Journal of Accounting Literature 34 (2015) 39–57

Contents lists available at ScienceDirect

Journal of Accounting Literature

journal homepage: www.elsevier.com/locate/acclit

Aggregate earnings and why they matter☆

Ray Ball [a], Gil Sadka [b,*]

[a] University of Chicago, Booth School of Business, United States
[b] University of Texas at Dallas, United States

ARTICLE INFO

Article history:
Received 6 January 2015
Accepted 10 January 2015
Available online 2 February 2015

Keywords:
Aggregate earnings
Asset pricing
Aggregate returns
Earnings–returns relation

ABSTRACT

The accounting literature has traditionally focused on firm-level studies to examine the capital market implications of earnings and other accounting variables. We first develop the arguments for studying capital market implications at the aggregate level as well. A central issue is that diversification makes equity investors at least partially and potentially almost completely immune to several firm-level properties of earnings by holding diversified portfolios. Diversification is particularly important when assessing the welfare consequences of random errors in accounting measurement (imperfect accruals) and, to the extent it is independent across firms, of deliberate manipulation (earnings management). Consequently, some firm-level metrics of association, timeliness, value relevance, conservatism and other earnings properties do not map easily into investor welfare. Similarly, earnings-related risk manifests itself to equity investors largely through systematic earnings risk (covariation with aggregate earnings and/or other macroeconomic indicators). We conclude that the design and evaluation of financial reporting must adopt at least in part an aggregate perspective. We then summarize the literature in accounting, economics and finance on aggregate earnings and stock prices. Our review highlights the importance of studying earnings at the aggregate level.

☆ We are grateful for helpful comments and suggestions from Bill Cready, Bjorn Jorgensen, Alon Kalay, Suresh Nallareddy, Jennifer Wu Tucker, Ronnie Sadka, and Lakshmanan Shivakumar.
* Corresponding author. Tel.: +1 972 883 5929.
E-mail addresses: ray.ball@chicagobooth.edu (R. Ball), gil.sadka@utdallas.edu (G. Sadka).

http://dx.doi.org/10.1016/j.acclit.2015.01.001

1. Introduction

The accounting literature has a long tradition of studying the relation between earnings and stock prices at the firm level, employing cross-sectional or pooled research designs. Consequently, important earnings properties including usefulness, timeliness, conservatism, value relevance, analyst forecast accuracy, trading volume, information asymmetry, and liquidity typically have been defined and measured from earnings–returns relations at the firm level. In contrast, the economics and finance literatures have an almost as long tradition of studying earnings and prices at the aggregate level, employing time series designs. Conceptually, "aggregate" variables are economy-wide totals, but in practice they are measured using market indexes of prices and earnings.[1] The two literatures have operated almost orthogonally; only recently has the accounting literature turned to aggregate-level work. We compare and contrast these approaches, describe why aggregate-level analysis is relevant to users of financial statement information and to policy makers, and summarize the principal results from aggregate-level research to date.

Studying earnings–returns relations at the firm level is relevant from the perspective of many financial statement users, including managers, boards, security analysts covering the stock, auditors, regulators and litigants. Firm-level results also are relevant for most contracting uses of earnings and earnings-related variables, including debt, compensation, supply, and licensing contracts. A considerable amount has been learned from what has become a very large firm-level literature, even though due to the complexity of economic activity there remains much work to be done.

Nevertheless, for diversified users of financial statement information, firm-level earnings–returns metrics do not map easily into welfare. Perhaps the most obvious example is the earnings–returns R^2 measure of association, due to Lev (1989). This metric does not aggregate from the firm to the portfolio level (Ball, Robin, & Sadka, 2008). In general, the portfolio-level earnings–returns R^2 increases in the number of stocks in the investor's portfolio, the effect being that firm-level metrics substantially under-estimate the importance of earnings to even modestly diversified investors. Other examples of the need to address diversification when assessing investor welfare are the consequences of random errors in accounting measurement (imperfect accruals) and, to the extent it is independent across firms, the consequences of deliberate manipulation (earnings management). We discuss these further below.

These examples illustrate the general point that the design and evaluation of financial reporting involves, at least in part, an aggregate perspective. There is a growing literature that takes this perspective. We summarize this literature below. Because the results obtained from aggregate earnings–returns studies are substantially different than at the firm level, we believe it is important for researchers and policy makers to be familiar with them.

Aggregate returns-earnings relations also are of considerable interest to economists and financial economists. A direct reason is that aggregate after-tax earnings are a component of US Gross Domestic Product that is both large (e.g., over 10% in Q2 2014) and variable (e.g., falling below 5% in Q4 2008).[2] Consequently, the nature of the relation between aggregate earnings and the economy is important to researchers and policy makers alike. For example, if aggregate earnings were strongly pro-cyclical, as was alleged to result from fair valuing banks' balance sheets during the Global Financial Crisis, they would add substantially to undiversifiable economic risk. Investors then would find that returns on even well-diversified share portfolios were correlated with returns on human capital, housing and other assets.

Another reason to study the relation between earnings and the aggregate economy is the issue of allegedly excess volatility in aggregate assets prices, in which price volatility frequently is benchmarked against aggregate earnings (e.g., Shiller, 2000). A dominant theme has been based on the Campbell–Shiller hypothesis that an excessive amount of the variation in asset prices is due to variation in expected returns, and that a correspondingly minor amount of the variation is due to shocks to aggregate earnings. We also summarize this literature.

[1] In this literature, indexes include both averages (equal- and value-weighted) across firms of scaled variables such as stock returns, and sums/totals. Samples typically are comprised of all publicly traded firms or important sub-samples such as the Standard and Poor's 500.

[2] http://research.stlouisfed.org/fred2/graph/?g=cSh

In this review of the aggregate-level literature, we ignore costly contracting based on accounting information and limit our attention to the equity market. This decision is because we are unaware of substantial contracting in which payoffs or other rights are a function of aggregate earnings. One reason for firm-level contracting on earnings is that insuring payoffs by pooling earnings risks across firms would substantially blunt the desired incentive effects of risky payoffs. For example, executives and other employees with earnings-based bonus schemes bear the firm-level risk to their bonus payoffs that arises from any errors in calculating earnings. While firm-level risk in principle would be largely diversifiable by entering a short contract on the bonus payoff with an investor selling insurance on many such contracts, this would defeat the purpose of having bonuses in compensation contracts in the first place, which is to provide earnings-based incentives. This presumably explains why insurance pooling of compensation payoffs is not observed as a common practice. Another reason for contracting on firm-level and not aggregate earnings is due to asymmetric effects of financial reporting. For example, consider an investor in debt issued by two firms. One firm understates earnings (and hence balance sheet ratios) by an amount that triggers lender rights (such as the right to restrict dividends). The other firm overstates earnings (and hence balance sheet ratios) by an equal amount, and no lender rights or payoffs are triggered. While the earnings errors aggregate to zero, the triggered effects do not. For these and perhaps other reasons, aggregate-level relations are inconsequential in most contracting contexts, so we focus on the equity market.

The paper proceeds as follows. Sections 2 and 3 discuss the implications of diversification for the time-series relation between earnings and returns. These sections also highlight why firm-level results may not hold at the aggregate-level and why it is difficult to draw certain capital market inferences using firm-level studies. Section 4 discusses the importance of aggregate earnings in understanding any potential risk in earnings. Section 5 summarizes the literature that examines the relation between aggregate earnings and aggregate returns. The section also describes the broad implications of these studies. Sections 6 and 7 summarize two potentially growing lines of research in accounting regarding the information content of aggregate earnings and cross-sectional earnings dispersion. Section 8 concludes.

2. Time-series versus cross-sectional analyses

Firm-level studies employ either cross-sectional or pooled regressions, whereas aggregate earnings studies employ time-series regressions (there is no cross section, by definition). These different empirical models have different economic interpretations.

A cross-sectional study compares different firms in the same period. In the context of the earnings–returns relation, it tests whether firms with higher earnings (or earnings growth) in period t have higher prices (or returns) in period t, compared to firms with lower earnings (or earnings growth) in the same period. Cross-sectional studies typically estimate regressions of returns on earnings, or estimate so-called "reverse" regressions of earnings on returns, or alternatively conduct analyses of portfolios formed by sorting on either earnings or some exogenous variable. Most firm-level cross-sectional studies, beginning with Ball and Brown (1968), document that a positive relation does hold. On average, firms with higher earnings growth have higher stock returns as well.

A firm-level time-series model examines whether, in periods when a firm's earnings are higher than their average level for that firm over the estimation period, stock returns also are higher than their firm-level average during the period. The evidence in Teets and Wasley (1996), Sadka and Sadka (2009), and Choi, Kalay, and Sadka (2014) shows that there is a positive firm-level association between earnings and returns in time-series as well as in cross-section, even though Teets and Wasley (1996) observe that this need not be the case. As we describe in the following section, this relation does not hold when the analysis is conducted at the aggregate level.

3. Diversification and its implications

Beginning with Ball and Brown (1968), the accounting literature documents robust positive associations between firm-level stock prices and firm-level earnings, and between firm-level returns and firm-level earnings growth and other measures of earnings surprise such as analyst forecast

errors. Generally, accounting earnings are "value relevant" in the sense that earnings expectations as well as earnings surprises affect stock prices. Consequently, this literature concludes that investors are interested in accounting earnings because they affect – or at least are correlated with – the value of their stock holdings.

