Are they all one?
An empirical examination of the construct validity of the firm-level conservatism proxies.

Current version: 3 February 2012

Key Words: conservatism, construct validity.
JEL classification: M41

Abstract

Conservatism enjoys a prominent role in the literature and has emerged as one of the most intensely researched topics over the last decade (Guay and Verrecchia 2007). This literature relies heavily on conservatism proxies to make inferences regarding the association between conservatism at the firm level and variables of interest. Whilst the use of multiple conservatism proxies is pervasive in the literature, there remains considerable conjecture regarding their ability to accurately gauge conservatism at the firm level. This is a pressing issue as drawing inferences from studies of accounting conservatism critically relies on the validity of conservatism proxies. Given the existence of multiple firm-specific proxies for a single construct and the debate regarding the validity of the multiple measures, an empirical examination of the construct validity of these measures is prudent and timely. This paper seeks to fill the gap.

Using a sample of U.S. listed firms for the period 1997 to 2007 this study empirically examines the construct validity of four firm-specific conservatism proxies that have emerged at the forefront of the literature: (i) Givoly and Hayn's (2000) negative accruals measure, (ii) Penman and Zhang's (2002) hidden-reserves measure, (iii) Beaver and Ryan's (2000) book-to-market measure and (iv) Khan and Watts' (2009) firm-year asymmetric timeliness measure, along four dimensions of validity: i) concurrent validity, ii) discriminant validity, iii) convergent validity and iv) predictive validity.

Importantly, what has emerged in the findings is that the alternative proxies exhibit different characteristics and show little consistency between them, thus suggesting that they do not capture the same construct. They are not all one. Hence, the alternative proxies cannot simply be considered substitutes. Further, the practice of using multiple proxies to enhance the validity of the results of a research paper is also questionable given the inconsistency between proxies.

1. Introduction

Conservatism enjoys a prominent role in the literature and has emerged as one of the most intensely researched topics over the last decade (Guay and Verrecchia 2007). This literature relies heavily on conservatism proxies to make inferences regarding the association between conservatism at the firm-level and variables of interest. Whilst the use of multiple conservatism proxies is pervasive in the literature, there remains considerable conjecture regarding their ability to accurately gauge conservatism at the firm-level. This is a pressing issue as drawing inferences from studies of accounting conservatism critically relies on the validity of conservatism proxies. Given the existence of multiple firm-specific proxies for a single construct and the debate regarding the validity of the multiple measures, an empirical examination of the construct validity of these measures is prudent and timely. This paper seeks to fill the gap. Hence, this purpose of this paper is to empirically examine the construct validity of four firm-specific accounting conservatism proxies using a sample of U.S. listed firms for the period 1997 to 2007. We ask the question: Are they all measuring the same thing?

Construct validity is examined using the validation criteria outlined by Wang *et al.* (2009). Specifically, this paper examines whether the proxies demonstrate the following dimensions of validity: i) concurrent validity, ii) discriminant validity, iii) convergent validity and iv) predictive validity. The four firm-specific conservatism proxies which have emerged at the forefront of the literature being: i) Givoly and Hayn's (2000) negative accruals measure (NA), ii) Penman and Zhang's (2002) hidden-reserves measure (HR), iii) Beaver and Ryan's (2000) book-to-market measure (BM) and iv) Khan and Watts' (2009) firm-year asymmetric timeliness measure (KW).

This paper extends Wang *et al.* (2009) by conducting a comprehensive empirical examination of the validity of four firm-specific accounting conservatism proxies along the four dimensions indicated. To ensure robust comparison can be made, the proxies are examined using the same sample of firms over the same period. With the exception of convergent validity, Wang *et al.* (2009) do not provide empirical evidence of the sub-validities. Rather, they provide a synthesis of the extant literature and from this, draw inferences regarding the validity of the constructs. Thus the proxies validity remain empirical untested.

The contribution of this paper is methodological. The literature provides little guidance on the appropriate choice of empirical proxy for conservatism but leaves the researcher to argue the suitability of their choice based on their perspective on accounting conservatism and the merits of the proxy relative to their experimental setting (Artiach and Clarkson 2011). Further, the researcher may chose to validate their result through the use of multiple proxies. This paper contributes to the extant literature by identifying if the proxies are substitutes and therefore can be used as alternative proxies.

The results are surprising. What has emerged in the findings is that the alternative proxies exhibit different characteristics and show little consistency between them, thus suggesting that they do not capture the same construct. Hence, the alternative proxies cannot simply be considered substitutes. Further, the practice of using multiple proxies to enhance the validity of the results of a research paper is also questionable given the inconsistency between proxies.

The remainder of the paper is structured as follows. Section 2 describes the setting and conservatism definition. Section 3 describes the construct validity methodology. Section 4 describes the conservatism proxies. Section 5 provides details of the sample selection process and sample characteristics including descriptive statistics. The results are presented in section 6. Section 7 concludes the paper.

2. Conservatism definition

Despite having a prominent role in financial reporting, the construct of conservatism does not have a well articulated and commonly agreed upon interpretation, which puts it in stark contrast to many other accounting constructs (Wang *et al.* 2009). In fact, there is no universally agreed upon definition of conservatism (Wang *et al.* 2009, Watts 2003a). In this regard, Guay and Verrechia (2007) note:

> "Because notions of conservatism have evolved somewhat informally from observations about various accounting policies and empirical regularities in financial data, offering a theory of conservatism is fraught with controversy. In particular, the accounting community has come to associate different empirical regularities with different attributes of conservatism. This has evolved to the point that often it is difficult to find consensus in the literature as to what conservatism means, and what role it serves within an efficient reporting system. "

Conservatism has traditionally been interpreted as the accountant's tendency to "*anticipate no profit, but anticipate all losses*" (Bliss 1924). Accordingly, the decision rule in practice has traditionally been to choose the lowest value amongst the alternative values for assets and the highest alternative for liabilities, and to recognise revenues later rather than sooner and expenses sooner than later (Watts and Zimmerman 1986). The only authoritative or "official" definition of conservatism is that which is contained in the accounting standards. Conservatism is defined in the initial IASB Framework for the Preparation and Presentation of Financial Statements (The Framework) published in 1989 as: *"the degree of caution in the exercise of the judgements needed in making the estimates required under conditions of uncertainty, such that assets or incomes are not overstated and liabilities or expenses are not understated" (IASB 1989, para. 37).*[1]

This definition of conservatism is consistent with FASB's Statement of Accounting Concept Two (1980), paragraphs 92 and 93, which states that there is a role for conservatism (meaning prudence) in financial reporting due to the inherent uncertainty of business activities:

> "There is a place for a convention such as conservatism – meaning prudence – in financial accounting and reporting, because business and economic activities are surrounded by uncertainty, but it needs to be applied with care. Since a preference "that possible errors in measurement be in the direction of understatement rather than overstatement of net income and net assets" introduces a bias into financial reporting, conservatism tends to conflict with significant qualitative characteristics, such as representational faithfulness, neutrality, and comparability…Conservatism in financial reporting should no longer connote deliberate, consistent understatement of net assets and profits. The Board emphasises that point because conservatism has long been identified with the idea that deliberate understatement is virtue."

Consequently, the definition of conservatism previously enunciated by the FASB and IASB has been replaced by that of prudence in The Framework: *"prudence is the inclusion of a degree of caution in the exercise of the judgements needed in making the estimates required under conditions of uncertainty, such that assets or incomes are not overstated and liabilities or expenses are not understated" (IASB 2009, para. 37).*

From these statements it is evident that regulators deemed conservatism to be an appropriate exercise of caution in financial reporting in instances of uncertainty. It is arguable that the definition of "prudence" currently adopted by regulators is consistent with the definition of conservatism contained in the initial IASB Framework. First, the 'undesirable' reporting behaviour described by regulators as

---

[1] Although a write-down in assets or overstatement of liabilities could be consistent with the concept of earnings management, earnings management cannot explain important parts of the conservatism literature, and it is not a plausible general explanation of the evidence of conservatism over long periods of time (Watts 2003b).

conservatism, is not consistent with the definition of conservatism offered by the initial IASB framework. Conservatism is not a deliberate misstatement of financial information; it is a principle of caution when estimates are required in uncertain circumstances. As such, there doesn't appear to be any distinction between the concept of conservatism and prudence as described by regulators. Second, the definition of prudence embodies the same principle as the initial IASB definition of conservatism. It highlights the need to exercise caution when making decisions under conditions of uncertainty, such that financial performance and position are not overstated. It appears that the definition of conservatism has changed in label only. Therefore, this paper adopts the IASB/FASB definition of conservatism as the benchmark on which to assess the various proxies.

However regulators have chosen to move away from conservative reporting to achieve neutrality of financial information. They argue that conservatism (or prudence) is likely to lead to a bias in reported financial position and performance, and is therefore not consistent with the qualitative characteristic of neutrality. A strong signal of this view was provided in May 2008 when the FASB/IASB issued an exposure draft under the IASB/FASB Joint Conceptual Framework entitled the 'Conceptual Framework for Financial Reporting: The Objective of Financial Reporting and Qualitative Characteristics and Constraints of Decision-Useful Financial Reporting Information'. This document signalled that regulators no longer viewed conservatism (or prudence) as a desirable qualitative characteristic of financial reporting:

> "…Being careful in the presence of uncertainty includes searching for additional information to reduce uncertainty, reflecting the uncertainty of a range of potential amounts in making an estimate, or selecting an amount from the midpoint of a range if a point estimate is required. Going beyond that to reflect conservative estimates of income and equity sometimes has been considered desirable to ensure that financial reports do not reflect excessive optimism or bias on the part of management. However, the Boards concluded that describing prudence or conservatism as a qualitative characteristic or a desirable response to uncertainty could conflict with the quality of neutrality because, even with the proscriptions of deliberate misstatement that appear in the existing frameworks, an admonition to be prudent is likely to lead to a bias in the reported financial position and financial performance. … Accordingly, the proposed framework does not include prudence or conservatism as desirable qualities of financial reporting information." (BC2.21).

Absent a universally accepted definition and the progressive change in the definition and preference enunciated by standards setters, it is perhaps not surprising that a number of different proxies for conservatism have developed within the literature. The following section provides an outline of the construct validity methodology and the tests that will be adopted to examine each conservatism proxy.

3. Construct validity methodology

Construct validity describes the strength of the link between a construct and its operational proxies. In essence, construct validity is a question of whether operational proxies are measuring what we want them to measure (Kwok and Sharp 1998). Construct validity methodology breaks construct validity into seven sub-validities (validation criteria): (i) convergent validity, (ii) concurrent validity, (ii) discriminant validity, (iv) predictive validity, (v) internal consistency (reliability), (vi) content validity and (vii) statistical

conclusion validity. Each of the sub-validities emphasises a particular aspect of construct validity.[2] This paper examines the conservatism proxies using four sub-validities: (a) concurrent validity, (b) discriminant validity, (c) convergent validity and (d) predictive validity. The remaining three sub-validities: content validity, internal consistency and statistical conclusion validity are not examined by this paper for reasons explained below.

A measure has *content validity* if it contains all content pertinent to its theoretical construct, usually by inspection of experts in the field (Wang *et al.* 2009). Although, content validity is explored to a limited extent by the sections of this paper which examine the various definitions and constructs of conservatism, a more detailed conceptual discussion and examination is an extensive endeavour in itself. Given the significant and ongoing uncertainty regarding the precise nature and definition of conservatism and the plethora of opinions on conservatism amongst experts (accountants, regulators and researchers), it seems a robust investigation on content validity would provide an avenue for future research.

