENT}}}{\mathrm{IC}_{t-1}} = \frac{(\mathrm{FCF}_t + \Delta V_t^{\mathrm{ENT}}) - \Delta\mathrm{DIFF}_t}{V_{t-1}^{\mathrm{ENT}} - \mathrm{DIFF}_{t-1}}. \tag{7.8}$$

To facilitate the discussion of the above equations, I begin by noting that if the accounting-measurement rules are unbiased and all projects have zero net present value $\forall t, \mathrm{DIFF}_t = 0, \mathrm{ROE}_t = (\mathrm{DIV}_t + \Delta V_t)/V_{t-1} \equiv \mathrm{ROE}_t^{\mathrm{UNB}}$ and $\mathrm{ROIC}_t = (\mathrm{FCF}_t + \Delta V_t^{\mathrm{ENT}})/V_{t-1}^{\mathrm{ENT}} \equiv \mathrm{ROIC}_t^{\mathrm{UNB}}$. That is, when there is no accounting bias and all projects have zero net present values, profitability ratios are unbiased and equal to the economic rate of return.

If $\mathrm{IC}_{t-1} > 0$ and $\Delta\mathrm{DIFF}_t = g_{\mathrm{DIFF}} \times \mathrm{DIFF}_{t-1}$, as shown in Lev *et al.* (2005), the results in the table shown below hold.[3] (If $B_{t-1} > 0, \mathrm{ROE}_t$ and $\mathrm{ROE}_t^{\mathrm{UNB}}$ can be substituted for ROIC_t and $\mathrm{ROIC}_t^{\mathrm{UNB}}$, respectively.)

	$g_{\mathrm{DIFF}} \leq \mathrm{ROIC}_t^{\mathrm{UNB}}$	$g_{\mathrm{DIFF}} > \mathrm{ROIC}_t^{\mathrm{UNB}}$
$\mathrm{DIFF}_t > 0 \ \forall \ t$	$\mathrm{ROIC}_t \geq \mathrm{ROIC}_t^{\mathrm{UNB}}$	$\mathrm{ROIC}_t < \mathrm{ROIC}_t^{\mathrm{UNB}}$
$\mathrm{DIFF}_t < 0 \ \forall \ t$	$\mathrm{ROIC}_t \leq \mathrm{ROIC}_t^{\mathrm{UNB}}$	$\mathrm{ROIC}_t > \mathrm{ROIC}_t^{\mathrm{UNB}}$

The above table shows that, when the accounting-measurement rules are conservative — that is, $\mathrm{DIFF}_t > 0 \ \forall \ t$, high (low) growth in expenditures

[3] Rajan *et al.* (2007) derive similar results in a more general setting.

that are accounted for conservatively implies relatively low (high) reported ROIC and negative (positive) residual income. Intuitively, there is a numerator effect that leads to lower ROIC and a countervailing denominator effect. When growth is high, the bias in reported earnings is high, the numerator effect dominates, reported ROIC is low and residual income is negative. The opposite occurs when growth is low.

The above results imply that conservatism reduces the usefulness of historical accounting ratios for developing forecasts of future accounting ratios. The reason for this is that growth rates in investment tend to fall as firms age. Hence, conservatism implies that early-stage firms have low (negative) profitability ratios (residual incomes) that are not informative about future profitability ratios (residual incomes), which are high (positive) in expectation. Hence, when accounting is conservative, value estimates that use forecasts of residual income that are based on historical accounting numbers will be inaccurate.

The empirical evidence shown in Monahan (2005) supports the above hypothesis. He adjusts reported earnings and equity book values for the conservative bias attributable to immediately expensing R&D expenditures. He then uses the time series of adjusted residual incomes to develop forecasts of future residual incomes, which he combines with adjusted equity book value to arrive at an estimate of contemporaneous equity value. He shows that these value estimates are more accurate than value estimates based on forecasts implied by the time series of unadjusted residual incomes and reported equity book value. Moreover, the improvement in accuracy is highest when past growth in R&D is high.

Penman and Zhang (2002) (PZ hereafter) also consider the interaction between conservatism and growth. They evaluate time-series plots of $\mathrm{ROIC}_{t-5}, \ldots, \mathrm{ROIC}_{t+5}$ for firm-years that have year t values of the variable Q_t in either the top or bottom tercile. Q_t is an estimate of the growth in the bias attributable conservative-accounting rules. PZ show that during years $t-5$ through t there is a steady downward trend in the average ROIC of firms with high values of Q_t but, in year $t+1$, average ROIC increases and remains high thereafter. On the other hand, firms with low values of Q_t have constant ROIC during years $t-1$ through t but, in year $t+1$, average ROIC decreases and remains low thereafter.

The evidence in PZ is important because it suggests that, when there is accounting conservatism and extreme growth, historical trends in ROIC are misleading. In particular, downward trends tend to be followed by increases and positive trends tend to be followed by decreases. Moreover, the evidence suggests that, during the time period studied by PZ, investors were misled. In particular, PZ show that an investment portfolio that is long (short) firms with values of Q_t in the top (bottom) decile earns negative size-adjusted returns in years $t-3$ through t and positive size-adjusted returns in years $t+1$ through $t+5$. These results are consistent with investors functionally fixating on historical trends in ROIC, and then being *predictably* surprised when these trends reverse.

The above discussion leads to the conclusion that, to the extent conservatism is equated with "too little" accounting, too little accounting is a bad thing in the sense that, when combined with growth, it causes historical accounting numbers to be less useful (and sometimes misleading) predictors of future accounting numbers.