To the extent that firm-level accounting earnings are involved in the price formation process, they are indirectly relevant to all investors, both active and passive. Price formation effects (as distinct from mere association) can occur through earnings surprises causing stock price revisions for the firm itself or via information spillovers to related firms. Spillover effects can occur between firms in an industry (Foster, 1981), between suppliers and their customers (Olsen & Dietrich, 1985) or due to the information in one firm's earnings about aggregate-level earnings affecting all firms (Brown & Ball, 1967). Price formation effects also can occur through audit-verified earnings realizations confirming and thus enhancing the credibility of other disclosures (Ball, 2001; Ball, Jayaraman, & Shivakumar, 2012; Gigler & Hemmer, 1998).

In addition, active investors may trade on earnings information. This includes investors who believe they are better able to predict future earnings and who thus expect they can profit from trading on their superior expectations. If their belief in their own superior forecasting ability is correct, these investors are able to buy stocks whose future earnings will positively surprise other investors with greater than random frequency, and sell (short) stocks whose future earnings are more likely to disappoint others. Since many studies document a positive association between earnings surprises and stock returns, such a trading strategy should yield positive average abnormal returns. In other words, using better foresight of earnings can allow an active investor to earn abnormal (risk adjusted) returns. If their belief in their own superior forecasting ability is not correct, these investors are unable to earn positive average abnormal returns from trading on earnings expectations. Regardless of whether active investors do possess superior ability to process accounting information, their belief in their own ability to do so makes firm-level earnings and earnings expectations relevant to their investment decisions. The trading strategies of active investors prohibit them from holding well diversified portfolios, and consequently their portfolio returns reflect effects other than aggregate market-wide events. Active investors trading on earnings choose less than fully diversified portfolios based on their expectations of earnings, and thus the values of their portfolios are affected by firm-level earnings realizations.

In contrast, many investors hold well-diversified portfolios that are not chosen based on expectations of firm-level earnings. For these investors, aggregate-level earnings and earnings expectations are directly value relevant when making investment decisions (for example, in deciding on the mix of bonds and stocks in their portfolios). From a welfare economics perspective, any evaluation of financial reporting therefore involves at least in part an aggregate perspective.

A related consideration is the fact that, by definition, the average investor holds the CAPM value-weighted market portfolio. By definition, the portfolios of all investors constitute an investment in the aggregate market portfolio and a set of bets against the aggregate. It follows that the sum across all investors of the bets on an individual security is zero. In other words, given the pricing of the aggregate portfolio, investors who are less than fully diversified are betting against each other. It follows that investors collectively affected by aggregate earnings, reinforcing the rationale for evaluating financial reporting at least in part from an aggregate perspective.

To illustrate the importance of diversification in context of the earnings–returns relation, consider a portfolio of n firms. Assume firm-level earnings surprise for firm i in period t, $\Delta X_{i,t}$, is independently and normally distributed with mean zero, $\Delta X_{i,t} \sim iidN(0, \sigma)$. Then, even if the firm-level relation between earnings growth and stock returns is positive, aggregate earnings growth might not affect prices because expected aggregate earnings growth converges to zero for high n. In the limit, if earnings variation is purely idiosyncratic, aggregate earnings contain zero actual and expected surprise, do not affect asset prices, and therefore contain no information that is value relevant for investors. This demonstrates that the well-known result that firm-level earnings affect relative asset prices does not imply that the average investor or a well-diversified investor finds earnings to be materially value relevant. More broadly, the functional form of value relevance at the portfolio level, including the sign of the earnings–returns relation, need not be the same as at the firm level. Clear evidence (summarized below) suggests that it is not.

3.1. Can we use firm-level metrics to draw conclusions about the aggregate?

For diversified users of financial statement information, firm-level earnings–returns metrics do not map easily into welfare. A notable example is the contemporaneous earnings–returns R^2 measure of association, which does not aggregate from the firm level to the investor's portfolio level (Ball, 2001). In general, the portfolio-level earnings–returns R^2 increases in the number of stocks in the investor's portfolio, the effect being that firm-level metrics substantially under-estimate the value of earnings information to even modestly diversified investors. Consider Lev (1989) call for financial reporting to be redesigned to increase the contemporaneous earnings–returns association, because the observed firm-level metric is in some sense too low. Even if successful in raising the firm-level association R^2, this proposal would be welfare-decreasing for diversified investors if any effect on the already high portfolio-level association was tiny, and did not compensate for any additional financial reporting costs incurred (Ball et al., 2008).

Another example of how firm-level results can be misleading indicators of welfare effects is provided by random errors in accounting accruals, such as errors in inventory counts, receivables allowances, marketable securities valuation, and depreciation rates. Ball and Brown (1969) observe that equity investors can diversify out of errors in accounting estimates to the extent they are uncorrelated across firms. Random errors in calculating earnings would reduce its signal-to-noise ratio at the firm level, thereby making individual-firm accounting earnings less useful, but their effect would be considerably attenuated at the level of a well-diversified portfolio. Any effect on earnings-price metrics would be attenuated on those same metrics when measured at the level of a well-diversified portfolio. Metrics derived from the relation between aggregate earnings and aggregate prices then would be more informative about the direct welfare effects on investors of random accounting errors, because the aggregate is a better proxy than the individual firm for a well-diversified portfolio. In contrast, accruals errors that are correlated across firms (such as those arising from using historical costs under conditions of inflation) are not diversifiable, and in principle their effects can be reflected in aggregate earnings–returns metrics.[3] The implication is that the design and evaluation of financial reporting systems necessarily involves addressing aggregate effects.

A similar example of the need to address diversification when assessing investor welfare is provided by earnings management. To the extent that deliberate manipulation of financial reporting is due to idiosyncratic firm-level events, its effects are independent across firms.[4] Consequently, the welfare effects of accounting manipulation are attenuated for diversified investors.

The discussion above demonstrates that we cannot necessarily infer the aggregate relation between earnings and returns from the results of firm level studies. One reason, noted above, is that some of the firm-level variation in earnings and returns is diversifiable. Another reason is that the relation between the idiosyncratic components of earnings and returns can be independent of the relation between aggregate earnings and aggregate returns, in which case the relation between aggregate earnings and aggregate returns is not directly evident from the firm-level relation. To illustrate this point, assume that firms' earnings have an aggregate component and an orthogonal firm-level component. That is, $\Delta X_{i,t} = \Delta X_t + \varepsilon_{i,t}$, where $\varepsilon_{i,t} \sim iid(0, \sigma)$ is the idiosyncratic component of earnings and ΔX_t is the aggregate/systematic component. Similarly, assume that firm-level stock returns, $R_{i,t}$, have an aggregate component and an orthogonal firm-level component. That is, $R_{i,t} = R_t + \mu_{i,t}$, where $\mu_{i,t} \sim iid(0, \sigma')$ is the idiosyncratic component of stock returns and R_t is the aggregate/systematic component. The relation between ΔX_t and R_t is clearly independent of the relation between $\varepsilon_{i,t}$ and $\mu_{i,t}$. Thus, the firm-level relation may or may not hold for the aggregate portfolio. Moreover, the relation between aggregate earnings and aggregate returns need not have the same sign as the relation between the corresponding idiosyncratic components of earnings and returns.

Indeed, the evidence described below is that contemporaneous annual aggregate earnings and aggregate returns are negatively correlated, in contrast with the robust positive firm-level association.

[3] Investor welfare would be indirectly affected via any relation between accounting errors and firms' investment, operating and financing decisions, or tax effects.

[4] It is possible that aggregate shocks affect the firms' incentives to manipulate earnings. In such cases, these manipulations can have an aggregate impact.

As the literature continues to examine the relation between aggregate accounting variables and stock returns, it documents additional cases where firm-level and aggregate-level relations differ. For example, Sloan (1996) documents (using a cross-sectional analysis) that firm-level accruals are negatively associated with future firm-level stock returns, but Hirshleifer, Hou, and Teoh (2009) show that this relation is reversed in the aggregate. That is, aggregate accruals are positively related to future aggregate stock returns.

A conclusion that emerges is that to determine whether earnings are value relevant for investors we must examine two questions at the aggregate level. First, do earnings and earnings surprises have systematic (aggregate) components? (Ball, Sadka, & Sadka, 2009; Brown & Ball, 1967). As demonstrated above, if earnings are independently distributed, firm-level earnings shocks may be diversified and hence will not be value-relevant for well-diversified investors. Second, if earnings contain systematic/undiversifiable components, how are those components priced? (Kothari, Lewellen, & Warner, 2006; Sadka, 2007; Sadka & Sadka, 2009). In other words, how do investors (and security prices) respond when realizations of aggregate earnings are different than their expectations? These questions cannot be answered using firm-level research designs.

4. Earnings and risk

Stock prices reflect discounted expected cash flows to investors. Thus, stock price behavior can be analyzed in terms of the behavior of two components: expected cash flows, and discount rates (known in the asset pricing literature as expected returns). Accounting earnings are an accruals-adjusted measure of firm' free cash flows, so earnings undoubtedly are related to stock prices because investors expect free cash flows to translate into future cash distributions to them.[5] Consistent with this expectation, the accounting literature documents a robust positive association between earnings and stock prices at the firm level.