Internal consistency (reliability) means that a measure is relatively free from measurement error, and therefore gives consistent results in repeated measures (Kwok and Sharp 1998). Although it is important for the proxies to give reliable results, internal consistency is only an important consideration if the major sub-validities indicate that the measure is valid for the following reason: if a proxy is internally consistent but lacks a sufficiently strong link with its theoretical construct, it is merely reliably wrong. Therefore, the focus of this paper is the sub-validities which provide evidence regarding the link between the proxies and the construct of conservatism. [3]

Statistical conclusion validity is wide ranging, including any statistical issues relating to the measurement of a construct (Wang *et al.* 2009). Cook and Campbell (1979) use the term statistical conclusion validity to refer to inferences about whether or not it is reasonable to presume covariation for a given a level of alpha and obtained variances. In essence, statistical conclusion validity means that the statistical conclusions derived from the proxies are valid. The examination of this validity is a also a detailed examination in itself. Further, it has received the most attention in the extant literature. As such, it is not examined in this paper.

A series of tests which examine the four sub-validities are applied on each of the firm-specific conservatism proxies. When assessing construct validity it is generally required that at least a majority (if not all) of the sub-validities are satisfied. However, the relevance of the various sub-validities may change depending on the construct and measures under investigation. Construct validity is a matter of

---

[2] The validation criteria have been developed by construct validity methodologists from many fields, including behavioural accounting research (Wang *et al.* 2009, Kwok and Sharp 1998).

[3] See Givoly, Hayn and Natarajan (2007) for some discussion of the internal consistency of Basu (1997) and BTM measures.

degree; therefore the extent to which a measure's validity is acceptable is ultimately a matter of judgement (Wang *et al.* 2009). The sub-validities examined and the associated tests are outlined below.

3.1 Concurrent validity

A measure has concurrent validity if, under predetermined conditions, it gives a score which it is theoretically predicted to give under those conditions. For example, we would expect the Altman Z-Score to indicate financial distress in firms which have filed for bankruptcy. Wang *et al.* (2009) provide a synthesis on the empirical evidence which examines concurrent validity though they do not empirical test concurrent validity. Further, the empirical evidence regarding concurrent validity relates to the aggregate measure of conservatism applied in the Basu (1997) methodology rather than the firm-specific proxies. Thus, to date, there is no empirical evidence regarding the concurrent validity of the conservatism proxies assessed in this paper.

To assess concurrent validity it is necessary to identify a given condition where conservatism does/does not exist. Such an examination requires a definition of conservatism and the identification of the aspects of conservatism which the respective measure purport to capture (Givoly *et al.* 2007). To confront this issue, given the general lack of agreement in the extant literature, we uses the definition of conservatism contained in The Framework as a foundation for its analysis: *Conservatism is "the degree of caution in the exercise of the judgements needed in making the estimates required under conditions of uncertainty, such that assets or incomes are not overstated and liabilities or expenses are not understated" (IASB 1989, para. 37).* According to this definition, management exhibit conservatism when they make cautious judgements in a response to conditions of uncertainty.

Guay and Verrecchia (2007) provide some support for a link between conservative accounting and uncertainty. They argue that the prevalence of conservative reporting can be explained by the discount which the market applies in response to uncertainty (absence of information). In essence, managers have a vested interest to disclose 'good' news whilst withholding 'bad' news. A conservative reporting system ensures that 'bad' news is disclosed to the market in a timely fashion. In combination, this results in full disclosure.

Therefore, one would expect firms operating under risky conditions to show a greater propensity for conservative accounting in order to reduce uncertainty. As such, the proxies should gauge higher levels of conservatism in firms which experience conditions of uncertainty. The degree of stock price volatility, measured as the standard deviation of daily stock returns is used as the proxy for degree of uncertainty.

In each year of the sample period, firms are broken into quintiles based on stock price volatility. Firms comprising the top (bottom) quintiles are classified as having high (low) levels of uncertainty respectively. A *t*-test for the difference in means is conducted for each proxy in each year to determine if the level of conservatism indicated by the proxies for the high uncertainty quintile is greater than that of

the low uncertainty quintile. A significant positive mean difference will indicate that the proxy has concurrent validity.

3.2 Discriminant validity

Discriminant validity means that a proxy is uncorrelated with another unrelated construct. For example, in management and psychology research, moral intelligence is considered a separate and distinct construct from self-esteem (Austin and Martin 2010). The tests of discriminant validity examine whether the measures are indeed measuring conservatism, or whether they are spuriously capturing another construct unrelated to conservatism. Discriminant validity is neither examined due to a lack of relevant evidence in the extant literature nor tests in the Wang *et al.* (2009) review. Thus, there is no empirical evidence regarding the discriminant validity of the conservatism proxies.

To assess this validity it is necessary to identify a construct that has no *direct* or *indirect* relationship with conservatism, either empirically or theoretically. In this regard, the construct of operational efficiency is selected. There are no theoretical or empirical grounds on which to expect a relationship between conservatism and efficiency, for which the sales-to-employee ratio is utilized as a proxy. Further, this ratio is not related to any of the firm characteristics identified in the extant literature as being associated with conservatism, including size, leverage, profitability and auditor conservatism. There is no theoretical reason for, nor empirical evidence of, a link between efficiency and auditor conservatism or leverage. This ratio is not associated with size because both the numerator and denominator are expected to increase with firm size. Further, there is no reason to believe that efficiency is systematically related to profitability. Therefore, this test examines whether the measures of conservatism are uncorrelated with the sales-to-employee ratio.

Discriminant validity is assessed using a test of correlation. For this test, the Pearson's and Spearman's Rank correlations between each of the conservatism proxies and the efficiency (sales-to-employee ratio) are calculated for each year of the sample. A low correlation between the level of efficiency and conservatism indicates that the measure has discriminant validity. If the correlation between efficiency and the conservatism proxies are found to be significant, be they positive *or* negative, it would suggest that the proxies are noisy and imprecise, and therefore lack discriminant validity.

3.3 Convergent validity

Convergent validity requires that measures of the same construct give results which are highly correlated with one another, or in other words – they converge. To assess convergent validity the Pearson's and Spearman's Rank correlation between each of the conservatism score proxies is calculated for each year of the sample.[4] A high correlation amongst the proxies is an indication of convergent validity.

---

[4] The Pearson's correlation coefficient relies on the assumption that the data is normally distributed, is not robust to outliers and is sensitive to the underlying distribution of the data. Spearman's correlation coefficient is non-parametric and

Convergent validity is the only sub-validity empirically examined by Wang *et al.* (2009) however, their tests include the Basu (1997) and Ball and Shivakumar (2005) aggregates measures. Because they are aggregate measures, it is not possible to calculate their correlations with the firm-specific measures. In line with Watts (2007) and Lara *et al.* (2009) they estimate cross-sectional regressions which are augmented by the firm-specific proxies, and use the estimated coefficients to determine the direction of the correlation. The regression coefficients and correlations are not comparable; as such they do not report the values, only the direction of the correlations. Further, they do not include the KW measure in their analysis, and they use the raw BTM ratio rather than the bias component of the BTM ratio. Their findings suggest that the aggregate measures of asymmetric timeliness are a distinct group which is negatively related to the firm-specific measures suggesting convergent validity issues with the proxies. They also find that the correlation amongst the firm-specific measures to be low, with none exceeding 0.2. Their results are consistent with the findings of Ahmed *et al.* (2000) who report a Pearson correlation coefficient between the BTM ratio and NA of 0.124, significant at the five percent level.[5]

3.4 Predictive validity

Predictive validity refers to the ability of a measure to predict what it should theoretically be able to predict. For example, the Altman Z-Score as a measure of the likelihood of default should be a predictor of actual bankruptcies. Predictive validity is similar to concurrent validity; however, it concerns the measures' ability to predict a *future* condition, whereas concurrent validity relates to a *current* condition. The aim of the tests of predictive validity is to determine whether the proxies detect conservatism in firms which exhibit characteristics which have been empirically shown to be associated with conservatism. Further, if the proxies are indeed capturing the same construct, it is expected that each would be associated with these firm characteristics in a consistent manner.

There is no systematic examination of the predictive validity of each of the empirical proxies in the extant literature and this validity is not examined in the Wang *et al.* (2009) review due to a lack of relevant evidence. Further to this point, there is no theoretical literature in which to appeal to in the design of these tests. As such, the tests of predictive validity rely on known associations documented in the empirical literature. This poses a joint-test problem. The tests conducted can only provide evidence of predictive validity to the extent that the firm characteristics examined are indeed associated with conservatism. Results which are contrary to predictions could be attributed to the choice of firm characteristics as opposed to the efficacy of the conservatism proxies. Therefore, the tests conducted in this paper can only provide circumstantial evidence of predictive validity. Importantly however, this test

---

therefore does not rely on distributional assumptions. Given that there are distributional issues in the sample data, both coefficients are computed.

[5] The Spearman correlation was not statistically significant.

will provide an insight into the relationship between the proxies by determining whether they are *consistently* associated with the firm characteristics examined.

The extant literature has identified a number of characteristics associated with conservatism; these include size, leverage, profitability, and auditor conservatism. Watts (1977) argues that losses arising from the overstatement of assets and income carry a higher cost compared to forgone gains because they are more observable and usable in the political process. As such, firms which face higher political costs have an incentive to adopt conservative accounting policies. Large firms are subject to greater public and regulatory scrutiny and therefore face higher political costs. Accordingly, size is a commonly used proxy for political costs. Evidence consistent with these arguments is provided by Watts and Zimmerman (1986) and Zmijewski and Hagerman (1981), whose findings suggest that large firms exhibit more conservative accounting. Ahmed *et al.* (2002) also find that conservatism is increasing with firm size. Therefore, the proxies should assign higher levels of conservatism to larger firms.

Watts (2003a) espouses contracting as an explanation for conservatism. According to this explanation, conservatism exists as an efficient contracting mechanism. According to Watts (2003a) conservatism mitigates moral hazard which arises from contracting parties having asymmetric information and payoffs, limited horizons and limited liability. Debt holders are one party which contracts with the firm. As debt holders are only concerned with the lower ends of the earnings and net asset distributions, they demand conservative financial information which shows that the minimum amount of net assets will cover their loan (Watts 2003a). This suggests that as leverage increases, accounting will become more conservative. Consistent with this explanation of conservatism, higher leverage is found to be associated with higher levels of conservatism. Ahmed *et al.* (2002) find that leverage induces conservatism by increasing the debt-contracting constraints of the firm. Pae (2007) finds that more highly levered firms are more conservative primarily through unexpected accruals. The findings of Watts and Zimmerman (1986) and Basu (1995) also suggest that conservatism increases with leverage. Therefore, the proxies should consistently indicate higher conservatism in more highly levered firms.

Conservatism is expected to be positively related to profitability for two reasons. First, more profitable firms are better placed to adopt conservative accounting policies. Conservative accounting policies would be relatively costly for low-profitability firms to implement. Second, evidence suggests that conservatism is associated with higher profitability. Ahmed and Duellman (2010) conjecture that the corporate governance mechanism provided by conservatism provides managers with less incentive to pursue negative NPV projects and increases the *ex post* monitoring of investments. As such, firms which exhibit conservative accounting experience higher profitability in the future. Therefore, more profitable firms should have higher levels of conservatism. However, the nature of the relationship between profitability and conservatism is not unambiguous. For instance, Ahmed *et al.* (2002) note that it is

difficult to predict whether the effects of non-operating accruals (in the case of NA) will dominate the expected positive relationship between conservatism and ROA. Similar logic follows for HR; higher values of hidden reserves imply higher R&D and advertising expenditure which in turn reduces net income.

The extant literature provides evidence which suggests that the firm's level of accounting conservatism is associated with the level of conservatism displayed by their auditor. Studies of auditor conservatism typically classify first-tier auditors as conservative. The findings of Watts and Zimmerman (1986), Basu (1995), Basu *et al*. (2001) and Krishnan (2005) all suggest that firms which engage a conservative auditor exhibit a higher degree of accounting conservatism.