7.4 Summary and segue to the next chapter

Perhaps not surprisingly, accounting measurement affects the usefulness of historical accounting numbers for developing forecasts of future earnings. This result is important because it has implications for what constitutes "good" accounting. By good, I mean accounting rules that imply that: (1) reported earnings have a strong, positive association with changes in value and (2) historical trends in profitability ratios are informative about future profitability ratios.

The main point of this section is that accounting-measurement rules that imply either too much or too little accounting are not, in the manner that I've defined it, good. There is no corner solution. Although cash-basis accounting implies that reported earnings are objectively measured and easily verified, it is too conservative. On the other hand, although accounting-measurement rules that require that all assets and liabilities be reported at their fair values are conceptually unbiased, these rules are not implementable. The fair values of most assets and liabilities are not observable and have to be estimated by managers, who will make errors and have incentives to manipulate reported numbers.

Hence, accruals attributable to changes in estimates of fair values may reflect the "true" change in value, measurement error or managerial manipulation. Hence, although fair-value accounting is conceptually appealing, it will result in extreme accruals that are uninformative about either the contemporaneous change in value or future earnings.

The above implies that there is no dominant accounting-measurement scheme; consequently, research should focus on evaluating the properties of different schemes. In this section, I discuss two properties: (1) the relation between earnings and changes in value and (2) the usefulness of historical accounting numbers for predicting the mean — that is, the expected value — of future earnings. The latter is motivated by the fact that, as discussed in Section 2, expected future earnings are a key determinant of value. However, uncertainty also affects value; and, as discussed in Penman (2016), conservative measurement rules are typically adopted when there is high uncertainty. Hence, within the context of both valuation and accounting measurement rules, expected earnings are not all that matter; rather, higher moments that reflect uncertainty matter too. With this in mind, I turn to the next section.

8

Forecasting the Higher Moments of Future Earnings

In the previous sections, I focus entirely on studies that either explicitly or implicitly develop and evaluate forecasts of the expected value — that is, the mean or first moment — of future earnings. In these sections, I ignore studies that develop approaches for forecasting higher moments. The primary reason for this is that, at present, these studies are rare. In my opinion, this is both unfortunate and an opportunity. Hence, in this section, I provide a brief overview of the issues related to forecasting higher moments. I begin by describing analytical results that imply that the higher moments of future earnings are relevant within the context of equity and debt valuation. I then discuss empirical issues that are unique to developing forecasts of higher moments and the extant empirical literature.

8.1 The economic relevance of the higher moments of future earnings

Higher moments can be put into two categories: (1) comoments and (2) total moments. An example of a comoment is the covariance of residual income and the stochastic discount factor. On the other hand, an example of a total moment is the variance of residual income.

Models based on the assumptions of classical finance — that is, rational agents who are symmetrically informed, have homogenous beliefs and preferences, and trade in unsegmented, frictionless capital markets — lead to the conclusion that, when valuing equities, only two moments matter: (1) the expected value of future earnings and (2) the covariances of future earnings with future stochastic discount factors. In particular, when the discount rate is either stochastic or time-varying, equity value at time zero equals (Feltham and Ohlson, 1999):

$$\begin{aligned} V_0 &= B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{EARN}_t - r^{\mathrm{f}}_{t-1,t} \times B_{t-1}]}{R^{\mathrm{f}}_{0,t}} \\ &\quad + \sum_{t=1}^{\infty} \mathrm{COV}_0(m_{0,t}, (\mathrm{EARN}_t - r^{\mathrm{f}}_{t-1,t} \times B_{t-1})). \end{aligned} \tag{8.1}$$

In Equation (8.1), $r^{\mathrm{f}}_{t-1,t}$ is the *uncertain* risk-free return from date $t-1$ to date t, $R^{\mathrm{f}}_{0,t} = (1 + r^{\mathrm{f}}_{0,t})$ and $m_{0,t}$ is the stochastic discount factor for period t.

Equation (8.1) demonstrates that value is a function of: (1) the present value of expected residual income discounted at the risk-free rate and (2) an additive risk adjustment that reflects the covariances between future residual incomes and future stochastic discount factors.[1,2] Hence,

---

[1]As discussed in Cochrane (2005), the consumption asset-pricing model implies that $m_{0,t} = \beta \frac{u^{`}(c_t)}{u^{`}(c_0)}$. β is an impatience parameter and $u^{`}(c_t)$ is the marginal utility of consumption in year t. Hence, $m_{0,t}$ reflects the marginal rate of substitution of consumption in t for consumption in period zero. This tradeoff depends on the outcome in year t, which is unknown at time zero. Hence, $m_{0,t}$ is a random variable. Moreover, assuming risk aversion, $u^{`}(c_t)$ is decreasing in consumption. Hence, $m_{0,t}$ is high (low) when aggregate consumption is low (high). For the typical or "average" firm, earnings are positively associated with aggregate consumption; consequently, for the typical or average firm the covariance term in Equation (8.1) is negative.

[2]This result is not unique to the residual income valuation model and it also obtains at the enterprise level. In particular, Christensen and Feltham (2009) show that the abnormal earnings growth valuation model can be expressed as a function of: (1) expected year-ahead earnings divided by $r^{\mathrm{f}}_{0,1}$; (2) the present value of subsequent expected abnormal earnings growth discounted at the risk-free rate; and, (3) an additive risk adjustment that reflects the covariances between future abnormal earnings growth and future stochastic discount factors. Moreover, abnormal earnings growth is defined as a function of the risk-free rate. Christensen and Feltham (2009) also show that the approach is valid at both the equity and enterprise level.

in order to estimate value, investors must estimate the covariance term. However, total moments do not matter.