While the association between expected earnings and investors' expectations of cash flows is relatively direct and well-understood, the relation between earnings and discount rates is not. A reasonable prior is that accounting earnings reflect firm risk. Revenues are affected by demand shocks and expenses are affected by shocks to factor costs and to productivity. Write-offs due to asset impairment charges and restructuring costs reflect shocks to the value of invested capital. It seems reasonable to expect earnings risk to affect the rate at which investors discount expected cash distributions out of earnings.

This reasoning is not sufficient, however, because when investors are diversified the relation between earnings and discount rates depends on whether earnings are in some sense *systematically* risky. Other things equal, firms with higher systematic risk in earnings, meaning that their earnings are unusually sensitive to the performance of the macro economy (in a mean-variance CAPM world, they have higher "earnings betas"), are expected to earn higher rates of return (i.e., have higher discount rates) and hence trade at a lower price per dollar of expected cash flows. As discussed in the section above, traditional firm-level measures of risk are insufficient to capture these effects because firm-level idiosyncratic volatility can be diversified. Similarly, having higher or lower profitability (at the firm-level and at the aggregate-level) does not necessarily indicate the level of risk to an investor. In other words, the relation between the level of profitability and the firm's discount rate is not clear ex ante and requires additional theoretical and empirical work.

While the relation between the level of profitability and discount rate is not clear, some theories do offer some insights. Under production based asset pricing (e.g., Cochrane, 1991, 1996; Lin & Zhang, 2013; Liu, Whited, & Zhang, 2009), profitability informs about risk because investment is endogenously determined by the discount rate. These models generally assume that return on investment is monotonically declining with investment and that the rate of return on capital (discount rate) is known. Thus, the return on investment declines for each additional dollar invested. Therefore, managers will invest until the firm earns a rate of return on investment equal to the required rate of

[5] In many national and state jurisdictions, firms are only allowed to pay dividends or make stock repurchases from their retained earnings. In the US, §6 of the Model Business Corporation Act (1971) gives firms the right to purchase their own shares to the extent of "unreserved and unrestricted earned surplus."

return on capital. Note that in these models the cost of capital is known based on the risk profile of the firm. So, since investment is endogenously determined the required rate of return on capital will equal the rate of return on investment (profitability), in equilibrium.

Production based asset pricing assumes that investments are endogenously determined. However, if the investment opportunity set and profitability is exogenously determined (for example by exogenous technological and productivity shocks), the relation between earnings and discount rates is less obvious ex ante. Discount rates then will not depend on the level of profitability, but rather on the sensitivity of profits to aggregate macroeconomic shocks. That is, the sensitivity of firm-level profits to aggregate shocks could be priced. For example, firms that are highly positively sensitive to the business cycle, performing well in periods of expansion but poorly during recessions, could be priced to earn investors a higher expected rate of return. If investors' risk preferences vary with the business cycle because they are more risk averse in recessions, firms that are sensitive to business cycles will have higher expected returns (higher discount rates) to compensate for this risk. This higher cost of capital will be required regardless of the average profitability of the firm or the variability of its earnings. In contrast, firms that are less sensitive to the business cycle provide a better hedge against cycles and will earn lower expected rates of return. Thus, if investments are not endogenously determined the relation between the level of profitability and discount rates (if it exists) is unclear.

The analysis points to two different aspects of the cross-sectional pricing of earnings risk. To study the price of firm-level systematic earnings risk in a cross-section, it is necessary to estimate both the aggregate pricing of earnings shocks (how they affect aggregate risk premiums) and firm-level earnings sensitivities to aggregate shocks (how firms' earnings commove with these aggregate shocks). Understanding these two questions would allow us to understand whether earnings risk is priced and, thus, better understand the relation between earnings and stock returns both at the aggregate and the firm levels.

In sum, stock prices have two major components: expectations of future cash flows and discount rates. To fully understand the relation between firm-level earnings and firm-level stock prices, it is necessary to examine aggregate earnings as well, because the firm-level discount rate is largely determined by the systematic risk embedded in the firm's earnings. Therefore, studying aggregate earnings and their association with stock prices is not just important for well diversified portfolios, but also for undiversified investors because the discount rate reflected in the stock price is driven at least in part by the firm's relation to the aggregate economy. As we note below, there has been relatively little research on this topic and hence the relation between earnings and discount rates is a potentially important topic for future empirical and theoretical research.

5. The aggregate earnings–returns literature

While the early accounting literature consists mainly of firm-level analysis, several studies do examine the more systematic components of earnings. For example, Brown and Ball (1967) provide evidence that firm-level earnings are a function of industry and aggregate earnings. Ball and Brown (1969) and Beaver, Kettler, and Scholes (1970) examine the relation between CAPM betas and earnings betas.

Renewed interest in aggregate earnings began with the curious finding that while earnings are positively related to contemporaneous stock returns in firm-level studies, the contemporaneous relation is negative for aggregate earnings growth and aggregate returns (Kothari et al., 2006; Sadka, 2007).[6] In this section we survey the research on contemporaneous, lead and lagged aggregate earnings–returns relations. Before doing so, we outline the important distinction between the earnings news and discount rate news components of returns, which forms the basis for alternative explanations of the observed results.[7]

[6] Gkougkousi (2014) finds similar results for the relation between bond returns and aggregate earnings. Anilowski, Feng, and Skinner (2007) examine the association between aggregate management guidance and aggregate stock returns.

[7] Research invariably focuses on annual observations of aggregate earnings and returns, in contrast with the firm-level literature in which quarterly or half-yearly frequencies commonly are studied.

It is worthwhile noting that studies show the negative relation between aggregate growth in net income and aggregate returns does not hold for the entire sample period (e.g., Sadka & Sadka, 2009; Patatoukas, 2014; Frederickson, Lyon, & Zolotoy, 2014). It appears that the change in the earnings–returns relation during the first decade this century is due largely to accounting shocks. One indication is that the change is less apparent in operating income than in net income. Operating income is substantially less affected by the goodwill write-offs during 2001–02 due to the introduction of impairment rules in SFAS 142. Also, the "great recession" affected aggregate earnings significantly after 2007 due to the extensive use of fair value accounting in financial institutions. In contrast, the pre-2000 sample shows a robust negative relation between historical cost earnings and aggregate stock returns (Jorgensen, Li, & Sadka, 2005). In unreported results, we find a negative association between aggregate earnings and aggregate returns when using operating income after depreciation as the earnings variable and excluding financial firms. We also find a positive association between stock returns and future growth in earnings, using annual data for the period 1950–2013. These findings support the conclusion that there generally is a negative contemporaneous relation between aggregate returns and aggregate historical-cost based earnings.[8]

5.1. A decomposition of the earnings–returns relation

To better understand the relation between aggregate earnings and returns, we employ the Campbell (1991) returns decomposition and characterize the earnings–returns relation in terms of the relation between earnings and three returns components:

$$\begin{aligned} R_t &= E_{t-1}[R_t] + (E_t - E_{t-1})\left[\sum_{j=0}^{\infty}\rho^j \Delta d_{t+j}\right] - (E_t - E_{t-1})\left[\sum_{j=1}^{\infty}\rho^j R_{t+j}\right] \\ &= E_{t-1}[R_t] + N_{cf,t} - N_{r,t} \end{aligned} \tag{1}$$

where R_t denotes returns (in logs) at time t, Δd_t denotes dividend growth (in logs) at time t, ρ is a deflator (the inverse of 1 plus the dividend yield), and $E[\cdot]$ is the expectation operator. Thus, the three components of returns are: expected returns, $E_{t-1}[R_t]$, changes in expected cash flows (cash-flow news, $N_{cf,t}$), and changes in expected returns (return news, $N_{r,t}$).

Firm-level studies of the relation between earnings and returns commonly estimate the term $cov(R_t, \Delta X_t)$, where ΔX_t denotes changes in earnings or earnings surprises, and conclude that $cov(R_{i,t}, \Delta X_{i,t}) > 0$. As noted above, the simple and straightforward interpretation of this positive relation between earnings changes and stock returns focuses on the information about cash flows to investors that is implied by earnings. In the Campbell taxonomy, the interpretation then is that $cov(N_{i,cf,t}, \Delta X_{i,t}) > 0$ and hence $cov(R_{i,t}, \Delta X_{i,t}) > 0$. From a theoretical standpoint, this positive association is largely expected. Assuming a CAPM economy, the cost of capital (expected return) is a function of the market risk premium and the firm's beta (co-variation with the market portfolio). In such an economy, if earnings surprises are unrelated to the firm's beta, they are unrelated to the cost of capital. In terms of Eq. (1), earnings surprises then are unrelated to $E_{t-1}[R_{i,t}]$ and $N_{i,r,t}$. Thus, to the extent that earnings are informative about the firm's performance and its ability to generate free cash flows, and thereby contain cash-flow news $N_{i,cf,t}$, the relation between earnings changes and stock returns at the firm level is expected to be positive.