To test predictive validity annual cross-sectional OLS regressions are estimated as indicated by equation (1):

$$CON_{it} = \alpha_0 + \alpha_1 SIZE_{it} + \alpha_2 LEV_{it} + \alpha_3 ROA_{it} + \alpha_4 AUD_{it} + \varepsilon_{it} \quad (1)$$

where CON_{it} is the conservatism proxy derived for firm *i* for fiscal year-end *t*. $SIZE_{it}$ is the proxy for political costs measured as the natural log of the market value of common equity for firm *i* for fiscal year-end *t*. LEV_{it} is the proxy for contracting constraints measured as the ratio of long-term debt to total assets for firm *i* for fiscal year-end *t*. ROA_{it} is the proxy for firm profitability measured as the return on total assets for firm *i* for fiscal year-end *t*. AUD_{it} is the proxy for auditor conservatism, measured as a dichotomous variable which equals 1 if the auditor of firm *i* in year *t* is a first-tier auditor, 0 otherwise.

Because time-series data exhibits autocorrelation Fama and Macbeth (1973) *t*-statistics are computed in order to assess the significance of the coefficients over the entire sample period. Under this methodology parameter estimates are obtained from annual regressions and the associated *t*-statistics are computed using the average coefficient for the sample period as indicated as $t = \overline{\alpha} / (s(\overline{\alpha}) / \sqrt{n})$ where *t* is the Fama-Macbeth (1973) *t*-statistic, $\overline{\alpha}$ is the average annual coefficient estimate, s($\overline{\alpha}$) is the standard deviation of annual coefficient estimates and *n* is the number of annual coefficient estimates. Significant coefficients with the predicted sign will indicate that a proxy has predictive validity. Additionally, the regression given by equation (1) is also estimated using the pooled sample 1997 to 2007. To overcome the issue of serially correlated error terms, these regressions utilise firm and year fixed effects.

3.5 Summary of construct validity methodology

A summary of the sub-validities examined and the associated empirical tests are provided below:

Summary of Sub-Validity Examinations

Sec	Validity	Description	Empirical Test
3.1	Concurrent	Under predetermined conditions, the proxies indicate a level of conservatism which they are theoretically predicted to give.	*t*-test for the difference in mean conservatism score for firms with 'high' and 'low' levels of uncertainty.
3.2	Discriminant	The proxies are uncorrelated with an unrelated construct.	Correlation between the proxies and the sales-to-employee ratio, a measure of efficiency.
3.3	Convergent	The proxies give results which are highly correlated with one another.	Correlation between the conservatism score indicated by each of the proxies.
3.4	Predictive	The proxies predict what they should theoretically be able to predict.	Annual cross-sectional regressions utilising firm characteristics associated with conservatism with Fama-Macbeth (1973) *t*-statistics and a pooled regression.

The nature of the tests for concurrent and discriminant validity make it possible to make a judgement as to whether the proxies demonstrate those aspects of validity. However, the results of the tests for predictive and convergent validity are not unambiguous and are interpreted with caution. Notwithstanding these tests do not provide clear cut evidence of their respective sub-validities, they do provide further insight into the results obtained.

At this stage it is important to note, unlike other constructs in the accounting and finance literature, conservatism does not have definitive benchmark with which to test the validity of the proxies. For example, in the literature examining the cost of equity capital, proxies can be judged with reference to known risk factors which have theoretical and empirical support.[6] In the absence of such a benchmark, it is not possible for the empirical tests to definitively identify which of the four proxies possess the greatest degree of construct validity. This is not, and cannot be a horse race. All that can be done is to provide evidence regarding how well each of the proxies demonstrates the four dimensions of validity, individually and collectively and to examine how consistent the measures are. The following section provides details of the four conservatism proxies.

4. Conservatism proxies

The four most prominent firm-specific proxies that have emerged within the literature are the Givoly and Hayn's (2000) negative accruals measure, Penman and Zhang's (2002) hidden reserves measure, Beaver and Ryan's (2000) bias component of the book-to-market ratio and Khan and Watts (2009) firm-specific adoption of Basu's asymmetric timeliness.

The scope of this paper is confined to the examination of the construct validity of the *firm-specific* measures of conservatism for several reasons. First, the aggregate and firm-specific measures were intended to answer different research questions. Aggregate measures, like Basu's (1997) asymmetric

[6] An example of such a study is Botosan and Plumlee (2005) who evaluate five proxies for the firm-specific cost of equity capital.

timeliness of earnings and Ball and Shivakumar's (2005) asymmetric timeliness of accruals to cash-flows, provide a proxy of conservatism for a group of firms (e.g. industry or country) at one point in time. Their purpose is to gauge whether a certain grouping of firms exhibits more or less conservatism compared to another, or whether conservatism has changed across time and are therefore not able to capture cross-sectional variation or the timing of changes in conservatism. Firm-specific measures serve this purpose; they produce scores unique to each firm and time period. Thus, firm-specific measures can characterise both cross-sectional and time-series variation in conservatism, allowing comparison between firms and across time periods (Khan and Watts 2009).

Second, the causes of conservatism are both time- and firm- specific; therefore, it is important that the conservatism literature possess strong firm-year-level measures of conservatism in order to examine time-series and cross-sectional variation in conservatism (Khan and Watts 2009). Further to this point, whilst it is feasible to aggregate firm-specific measures, it is difficult to create an accurate firm-specific proxy from an aggregate measure. As such, it can be argued that firm-specific measures are more versatile as they can be used to answer questions in aggregate and at the firm-level. Therefore, the validity of the firm-specific measures is considered a more pressing issue for researchers moving forward.

Third, aggregate and firm-specific measures have fundamentally different characteristics. Due to the nature of the aggregate measures, it is not possible to test all of the aspects of construct validity outlined in Wang *et al.* (2009). In addition, the same tests cannot be used to collectively examine aggregate and firm-specific measures. One cannot conclude that combining the results for the firm-specific measures and comparing these to the results of the aggregate measures will result in a valid evaluation. Therefore, it cannot be concluded that tests of construct validity designed to examine aggregate measures can be directly and validly compared to those for the firm-specific proxies.

Finally, most of the evidence provided by the extant literature which is relevant to the dimensions of construct validity pertains to Basu's (1997) measure.[7] Khan and Watts (2009) is a firm-level application of the Basu (1997) measure of the asymmetric timeliness of earnings.

The following section provides a brief discussion of the measurement for the four firm-specific proxies used in this analysis. For an extensive review of each of these proxies, see Artiach and Clarkson (2011). Further, the extant literature is supported by a number of review articles including Watts (2003a, b), Beaver and Ryan (2005), Ryan (2006), Givoly *et al.* (2007) and Wang *et al.* (2009) which provide an extensive summary of the literature.

4.1 Conservatism - negative-accruals measure (*NA)*

Givoly and Hayn (2000) characterise existing interpretations of conservatism as imprecise. They argue that because the sum of reported earnings over the life of the firm are the same regardless of

[7] For example: Dietrich *et al.* (2007), Patatoukas and Thomas (2009), and Givoly *et al.* (2007).

accounting choice, what is considered conservative in one accounting period may lead to 'non-conservative' results in subsequent periods (Givoly and Hayn 2000). They provide a specific definition of conservatism which aims to emphasise the multi-period dimensionality of conservatism. They define conservatism as *"a selection criterion between accounting principles that leads to the minimisation of cumulative reported earnings by slower revenue recognition, faster expense recognition, lower asset valuation, and higher liability valuation". (Givoly and Hayn 2000, p. 292)*

In the long-term, accounting performance (earnings) converge to the 'true' performance of the firm, measured as the cash flow from operations. This implies that any departures of earnings from cash flows is temporary and mean reverting. Givoly and Hayn (2000) posit that changes in the time-series properties of earnings and cash flows, namely the accumulation of negative non-operating accruals, are consistent with accounting conservatism. Without intervention, accruals can typically be expected to reverse over time, and thus, persistence in the level of cumulative negative accruals should reflect a conservatism bias. Givoly and Hayn (2000) argue that negative non-operating accruals are increasing with the level of accounting conservatism.

Total accruals are broken into two categories: operating and non-operating. Operating accruals arise from the day-to-day business of the firm, and accumulate with growth in the firm's operations (Givoly and Hayn 2000). Operating accruals are primarily a reflection of the underlying economics of the firm. The remaining component of total accruals is non-operating accruals. Non-operating accruals include items such as loss and bad debt provisions, restructuring charges, asset write-downs, the accrual and capitalisation of expenses, and the deferral and subsequent recognition of revenues (Givoly and Hayn 2000, p. 304). Although GAAP mandates treatment of some non-operating accruals, their measurement and timing are most often subject to management discretion. As such, an accumulation of non-operating accruals is indicative of conservative accounting.[8] NA is computed as indicated in equation (2):

$$NA_{it} = -1 \times \left(\frac{1}{N} \sum_{j=1}^{N} \frac{NOACC_{it}}{TA_{it}} \right) \qquad (2)$$

where *i* is the firm identifier; *j* is the selected accumulation period; NA_{it} is the accruals-based conservatism proxy derived for firm *i* for fiscal year end *t.* $NOACC_{it}$ is non-operating accruals calculated as total accruals (*TAC*) minus operating accruals (*OAC*) for firm *i* for fiscal year-end *t* determined using the direct approach (cashflow method) consistent with Dechow and Dichev (2002) and Givoly and Hayn (2002).[9] TA_{it} is the total assets for firm *i* at fiscal year-end *t.* TAC is total accruals (before depreciation)

---

[8] The literature has also computed this proxy using total accruals as opposed to non-operating accruals, for example Ahmed *et al.* (2002). Givoly and Hayn (2000) argue against using total accruals as they will reflect firm growth, injecting noise into the measure. As such, non-operating accruals as opposed to total accruals are used in this paper consistent with Givoly and Hayn (2000).

[9] The alternative method for estimating accruals is to use the indirect approach which uses the balance sheet. This approach is necessary for sample pre- 1988 when the Statement of Cashflow was unavailable. This approach is prone to measurement

calculated as Net Income before Extraordinary Items + Depreciation - Cash flow from operations. OAC is operating accruals calculated as ΔAccounts Receivable + ΔInventories - ΔAccounts Payable - ΔTaxes Payable + Δ Other Assets and Liabilities. The NA measure is calculated as the average of the ratio of non-operating accruals to total assets, which is multiplied by -1 to give a positive measure of conservatism.

The NA measure requires non-operating accruals to be accumulated over a period of time. A practical issue arising from this is what a sufficient period of accumulation is and how to standardise the base year across firms. Several alternatives are identified in the literature. Ahmed *et al.* (2002) and Francis *et al.* (2004) choose six years as a sufficient period of time to identify persistence in the accumulation of negative accruals.[10] This paper uses a six year accumulation period as the core measure of NA. To ensure that firms in the sample have a common base year, it is required that the firm have available a value for total assets and all data required to compute non-operating accruals in each of the six years (year t through t_{-5}).

The validity of the NA measure has not been empirically examined in any significant way in the extant literature. As such, whether there are biases in the measure is unknown (Wang *et al.* 2009).

4.2 Conservatism – hidden reserve measure (*HR)*

Penman and Zhang (2002) characterise conservatism as the biased application of historical cost accounting. Their definition follows from the arguments provided by Gjesdal (1999). Gjesdal (1999) classifies accounting as conservative when it carries assets at a value which yields an accounting rate of return exceeding the internal rate of return. Penman and Zhang (2002) conjecture that a conservative accounting system would carry an asset with a value equal to historical cost (implying a zero NPV investment) at less than historical cost. They argue that this is the case with the immediate expensing of R&D and advertising costs.