Recent analytical results, however, imply that higher *total* moments are relevant for equity valuation. For example, Pastor and Veronesi (2003) develop a model in which equity value is an increasing function of long-run equity book value. This implies that equity value is an increasing, convex function of the growth rate of equity book value, which is increasing in ROE. Consequently, given Jensen's inequality, Pastor and Veronesi (2003) show that equity value is increasing in the variance of ROE. Johnson (2004) uses option-pricing theory to arrive at a similar result for levered firms.

Another set of studies argue that the assumptions made in classical finance that imply that investors hold fully diversified portfolios are too restrictive. For example, Merton (1987) shows that investors who trade in segmented markets cannot fully diversify. Consequently, these investors are exposed to idiosyncratic risk and equity value is a decreasing function of volatility. Alternatively, investors may choose not to diversify so that they can gain exposure to certain "lottery" stocks — that is, stocks with positively skewed payoffs. For instance, Brunnermeier *et al.* (2007) show that when optimistic investors "hold beliefs that optimally trade off the *ex ante* benefits of anticipatory utility against the *ex post* costs of basing investment decisions on biased beliefs," these investors choose to under-diversify so that they can hold a portfolio of stocks with positively skewed payoffs. Consequently, equity value is increasing in the skewness of earnings. A similar result is obtained by Barberis and Huang (2008), who assume that investors behave according to cumulative prospect theory. Finally, Mitton and Vorkink (2007) show that, when rational investors have heterogeneous preferences for skewness, equity value is increasing in the skewness of future earnings.[3]

[3]It is important to note that the analytical results in Merton (1987), Brunnermeier *et al.* (2007), Mitton and Vorkink (2007), and Barberis and Huang (2008) are based on two-period models (some authors refer to these as one-period models with two dates). Agents invest in period one and in period two they receive a terminal dividend and consume. Hence, although these papers often refer to the moments of returns, they also relate to the moments of either earnings or dividends. The reason for this is that, in a two-period model, the terminal price equals the terminal dividend, which, in turn, equals the terminal earnings. More generally, infinite-period models lead

Higher total moments are also potentially relevant in the context of debt valuation. As shown in Black and Scholes (1973) and Merton (1974), debt values are decreasing in asset volatility. This result is obtained by assuming that asset returns are normally distributed. Hence, higher moments do not matter in these models. However, if the assumption of normality is relaxed, higher moments are likely also relevant (e.g., Dynkin *et al.*, 2007). That is, given debt holders face relatively high exposure to downside risk while benefitting little from positive shocks (e.g., Easton *et al.*, 2009), they should assign lower values to debt instruments issued by firms with future earnings that exhibit negative skewness or a higher likelihood of extreme outcomes (i.e., positive kurtosis).

The above provides clear motivation for developing and testing different approaches for forecasting higher moments. As discussed above, very few studies do this but there are a few, which I discuss below.

8.2 Empirical issues and empirical evidence

When forecasting higher moments of earnings, the researcher makes the same choices as she makes when forecasting the mean of earnings — that is, she chooses what to predict, the predictors, the estimator, the estimation sample and how to evaluate the predictions. Hence, the discussion in Section 6 remains relevant. However, there are some unique issues that arise when forecasting higher moments

8.2.1 Scale

The moments of unscaled earnings are not homogeneous across firms or time. Moreover, the relation between scale and higher moments is non-linear. Given the complexity of modeling non-linear relations, models of scaled earnings metrics are more tractable and will generate forecast errors that are more comparable across firms and time than models of unscaled metrics. Hence, unless there is a compelling reason, the default choice is to use a scaled earnings metric such as ROE, ROIC, etc.

---

to the conclusion that equity prices are a function of the moments of the future earnings. Hence, it is the moments of earnings that matter.

8.2.2 Defining the stochastic discount factor

In order to forecast the covariance term in Equation (8.1), the stochastic discount factor, $m_{0,t}$, must be defined. The issue is that, as discussed in Chapter 20 of Cochrane (2005) and Sections 6 and 7 of Campbell *et al.* (1996), at present, there is no consensus regarding how economic agents make risk-return tradeoffs. The four-factor model inspired by Fama and French (1993) and Carhart (1997) is popular but it is not based on a well-accepted theory of capital-market equilibrium. On the other hand, consumption-based approaches are derived from first principles but do not perform well empirically.

Given the above, it is not surprising that the two extant studies that attempt to directly estimate the covariance term in Equation (8.1) take different approaches. These two studies are by Nekrasov and Shroff (2009) (NS hereafter) and Bach and Christensen (2016) (BC hereafter). NS assume a linear factor model in which the factors are: (1) accounting beta; (2) the beta of the size factor in ROE; and, (3) the beta of the book-to-market factor in ROE. They also assume that the covariance between the stochastic discount factor and residual ROE (i.e., $\frac{\mathrm{EARN}_t}{B_{t-1}} - r^{\mathrm{f}}_{t-1,t}$) is constant over the forecast horizon. These assumptions imply that the covariance term in Equation (8.1) is a time-invariant weighted sum of the covariances of residual ROE and the three factors mentioned above. As discussed above, this approach is *ad hoc* but it has the advantage of being consistent with extant empirical results (e.g., Fama and French, 1993).

BC, on the other hand, implement an equilibrium approach that is based on the consumption capital asset pricing model, CCAPM. To do this, they make assumptions about the time-series properties of residual ROE and aggregate consumption. They then use estimates of the time-series parameters implied by these assumptions to arrive at a time-varying covariance term. As discussed above, this approach has theoretical support; however, there is very little empirical support for the CCAPM.