However, the Campbell return decomposition shows that the relation between earnings changes and stock returns is more complex than a focus on cash flows alone suggests. The complexity manifests itself more strongly at the aggregate level. Hecht and Vuolteenaho (2006) note that the relation between returns and variables such as earnings also involves a relation with the remaining components of returns. Formally:

$$cov(R_t, \Delta X_t) = cov(E_{t-1}[R_t], \Delta X_t) + cov(N_{cf,t}, \Delta X_t) - cov(N_{R,t}, \Delta X_t). \tag{2}$$

The decomposition in Eq. (2) is helpful when interpreting the contemporaneous, lagged and lead relations that are observed between aggregate earnings and returns, to which we now turn.

[8] This conclusion is further supported by the robust negative relation between aggregate returns and aggregate GDP.

5.2. Aggregate earnings and contemporaneous aggregate returns

In contrast with firm-level studies, aggregate (market-level) analyses show that the contemporaneous relation between aggregate earnings changes and aggregate stock returns is negative, $cov(R_t, \Delta X_t) < 0$. To interpret this result, the aggregate earnings–returns literature uses the decomposition in Eq. (2) as the accepted starting point. This equation suggests that the explanation for the negative association between aggregate earnings and aggregate returns lies in the relation between aggregate earnings and discount rates (expected returns).

Indeed, asset-pricing theories suggest several reasons why aggregate earnings will be related to the cost of capital (discount rate). Chen (1991) points out that consumption based asset pricing models imply that the state of the economy will be related to the cost of capital. For example, higher expected earnings, which increases investors' wealths, should lower their risk aversion and the cost of capital. Also, earnings increases can increase the demand for investments and raise asset prices, lowering the cost of capital. Under production based asset pricing, earnings can inform investors about the risk associated with firms' investments. If higher profitability suggests that managers took on riskier projects, the market risk premium can increase as a result. Thus, the relation between earnings and returns is more complex at the aggregate level than typical firm-level studies assume.[9]

The literature presents two alternative hypotheses for the negative contemporaneous relation between earnings changes and stock returns. Several papers such as Cread and Gurun (2010), Kothari et al. (2006), and Patatoukas (2014) hypothesize that the negative relation is due to changes in expected returns, or return news, and hence $cov(N_{R,t}, \Delta X_t) > 0$. This hypothesis suggests that investors react negatively to positive aggregate earnings surprises by increasing the cost of capital. Following Choi et al. (2014), we refer to this hypothesis as the return news hypothesis. Other studies such as Chen (1991), Sadka (2007), and Sadka and Sadka (2009), suggest the negative relation results from a negative relation between expected aggregate earnings and expected aggregate returns; $cov(E_{t-1}[R_t], \Delta X_t) < 0$, because $cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t]) < 0$. This hypothesis implies that investors demand low rates of return when they expect high aggregate future profits and vice versa. We refer to this hypothesis as the expected earnings hypothesis. The difference between the two hypotheses involves the extent to which aggregate earnings changes represent aggregate earnings surprises (as distinct from being predictable or expected).

To better understand the implication of earnings predictability, we decompose earnings changes into expected and unexpected changes as follows:

$$\Delta X_t = E_{t-1}[\Delta X_t] + (E_t - E_{t-1})[\Delta X_t] \equiv E_{t-1}[\Delta X_t] + UE_{X,t}.$$

Next, since expected earnings cannot be correlated with either cash-flow news or return news (or any news for that matter), and since expected returns cannot be correlated with earnings surprises (following the same logic), the decomposition of the earnings returns relation can be rewritten as:

$$\begin{aligned} cov(R_t, \Delta X_t) &= cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t]) + cov(N_{cf,t}, UE_{X,t}) - cov(N_{R,t}, UE_{X,t}) \\ &= cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t]) + cov(N_{cf,r,t}, UE_{X,t}) \end{aligned} \quad (3)$$

If aggregate earnings changes consist entirely of earnings surprises, $cov(E_{t-1}[R_t], \Delta X_t) = 0$, because expected earnings are a constant, i.e., $E_{t-1}[\Delta X_t] = Const$. The earnings–returns relation then is determined by how investors respond to the news/surprise component of earnings, $cov(N_{cf,R,t}, UE_{x,t})$. However, if earnings are predictable to some degree, the earnings–returns relation also depends on the relation between expected earnings and expected returns, $cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t])$. In this case, the empirical relation between returns and earnings changes captures the relation between returns and both of the underlying earnings components (expected and unexpected earnings).

The return news hypothesis focuses on the relation between earnings news and unexpected stock returns, $cov(N_{cf,R,t}, UE_{x,t})$. In other words, the hypothesis addresses how investors change the required

[9] Here we are assuming that one can fully control for firm beta. Otherwise, if earnings are associated with the firm's beta, the firm-level relation between earnings and returns is as complex as the aggregate-level relation.

rate of return (discount rate) in response to an earnings surprise. Formally, this hypothesis proposes that $cov(N_{cf,R,t}, UE_{x,t}) < 0$. The relation between earnings news and cash flow news is likely to be positive, $cov(N_{cf,t}, UE_{x,t}) > 0$, because positive earnings surprises are not likely to represent declining cash flows (an assumption supported by empirical evidence). Therefore, this hypothesis suggests the relation between earnings news and return news must be positive, $cov(N_{R,t}, UE_{x,t}) > 0$, such that $[-cov(N_{R,t}, UE_{x,t})] < 0$. In other words, investors increase the rate of return they demand on their investment when they receive positive earnings news. Consistent with this hypothesis, Patatoukas (2014) finds that aggregate earnings growth is positively associated with the aggregate implied cost of capital.

One possible explanation for this relation is as follows. Suppose aggregate earnings changes are largely unpredictable and provide a valid measure of aggregate earnings surprise. The return news hypothesis implies that investors update their beliefs about the risk of the aggregate stock of capital investment in the economy following this surprise. For example, investors attribute positive unexpected growth in earnings to aggregate investment being riskier than earlier expected. This realization causes investors to increase the rate of return required as compensation for the increased level of risk in the economy. As a result, investors react negatively and prices decline following the unexpected increase in aggregate earnings (earnings news). This results in a negative contemporaneous relation between aggregate earnings changes and aggregate stock returns. It is also important to note that the price decline due to the increase in discount rate must be sufficiently large to offset the price increase due to any potentially positive cash flow news embedded in the earnings surprise.

Another possible explanation is that the positive relation between earnings surprises and returns news is due to a positive relation between earnings changes and inflation, and hence the risk free rate (Gallo, Hann, & Li, 2013; Shivakumar, 2007; Shivakumar & Urcan, 2014). These studies document that controlling for some measures of Federal Reserve policies, the relation between earnings and returns becomes positive.

In contrast to the return news hypothesis, the expected earnings hypothesis focuses on the relation between expected earnings and expected returns, proposing that $cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t]) < 0$. While the return news hypothesis focuses on investors' collective response to aggregate earnings surprises, the expected earnings hypothesis examines how expected future aggregate earnings impact the cost of capital. This hypothesis suggests aggregate earnings are predictable, and thus aggregate earnings changes are not an ideal measure of aggregate earnings surprises (news). The expected earnings hypothesis further suggests investors become less risk averse when they expect (forecast) higher earnings in the future. The intuition is as follows. Suppose aggregate earnings changes are largely predictable. The hypothesis posits that investors lower the rate of return they require to hold risky assets (risk premium) when they expect an improvement in future economic conditions, and increase it when they expect a deterioration. This idea is consistent with consumption-based asset pricing. In these models, investors generally become less risk averse, and demand lower returns for their investments, when they expect to be wealthier. If aggregate earnings changes indeed are predictable, expected returns should be lower during periods of high aggregate earnings growth. Therefore, the contemporaneous relation between aggregate earnings changes and aggregate returns is negative.

Consistent with the expected earnings hypothesis, Sadka and Sadka (2009) and Ball et al. (2009) provide evidence that aggregate earnings are highly predictable. Furthermore, Chen (1991), Sadka (2007), Choi et al. (2014) and Bonsall et al. (2013) provide evidence that expected earnings are negatively related to stock returns. These findings are consistent with the hypothesis that investors demand lower rates of return when they expect better economic conditions, consistent with consumption-based asset pricing. In addition, the studies also document that measures of aggregate earnings surprises such as forecast errors and forecast revisions are positively related to aggregate returns.[10] These studies suggest that investors react positively to positive aggregate earnings surprises, which is consistent with extant firm-level evidence and the expected earnings hypothesis.

[10] He and Hu (2014) use international evidence to show that in countries with more transparent disclosures the earnings–returns relation becomes less positive (more negative).

5.3. Aggregate returns predict aggregate earnings

An additional result highlighted in prior studies (e.g., Ball et al., 2009; Kalay, Nallareddy, & Sadka, 2014; Sadka & Sadka, 2009) is that aggregate returns are positively related to future (at least one year ahead) earnings growth. Again this finding has two possible interpretations. One explanation for the positive association is straightforward and implies that aggregate earnings growth can be forecast well in advance (earlier than the firm-level literature would imply) and investors trade their expectations of aggregate earnings into prices in advance of the actual earnings outcomes. In terms of Eq. (2) above, this means that in year $t-1$ investors revise their expectations of earnings in year t ($cov(N_{cf,t-1},\ \Delta X_t)>0$). This interpretation is consistent with rational expectations and market efficiency. The second explanation for the positive lagged association but negative contemporaneous association is that investors increase (decrease) the required rate of return when they expect lower (higher) future earnings growth. In terms of Eq. (3) above, this explanation is that the relation between expected earnings and expected returns is negative, $cov(E_{t-1}[R_t], E_{t-1}[\Delta X_t])<0$. For example, if in year $t-1$ investors revise upward their expectations of earnings growth for year t, under this explanation expected returns (discount rates) for year t decline, $N_{R,t}<0$, causing year $t-1$ prices and returns to rise, and year t expected returns to fall. These explanations are consistent with the expected returns hypothesis.