The basis of the Penman and Zhang (2002) measure is that accounting conservatism generates hidden reserves (also called cookie jars), and therefore, the higher the value of those hidden reserves the more conservative a firm's accounting (Penman and Zhang 2000). The estimated hidden reserves results from the LIFO, R&D and advertising reserves.[11] Penman and Zhang (2002) choose these investments because the accounting treatment of R&D and advertising is dictated by regulators and in the case of LIFO; management is unlikely to change their accounting policy from one period to another. As such, these items are isolated from management discretion once the expenditure occurs.

---

error, particularly in the case where firms have discontinued operations or merger and acquisition activity (Hribar and Collins 2002).

[10] Ahmed *et al.* (2002) suggest that accumulation periods as short as three years may be sufficient. Recent papers by Lara *et al.* (2009) and Ahmed and Duellman (2007) use a three year average of non-operating accruals centred on time t (i.e. year t-1 through t+1). These alternative operationalisations are use in sensitivity analysis.

[11] An alternative methodology for estimating hidden reserves is proposed by Ahmed *et al.* (2000), who proxy hidden reserves using two ratios: R&D expense/sales and advertising expense/sales. This methodology has not been adopted by the literature, and is therefore not examined in this paper.

For the proxy to be computed, firms are required to have available at least one of the three components of hidden reserves (LIFO reserve, R&D reserve or advertising reserve) and must have all necessary inputs to calculate net operating assets. For the purposes of computing the R&D and advertising reserves it is required that a firm have at least one valid observation for R&D/advertising expense during the useful life period. Firms are only eliminated from the sample if they have no valid observations for the entire useful life period. Firm-years which do not fulfil these criteria are eliminated. HR is calculated as specified in equation (3):

$$HR_{it} = \frac{ER_{it}}{NOA_{it}} \tag{3}$$

where HR_{it} is the conservatism proxy derived for firm *i* for fiscal year end *t.* ER_{it} is the estimated hidden reserves created by accounting conservatism, for firm *i* at fiscal year-end *t*, calculated as the inventory reserve plus the R&D reserve plus the advertising reserve. The inventory reserve is measured as the value of the LIFO reserve reported under US GAAP.[12] The R&D reserve is calculated by capitalising all R&D expenses from the income statement and amortising the balance using the industry coefficients estimated by Lev and Sougiannis (1996), which represent the estimated average industry rates of amortisation.[13] The advertising reserve is calculated by capitalising all advertising expenses from the income statement and amortising the balance using the SOYD method over two years.[14] NOA_{it} is net operating assets calculated as: Common Equity (Common Equity + Preferred Treasury Stock - Preferred Dividends in Arrears) + Financial Obligations (Debt in Current Liabilities + Total Long-Term Debt + Preferred Stock - Preferred Treasury Stock + Preferred Dividends in Arrears) - Financial Assets (Cash and Short-Term Investments + Investments and Advances - Other) + Minority Interest (Minority Interest on Balance Sheet). The higher the value of HR, the more conservative a firm's accounting. Therefore, HR is a positive measure of conservatism.

4.3 Conservatism - Book-to-Market Measure (*BM*)

This measure has its origins in the Feltham and Ohlson (1995) Residual Income Valuation Model. Feltham and Ohlson (1995) describe conservative accounting as an expectation that reported net assets will be less than market value in the long-run; and under the assumption of clean surplus accounting this implies lower average net incomes (Penman and Zhang 2002, Ball and Shivakumar 2005). *Ceteris paribus*, conservative accounting tends to bias the book value of the firm downwards relative to its 'true' economic value, resulting in a low book-to-market ratio, or equivalently a high market-to-book ratio.

---

[12] The LIFO reserve represents the difference between the carrying value of inventory between LIFO and FIFO inventory methods and is disclosed in the notes to the firm's financial statements.

[13] Consistent with Penman and Zhang (2000), a second specification of hidden reserves is calculated by amortising the capitalised R&D expenditure using the sum-of-the-year's digits (SOYD) method over five years.

[14] This is consistent with evidence from Hall (1993) and Bubblitz and Ettredge (1989) that the useful life of advertising is short, typically spanning one to two years.

Beaver and Ryan (2000) refine this idea by decomposing the book-to-market ratio ratio into two components: the bias component and the lag component. They argue that the bias component is a more accurate measure of conservatism compared to the raw book-to-market ratio. The bias component captures the portion of the difference between the book and market value of equity which persists.

The bias component of the book-to-market is estimated using a fixed effects regression. As indicated in equation (4), in each year the firm's book-to-market ratio is regressed on current and six lagged annual security returns along with fixed firm and time effects.

$$BTM_{it} = \alpha_t + \alpha_i + \sum_{j=0}^{6} \beta_j R_{t-j,i} + \varepsilon_t \qquad (4)$$

Where BTM_{it} is the book-to-market ratio measured as the book value of common equity divided by the market value of common equity for firm *i* at fiscal year *t*. α_t is the year-to-year variation in the BTM common to the sample firms. α_i is the bias component of BTM for firm *i*. $R_{t-j,i}$ is the return on equity over each of the six preceding years for firm *i* for fiscal year *t-1*.[15] The individual firm effects (α_i) represent the conservatism proxy for each firm. A low coefficient implies a higher degree of accounting conservatism (the more the book value of assets is depressed). For consistency with each of the other measures, the BM proxy is multiplied by -1 to create a positive measure of conservatism.

4.4 Conservatism - Khan and Watts (2009) Firm-Year AT Measure (*KW*)

The Khan and Watts (2009) measure is a derivative of the most widely applied conservatism proxy, the Basu (1997) Asymmetric Timeliness (AT) measure. As such, it relies on the same fundamental definition and interpretation of conservatism as Basu (1997). Watts (2003a) notes that conservatism does not require all revenue cash flows to be received before they are recognised; rather, to be recognised those cash flows must be verifiable. Basu (1997) embraces this notion of verifiability and interprets conservatism as "*the accountants' tendency to require a higher degree of verification to recognise good news as gains than to recognise bad news as losses" (Basu 1997, p.7).*[16]

The implication of Basu's (1997) interpretation is that earnings will reflect bad news more quickly than good news. As such, conservatism results in the asymmetric timeliness of earnings. Khan and Watts' (2009) measure relies on the Basu (1997) coefficient as an indication of conservatism. In order to transform the asymmetric timeliness measure into a firm-specific proxy Khan and Watts assume that the asymmetric timeliness of earnings represents a flow measure of conservatism. A firm-specific conservatism proxy is generated by specifying the calculated Basu (1997) coefficient as a linear function of firm-specific characteristics (size, MTB ratio and leverage) which vary with conservatism.

---

[15] Following Beaver and Ryan (2000) annual stock returns are holding period returns adjusted for dividends and stock splits. To ensure reliable estimation of the firm-effects, four time-series observations are required (Beaver and Ryan 2000). Therefore, the regression is estimated in each year using data for the current and preceding three years. Firms that do not have the data required to calculate all variables and those without four time-series observations are deleted.

[16] This interpretation of conservatism is also adopted by Ball and Shivakumar (2005).

The KW measure is computed in two stages. First, a cross-sectional regression is estimated annually as indicated in equation (5). Second, the estimated lamda (λ) coefficients obtained from the cross-sectional regression are combined with each firm's size, MTB ratio and leverage to obtain the KW measure as shown in equation (6).

$$X_i = \beta_0 + \beta_1 DR_i + R_i\,(\mu_1 + \mu_2 Size_i + \mu_3 MTB_i + \mu_4 Lev_i) + DR_i R_i(\lambda_1 + \lambda_2 Size_i + \lambda_3 MTB_i + \lambda_4 Lev_i)$$
$$+ D_i(\alpha_1 Size_i + \alpha_2 MTB_i + \alpha_3 Lev_i) + (\delta_1 Size_i + \delta_2 MTB_i + \delta_3 Lev_i) + \varepsilon_i \quad (5)$$

$$KW = \lambda_1 + \lambda_2 Size + \lambda_3 MTB + \lambda_4 Lev \quad (6)$$

Where X_i is earnings measured as net income before extraordinary items scaled by lagged market value of equity for firm *i* year *t*. R_i is the buy-and-hold stock market return for firm *i* from nine months before fiscal year end *t* to three months after fiscal year-end *t*. $Size_i$ is the natural log of the market value of equity for firm *i* at fiscal year-end *t.* MTB_i is the market-to-book ratio measured as the market value of common equity divided by the book value of common equity for firm *i* at fiscal year-end *t*. Lev_i is leverage defined as long-term plus short-term debt scaled by market value of equity for firm *i* at fiscal year-end *t.* DR_i is a dummy variable that is equal to 1 if the stock market return for firm *i* year *t* is negative (bad news), and equal to 0 if the stock market return for firm *i* year *t* is non-negative (good news).[17] ($R_{it} < 0 = 1$; $R_{it} \geq 0 = 0$). The value of KW is increasing with conservatism; as such it is a positive measure of accounting conservatism.

5. Sample selection and characteristic

The construct validity of the conservatism proxies is examined using a sample of U.S. listed firms for the period 1997 to 2007. To ensure that the empirical tests give each of the proxies an equal opportunity to demonstrate their validity, it is necessary to use the same sample of firms to assess each proxy. Therefore, the aim of the sample selection process is to identify a common sample in which all proxies are represented. To do so, a sample is derived in each year by eliminating firm-years from the population of COMPUSTAT firms without sufficient data to calculate all proxies and firm characteristics. This process yields a preliminary sample comprising 3,822 firm-year observations.[18]

An initial investigation of this sample uncovered the presence of outliers. To confront the issue, a search is conducted to identify observations which are considered unreasonable or anomalous. First, in the spirit of Beaver and Ryan (2000) firm-year observations with a BTM ratio less than zero or exceeding four are eliminated from the sample. Next, firms with ROA greater than 100 percent in either direction are also deleted. These filters result in the removal of 102 firm-year observations from the sample,

[17] Consistent with Basu (1997), the stock market return is the buy-and-hold return on firm *i*, computed using the CRSP monthly stock file. Khan and Watts (2009) are unclear on the timing of the measurement of size and leverage. It is however, explicitly stated that MTB is calculated at fiscal year-end. It has been inferred that size and leverage are also measured at fiscal year-end.

[18] The HR measure results in considerable sample attrition.

resulting in a final sample of 3,720 firm-year observations. The sample breakdown by year and industry is provided in Table 1.

The number of firms is increasing over the sample period; with an almost monotonic increase in each year's representation in the overall sample. There is no year which dominates the sample with 2006 having the largest percentage at 11.4 percent, and 1998 the smallest covering 5.6 percent of firm-year observations. In total, twenty-three (of twenty-four) industry groups are represented in the sample as classified using the GICS industry group classifications.[19] Two groups, capital goods and technology hardware and equipment are largest comprising 13.9 and 13.4 percent of firm-year observations respectively. Four industries represent less than one percent of the sample: diversified financial and insurance with 0.3 percent and telecommunications and utilities with 0.6 percent of firm-year observations respectively. Overall, the sample is relatively well distributed across industries.

Table 2 provides the descriptive statistics of the common firm characteristics, test variables and conservatism proxies for the pooled sample. The non-parametric Kruskal-Wallis ANOVA reveals that medians for all firm characteristic variables are statistically different across all years in the sample. These results suggest that the nature of the firms which comprise the sample is changing over time which should provide a relatively heterogeneous sample and thus increasing the generalisability of the results.

There are significant distributional issues in the conservatism variables, with the proxies exhibiting high levels of skewness and kurtosis in the majority of years. To deal with the distributional issues, the firm-year observation data is truncated at the top and bottom one percent for each of the conservatism variables. This results in the deletion of 354 firm-year observations, leaving a sample of 3,366 firm-year observations. The descriptive statistics for the conservatism variables truncated at 1 percent are provided in Table 2. This sample is used as the primary sample for the subsequent analyses.[20] Truncating the sample at the 1 percent level results in much more reasonable distributions without a significant reduction in sample size. The NA measure appears well distributed across all years of the sample. Mean and median values of NA are statistically different in every year of the sample. NA appears to experience a significant increase in 2001 and a smaller increase in 2002, remaining relatively constant for the remaining years of the sample.[21] The untabulated results of *F*-test for the difference in means and a Kruskal-Wallis test for the difference in medians indicate that the mean values of both HR and KW do not

---

[19] During the sample period no firms from the banking industry had all the required data items available on COMPUSTAT, and therefore the industry is not represented in the sample.