The studies by NS and BC are important for three reasons. First, empirical results in both studies imply that estimates of value that are based on finite-horizon versions of Equation (8.1) — that is, estimates that "put risk in the numerator" — are considerably more accurate that estimates that are based on finite-horizon versions of Equations (2.8) — that is, estimates that "put risk in the denominator."

Second, despite the fact that NS and BC make different assumptions about the stochastic discount factor, their value estimates are similar in terms of accuracy. BC interpret this result as follows pp. 6–7: "This suggests that making the risk adjustments in the numerator rather than through a risk-adjusted cost of equity in the denominator is the key to improving valuation accuracy, and maybe not so much whether the risk adjustments are based on an *ad hoc* factor model or they are based on an equilibrium approach using, for example, the CCAPM." This interpretation seems reasonable. However, it leaves the impression that risk adjustments play a relatively minor role in equity valuation. Although possible, this is unlikely. An alternative view is that, at present, our understanding of how to make these adjustments is so rudimentary that both *ad hoc* and equilibrium approaches are equally inaccurate.

Finally, the tests of relative valuation accuracy used by both NS and BC are interesting because they suggest a complement to standard empirical approaches for evaluating models of risk. As discussed in Elton (1999), because the news component embedded in realized returns is so large, temporal and cross-sectional variation in realized returns is a noisy and potentially biased indicator of variation in expected returns. Hence, standard approaches that evaluate the relation between proposed risk measures and realized returns have low power and have led to a "zoo of new factors" (Cochrane, 2011, p. 1027). However, the relative-valuation-accuracy tests used in NS and BC are based on prices not returns. This is an advantage because prices reflect levels of expectations not news resulting from revisions in expectations

8.2.3 Choice of estimator

Estimating the covariances in Equation (8.1) is not particularly complicated. For example, if the stochastic discount factor is defined by reference to a linear factor model, OLS regressions can be used to estimate the factor betas, and then the in-sample estimates can be converted into out-of-sample forecasts. OLS regressions, however, are not appropriate for estimating higher total moments. Rather, the natural choice is to use quantile regressions to forecast the quantiles of the distribution of earnings, and then use the out-of-sample forecasts of the quantiles to arrive at a forecast of the moments of interest. For example, as discussed in Chang *et al.* (2017), a forecast of the conditional variance of ROE in year $T+1$, $\mathrm{var}(\mathrm{ROE}_{T+1}|\cdot)$, is computed as follows:

$$\mathrm{var}(\mathrm{ROE}_{T+1}|\cdot) = \frac{1}{Q}\sum_{q\in Z}(\mathrm{quant}_q(\mathrm{ROE}_{T+1}|\cdot))^2 - (e[\mathrm{ROE}_{T+1}|\cdot])^2. \tag{8.2}$$

In Equation (8.2), $\mathrm{quant}_q(\mathrm{ROE}_{T+1}|\cdot)$ is the out-of-sample forecast of the qth quantile of ROE_{T+1}, q is an element of the set $Z = \{a, \ldots q, \ldots z\}$, which is an ordered sequence of Q numbers that fall between zero and one — that is, $q \in Z = \{a, \ldots q, \ldots z\} \subset (0,1)$; and, $e[\mathrm{ROE}_{T+1}|\cdot]$ is the out-of-sample forecast of the mean of return on equity. ($e[\mathrm{ROE}_{T+1}|\cdot]$ is also based on the out-of-sample forecasts of the Q quantiles.)

Two studies that use quantile regressions to forecast and evaluate the higher moments of future earnings are: (1) Konstantinidi and Pope (2016) (KP hereafter) and (2) Chang *et al.* (2017). Both studies estimate in-sample quantile regressions of ROE on lagged accounting numbers, and then use the out-of-sample forecasts of the quantiles to arrive at out-of-sample forecasts of the higher moments of earnings. When doing this, KP use an *ad hoc* approach to estimate the quantile function and they use heuristic-based estimates of the moments — for example, they use the inter-quartile range as their estimate of the standard deviation. KP then evaluate the relation between the out-of-sample forecasts of these heuristics and seven "outcome" variables. They show that the moments are associated with the "risk in earnings." For example, they show that equity analysts' risk ratings are increasing in the inter-quartile range, skewness and kurtosis of ROE.

The study by Chang *et al.* (2017) differs from KP's in three key ways. First, Chang *et al.* (2017) develop a general approach for estimating both the quantile function and the moments. This approach yields consistent — that is, asymptotically unbiased — estimates of the higher moments. Second, as discussed below, Chang *et al.* (2017) develop an approach for evaluating the absolute and relative construct validity of forecasts of higher moments. They then use this approach to show that their forecasts are reliable on both an absolute and relative basis. In particular, their forecasts of the variance and skewness (kurtosis) are more reliable than (as reliable as) the forecasts generated by all the alternative approaches that they consider, including the approach used by KP.

Finally, Chang *et al.* (2017) show that the higher moments of earnings are priced. They find that, consistent with the results in Pastor and Veronesi (2003) and Johnson (2004), equity prices are increasing in the standard deviation of future earnings. However, debt prices are decreasing in the standard deviation of future earnings, which is consistent with predictions made by Black and Scholes (1973) and Merton (1974). Chang *et al.* (2017) also find that both equity and debt prices are increasing (decreasing) in the skewness (kurtosis) of future earnings. These results are consistent with analytical results in Brunnermeier *et al.* (2007), Mitton and Vorkink (2007), and Barberis and Huang (2008) as well as practitioner arguments described in Dynkin *et al.* (2007).