The positive association between stock returns and future earnings growth is important because the long-term relation between earnings growth and stock returns cannot be negative, given the data. Consider the case where earning growth is negatively related to contemporaneous stock returns and is uncorrected with past or future stock returns. In this case, since in our time period earnings have largely grown and the association with stock returns are negative, this contemporaneous relation implies that the earnings-to-price ratio will increase over time. This is because when earnings grow, prices must grow at a slower rate or even decline, and the earnings-to-price ratio will increase over time. However, no long term increase is observed as the earnings-to-price ratio, which seems relatively stationary. Therefore, the relation between earnings growth and past, present, and future stock returns cannot be negative. This implies that the earnings growth must be positively related to either past or future stock returns.

The expected earnings and earnings news hypotheses both imply that such as positive relation exists. The expected earnings hypothesis directly implies that stock returns are positively associated with future earnings growth. Consistent with this hypotheses, Sadka and Sadka (2009) and Ball et al. (2009) document that the positive relation between earnings growth and past stock returns is larger in magnitude than the negative relation between contemporaneous earnings and returns. In contrast, the earnings news hypothesis implies that earnings growth is positively related to future stock returns. This positive association between earnings growth and future discount rates must be sufficiently large to offset any positive cash flow news in earnings, such that the overall contemporaneous relation between earnings growth and stock returns is negative. In sum, both hypotheses imply that the overall relation between earnings growth and long-run stock returns is positive.

5.4. Aggregate returns are predictable

The behavior of aggregate earnings also provides insights into the finance literature debate on whether aggregate returns are predictable. Early studies such as Fama and French (1988) and Campbell and Shiller (1988a, 1988b) document that the aggregate dividend yield predicts aggregate returns, leading to the initial conclusion that stock returns and equity risk premiums are predictable. However, Goyal and Welch (2003, 2007) document that these prediction models do not perform well in out of sample tests, suggesting that it is not clear whether returns are predictable. Goyal and Welch (2007) conclude: "Therefore, although it is possible to search for, to occasionally stumble upon, and then to defend some seemingly statistically significant models, we interpret our results to suggest that a healthy skepticism is appropriate when it comes to predicting the equity premium, at least as of early 2006. The models do not seem robust." In a later study, Cochrane (2008) makes the distinction between return predictability and the ability to estimate reliably a point estimate of the coefficient on

a specific variable used in prediction. The latter is a difficult task. It is indeed difficult to identify and employ an empirical model that reliably estimates expected future returns.

Nevertheless, the negative relation observed between contemporaneous aggregate earnings and returns does imply that aggregate returns are predictable. To understand why, consider the case where returns are unpredictable. In this case, expected returns are a constant, $E_{t-1}[R_t] = k$, and thus investors do not alter their expected returns, $N_{R,t} = 0$. It is easy to see from Eq. (2) that in such a case the relation between earnings growth and stock returns is determined only by the cash flow information embedded in earnings growth. This is because variation in returns then is only a function of changes in expected cash flows.

Continuing with this line of reasoning, the negative relation between earnings and returns implies that higher earnings cause investors to reduce prices because they expect lower future cash flows. Such a reaction by investors is consistent with investor irrationality, such that investors misinterpret the information in earnings growth or have inefficient expectations. However, prior studies document that stock returns are good predictors of earnings both at the firm level and the aggregate level. In addition, the dividend yield is a good predictor of future GDP and earnings growth (Chen, 1991; Sadka, 2007). Therefore, little evidence supports investor irrationality with respect to aggregate earnings. A logically feasible alternative explanation for the negative information in earnings growth is that higher current earnings imply lower future earnings to such an extent that the overall value of earnings indeed is reduced. For example, an earnings increase by $100 in the current period could imply a $200 decline in the next period. However, studies do not find evidence of a strong negative serial covariance, so higher earnings surprises represent "good" cash flow news, inconsistent with this explanation.[11]

Since earnings surprises represent positive cash flow news, it follows that the negative relation between earnings growth and stock returns must be driven by the other components of the variation in stock returns, namely expected returns and changes in expected returns. Thus, the only plausible explanation for the negative relation between earnings growth and stock returns implies that aggregate returns are predictable and are related to aggregate earnings.

5.5. Aggregate earnings are priced

The literature on the aggregate earnings–returns relation has implications for the pricing of earnings risk as well. The discussion above illustrates why the negative contemporaneous relation between aggregate earnings and returns implies that returns are time varying and predictable. The analysis also implies that stock returns vary with aggregate earnings. Thus, it suggests that variation in aggregate earnings is priced because it affects the risk premium.

The pricing of aggregate earnings is a critical component in the pricing of earnings risk in the cross-section of expected stock returns. Earnings risk is priced in the cross-section only if two conditions hold. First, aggregate earnings must be related to the discount rate. Second, different firms must have different sensitivities (covariations) to these aggregate earnings shocks.

Consider two firms differing in the sensitivity of their earnings to aggregate earnings shocks. One firm's earnings is highly sensitive to aggregate earnings shocks, so it has a high earnings beta. The second firm's earnings are largely independent of aggregate earnings in the economy, so it has a low earnings beta. If aggregate earnings variation does not relate to the discount rate, then all else equal, these firms will have the same cost of capital. This is because the relation between these different covariances/sensitivities (earnings betas) and discount rates depend on whether aggregate earnings shocks are priced themselves. Thus, the evidence that aggregate earnings are priced satisfies one of the necessary conditions for the pricing of systematic earnings risk in the cross-section of expected stock returns.

5.6. Earnings risk and the cross-section of expected stock returns

As noted, few papers study the implications of systematic earnings or cash flow risk for stock prices. The available evidence suggests that earnings risk is indeed priced. Some studies document that the

[11] Note that accounting earnings by definition reflect the firm's ability to distribute cash flows. Specifically, earnings represent past current and future cash inflows that the firm has "earned" during the period. Thus, the firm is then free to either invest or distribute these earnings.

book-to-market return spread is associated with systematic earnings risk (Campbell & Vuolteenaho, 2004; Fama & French, 1995). Ball et al. (2009) find that systematic earnings risk can help explain several known anomalies, including the value/growth premium and earnings momentum. Using analyst forecast revisions, Da and Warachka (2009) find evidence that cash-flow risk is priced in the cross-section of expected stock returns. Khan (2008) decomposes market returns to cash-flow risk and returns risk following Campbell and Vuolteenaho (2004). The study finds that such a decomposition explains a portion of the accruals anomaly (Sloan, 1996). In sum, while the empirical evidence to date is scant, these studies suggest that studying systematic earnings risk is a promising avenue for further research.

We note some caveats to using earnings to identify systematic risk factors. First, the distribution of earnings is not well behaved, with relatively frequent large negative shocks. One reason for the left skew in earnings is that negative shocks include large non-recurring items (e.g., restructuring charges, asset write-offs, loss provisions) that are measured differently and are more transitory than the recurring components of earnings. Left skew makes it more difficult to identify earnings covariation. Moreover, the skewed distribution means that one needs to aggregate into fewer portfolios to get a well behaved time-series of portfolio-level earnings that is affected less by large negative shocks. Second, accounting earnings is a low-frequency (quarterly or annual) variable, which lowers the number of observations/periods. Note that the increase in observations obtained by employing quarterly data comes with a cost because fourth quarter earnings are distributed differently that the first three quarters.

While the caveats listed above can be significant drawbacks to using earnings rather than returns to identify risk factors, there are several advantages. One advantage is that it is easier to relate aggregate earnings shocks to contemporaneous macroeconomic conditions. In contrast, returns reflect shocks to expected future quantities, and hence are less efficient in identifying macroeconomic factors. For example, Ball et al. (2009) document a high association between the systematic components of earnings and real GDP growth. Using earnings can also help identify firm characteristics that will result in higher/lower sensitivity to aggregate shocks. The finance literature been struggling to identify the macroeconomic factors that are associated with known empirical factors such as the Fama French HML, so we believe it is worth exploring further whether earnings risk can provide better economic insights.

6. Earnings are a leading indicator of economic activity

Accounting variables are primarily historical, in the sense that accounting measurement methods generally reflect past transactions and events. In accounting parlance, earnings are "recognized" only when the earnings process has been completed, which means that accounting earnings generally are measured at the point when products or services are delivered to customers. Because information about expected transactions becomes available to investors before the point of accounting measurement, earnings growth is to a degree predictable, and expectations of future earnings growth are reflected in current stock prices. This is the case at both the firm level and, even more so, at the aggregate level. Thus, accounting earnings are generally viewed as being predictable, but not as predictors of future investment or future growth.