[20] In robustness testing, data truncation occurs for values of conservatism falling outside three standard deviations from the mean with at least the largest and smallest values deleted. This yields an elimination of 290 firm-year observations, leaving a sample of 3,430 firm-year observations.

[21] Ahmed *et al.* (2002) report a mean of 0.004, median of 0.003 and standard deviation of 0.031 for their pooled sample which runs from 1993 to 1998. The mean, median and standard deviation in this examination appears smaller. However, comparisons are difficult given that the sample for Ahmed *et al.* (2002) only comprises 704 firms and uses total accruals as opposed to non-operating accruals in the numerator. So fundamentally, the measures are not comparable. Givoly and Hayn (2000) only report the ratio of non-operating accruals to total assets. For their pooled sample running from 1951 to 1980 they report a mean of -0.016 and median of -0.003. For the subsequent sample period extending from 1981 to 1998 the mean and median are lower at -0.047 and -0.037 respectively.

change significantly over the sample period with *p*-values of 0.828 and 0.303 respectively. Test of medians indicate that only the HR proxy is not statistically different across years with a *p*-value of 0.603.

6. Results

This section presents the results for each of the validity tests as explained in section 3 on each of the conservatism proxies as identified in section 4.

6.1 Concurrent Validity

Given the adopted definition of conservatism, it is expected that conservatism is likely to be greater in firms which experience high uncertainty. The tests of concurrent validity test this proposition by determining whether there is a significant difference in the mean conservatism score given by each of the proxies for firms in the 'high' and 'low' uncertainty quintiles. Table 3 provides the comprehensive results of the tests of concurrent validity for each year of analysis. A summary of results is provided below indicating for each proxy, the number of statistically significant positive and negative mean differences over the 11 year sample period 1997 to 2007. (For completeness the number of positive and negative mean differences that are not significant are also reported).

Summary of the Results for Concurrent Validity – Test of Mean Differences over 11 year sample.

Variable	Predicted Sign	Significant +ve Mean Diff	Significant -ve Mean Diff	Not Significant +ve Mean Diff	Not Significant -ve Mean Diff
NA	+	6	1	3	1
HR	+	6	0	3	2
BM	+	2	9	0	0
KW	+	9	2	0	0

With respect to NA, there is no positive statistical difference between the 'high' and 'low' uncertainty quintiles in the first five years of the sample indicating limited concurrent validity. However, in 2000, contrary to expectations NA exhibits a significant negative mean difference with a *p*-value of 0.058 (one-tailed). The mean difference is positive but not significant in 2001. From 2002 onwards, there is a significant difference in NA for the two samples, with *p*-values not exceeding 0.03 (two-tailed). The mean difference is always positive and therefore consistent with predictions. Therefore, in the latter half of the sample, consistent with the definition of conservatism, NA indicates higher conservatism in firms experiencing conditions of uncertainty.

There is a significant difference in HR for the two groups in six of eleven years. All mean differences are in the predicted (positive) direction. This suggests, that like NA, HR indicates greater conservatism in 'high' uncertainty firms. However, in contrast to NA, the positive mean differences in HR are irregular. The difference is observed in each of the years 1999-2002. The other two instances of significance occur in 1997 and 2004. So, although HR does distinguish between the two groups in the way in which we would expect in the majority of years, it doesn't appear to do so with the same level of consistency as NA.

BM is statistically different for the two groups in every year of the sample. However, in nine of eleven years the direction of the difference is opposite to predictions. A positive mean difference only occurs in two years (2005 and 2006). Therefore, although BM does consistently discriminate between the two groups, it doesn't do so correctly. These results suggest that in most cases, BM indicates higher levels of conservatism in firms with low uncertainty, which is contrary to the definition of conservatism adopted by this paper and with what theory suggests.

There is a statistically significant difference in KW for the two groups in all years of the sample. The mean difference is positive and therefore consistent with predictions in nine of those years. The results suggest that overall KW indicates higher levels of conservatism for firms in the 'high' uncertainty quintile, which is consistent with the definition of conservatism adopted. There is a negative mean difference in two of the eleven years, both occurring early in the sample (1997 and 2001). Therefore, like NA, it appears that KW performs better in the later years of the sample.

To check for robustness of the concurrent validity results two additional analyses are conducted. First the tests are conducting use the sample with each conservatism variable truncated at three standard deviations. The results are qualitatively similar and are therefore not reported.

Second, instead of breaking firms into quintiles, in each year, firms were classified as having 'high' or 'low' uncertainty if they were in the top and bottom 10 and 30 percent of stock price volatility respectively. The untabulated results remained unchanged when 30 percent is used as the alternative cut-off point. When 10 percent is used as the cut-off point, the results do not change with respect to BM. KW exhibits a very marginal decrease in performance, indicating a significant positive difference in eight instead of nine years. NA appears to have greater power to discriminate correctly between the two groups when the more stringent cut-off is applied. There is a significant and positive mean difference in NA for eight of eleven years. The negative mean difference observed for 2000 for the other cut-off points is no longer significant. HR by contrast performs more poorly. HR achieves one less significant positive mean difference. Further, two negative mean differences which were found to be insignificant under the other specifications, are significant when using the 10 percent cut-off. These results may suggest that NA has a superior level of concurrent validity compared to HR.

Overall, the results suggest that three of the four proxies, NA, HR and KW, all have some degree of concurrent validity, whereas BM does not appear to have concurrent validity. There is statistical difference for KW in all eleven years. However, in two of those years, the direction of the difference is opposite to predictions. The results for KW are qualitatively similar for all alternative cut-off points. BM gives results which are inconsistent with the definition of conservatism adopted by this paper under all test specifications. Based on the primary results, NA and HR are difficult to separate. Both are statistically different in the predicted direction in the majority of years. NA exhibits a negative mean difference during one year of the sample, which would seem to indicate that HR is a marginally better performer.

However, NA appears to be more consistent, exhibiting a difference in the predicted direction in consecutive years from 2002.

6.2 Discriminant Validity

The test of discriminant validity seeks to determine whether the conservatism proxies are correlated with efficiency (measured as the sales-to-employee ratio), a construct theoretically and empirically unrelated to conservatism. A significant correlation of any sign between the proxies and efficiency will indicate that the proxies are noisy and imprecise, and therefore lack discriminant validity. Table 4 provides the results for the discriminant validity tests.

Summary of the Results for Discriminant Validity – Significant Correlations over 11-year sample.

Variable	Pearson's Years Significant	Spearman's Years Significant
NA	2	1
HR	0	2
BM	3	3
KW	5	7

Given the distributional issues identified in the data, Spearman's correlation coefficients are relied upon. NA exhibits a significant correlation with the sales-to-employee ratio in only one year (2000) with a correlation coefficient of 0.101 (*p*-value 0.097). This suggests that NA exhibits a very weak correlation with efficiency, indicating that the proxy has discriminant validity. Pearson's correlation coefficients indicate a similar conclusion, showing significant correlations of -0.115 in 2001 and -0.134 in 2007.

With respect to HR, Spearman's correlation coefficients are statistically significant in two years (2005 and 2006) revealing correlations coefficients of 0.119 and 0.110 respectively. Both coefficients have a similar magnitude compared to those for NA. None of the Pearson's correlation coefficients are significant at the 10 percent level. These results suggest that HR has discriminant validity, with the proxy exhibiting only a weak (if any) correlation with the sales-to-employee ratio.

The Spearman's correlation between BM and the sales-to-employee ratio is statistically significant for the first three years of the sample, measuring -0.182, -0.148 and -0.123 respectively. Pearson's correlation coefficients are significant in 1997, 2005 and 2006, measuring -0.153, -0.087 and -0.111 respectively. These results suggest that BM does posses discriminant validity. However, the frequency and magnitude of these coefficients is marginally higher than those for NA and HR, suggesting that BM may exhibit comparatively weaker discriminant validity.

The KW proxy has a negative and statistically significant correlation with the sales-to-employee ratio in seven out of eleven years. The correlations are all negative, and range from -0.137 to -0.213. These correlations are larger and much more frequent compared to those observed for the other proxies. Pearson's correlation coefficients tell a similar story. They are significant in five years, but have a slightly lower magnitude ranging from -0.106 to -0.158. These results suggest that there is a systematic

correlation between KW and efficiency. Therefore it appears that KW is imprecise and captures considerable noise. Therefore, the results indicate that KW lacks discriminant validity.

For robustness, the tests of discriminant validity were also undertaken using the sample with conservatism variables truncated at three standard deviations from the mean. The results are qualitatively similar and are therefore not reported.

On the basis of Spearman's correlation coefficients, it appears that NA has superior discriminant validity, with only one significant correlation with the sales-to-employee ratio. HR and BM show some discriminant validity, exhibiting correlation in only a small minority of years. The results suggest that KW does not have discriminant validity, exhibiting a significant correlation with efficiency in the majority of years.

6.3 Convergent Validity

The test of convergent validity seeks to determine whether the proxies demonstrate a high level of correlation, or in other words, converge. A significant correlation with a high magnitude and the predicted sign will indicate that the proxies have convergent validity.

Detailed convergent validity results by year are given by the correlation matrices provided in Table 5. A summary of the results are provided below. The left-hand side of each cell indicates the number of significant positive coefficients whilst the right-hand of each cell shows the number of significant negative correlations coefficients over the 11-year sample period. Coefficients significant at 10% level or better are reported. The results above the diagonal pertain to Spearman's correlation coefficients with the Pearson's correlation coefficients below the diagonal. Due to distributional issues identified in the data, Spearman's correlation coefficients are discussed below.

Summary of the Results for Convergent Validity – Significant Correlations over 11-year sample.

Variable	NA		HR		BM		KW	
NA			+2		+2	-2	+3	-2
HR	*+1*				+7		+5	-3
BM	*+0*	-5	+5	-1			+2	-3
KW	*+5*	-1	+1	-1	+3	-3		

Overall, there is limited consistency in correlations across years, and it is difficult to identify persistent trends. NA is not significantly correlated with any other measure for three years of the sample when using the Spearman correlation coefficients (this is the same when relying on Pearson's coefficients). NA shows a statistically significant correlation with KW in five years though the result is inconsistent with three years positive and two years negative years. NA is significantly correlated with BM in four years; however, the sign of the correlation is only consistent with predictions in two of those years. Whilst the nature of the relation between NA and BM is not clear, given the even split between positive and negative correlation, the predicted positive correlations have a lower magnitude. There is a

statistically significant correlation between NA and HR in two years. On both occasions the correlation has the predicted positive sign. However, the correlations are low, measuring only 0.098 and 0.096 respectively. Pearson's correlation coefficients tell a similar story, indicating that NA is only significantly positively correlated with HR on one occasion.

HR exhibits a higher correlation with BM and KW compared to NA. Consistent with predictions, HR has a significant positive correlation with BM in seven of eleven years with Spearman coefficients ranging from 0.100 to 0.242. Pearson's correlation coefficients are significant in six years, with a positive sign in five of those years. The magnitude of the coefficients is qualitatively similar. HR has a significant Spearman correlation with KW in eight of the years however five years have positive coefficients and three have negative. The magnitude of the coefficients range from -0.190 to 0.262. The Pearson's correlation coefficients are only significant in two years, 1999 with a positive coefficient of 0.223 and 2007 with a negative coefficient of -0.137.