The studies by KP and Chang *et al.* (2017) are important for three reasons. First, the evidence in Chang *et al.* (2017) shows that it is possible to develop reliable forecasts of the higher moments of future earnings. This result is not obvious. Higher moments relate to rare events and annual earnings are a low-frequency variable. Hence, whether historical data will yield reliable estimates of higher moments is an empirical question. Second, KP and Chang *et al.* (2017) show that historical accounting data are informative about the higher moments of future earnings. Hence, these studies shed additional light of the role that accounting numbers play in the financial statement analysis process. Finally, the results in KP and Chang *et al.* (2017) show that the higher moments of earnings matter. KP show that higher moments are related to various risk proxies; and, Chang *et al.* (2017) show that the higher moments of earnings are priced by investors and that the manner in which they are priced is consistent with predictions made by extant analytical models.[4]

8.2.4 Evaluating forecasts of higher moments

Higher moments are not observable *ex post*. Hence, there is no direct way to evaluate their accuracy. That said, there are indirect ways. With regards to estimates of the covariance term in Equation (8.1), valuation-based tests like those described in NS and BC can be used. For example, a single estimate of the term $B_0 + \sum_{t=1}^{\infty} \frac{E_0[\mathrm{EARN}_t - r^{\mathrm{f}}_{t-1,t} \times B_{t-1}]}{R^{\mathrm{f}}_{0,t}}$ in Equation (8.1) can be developed, the different covariance terms implied by the different forecasting approaches can be added to this estimate to arrive at different value estimates, and then the different estimates of value can be compared to contemporaneous equity market value.

Regarding forecasts of higher moments, Chang *et al.* (2017) show that realized industry-level moments can be used to assess the accuracy of firm-level forecasts. This approach relies on the law of total moments, which is described in Brillinger (1969). To understand the approach better, it is useful to consider the standard deviation, which per the law of total variance, can be re-written as follows. (See Appendix A of Chang *et al.* (2017) for a discussion of how to use the law of total moments to evaluate forecasts of skewness and kurtosis.)

$$\sqrt{\mathrm{VAR}(\mathrm{ROE}_{T+1}|\cdot)} \equiv \sqrt{\mathrm{VAR}(E[\mathrm{ROE}_{T+1}|\cdot]) + E[\mathrm{VAR}(\mathrm{ROE}_{T+1}|\cdot)]}. \tag{8.3}$$

In Equation (8.3), $\mathrm{VAR}(\mathrm{ROE}_{T+1}|\cdot)$ and $E[\mathrm{ROE}_{T+1}|\cdot]$ denote the conditional variance and conditional expectation of ROE_{T+1}, respectively.

[4]Similar to Chang *et al.* (2017), KP also evaluate bond prices. Like Chang *et al.* (2017), they show that bond prices are decreasing in the inter-quartile range and kurtosis of future earnings. However, KP find no association between skewness and bond prices. This implies that bond investors either do not care about whether they hold bonds issued by firms with negatively skewed ROE — that is, firms that have a relatively high likelihood of reporting large accounting losses. Given the asymmetric payoff structure of bonds, this is unlikely. The results in Chang *et al.* (2017), on the other hand, imply that bond investors put a positive price on skewness, which, in turn, implies that investors put lower (higher) prices on bonds issued by firms that are more likely (less likely) to report large accounting losses.

Equation (8.3) implies the following approach. First, develop firm-year forecasts of the mean and the variance of ROE_{T+1}. For example, Chang *et al.* (2017) evaluate forecasts from their quantile-based approach as well as forecasts obtained from several alternative approaches. Second, for each industry year and each estimation approach, use Equation (8.3) to convert the firm-level forecasts into an industry-level forecast. To do this, compute the sum of: (1) the variance of the firm-level forecasts of the mean of ROE_{T+1} and (2) the mean of the firm-level forecasts of the variance of ROE_{T+1}. Finally, compare the industry-level forecasts to the realized within-industry standard deviation of ROE_{T+1}.

8.3 Summary

There is compelling analytical evidence that the higher comoments and total moments of future earnings are relevant for valuation. Hence, empirical evidence regarding how to forecast these moments, how to evaluate these forecasts and the role that they play in determining value is needed. At present, empirical studies of the higher moments of earnings are rare. The evidence in KP and Chang *et al.* (2017) implies that the higher total moments of earnings can be reliably estimated and that these moments are related to alternative risk proxies and value. Nonetheless, there is a lot that is not known about how to forecast the higher total moments of earnings, how to evaluate these forecasts and their economic relevance. Moreover, the evidence regarding forecasts of the covariances between future stochastic discount factors and future residual incomes is ambiguous. Although NS and BC make very different assumptions when estimating these covariances, their value estimates are similar in terms of accuracy. One interpretation of this result is that the two approaches are equally inaccurate. Hence, more study is needed. This is neither surprising nor a criticism of NS and BC. The manner in which investors make risk-return tradeoffs is a crucial yet unresolved issue in financial economics.

9

Summary

Expected future earnings and the higher moments of future earnings are the determinants of the value of a company and the securities it issues. This statement is not mere conventional wisdom. Rather, it is based on compelling analytical and empirical evidence. This is important because it implies that forecasting the moments of future earnings is *the* activity that defines valuation. It also implies that, *if* historical accounting numbers are useful within the context of valuation, it is because they can be used to develop accurate and precise forecasts of the moments of future earnings. A basic premise underlying financial statement analysis is that they can. However, the veracity of this premise is ultimately an empirical question; and, at present, the empirical evidence is ambiguous.