Nevertheless, accounting variables can be leading indicators for some aspects of economic activity because they reflect real events in firms and in their product and factor markets. An obvious mechanism for this is that earnings reflect past return on investment, which can be informative of the profitability of new investments (both for the firm itself, its competitors, its suppliers, and firms generally), depending on the degree of persistence over time in the rate of return on assets. For example, unexpected positive earnings growth can spur additional investment in the economy. Further, from a financing perspective, positive earnings growth increases the supply of equity capital available to invest and thus, given firms' aggregate demand for new investment, reduces capital costs.[12]

[12] This holds regardless of whether the firm retains the additional profits for internal reinvestment, or distributes the earnings to shareholders.

Several recent studies find consistent evidence that earnings are indeed a leading indicator, particularly for investment. Kothari, Lewellen, and Warner (2014) document that earnings growth predicts future investments for up to five quarters after the earnings are disclosed. Arif (2012) finds that corporate investment is higher when analysts are more optimistic in their earnings forecasts. These studies further document that aggregate investment is negatively related to future performance (see also Arif & Lee, 2014). Shivakumar and Urcan (2014) conclude that aggregate earnings growth is correlated with aggregate investment by firms in subsequent quarters, as well as with future shocks to producer prices. They find no equivalent correlation of aggregate earnings growth with either future consumption or consumer prices. In sum, real investment is correlated with both current and expected profitability, highlighting the role of earnings as an information variable in investment decisions.

Accounting earnings also are informative for monetary policy. Aggregate earnings are associated with future inflation as well as money supply and other Federal Reserve actions (Crawley, 2014; Gallo et al., 2013; Patatoukas, 2014; Shivakumar, 2007). The Shivakumar and Urcan (2014) results described above suggest the relation between earnings growth and future inflation is due largely to shocks to the producer price index (PPI). Crawley (2014) extends prior studies and documents that incorporating the implications of accounting conservatism on earnings significantly increases the explanatory power of earnings with respect to monetary policy. Finally, Konchitchki (2011) documents the cross-sectional implications of inflation that arise from the use of nominal accounting.

Konchitchki and Patatoukas (2014a) and Konchitchki and Patatoukas (2014b) document that employing standard tools of financial statement analysis can help predict GDP growth. The first study documents that an aggregate measure of earnings scaled by sales (an aggregate profit margin) helps predict future GDP growth. The second study employs standard measures, such as return on net operating assets (RNOA), asset turnover and profit margins, to predict future GDP growth. Shivakumar and Urcan (2014) further document that inefficient forecasts of the producer price index (PPI) result in inefficient forecasts of GDP growth. Investors and policy makers use GDP forecasts, yet these studies document that macroeconomic forecasts do not fully reflect the information in earnings.[13] Nallareddy and Ogneva (2014) find that earnings and earnings dispersion also predict errors in GDP estimates from the Bureau of Economic Analysis (BEA).

Overall, this stream of literature highlights the usefulness of accounting variables in predicting macroeconomic activity and macroeconomic indicators. Using publicly available accounting information can improve forecasts and expectations of macroeconomic shocks, and thus can inform policy makers and help in generating more timely and more efficient policy changes.

7. The sectoral shift theory and earnings dispersion

The sectoral shift hypothesis (Lilien, 1982; Lucas & Prescott, 1974) proposes that unemployment is driven in part by cross-sectional shifts across sectors. The hypothesis links cross-sectional variation in firm-level and industry-level performance to aggregate performance. Consider for example an economy with two sectors. One of the sectors is in decline while the other is growing. In such an economy, employees will migrate from the poor performing sector to the growing sector, in search of better job opportunities. Due to frictions in the labor market, such as search costs, retraining costs, and relocation costs including geographical preferences, migration is incomplete and unemployment results. Consistent with this hypothesis, Lilien (1982) finds that cross-sectional dispersion in sectoral performance explains much (if not most) of the variation in aggregate unemployment. Thus, this literature establishes a causal link between performance in the cross-section and the macroeconomy.

One determinant of cross-sectional dispersion in sectoral performance is earnings dispersion. From an asset-pricing standpoint, excluding frictions, cross-sectional dispersion in earnings should not affect aggregate stock prices because dispersion per se does not affect aggregate earnings. As we discussed above, firm-level variation is somewhat diversifiable. Yet, consistent with the sectoral shift hypothesis, several recent studies document that cross-sectional earnings dispersion (in both earnings changes and forecasted earnings) is significantly associated with aggregate stock returns (e.g., Johnson

[13] Several studies (Darrough & Russell, 2002; Hann, Ogneva, & Sapriza, 2014) suggest that macroeconomic forecasters are more efficient in forecasting aggregate shocks compared with firm-level analysts (using an aggregated firm-level forecasts).

& Lee, 2014; Jorgensen, Li, & Sadka, 2012; Ma, 2011; Maio, 2013).[14] Kalay et al. (2014) document a positive relation between cross-sectional earnings dispersion and industrial production, with the relation being more pronounced during periods of low aggregate earnings performance. This literature also documents that unemployment is explained more by dispersion in earnings than by dispersion in returns. As noted above, earnings primarily reflect contemporaneous shocks to production and demand, whereas returns primarily reflect shocks to expected future quantities, the effect being that earnings provide a better proxy for short term shocks to the labor market. This literature highlights the usefulness of accounting variables as measures of contemporaneous economic performance.

Consistent with the evidence on aggregate earnings, cross-sectional analyst-forecast dispersion provides a leading indicator for certain macroeconomic indicators, namely unemployment and industrial production. Similar to aggregate earnings, macroeconomic forecasts fail to fully incorporate the information content in earnings forecast dispersion. This inefficiency leads to predictable forecasting errors, which in turn can yield inefficient investment and ineffective policies. One role for accounting research could be to better inform macroeconomic forecasters about the information embedded in accounting variables, potentially resulting in more accurate and efficient forecasts.

8. Efficiency, rationality and excess asset price volatility

We noted above that aggregate earnings predict aggregate returns. This result parallels the firm-level evidence of earnings predicting returns, known as "post earnings announcement drift," that has been evident since Ball and Brown (1968). As is the case in the firm-level literature, interpreting the aggregate-level result is problematic because it is consistent with several interpretations, including both rational pricing (aggregate earnings are correlated with equilibrium expected returns) and irrational pricing (market under-reaction to earnings growth).

This dilemma is evident in the important but controversial issue of excess asset price volatility, in which aggregate earnings play a central role. The issue is whether aggregate asset prices exhibit substantial deviations from fundamental values and hence vary too much, and what if anything economic policy makers should do to prevent or correct this. The role of aggregate earnings in this literature is to provide a credible benchmark for assessing fundamental values.

In popular usage, and particularly since the recent Global Financial Crisis, this issue typically is discussed in terms of so-called price "bubbles." The notion of a bubble is loosely described as an episode in which aggregate asset prices increase suddenly and substantially, and then suddenly collapse. Episodes that are said to correspond to this notion include the 1637 Dutch Tulip Mania, the South Sea Company Bubble and the Mississippi Company Bubble of 1720, the Railway Mania of the 1840s, the 1926 Florida Land Bubble, the stock market collapse of 1929 that preceded the Great Depression, the 1997–2000 NASDAQ dot-com boom and bust, and the 2009 stock market collapse in the Global Financial Crisis.

Researching such episodes presents several challenges to economists and policy makers. First, price increases followed by decreases can be naturally occurring phenomena. For example, the random walk model of asset price behavior over time (Samuelson, 1965) implies that any sequence of price increases is as likely to be followed by reversals as by further increases. How does one identify an event in which prices are in some sense too high from one that is a natural consequence of variation in supply and demand for assets over time? Second, "bubbles" typically are observed only ex post. How does one identify asset over-pricing as it occurs, as distinct from in hindsight?

To address these problems, economists typically define bubbles in a different fashion than in popular usage, where a bubble is an increase followed by a decrease. They model them as episodes in which aggregate asset prices substantially exceed "fundamental value" and thus are susceptible to collapse. This is where aggregate earnings play a role. First, aggregate earnings can be used as a proxy for the unobservable (and even undefined) construct of fundamental value. "Bubbles" then are identified by prices increasing out of line with earnings. Second, earnings are observable at the time, not simply in hindsight. Thus, if prices are out of line with current earnings, in theory one might be able to identify bubbles as they develop.

[14] Loungani, Rush, and Tave (1990) find that dispersion in stock prices is associated with higher unemployment. Maio examines dispersion in returns rather than earnings.

However, this merely kicks the problem further down the road. How does one know that prices are out of line with earnings? The approach researchers have taken to this problem is very simple: prices are out of line with earnings when the aggregate price/earnings ratio is "high" by historical norms. But this approach has its own problems. High price/earnings is equivalent to low earnings yield, so how does one distinguish a "bubble" from a period in which asset yields in general are below historical norms?