BM is significantly correlated with KW in five of eleven years. The Spearman correlation coefficients are negative in three years, and exhibit a relatively high magnitude ranging from -0.306 to -0.516. The positive coefficients exhibit a greater range, measuring 0.110 and 0.363. Pearson's correlation coefficients are significant in six years (three positive, three negative). The coefficients are of a similar magnitude ranging from -0.454 to 0.168.

KW is correlated with at least one of the alternative proxies in each year of the sample (all but one using Pearson's correlation coefficients). It appears to be most correlated with the HR measure, with significant correlations in eight years, the highest correlation being 0.262. KW is correlated with BM and NA in five and four years respectively. The correlation between KW and BM is the strongest uncovered ranging between 0.363 and -0.516. The strongest period of correlation between the two proxies occurs from 2000 to 2002, and briefly again in 2005. It is difficult to ascertain the relation between KW and the other proxies, given that the number of positive and negative coefficients is similar.

To check for robustness the convergent validity tests are also conducting use the sample with the conservatism variables truncated at three standard deviations. The results are qualitatively similar and are therefore not reported.

In conclusion, the results do not provide convincing evidence of convergent validity. Overall it appears that the correlation amongst the firm-specific proxies is relatively low. HR and BM appear to be the most correlated proxies, with significant correlations in the predicted direction occurring in seven of eleven years. KW shares a similar correlation with all the other proxies. The nature of its relation however is difficult to ascertain, as instances of negative and positive correlation appear almost equal. NA appears to be the least correlated proxy, indicating that it is somewhat separate from its alternatives. With respect to NA, HR and BM the findings appear qualitatively similar to those of Wang *et al.* (2009) who find the correlation amongst those proxies to be low.

6.4 Predictive Validity

The tests of predictive validity determine whether the conservatism proxies are able to predict what they should theoretically be able to predict. This is empirically tested by regressing the conservatism proxies on various firm characteristics which are theoretically and empirically associated with conservatism. These include: size, leverage (LEV), profitability (ROA) and auditor conservatism (AUD). Predictive validity is indicated by a significant coefficient with the predicted sign. A second and important question answered by these tests is whether the proxies are associated with the chosen characteristics in a consistent manner. If so, this would indicate that the proxies are capturing the same construct.

The maximum correlation between the independent variables (Size, LEV, ROA and AUD) included in the regression in equation (1) is 0.547. This reflects an increasing correlation between AUD and Size in the later years of the sample. In most instances, the correlation between the other independent variables is significantly lower and do not exceed 0.42.[22] Given that none of the correlations exceeds 0.8, the threat of serious multicollinearity is low (Gujarati 1995).

Predictive validity is analysed from three perspectives: i) the significance of the annual estimated coefficients, ii) Fama-Macbeth (1973) *t*-statistics and iii) pooled regression with fixed firm and year effects. A summary of the annual coefficients and the average coefficients and associated Fama-Macbeth (1973) *t*-statistics.

6.4.1 Annual Estimated Coefficients

The summary results of the annual regressions are presented below the number of coefficients greater than zero (positive) on the left side of the column, and the number of coefficients less than zero (negative) on the right side of the column. The number of significant coefficients[23] are indicated in parenthesis.[24] With respect to the annual coefficients, all proxies exhibit a different relationship with size, leverage and profitability.

Summary of the Results for Predictive Validity – Annual OLS Regressions over 11-year sample.

Variable	Size		Leverage		Profitability		Auditor	
NA	2+ *(0)*	9- *(2)*	11+ *(5)*	0- *(0)*	0+ *(0)*	11- *(7)*	4+ *(0)*	7- *(1)*
HR	7+ *(2)*	4- *(0)*	4+ *(1)*	7- *(4)*	4+ *(4)*	7- *(4)*	6+ *(0)*	5- *(1)*
BM	10+ *(9)*	1- *(0)*	6+ *(0)*	5- *(1)*	6+ *(3)*	5- *(2)*	7+ *(0)*	4- *(1)*
KW	2+ *(2)*	9- *(8)*	3+ *(3)*	8- *(5)*	4+ *(1)*	7- *(4)*	2+ *(0)*	9- *(2)*

Size is predicted to have a positive association with conservatism. NA bears the weakest association with Size with only two of the 11 years indicating a significant though negative association. KW exhibits a much stronger association with size with ten of the 11 years indicating a significant association, however contrary to predictions; eight of the coefficients are significant and negative. The strongest association is

[22] Untabulated results are available from the corresponding author.
[23] Significant coefficients are determined as coefficients $\pm$1.645.
[24] The individual annual regressions are not tabulated given the extensive nature. These are available from the authors.

evidenced by BM where nine of the 11 years have a significant positive association. HR has a weak association with two years significant positive coefficients. From the result, it is difficult to suggest an overriding direction of the association between conservatism and size. What the results do indicate however is that the proxies are not related to size in a consistent manner as would be expected if the proxies were capturing the same construct.

Leverage is predicated to have a positive association with conservatism. NA is the only proxy to indicate a relationship with leverage in the predicted direction however the association is not strong with only five of the 11 significant. HR indicates one significant positive and four significant negative coefficients. BM has no relationship with leverage with only one negative coefficient in the 11 years. KW displays a significant positive association with leverage in 3 years however in five years the association is significant and negative. In conclusion, like size, it appears that the proxies do not share a consistent relationship with leverage. The annual estimated coefficients suggest that NA is the only proxy which is increasing with leverage as predicted though the relationship is evident in less than half of years.

Profitability is predicated to be positively associated with conservatism though, as indicated in section 3.4 there is evidence in the literature of a negative association. NA records seven negative significant coefficients, HR and KW indicate four and BM has two. KW records one significant positive coefficient, BM records three significant positive and HR records four significant positive coefficients. Hence, HR has as many significant positive as negative. In summary, the proxies do not appear to be consistently related to profitability. All proxies with the exception of BTM exhibit a stronger negative relationship to ROA. For example, all seven of the significant coefficients estimated for NA are negative. A plausible explanation for this result is that the effect of non-operating accruals dominates the expected positive relationship between NA and ROA (Ahmed *et al.* 2002).

Auditor is predicted to have a positive association with conservatism. The results of the annual regressions are very weak with NA, HR and BM recording only one negative and no positive significant coefficients whilst KW records only two negative significant coefficients. Although the sign of the significant annual estimated coefficients are contrary to predictions, the proxies do appear to be consistent in their association with auditor conservatism.

Overall, what is evident from the annual regressions is that there is no strong evidence in the sample supporting the predicted relationship between conservatism and size, leverage, profitability or auditor. However, what the annual regressions do highlight is that the proxies are not consistently associated with the variable of interest. Lack of consistency can be interpreted as a lack of predictive validity.[25]

[25] In robustness tests, the annual cross-sectional regressions are also estimated using the sample with conservatism truncated at three standard deviations. The results are qualitatively similar and are therefore not reported.

6.4.2 Fama-Macbeth t-Statistics

Fama and Macbeth (1973) estimate *t*-statistics in a two-step process. First, parameter estimates are obtained from annual cross-sectional regressions. Second, the associated *t*-statistics are computed using the average estimated coefficient for the sample period. The average annual coefficient and Fama-Macbeth (1973) *t*-statistics are reported in Table 6 Panel A.

Size yields two significant coefficients; BM is found to be positive and significantly related to size as predicted, however contrary to predictions, NA yields a significant negative association. HR and KW are not significantly related. Consistent with predications, leverage is only positive and significant for NA. The only significant relationship for profitability is a significant negative association with NA. This result is consistent in Ahmed *et al.* (2000) with the prediction that the non-operating accruals dominates the expected positive relationship and hence evidences a negative relationship. This relationship is not replicated by the other conservatism proxies. The results for auditor remain weak with KW indicating a significant negative coefficient contrary to expectations.[26] Overall, the Fama-Macbeth result reveals that NA yields results consistent with prediction for Leverage and profitability though a result contrary to prediction for Size. HR reveals no predictive results, BM has a predicted relationship with size and KW has a contrary to predicted relationship with auditor. The overall results are week and provide very little support for the predictive ability of the constructs and inconsistency between measures.

6.4.3 Pooled Regression

The results of the pooled regression with fixed firm and year effects are given by Table 6 panel B.[27] Size is a positive and significant for BM, a negative and significant for NA, and not related to HR or KW. The result again reveals inconsistency between measures.

Despite loading significantly in the Fama-Macbeth (1973) regressions for at least NA, the leverage coefficient is not significant in the pooled regression estimates for NA, HR or KW. The predicted positive significant relation is evident for BM. Consistent with the Fama-Macbeth (1973) results; ROA continues to exhibit a negative significant relationship with NA. This result is replicated for both the HR and KW result but not BM. This provides further evidence the negative relationship between conservatism and profitability may be the more correct specification consistent with Ahmed *et al.* (2002). Each of the conservatism proxies continues to show a consistent lack of predictive ability with regards to auditor with no significant coefficients evident in the pooled regressions.

In summary, the pooled regression results for leverage and profitability do provide evidence of a consistency between NA, HR and KW but indicate an inconsistency with BM. Based on the concurrence of the results to predictions it appears that BM has the highest degree of predictive validity when relying

---

[26] Robustness checks of the Fama-Macbeth (1973) *t*-statistics using the sample with conservatism truncated at three standard deviations yield results qualitatively similar and are therefore not reported.

[27] In robustness tests, the pooled regressions are also estimated using the sample with conservatism truncated at three standard deviations. The results are qualitatively similar and are therefore not reported.

on the results of the pooled regression, though clearly inconsistent with the other proxies. Finally, each is consistently poor in the predictive relationship with auditor.

It should be noted that the tests conducted cannot provide definitive evidence of predictive validity and an accurate interpretation of the results obtained is difficult. First, it is not possible to know whether the different results obtained using the various methodologies are due to methodological differences or otherwise. Second, as discussed in Section 3.4 the tests of predictive validity are inherently a joint-test. As such, results contrary to predications can be attributed to both the efficacy of the proxies and the firm characteristics chosen.

When considering the concurrence of the results with predictions and accepting a negative relationship with ROA, it appears that NA and BM may have the greatest predictive validity. HR and KW appear to be less consistent with predictions, as such is seems that they have a lesser degree of predictive validity. However, these judgements are made with caution due to the lack of a definitive benchmark and as such do not provide robust evidence of construct validity. The primary aim of conducting this test is to provide an understanding of the relationship amongst the alternative proxies. Therefore, the take away from these results is that the proxies are not associated with these firm characteristics in a consistent manner. As such, it appears that some or all the proxies are either (a) capturing something other than conservatism or (b) operationalising different constructs of conservatism.

6.5 Summary of results

The final requirement is to synthesis the individual results to endeavour to make a decision regarding the overall effectiveness of the proxies. As indicated in 3.5, this cannot be a definitive horse-race where a winner can be declared. Further, there is no methodology by which to synthesis the result, so this is left to the researcher judgement.

Therefore, to determine the degree of construct validity, each proxy is assigned a rank from one to four, with one representing the best performance and four the worst. From these, the average rank for each proxy is computed to determine their standing in relation to the other proxies. Table 8 provides the rankings for each test of sub-validity. To enhance the readability of the table, a short description of the sub-validity is also provided. Preliminary average ranks are computed on the basis of the tests of concurrent and discriminant validity. These tests are the most definitive (least unambiguous), and therefore the ranking of the proxies is more reliable. The tests of predictive and convergent validity are not unambiguous, and are interpreted in light of the other results obtained. For comparison, average ranks are recomputed including tentative rankings for predictive validity. Due to the nature of the test of convergent validity it is not possible to provide rankings; the results are merely interpreted in light of the rankings in the other tests of construct validity.