The ambiguity described above is an important gap in the literature and the source of a number of important research questions. In this section, I discuss a few of these questions. Obviously, this is not an exhaustive discussion. The limits of my imagination prevent me from identifying important questions that are obvious to others. Hence, the discussion below relates only to the broad questions that I find interesting and important.

The first, and, in my opinion, most-fundamental question, relates to the relative accuracy of forecasts obtained from the random-walk model. The fact that the random-walk model is the best ARIMA model is not surprising upon reflection. However, the fact that the performance of random-walk model is similar to the performance of extant panel-data approaches is surprising. Panel-data approaches have many *a priori* advantages; and, the random-walk model is inconsistent with standard economic assumptions, accounting practice and the manner in which financial statement analysis is practiced and taught. Hence, additional research on why the random-walk model performs so well is needed. One possibility is that forecasting is too contextual to be described by econometric models. Although this alternative cannot be dismissed out of hand, it cannot be the default answer. After all, the objective of academic research is to bring order out of confusion. Moreover, if the random-walk model is the best academics can do, the relevance of the entire literature on forecasting and financial statement analysis is called into question.

A second question of importance relates to the role of peer analysis in the earnings forecasting process. Extant empirical evidence suggests that approaches that use economy-wide data are as accurate as approaches that use industry-specific data to train the estimator. This evidence is interesting but it is not the final word. Absence of evidence is not evidence of absence. In particular, these results do not imply that identifying a company's peers and benchmarking the company against them is unimportant. Rather, this process may be different from and more complicated than what is assumed in extant studies. For example, grouping on characteristics other than industry (e.g., age, size, etc.) may lead to a set of peers that are more comparable. In addition, the optimal set of predictors may vary systematically across the groups of peers.

The issue of accounting measurement is a third source of questions. The extant evidence regarding the effect of conservative accounting on the usefulness of historical accounting numbers for developing forecasts of future earnings is interesting but likely incomplete. As discussed in Penman (2016), one reason that certain expenditures are accounted for conservatively is that they are made on projects that involve high uncertainty — for example, research and development, customer acquisition, etc. This, in turn, implies that conservative accounting-measurement rules may have implications for the relation between historical accounting numbers and the higher moments of future earnings. Moreover, conservative accounting-measurement rules may lead to accounting numbers that are more informative about the higher moments of future earnings.

A more general issue related to the previous point is the forecasting of higher moments. Although the extant results are promising, there is a lot more that can be learned about how to forecast higher moments, how to validate the forecasts and their role in determining value. Further evaluation of these issues will shed additional light on the valuation role of earnings and the pricing of uncertainty.

Acknowledgements

I benefited greatly from comments given to me by Peter Easton, Sujesh Nambiar and Stephen Penman.

References

Abarbanell, J. S. and B. J. Bushee (1997). "Fundamental analysis, future earnings, and stock prices". *Journal of Accounting Research*. 35(1): 1–24.

Albrecht, W. S., L. L. Lookabill, and J. C. McKeown (1977). "The time-series properties of annual earnings". *Journal of Accounting Research*. 15(2): 226–244.

Arrow, K. J. (1964). "The role of securities in the optimal allocation of risk-bearing". *The Review of Economic Studies*. 31(2): 91–96.

Bach, C. and P. O. Christensen (2016). *Consumption-based Equity Valuation*. Copenhagen Business School. Journal article.

Baginski, S. P. and J. M. Wahlen (2003). "Residual income risk, intrinsic values, and share prices". *The Accounting Review*. 78(1): 327–351.

Ball, R. and P. Brown (1968). "An empirical evaluation of accounting income numbers". *Journal of Accounting Research*. 6(2): 159–178.

Ball, R. and R. Watts (1972). "Some time series properties of accounting income". *The Journal of Finance*. 27(3): 663–681.

Barberis, N. and M. Huang (2008). "Stocks as lotteries: The implications of probability weighting for security prices". *American Economic Review*. 98(5): 2066–2100.

Basu, S. (1997). "The conservatism principle and the asymmetric timeliness of earnings". *Journal of Accounting & Economics*. 24: 3–37.

Beaver, W. H. (1989). *Financial Reporting: An Accounting Revolution.* second edition. Prentice-Hall International, Inc.

Beaver, W. H. and J. S. Demski (1979). "The nature of income measurement". *The Accounting Review.* 54(1): 38–46.

Beaver, W. H. and D. Morse (1978). "What determines price-earnings ratios?" *Financial Analysts Journal.* 34(4): 65–76.

Black, F. and M. Scholes (1973). "The pricing of options and corporate liabilities". *Journal of Political Economy.* 81(3): 637–685.

Box, G. E. P. (1979). "Robustness in the strategy of scientific model building". In: *Robustness in Statistics.* Ed. by R. Launer and G. N. Wilkinson. Academic Press. Elsevier Inc.

Bradshaw, M. T., M. S. Drake, J. N. Myers, and L. A. Myers (2012). "A re-examination of analysts' superiority over time-series forecasts of annual earnings". *Review of Accounting Studies.* 17: 944–968.

Bradshaw, M. T., Y. Ertimur, and P. O'Brien (2016). "Financial analysts and their contribution to well-functioning capital markets". *Foundations and Trends in Accounting.* 11(3): 119–191.

Bradshaw, M. T. and R. G. Sloan (2002). "GAAP versus The Street: An empirical assessment of two alternative definitions of earnings". *Journal of Accounting and Research.* 40(1): 41–66.

Brier, G. W. (1950). "Verification of forecasts expressed in terms of probability". *Monthly Weather Review.* 78(1): 1–3.