In an attempt to persuade readers that asset bubbles do in fact occur, researchers have sought to demonstrate that prices are more volatile than can be explained by variation in earnings. The Nobel Prize winner Robert Shiller published the first edition of his extremely influential book Irrational Exuberance in 2000. The title was motivated by the brief but notorious use of that term by the then Fed Chairman Alan Greenspan in a long and otherwise unheralded 1996 speech. The first edition opened with a graph (Shiller, 2000, Fig. 1.1) of real (inflation-adjusted) prices and earnings over 1871–2000, more than a century, and described the figure as showing that the "dramatic increase in prices since 1982 is not matched in real earnings growth. Looking at the figure, no such spike in earnings growth occurs in recent years. Earnings in fact seem to be oscillating around a slow, steady growth path that has persisted for over a century" (Shiller, 2000, p. 6). The second edition updated the data to 2005. While it added the qualifying observation "The boom and crash in the years surrounding the peak in 2000 is clearly related to the behavior of earnings," it continued to conclude: "But historically the earnings movements were generally less dramatic than the stock price movement. Earnings in fact seem to be oscillating around a slow, steady growth path that has persisted for over a century" (Shiller, 2005, p. 4).

Eyeballing these graphs leaves the definite impression that earnings have been on a steady growth path without the dramatic movements that occur in price. But the earning series should look much less volatile than the price series, because the average price observation is approximately 16 times the average earnings observation. The appearance of comparatively low earnings volatility has been exaggerated by plotting earnings against a scale that is only one-quarter of the scale selected for price (0–450 versus 0– 1800). Modifying the scale can make the earnings series to appear just as volatile as price. The rhetorical question "which graph/scale is correct?" indicates how difficult it is to specify an earnings-based test of whether aggregate asset prices exhibit substantial deviations from fundamental values.

A more scientific approach to the issue was initiated by Shiller (1981). Subsequent work is based on the Campbell and Shiller (1988a)) and Campbell (1991) Eq. (1) decomposition of aggregate returns into two components, shocks to expected future cash flows (or earnings) and shocks to expected future discount rates. Aggregate stock returns are expected to be increasing in shocks to expected future earnings and decreasing in shocks to expected future discount rates. The early literature concluded that most price volatility is due to discount rate shocks, and little is due to aggregate expected cash flows. Because cash flow shocks are assumed to measure changes in fundamental values, their seemingly unimportant role in explaining aggregate price volatility has been interpreted as indicating that prices depart substantially from fundamental values and hence are excessively volatile: that is, that asset price "bubbles" occur. This interpretation remains controversial.

Controversy resolves around two issues. First, people have different priors about whether securities generally are efficiently priced or mispriced due to limits to arbitrage together with behavioral factors such as overconfidence and herding (Barberis & Thaler, 2003). Second, different research designs lead to radically different conclusions. This stems from the fact that both components of stock returns (shocks to expected future earnings and shocks to expected future discount rates) are unobservable, so researchers seeking to estimate their relative importance are required to adopt some strategy to separately identify them. Imposing different strategies reverses the conclusions.

The strategy of Campbell and Shiller (1988a, 1988b) and Campbell (1991) is to model discount rate shocks and then back out cash flow shocks as the residual amount of returns that is not explained by the discount rate modeling. Chen and Zhao (2009) point out that the discount rate modeling is weak, and the residual method dumps all the model error into the estimate of cash flow shocks. To demonstrate the point, they simply reverse the strategy, first modeling cash flow shocks and backing out discount rate shocks as the residual. This reverses the conclusion. Now, discount rate shocks are seemingly unimportant contributors to aggregate price volatility, which appears to be largely due to changes in fundamental value in the form of cash flow shocks, and thus prices do not seem excessively volatile. Just as two versions of the Shiller (2005) graph of the same data make prices look substantially more volatile than earnings and alternatively about equal in volatility, two different strategies for

estimating the unobservable components of returns make prices appear excessively volatile and not excessively volatile. These contradictory conclusions – from the same data – demonstrate how unresolved is the issue of excess price volatility.

One feature of this literature is that non-accountant researchers typically use dividends as the cash flow variable. A priori, one would view earnings as the primitive variable, and dividends as a derivative. Legally, dividends are distributions of earnings. Economically, Miller and Modigliani (1961) famously argue that, given earnings, dividends are irrelevant for security pricing. A large proportion of firms do not pay dividends (DeAngelo, DeAngelo, & Skinner, 2004; Fama & French, 2001). When dividends are paid they are a lagged and smoothed function of earnings (Fama & Babiak, 1968; Lintner, 1956), and frequently do not change annually, the effect being that they are less informative and exhibit lower volatility than earnings. The choice of dividends as the proxy for cash flows loads the dice in favor of concluding that aggregate discount rate shocks dominate aggregate price changes.

The basic intuition behind the irrationality argument is as follows. If price changes are unrelated to the cash flows (dividends), then changes in prices are inefficient and/or irrational. Campbell and Shiller (1988a, 1988b)) find that price movement are not associated with changes in expected dividends. Specifically, fluctuations in price multiples, such as the dividend yield do not predict future changes in dividends. However, as we note above dividends are less informative than other measures of cash flows. Consistently, Sadka (2007) shows that earnings rather than dividends explain much of the volatility in the dividend yield. Fama (1990) and Schwert (1990) use industrial production as a measure of aggregate earnings and find that fluctuations in current and future industrial production (as well as future returns) explain a large portion of the variation in stock returns.

Another limitation of this modeling arises from the correlation between cash flow shocks and discount rate shocks, which makes it impossible to completely parse price movements into one or the other. Further, a simplified and unrealistic lag structure commonly is implied for earnings. That prices lead earnings has been documented at the firm level since Ball and Brown (1968) but, as we noted above, the relation at the aggregate level is more complex, and aggregate earnings are highly predictable.

The substantial effect of research design choice on the results obtained, together with differences in priors about market efficiency, suggests the controversy about the existence and identifiability of aggregate asset price "bubbles" seems unlikely to be resolved in the foreseeable future.

9. Conclusions

Our survey highlights the contribution of aggregate studies to understanding the role of earnings and other accounting variables in the capital markets. The accounting literature traditionally has focused on firm-level studies. We develop the arguments for studying capital market implications at the aggregate level as well. A central difference between these approaches results from the fact that investors can diversify out of some firm-level effects of earnings by holding diversified portfolios. Thus firm-level earnings–returns metrics may not be relevant for well diversified investors and may be misleading indicators of the welfare effects of accounting information.

The Campbell (1991) return decomposition shows that the relation between earnings changes and stock returns is more complex than typical firm-level studies assume. While this complexity manifests itself more strongly at the aggregate level, it is present at the firm level as well, even if not so obviously. The finding in the aggregate literature that highlights this point most effectively is the negative association between contemporaneous aggregate earnings and returns, which requires a more complex explanation for the earnings–returns relation than typically is invoked in the firm-level literature.

The effect of risk on the pricing of firm-level earnings provides an additional reason for paying attention to aggregate issues. In theory, any risk in earnings should manifest itself through systematic earnings risk (covariation with aggregate earnings and/or other macroeconomic indicators). Therefore, to understand the pricing of firm-level earnings, it is necessary understand how earnings covary with aggregate earnings shocks and/or other macroeconomic shocks. Consequently, firm-level earnings variation is priced to the extent that it covaries with aggregate shocks that are priced by investors.

Our survey also points to some research topics that could benefit from further exploration. In that context, we note that accounting earnings, viewed as an economic variable, possesses several research advantages. Chief among those advantages is that earnings shocks are better measures of current

period shocks to production and demand, in contrast with stock returns which incorporate shocks to expected future quantities.

In sum, we view the aggregate earnings literature as a comparatively recent but growing body of work with a promising future.

References

Anilowski, C., Feng, M., & Skinner, D. J. (2007). Does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. *Journal of Accounting and Economics, 44*, 36–63.

Arif, S. (2012]). Aggregate investment and its consequences. *Working paper*.

Arif, Salman, & Lee, C. M. C. (2014]). Aggregate investment and investor sentiment. *Review of Financial Studies* in press.

Ball, R. (2001]). Infrastructure requirements for an economically efficient system of public financial reporting and disclosure. *Brookings-Wharton Papers on Financial Services,* (1), 127–169.

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

Ball, R., & Brown, P. (1969]). Portfolio theory and accounting. *Journal of Accounting Research, 7*, 300–323.

Ball, R., Jayaraman, S., & Shivakumar, L. (2012). Audited financial reporting and voluntary disclosure as complements: A test of the confirmation hypothesis. *Journal of Accounting and Economics, 53*(1-2), 136–166.

Ball, R., Robin, A., & Sadka, G. (2008). Is financial reporting shaped by equity markets or by debt markets? An international study of timeliness and conservatism. *Review of Accounting Studies, 13*, 168–205.

Ball, R., Sadka, G., & Sadka, R. (2009). Aggregate earnings and asset prices. *Journal of Accounting Research, 47*, 1097–1133.

Barberis, N., & Thaler, R. (2003]). A survey of behavioral finance. In G. Constantinides, M. Harris, & R. Stulz (Eds.), *Handbook of the economics of finance, Volume 1, Chapter 18* (pp. 1053–1128). Amsterdam: Elsevier.

Beaver, W. H., Kettler, P. C., & Scholes, M. S. (1970). The association between market-determined and accounting-determined risk measures. *The Accounting Review, 45*, 654–682.

Bonsall, S., Bozanic, Zahn, & Fischer, Paul E.. (2013). What do management earnings forecasts convey about the macroeconomy? *Journal of Accounting Research, 51*, 225–266.

Brown, P., & Ball, R. (1967]). Some preliminary findings on the association between the earnings of a firm, its industry, and the economy. *Journal of Accounting Research, 5*, 55–77.