KW clearly possesses the greatest degree of concurrent validity, showing a positive mean difference between the 'high' and 'low' uncertainty groups in nine of eleven years. NA and HR perform similarly and accordingly both receive a rank of two. BM exhibits little concurrent validity, incorrectly differentiating between the groups in nine of eleven years. *Ex post*, this is not entirely surprising given the tendency of BM to suffer from confounding bias. It is plausible that confounding problems caused BM's poor performance in the concurrent validity test.

NA, HR and BTM all exhibit some degree of discriminant validity. However, based on the number of significant correlations with efficiency, NA appears to have the greatest discriminant validity. HR and BM exhibit somewhat higher correlation with efficiency and are assigned ranks of two and three respectively. KW performs poorly in the tests, exhibiting a significant correlation with efficiency in the majority of years (seven). This result suggests that KW does not possess discriminant validity; as such it is assigned a rank of four.

Based on the preliminary ranks it appears that NA and HR appear to have the greatest degree of construct validity, with NA achieving the highest rank. KW and BM are the worst performers, reflecting their lack of discriminant and concurrent validity respectively.

The tentative rankings assigned for predictive validity confirm NA's superior ranking. Although HR's average rank falls in response to its performance in the predictive validity tests, however, it maintains its lead over BM and KW. The predictive validity results move BM and KW into equal third place. Therefore, although the tests conducted cannot provide definitive evidence of predictive validity, they appear to confirm the results of concurrent and discriminant validity.

The convergent validity results and trend regressions are used to identify patterns and to provide further insight into the results obtained in the other empirical tests. Given KW's poor discriminant validity, its correlation with the other proxies is not surprising and cannot be interpreted as convergent validity. KW shares a relatively equal correlation with each of the alternative proxies; in this regard the results do not provide further insight into the relationship amongst the proxies.

HR and BM appear to be the most highly correlated measures. This would seem to suggest that these proxies have some convergent validity. However, the magnitude of the correlation coefficients are not particularly high, as such, it cannot be considered definitive evidence of convergent validity. Further, given BM's poor concurrent validity, a strong correlation between BM and HR cannot be considered evidence of the construct validity of these measures. To the contrary, it casts doubt over the validity of the HR proxy.

NA exhibits the lowest correlation with the other measures. Given the poor performance of BM and KW in terms of concurrent and discriminant validity respectively, these results may indeed provide additional confidence in the NA measure. Perhaps the most interesting observation from these results is the low level of correlation between the highest performing proxies, HR and NA.

The divergence between NA and HR is mirrored by the different trends in conservatism detected by the proxies. Whilst NA suggests that conservatism amongst the sample firms is increasing with time, HR would seem to suggest that conservatism is decreasing. Further to this point, BM and KW both indicate that conservatism is increasing, suggesting that HR may be capturing something different from the alternative proxies. However, as only NA exhibits a statistically significant trend, this evidence is cannot be considered more than circumstantial.

In sum, the evidence provided in this paper appears to suggest that NA may possess the greatest degree of construct validity. Although, HR is a close second based on preliminary rankings it appears to be strongly correlated with BM, a proxy which does not possess concurrent validity. Further, HR appears to be separate from NA, exhibiting a low correlation and potentially indicating a different trend in conservatism amongst the sample firms.

The superior ranking of NA over HR is emphasised by two further considerations. First, compared to NA, HR is not easy to implement. It requires several hard to find data items, generally resulting in significant sample attrition. Second, it can be argued that HR captures the level of conservatism due to the mandatory nature of accounting standards, not discretionary conservatism (Artiach 2009).

BM's poor concurrent validity and KW's poor discriminant validity may indicate that they possess a much lower degree of construct validity compared to the alternative proxies. As such, the results suggest that they may not reliable proxies, and highlight the need for caution in the interpretation of research which uses these proxies.

Factor Analysis

Given the degree of inconsistent evident in the proxies, factor analysis is undertaken to determine if the proxies converge on a uni-dimensional underlying factor to try to establish if there is a commonality between the proxies. The results of the factor analysis indicated that the four proxies do not have a uni-dimension underlying the variable. This suggests there is no single commonality between these proxies indicative of an underlying construct.[28]

On balance however, as previously stated, there can be no unambiguous winner in the tests of construct validity conducted. What can be definitively concluded is that the alternative proxies are not capturing the same thing. Whether the proxies are capturing different constructs of conservatism or if they simply fail to accurately gauge conservatism is unknown.

7. Conclusion

Conservatism's influence on accounting theory and practice is enduring. It enjoys a prominent role in the accounting literature and has emerged as one of the most intensely researched topics in the accounting literature over the last decade (Guay and Verrecchia 2007). Such research relies heavily on

[28] The factor analysis reveals the following statistics for each of the proxies: NA coefficient 0.002, *p*-value 0.919, HR -0.012, *p*-value 0.920; BM 0.059, *p*-value 0.919, KW 1.008, *p*-value 0.919, Goodness of Fit Index 0.998, Chi-square 15.529 (<0.001).

conservatism proxies. The purpose of this paper is to provide a comprehensive empirical evaluation regarding the construct validity of four firm-specific conservatism proxies that have emerged in the extant literature. Four dimensions of construct validity are examined; concurrent, discriminant, convergent and predictive validity. In the absence of a definitive benchmark for accounting conservatism, it is not possible to make robust conclusions regarding the construct validity of the alternative proxies. As such, the evidence provided is preliminary.

Despite the variety of proxies available, to date there has been a deficiency of studies which examine the validity of the alternative proxies. As such, there is little guidance available to the researcher regarding their choice of conservatism proxy. The research is left with the choice to use multiple proxies or justify based on their particular research objective the adopted proxies. Without evidence regarding construct validity, the researcher's decision can be neither theoretically, empirically or methodology guided. By providing evidence regarding the efficacy of the alternative conservatism proxies this paper aims to provide guidance to researchers and potentially drive more insightful research into the important topic of accounting conservatism. Further, by examining the relationship between the conservatism proxies, this paper aims to provide insight into whether the proxies are capturing one or varying constructs of conservatism and hence justify or otherwise the common practice of use of multiple proxies.

The findings suggest that the firm-specific proxies examined have varying degrees of construct validity. Based on the evidence provided by the empirical tests conducted, it appears that NA exhibits the highest degree of construct validity. HR performs relatively well in the initial testing, but falls short of NA in the subsequent examinations. BM and KW appear to have the lowest level of validity, exhibiting limited concurrent and discriminant validity respectively.

The one definitive conclusion which can be drawn from the results is that the alternative proxies available to the researcher are not measuring the same construct. They are not all one. Therefore, the proxies cannot simply be considered as substitutes. This in itself is an important conclusion. Whether the proxies fail to capture conservatism or are capturing varying constructs of conservatism is unclear. Unlike other areas of accounting research, conservatism does not have definitive benchmark with which to gauge the efficacy of the proxies. In the absence of such a benchmark, it is not possible to definitely identify which of the four proxies possess the greatest degree of construct validity. Specifically, the tests of predictive and convergent validity are not unambiguous and should be interpreted with caution. Only a resolution on the definition of conservatism will assist in determining which proxy is the strongest measure of accounting conservatism.

This paper contributes to the literature primarily by conducting a systematic empirical examination of the construct validity of the four firm-specific conservatism proxies. This sheds light on the relationship amongst the various alternative proxies and the potential existence of multiple constructs of

conservatism. In doing so, this may provide guidance to the researcher in their choice of conservatism proxy.

Given the complex nature of this investigation, like Wang *et al.* (2009) this leaves many unanswered questions regarding the validity of conservatism proxies. As such, this provides fertile ground for further research into the construct validity of conservatism proxies. Research which makes progress on defining the precise nature and definition of conservatism would contribute to a further understanding of the validity of conservatism proxies by providing a more definitive benchmark with which to judge the alternatives. Further, such a process could result in the design of new proxies with greater construct validity. Given the inferences made by studies of accounting conservatism, the identification of a strong construct would seem a worthwhile research endeavour.

Reference List

Ahmed, A. S., & Duellman, S. (2007). Accounting conservatism and board of director characteristics: an empirical analysis. *Journal of Accounting and Economics , 43* (2/3), 411-437.

Ahmed, A. S., & Duellman, S. (2010). Evidence on the role of accounting conservatism in monitoring manager's investment decisions. *Accounting and Finance , Forthcoming*.

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

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

Artiach, T. C. (2009). The effect of financial reporting conservatism and disclosure on the cost of equity capital. PhD Thesis, The University of Queensland, Brisbane.

Artiach, T. C., & Clarkson, P. M. (2011). Disclosure, conservatism, and the cost of equity capital: a review of the foundation literature. *Accounting and Finance, 51(1) 2-49*.

Austin, B., & Martin, D. E. (2010). Validation of the moral competency inventory management instrument; content, construct, convergent and discriminant approaches. *Management Research Review , 33* (5), 437-451.

Ball, R., & Shivakumar, L. (2005). Earnings quality in UK private firms: comparative loss recognition timeliness. *Journal of Accounting and Economics , 39* (1), 83-128.

Basu, S. (1995). Conservatism and the asymmetric timeliness of earnings. PhD Thesis, University of Rochester, New York.

Basu, S. (1997). The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics , 24* (1), 3-37.

Basu, S., Hwang, L. S., & Jan, C. (2001). Differences in conservatism between big eight and non-big eight auditors. Working Paper, Baruch College and California State University-Hayward.

Beaver, W. H., & Ryan, S. G. (2000). Biases and lags in book value and their effects on the ability of the book-to-market ratio to predict book return on equity. *Journal of Accounting Research , 38* (1), 127-148.

Beaver, W. H., & Ryan, S. G. (2005). Conditional and Unconditional Conservatism: Concepts and Modelling. *Review of Accounting Studies , 10* (2/3), 269-309.

Bliss, J. H. (1924). *Management through accounts.* New York, NY: The Ronald Press Co.

Botosan, C. A., & Plumlee, M. A. (2005). Assessing alternative proxies for expected risk premium. *The Accounting Review , 80* (1), 21-53.

Bubblitz, B., & Ettredge, M. (1989). The information in discretionary outlays: advertising, research, and development. *The Accounting Revew , 64* (January), 108-124.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design & analysis issues for field settings.* USA: Houghton Mifflin Company.

Dechow, P.M. and Dechiv I.D. (2002). The quality of accruals and earnings: the role of acrual estimation errors, *The Accounting Review*, 77, 35-59.

Dietrich, J. R., Muller III, K. A., & Riedl, E. J. (2007). Asymmetric timeliness tests of accounting conservatism. *Review of Accounting Studies , 12* (1), 95-124.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy , 81*, 607-636.

Feltham, G., and J. Ohlson, (1995). Valuation and clean surplus accoutning for operating and finanical activities, *Contemporary Accounting Research*, 11(2), 689-731.

Financial Accounting Standards Board. (1980). *Statement of Financial Accounting Concepts. No. 2. Qualitative Characteristics of Accounting Information.* SFAC 2: 1980. Financial Accounting Standards Board, USA.

Francis, J., LaFond, R., Olsson, P. M., & Schipper, K. (2004). Cost of equity and earnings attributes. *The Accounting Review , 79* (4), 967-1010.

Givoly, D., & Hayn, C. (2000). The changing time-series properties of earnings, cash flows and accruals: has financial reporting become more conservative? *Journal of Accounting and Economics , 29* (3), 287-320.

Givoly, D., Hayn, C., & Natarajan, A. (2007). Measuring reporting conservatism. *The Accounting Review , 82* (1), 65-106.

Gjesdal, F. (1999). A steady state growth valuation model: a note on accounting and valuation. Working paper, Norwegian School of Economics, Bergen, Norway.

Guay, W., & Verrecchia, R. E. (2007). Conservative disclosure. Working Paper, University of Pennsylvania.

Gujarati, D. N. (1995). *Basic Econometrics* (3rd ed.). Sydney: McGraw-Hill Inc.

Hall, B. (1993). Industrial research during the 1980s: did the rate of return fall? *Brookings Papers on Economic Activity, Microeconomics* , 289-344.