Brillinger, D. (1969). "The calculation of cumulants via conditioning". *Annals of the Institute of Statistical Mathematics.* 21(1): 215–218.

Brooks, L. D. and D. A. Buckmaster (1976). "Further evidence on the time series properties of accounting income". *Journal of Finance.* 31(5): 1359–1373.

Brown, L. D. (1993). "Earnings forecasting research: Its implications for capital markets research". *International Journal Forecasting.* 9: 295–320.

Brown, L. D., R. L. Hagerman, P. A. Griffin, and M. E. Zmijewski (1987). "An evaluation of alternative proxies for the market's assessment of unexpected earnings". *Journal of Accounting and Economics.* 9: 159–193.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007). “Optimal beliefs, asset prices, and the preference for skewed returns”. *American Economic Review*. 97(2): 159–165.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1996). *The Econometrics of Financial Markets*. second edition. Princeton University Press.

Carhart, M. M. (1997). “On persistence in mutual fund performance”. *Journal of Finance*. 52(1): 57–82.

Chang, W.-J., S. J. Monahan, A. Ouazad, and F. P. Vasvari (2017). “The higher moments of future earnings”. Working Paper.

Choi, B. C. K. (1998). “Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test”. *American Journal of Epidemiology*. 148(11): 1127–1132.

Christensen, P. O. and G. A. Feltham (2009). “Equity valuation”. *Foundations and Trends in Accounting*. 4(1): 1–49.

Cochrane, J. H. (2005). *Asset Pricing*. Ed. by second edition. Princeton University Press.

Cochrane, J. H. (2011). “Discount Rates”. *Journal of Finance*. 66(4): 1047–1108.

Cook, N. R. (2007). “Use and misuse of the receiver operating characteristics curve in risk prediction”. *Circulation*. 115: 928–935.

Dechow, P. M., A. P. Hutton, and R. G. Sloan (1999). “An empirical assessment of the residual income valuation model”. *Journal of Accounting and Economics*. 26: 1–34.

Dynkin, L., A. Gould, J. Hyman, V. Konstantinovsky, and B. Phelps (2007). *Quantitative Management of Bond Portfolios*. Princeton University Press.

Easton, P. (2016). “Financial reporting: An enterprise operations perspective”. *Journal of Financial Reporting*. 1(1): 143–151.

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

Easton, P. D. and S. Monahan (2016). “Review of recent research on improving earnings forecasts and evaluating accounting-based estimates of the expected return on equity capital”. *Abacus*. 52(1): 35–58.

Easton, P. D., S. J. Monahan, and F. P. Vasvari (2009). “Initial evidence on the role of accounting earnings in the bond market”. *Journal of Accounting Research*. 47(3): 721–766.

Elton, E. J. (1999). “Expected return, realized return and asset pricing tests”. *Journal of Finance*. 54(4): 1199–1220.

Fairfield, P. M., S. Ramnath, and T. L. Yohn (2009). “Do industry-level analyses improve forecasts of financial performance?” *Journal of Accounting Research*. 47: 147–178.

Fairfield, P. M., R. J. Sweeney, and T. L. Yohn (1996). “Accounting classification and the predictive content of earnings”. *The Accounting Review*. 71(3): 337–355.

Fairfield, P. M. and T. L. Yohn (2001). “Using asset turnover and profit margin to forecast changes in profitability”. *Review of Accounting Studies*. 6: 371–385.

Fama, E. F. (1991). “Efficient capital markets: II”. *The Journal of Finance*. 46(5): 1575.

Fama, E. F. and K. R. French (1993). “Common risk factors in the returns on stocks and bonds”. *Journal of Financial Economics*. 33(1): 3–56.

Fama, E. F. and K. R. French (2000). “Forecasting profitability and earnings”. *The Journal of Business*. 73(2): 161–175.

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

Feltham, G. A. and J. A. Ohlson (1999). “Residual earnings valuation with risk and stochastic interest rates”. *The Accounting Review*. 74(2): 165–183.

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

Gao, Z., J. A. Ohlson, and A. J. Ostaszewski (2013). “Dividend policy irrelevancy and the construct of earnings”. *Journal of Business Finance & Accounting*. 40(5–6): 673–694.

Garman, M. B. and J. A. Ohlson (1980). “Information and the sequential valuation of assets in arbitrage-free economics”. *Journal of Accounting Research*. 18(2): 420–440.

Gerakos, J. and R. B. Gramacy (2013). *Regression-based Earnings Forecasts*. University of Chicago Booth School of Business. Working Paper.

Hanley, J. A. and B. J. McNeil (1982). “The meaning and use of the area under a receiver operating characteristic (ROC) curve”. *Radiology*. 143(1): 29–36.

Harsanyi, J. C. (1990). “Interpersonal Utility Comparisons”. In: *Utility and Probability*. Ed. by J. Eatwell, M. Milgate, and P. Newman. The New Palgrave. Palgrave Macmillan, London.

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

Hayek, F. A. (1945). “The use of knowledge in society”. *The American Economic Review*. 35(4): 519–530.

Hou, K. and D. T. Robinson (2006). “Industry concentration and average stock returns”. *Journal of Finance*. 61(4): 1927–1956.

Hou, K. and M. A. van Dijk (2008). “Resurrecting the size effect: Firm size, profitability shocks, and expected stock returns”. Working Paper.

Hou, K., M. A. van Dijk, and Y. Zhang (2012). “The implied cost of capital: A new approach”. *Journal of Accounting Economics*. 53: 504–526.