Campbell, J. Y., & Shiller, R. J. (1988a]). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies, 1*, 195–227.

Campbell, J. Y., & Shiller, R. J. (1988b]). Stock prices, earnings, and expected dividends. *Journal of Finance, 43*, 661–676.

Campbell, J. Y. (1991]). A variance decomposition for stock returns. *Economic Journal, 101*, 157–179.

Campbell, J. Y., & Vuolteenaho, T. (2004]). Bad beta, good beta. *American Economic Review, 94*, 1249–1275.

Chen, Long, & Zhao, Shelly. (2009]). Return decomposition. *Review of Financial Studies, 22*, 5213–5249.

Chen, Nai-Fu (1991]). Financial investment opportunities and the macroeconomy. *Journal of Finance, 46*, 529–554.

Choi, Jung Ho, Kalay, A., & Sadka, G. (2014). *Earnings news, expected earnings and aggregate stock returns*. Columbia University Working paper.

Cochrane, J. H. (1991]). Production based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance, 46*, 209–237.

Cochrane, J. H. (1996]). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy, 104*, 572–621.

Cochrane, J. H. (2008]). "The dog that did not bark A" defense of return predictability. *Review of Financial Studies, 21*, 1533–1575.

Cread, W. M., & Gurun, U. G. (2010]). Aggregate market reaction to earnings announcements. *Journal of Accounting Research, 48*(2), 289–334.

Crawley, M. (2014]). *Macroeconomic consequences of accounting: The effects of accounting conservatism on macroecnomic indicators and the money supply*. Working paper.

Da, Zhi, & Warachka, M. C. (2009]). Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. *Journal of Financial Economics, 94*, 448–468.

Darrough, M. N., & Russell, T. (2002]). A positive model of earnings forecasts: Top down versus bottom up. *The Journal of Business,* 127–152.

DeAngelo, H., DeAngelo, L., & Skinner, D. J. (2004). Are dividends disappearing? Dividend concentration and the consolidation of earnings. *Journal of Financial Economics, 72*, 425–456.

Fama, E. F. (1990]). Stock returns, expected returns, and real activity. *Journal of Finance, 45*, 1089–1108.

Fama, E. F., & Babiak, H. (1968]). Dividend policy: An empirical analysis. *Journal of the American Statistical Association, 63*, 1132–1161.

Fama, E. F., & French, K. R. (1988]). Dividend yields and expected stock returns. *Journal of Financial Economics, 22*, 3–25.

Fama, E. F., & French, K. R. (1995]). Size and book-to-market factors in earnings and returns. *Journal of Finance, 50*, 131–155.

Fama, E. F., & French, K. R. (2001]). Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics, 60*, 3–43.

Frederickson, J. R., Lyon, J. D., & Zolotoy, L. (2014). *Aggregate earnings news and stock market returns: The good, the ugly, and the state dependent*. Working paper University of Melbourne.

Foster, G. (1981]). Intra-industry information transfers associated with earnings releases. *Journal of Accounting and Economics, 3*, 201–232.

Gigler, F., & Hemmer, T. (1998]). On the frequency, quality, and informational role of mandatory financial reports. *Journal of Accounting Research, 36*, 117–147.

Gallo, L. A., Hann, R. N., & Li, Congcong. (2013). *Aggregate earnings surprises, monetary policy, and stock returns*. Working paper University of Maryland.

Gkougkousi, X. (2014]). Aggregate earnings and corporate bond market. *Journal of Accounting Research, 52*, 75–106.

Goyal, A., & Welch, I. (2003]). Predicting the equity premium with dividend ratios. *Management Science, 49*, 639–654.

Goyal, A., & Welch, I. (2007]). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies, 21*, 1455–1508.
Hann, R. N., Ogneva, M., & Sapriza, H. (2014). *Forecasting the macroeconomy: Analysts versus economists*. Working Paper.
He, Wen, & Hu, Maggie Rong. (2014]). Aggregate earnings and market returns: International evidence. *Journal of Financial and Quantitative Analysis, 49*(4), 879–901.
Hecht, P., & Vuolteenaho, T. (2006]). Explaining returns with cash-flow proxies. *Review of Financial Studies, 19*, 159–194.
Hirshleifer, D., Hou, Kewei, & Teoh, Siew Hong. (2009). Accruals, cash flows, and aggregate stock market returns. *Journal of Financial Economics, 91*, 389–406.
Johnson, T., & Lee, J. (2014]). On the systematic volatility of unpriced earnings. *Journal of Financial Economics, 114*(1), 84–104.
Jorgensen, B., Li, J., & Sadka, G. (2005). *Are accounting standards diversifiable?*. Unpublished manuscript.
Jorgensen, B., Li, J., & Sadka, G. (2012). Earnings dispersion and aggregate stock returns. *Journal of Accounting and Economics, 53*(1–2), 1–20.
Kalay, A., Nallareddy, S., & Sadka, G. (2014). *Conditional earnings dispersion, the macroeconomy and aggregate stock returns*. Working Paper.
Khan, M. (2008]). Are accruals mispriced? Evidence from tests of an intertemporal capital asset pricing model. *Journal of Accounting and Economics, 45*, 55–77.
Konchitchki, Y. (2011]). Inflation and nominal financial reporting: Implications for performance and stock prices. *The Accounting Review, 86*(3), 1045–1085.
Konchitchki, Y., & Patatoukas, P. N. (2014a]). Accounting earnings and gross domestic product. *Journal of Accounting and Economic, 57*(1), 76–88.
Konchitchki, Y., & Patatoukas, P. N. (2014b]). Taking the pulse of the real economy using financial statement analysis: Implications for macro forecasting and stock valuation. *The Accounting Review, 89*(2), 669–694.
Kothari, S. P., Lewellen, J. W., & Warner, J. B. (2006). Stock returns, aggregate earnings surprises, and behavioral finance. *Journal of Financial Economics, 79*, 537–568.
Kothari, S. P., Lewellen, J. W., & Warner, J. B. (2014). *The behavior of aggregate corporate investment*. Working Paper.
Lev, B. (1989]). On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research. *Journal of Accounting Research, 27*(supplement), 153–192.
Lilien, D. M. (1982]). Sectoral shifts and cyclical unemployment. *The Journal of Political Economy, 90*, 777–793.
Lin, Xiaoji, & Zhang, Lu. (2013]). The investment manifesto. *Journal of Monetary Economics, 60*, 351–366.
Lintner, J. (1956]). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *American Economic Review, 46*, 97–113.
Liu, L. X., Whited, T. M., & Zhang, Lu. (2009). Investment-based expected stock returns. *Journal of Political Economy, 117*, 1105–1139.
Loungani, P., Rush, M., & Tave, W. (1990). Stock market dispersion and unemployment. *Journal of Monetary Economics, 25*, 367–388.
Lucas, R. E., & Prescott, E. C. (1974]). Equilibrium search and unemployment. *Journal of Economic Theory, 7*(2), 188–209.
Ma, G. (2011]). *Analyst forecast dispersion and aggregate stock returns*. University of Texas at Dallas Working Paper.
Maio, P. F. (2013]). *Return Dispersion and the Predictability of Stock Returns*. Hanken School of Economics Working Paper.
Miller, M. H., & Modigliani, F. (1961]). Dividend policy, growth, and the valuation of shares. *Journal of Business, 34*, 235–264.
Nallareddy, S., & Ogneva, M. (2014]). *Predicting restatements in macroeconomic indicators using accounting information*. Working paper.
Olsen, C., & Dietrich, J. R. (1985]). Vertical information transfers: The association between retailers' sales announcements and suppliers' security returns. *Journal of Accounting Research, 23*(Supplement), 144–169.
Patatoukas, P. N. (2014]). Detecting news in aggregate accounting earnings: Implications for stock market valuation. *Review of Accounting Studies, 19*, 134–160.
Sadka, G. (2007]). Understanding stock price volatility: The role of earnings. *Journal of Accounting Research, 45*(1), 199–228.
Sadka, G., & Sadka, R. (2009]). Predictability and the earnings–returns relation. *Journal of Financial Economics, 94*, 87–106.
Samuelson, P. A. (1965]). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review, 6*, 41–49.
Schwert, W. G. (1990]). Stock returns and real activity: A century of evidence. *Journal of Finance, 45*, 1237–1257.
Shiller, R. J. (1981]). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review, 71*, 421–436.
Shiller, R. J. (2000]). *Irrational exuberance* (1st Ed.). Princeton University Press: Princeton, NJ.
Shiller, R. J. (2005]). *Irrational exuberance* (2nd Ed.). Princeton, NJ: Princeton University Press.
Shivakumar, L. (2007]). Aggregate earnings, stock market returns and macroeconomic activity: A discussion of does earnings guidance affect market returns? The nature and information content of aggregate earnings guidance. *Journal of Accounting and Economics, 44*, 64–73.
Shivakumar, L., & Urcan, O. (2014]). *Why do aggregate earnings shocks predict future inflation shocks?* Working paper.
Sloan, R. G. (1996]). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review, 71*, 289–315.
Teets, W. R., & Wasley, C. E. (1996]). Estimating earnings response coefficients: Pooled versus firm-specific models. *Journal of Accounting and Economics, 21*, 279–295.