Johnson, T. C. (2004). “Forecast dispersion and the cross section of expected returns”. *Journal of Finance*. 59(5): 1957–1978.

Joos, P. and G. A. Plesko (2005). “Valuing loss firms”. *The Accounting Review*. 80(3): 847–870.

Kennedy, P. (1992). *A Guide to Econometrics*. third edition. The MIT Press.

Konstantinidi, T. and P. F. Pope (2016). “Forecasting risk in future earnings”. *Contemporary Accounting Research*. 33(2): 487–525.

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

Lev, B., B. Sarath, and T. Sougiannis (2005). “R&D reporting biases and their consequences”. *Contemporary Accounting Research*. 22(4): 977–1026.

Lev, B. and S. R. Thiagarajan (1993). “Fundamental information analysis”. *Journal of Accounting and Research*. 31(2): 190–215.

Li, K. and P. Mohanram (2014). "Evaluating cross-sectional forecasting models for implied cost of capital". *Review of Accounting Studies*. 19(3): 1152–1185.

Lipe, R. C. (1986). "The information contained in the components of earnings". *Journal of Accounting Research*. 24: 37–64.

Lipe, R. and R. Kormendi (1994). "Mean reversion in annual earnings and its implications for security valuation". *Review of Quantitative Finance and Accounting*. 4: 27–46.

McKinsey & Company Inc., T. Koller, M. Goedhart, and D. Wessels (2015). *Valuation: Measuring and Managing the Value of Companies*. 6th. John Wiley and Sons.

Merton, R. C. (1974). "On the pricing of corporate debt: The risk structure of interest rates". *Journal of Finance*. 29(2): 449–470.

Merton, R. C. (1987). "A simple capital market equilibrium with incomplete information". *Journal of Finance*. 42(3): 483–510.

Miller, M. H. and F. Modigliani (1961). "Dividend policy, growth, and the valuation of shares". *The Journal of Business*. 34(4): 411–433.

Mitton, T. and K. Vorkink (2007). "Equilibrium underdiversification and the preference for skewness". *Review of Financial Studies*. 20(4): 1255–1288.

Modigliani, F. and L. H. Miller (1958). "The cost of capital, corporation finance and the theory of investment". *The American Economic Review*. 48(3): 261–297.

Monahan, S. J. (2005). "Conservatism, growth and the role of accounting numbers in the fundamental analysis process". *Review of Accounting Studies*. 10: 227–260.

Myers, J. N. (1999). "Implementing residual income valuation with linear information dynamics". *The Accounting Review*. 74(1): 1–28.

Nekrasov, A. and P. K. Shroff (2009). "Fundamentals-based risk measurement in valuation". *The Accounting Review*. 84(6): 1983–2011.

Nichols, D. C. and J. M. Wahlen (2004). "How do earnings numbers relate to stock returns? A review of classic accounting research with updated evidence". *Accounting Horizons*. 18(4): 263–286.

Nissim, D. and S. H. Penman (2001). "Ratio analysis and equity valuation: From research to practice". *Review of Accounting Studies*. 6: 109–154.

Nissim, D. and S. H. Penman (2003). "Financial statement analysis of leverage and how it informs about profitability and price-to-book ratios". *Review of Accounting Studies*. 8: 531–560.

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

Ohlson, J. A. and Z. Gao (2006). "Earnings, earnings growth and value". *Foundations and Trends in Accounting*. 1(1): 1–70.

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

Ou, A. J. (1990). "The information content of nonearnings accounting numbers as earnings predictors". *Journal of Accounting Research*. 28(1): 144–163.

Pastor, L. and P. Veronesi (2003). "Learning about profitability". *Journal of Finance*. 58(5): 1749–1790.

Penman, S. H. (2005). "Discussion of "On accounting-based valuation formulae" and "Expected EPS and EPS growth as determinants of value"". *Review of Accounting Studies*. 10: 367–378.

Penman, S. H. (2016). "Valuation: Accounting for risk and the expected return". *Abacus*. 52(1): 106–130.

Penman, S. H. and T. Sougiannis (1998). "A comparison of dividend, cash flow, and earnings approaches to equity valuation". *Contemporary Accounting Research*. 15(3): 343–383.

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

Rajan, M. V., S. Reichelstein, and M. T. Soliman (2007). "Conservatism, growth, and return on investment". *Review of Accounting Studies*. 12: 325–370.

Ramakrishnan, R. T. S. and J. K. Thomas (1992). "What matters from the past: Market value, book value, or earning? Earnings valuation and sufficient statistics for prior information". *Journal of Accounting, Auditing & Finance*. 7(4): 423–464.

Ramnath, S., S. Rock, and P. B. Shane (2008). "Financial analysts' forecasts and stock recommendations: A review of the research". *Foundations and Trends in Finance*. 2(4): 311–421.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna (2005). "Accrual reliability, earnings persistence and stock prices". *Journal of Accounting & Economics*. 39: 437–485.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna (2006). "The implications of accounting distortions and growth for accruals and profitability". *The Accounting Review*. 21(3): 713–743.

Sloan, R. G. (1996). "Do stock prices fully reflect information in accruals and cash flows about future earnings?" *The Accounting Review*. 71(3): 289–315.

Soliman, M. T. (2008). "The use of dupont analysis by market participants". *The Accounting Review*. 83(3): 823–853.

Watts, R. L. and R. W. Leftwich (1977). "The time series of annual accounting earnings". *Journal of Accounting Research*. 15(2): 253–271.

Williams, J. B. (1997). "The theory of investment value". *Fraser Publishing Company*: 1–613.