Hribar, P., & Collins, D. W. (2002). Errors in estimating accruals: implications for empirical research. *Journal of Accounting Research , 40* (1), 105-134.

IASB. (1989). Framework for the preperation and presentation of financial statements.

IASB. (2009). Framework for the preperation and presentation of financial statements.

International Accounting Standards Board and Financial Accounting Standards Board. (2008). *Exposure Draft No: 1570-100, Conceptual Framework for Financial Reporting: The Objective of Financial Reporting and Qualitative Characteristics and Constraints of Decision-Useful Financial Reporting Information.* ED1570-100:2008.

Khan, M., & Watts, R. L. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of Accounting and Economics , 48* (2/3), 132-150.

Krishnan, V. G. (2005). The association between Big 6 auditor industry expertise and the asymmetric timeliness of earnings. *Journal of Accounting, Auditing and Finance , 20* (3), 209-228.

Kwok, W. C., & Sharp, D. J. (1998). A review of construct measurement issues in behavioural accounting research. *Journal of Accounting Literature , 17*, 137-174.

Lara, J. M., Osma, B. G., & Penalva, F. (2009). Accounting conservatism and corporate governance. *Review of Accounting Studies , 14* (1), 161-201.

Lara, J. M., Osma, B. G., & Penalva, F. (2009). The economic determinants of conditional conservatism. *Journal of Business Finance and Accounting , 36* (3-4), 327-336.

Lev, B., & Sougiannis, T. (1996). The capitalisation, amortisation, and value-relevance of R&D. *Journal of Accounting and Economics , 21*, 107-138.

Pae, J. (2007). Unexpected accruals and conditional accounting conservatism. *Journal of Business Finance & Accounting , 34* (5-6), 681-704.

Patatoukas, P., & Thomas, J. (2009). Evidence of conditional conservatism: fact or artefact? Working Paper, Yale University.

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

Ryan, S. G., (2006). Identifying conditional conservatism. *European Accounting Review*, 15(4), 511-525.

Wang, R. Z., Ó hÓgartaigh, C., & van Zijl, T. (2009). Measures of accounting conservatism: a construct validity approach. *Journal of Accounting Literature , 28*, 165-203.

Watts, R. L. (2003a). Conservatism in accounting part I: explanations and implications. *Accounting Horizons , 17* (3), 207-221.

Watts, R. L. (2003b). Conservatism in accounting part II: evidence and research opportunities. *Accounting Horizons , 17* (4), 287-301.

Watts, R. L. (1977). Corporate financial statements, a product of the market and political process. *Australian Journal of Management , 2* (April), 53-75.

Watts, R. L., & Zimmerman, J. L. (1986). *Positive accounting theory.* Sydney: Prentice Hall of Australia Pty. Limited.

Zmijewski, M., & Hagerman, R. (1981). An income strategy approach to the positive theory of accounting standard setting/choice. *Journal of Accounting and Economics , 3*, 129-149.

Table 1: Final Sample by Industry and Year

GICS Industry	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Σ	Σ%
Energy	9	13	21	22	23	17	19	24	22	21	23	214	5.8
Materials	21	20	26	26	27	25	28	33	33	33	31	303	8.1
Capital Goods	33	33	40	46	46	48	54	56	57	55	48	516	13.9
Commercial & Professional Services	12	14	19	22	18	21	20	16	14	20	23	199	5.3
Transportation	3	1	2	4	4	5	6	6	7	8	7	53	1.4
Automobiles	3	4	2	3	5	3	4	7	7	7	8	53	1.4
Consumer Durables	16	18	29	27	33	33	37	26	24	25	24	292	7.8
Consumer Services	5	5	6	3	6	7	7	6	6	10	9	70	1.9
Media	3	3	3	3	6	5	9	9	10	7	7	65	1.7
Retailing	6	5	12	10	13	18	20	22	22	22	22	172	4.6
Food & Staples Retailing	1	1	4	4	5	5	5	3	3	4	3	38	1.0
Food, Beverage & Tobacco	10	8	9	9	9	10	13	10	12	14	12	116	3.1
Household & Personal Products	8	8	8	8	9	10	10	9	9	12	10	101	2.7
Health Care	11	12	21	22	27	33	34	37	35	36	32	300	8.1
Pharmaceuticals, Biotechnology	10	10	13	19	18	20	26	26	29	35	40	246	6.6
Banks	-	-	-	-	-	-	-	-	-	-	-	0	0.0
Diversified Financials	-	1	1	-	1	1	1	1	1	2	1	10	0.3
Insurance	2	2	2	1	1	1	1	1	1	-	-	12	0.3
Real Estate	3	3	3	3	3	3	3	4	4	3	5	37	1.0
Software & Services	9	8	14	13	12	17	22	24	26	25	24	194	5.2
Technology Hardware & Equipment	39	33	40	37	39	45	53	55	53	50	53	497	13.4
Semiconductors & Equipment	7	7	14	13	14	14	19	22	27	27	23	187	5.0
Telecommunication	-	-	1	-	2	1	1	3	5	6	4	23	0.6
Utilities	1	1	1	1	1	2	3	3	3	3	3	22	0.6
Total	**212**	**210**	**291**	**296**	**322**	**344**	**395**	**403**	**410**	**425**	**412**	**3,720**	**100.0**
Total %	**5.7**	**5.6**	**7.8**	**8.0**	**8.7**	**9.2**	**10.6**	**10.8**	**11.0**	**11.4**	**11.1**	**100.0**	

Table 1 provides the final sample break-down by industry and year with totals reported by number and percentage.

Table 2: Descriptive Statistics - Firm Characteristics (Pooled Sample)

Variable	Size ($M)	Ln(Size)	ROA	BTM	LEV	Volatility	AUD	Sales/Emp	NA	HR	BM	KW
Mean	3317.090	5.828	0.030	0.615	0.130	0.032	0.802	311.196	-0.016	0.225	-0.281	0.158
Median	335.806	5.817	0.051	0.490	0.080	0.028	1.000	195.914	-0.016	0.076	-0.688	0.115
Std Dev	15098.730	2.119	0.127	0.462	0.148	0.018	0.399	503.271	0.072	1.743	1.955	0.398
Min	1.101	0.096	-0.867	0.000	0.000	0.006	0.000	2.000	-0.321	-25.875	-7.718	-1.526
Max	467092.880	13.054	0.610	3.676	0.905	0.222	1.000	17210.620	0.297	19.243	23.043	7.666
Skew	14.657	0.233	-2.639	1.997	1.205	2.384	-1.513	14.354	0.289	-5.036	3.974	8.171
Kurt	332.571	-0.350	12.228	5.876	1.060	10.620	0.289	397.903	1.714	89.352	24.969	116.746

Table 2 provides the descriptive statistics for common firm characteristics and conservatism proxies computed using the sample truncated at 1% for the pooled sample of 3366 firm-year observations for the period 1997 to 2007. Size ($m) is the market value of common equity in $millions. Ln(Size) is the natural log of the market value of common equity. ROA is return on assets measured as net income before extraordinary items divided by total assets. BM is the book-to-market ratio calculated as the book value of common equity divided by the market value of common equity. LEV is leverage measured as long-term debt divided by total assets. Volatility is the standard deviation of daily stock returns. AUD is a dichotomous variable which equals 1 if the firm uses a first-tier auditor, 0 otherwise. Sales/Emp is the sales-to-employee ratio calculated as total sales divided by the number of employees. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure. The top and bottom 1% of distribution have been truncated

Table 3: Results for the Test of Concurrent Validity by Year

1997

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	-0.029	-0.034	+	0.005	0.359	0.720
HR	0.263	0.130	+	0.134	1.768	0.081
BM	-0.921	-0.554	+	-0.367	-4.966	0.000
KW	-0.200	-0.067	+	-0.133	-2.006	0.048

1998

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	-0.041	-0.032	+	-0.009	-0.575	0.567
HR	0.141	0.098	+	0.043	0.105	0.917
BM	-0.969	-0.684	+	-0.285	-3.563	0.001
KW	0.096	0.028	+	0.068	1.399	0.166

1999

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	-0.032	-0.033	+	0.001	0.085	0.932
HR	0.616	0.119	+	0.497	3.802	0.000
BM	-1.027	-0.715	+	-0.311	-3.759	0.000
KW	0.485	0.186	+	0.299	6.029	0.000

2000

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	-0.053	-0.034	+	-0.019	-1.584	0.116
HR	0.642	0.132	+	0.511	2.094	0.039
BM	-1.044	-0.764	+	-0.280	-3.714	0.000
KW	0.126	0.034	+	0.091	6.518	0.000

2001

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	-0.022	-0.036	+	0.015	1.020	0.310
HR	0.852	0.102	+	0.750	2.108	0.037
BM	-0.919	-0.783	+	-0.136	-1.767	0.080
KW	-0.077	0.209	+	-0.286	-9.246	0.000

2002

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.023	-0.017	+	0.040	3.483	0.001
HR	0.953	0.134	+	0.819	4.085	0.000
BM	-1.136	-0.769	+	-0.368	-3.862	0.000
KW	0.365	0.134	+	0.231	5.434	0.000

2003

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.032	-0.008	+	0.040	3.089	0.002
HR	0.294	0.137	+	0.157	0.435	0.664
BM	-1.098	-0.701	+	-0.397	-4.102	0.000
KW	0.995	0.157	+	0.838	4.722	0.000

2004

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.025	-0.012	+	0.037	3.136	0.002
HR	0.835	0.117	+	0.718	2.827	0.005
BM	-1.757	-1.172	+	-0.585	-3.730	0.000
KW	0.180	0.101	+	0.078	9.725	0.000

2005

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.027	-0.010	+	0.037	3.838	0.000
HR	0.247	0.070	+	0.177	0.788	0.432
BM	1.224	0.746	+	0.478	2.082	0.039
KW	0.295	0.056	+	0.239	12.099	0.000

2006

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.018	-0.007	+	0.025	2.196	0.030
HR	-0.346	0.133	+	-0.480	-0.908	0.365
BM	3.280	1.543	+	1.736	3.192	0.002
KW	0.113	-0.007	+	0.119	10.544	0.000

2007

Variable	High	Low	Pred Sign	Mean Diff	*t*-statistic	*p*-value
NA	0.018	-0.012	+	0.030	2.547	0.012
HR	-0.223	0.147	+	-0.371	-0.676	0.500
BM	-0.699	-0.571	+	-0.128	-2.386	0.018
KW	0.268	0.092	+	0.176	9.004	0.000

Table 3 provides the results of the *t*-tests for difference in means by year described in Section 3.1. High and Low are the mean conservatism score for the 'high' and 'low' uncertainty samples respectively. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure. The *t*-statistics and *p*-values using a two-tailed test of significance are reported.

Table 4: Results for the Tests of Discriminant Validity Correlations by Year

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	# Sig @ 10%
NA	0.044	0.012	0.017	-0.073	**-0.115**	0.018	0.001	-0.036	-0.085	-0.002	**-0.134**	2
	(0.541)	*(0.871)*	*(0.776)*	*(0.237)*	***(0.051)***	*(0.755)*	*(0.985)*	*(0.496)*	*(0.106)*	*(0.963)*	***(0.010)***	
	0.030	0.056	0.055	**0.101**	-0.024	0.079	0.018	-0.082	-0.036	0.018	-0.074	**1**
	(0.684)	*(0.441)*	*(0.364)*	***(0.097)***	*(0.689)*	*(0.162)*	*(0.739)*	*(0.118)*	*(0.488)*	*(0.726)*	*(0.152)*	
HR	0.091	0.012	-0.052	-0.028	0.034	-0.059	-0.028	-0.043	0.000	0.006	-0.041	0
	(0.209)	*(0.864)*	*(0.393)*	*(0.653)*	*(0.565)*	*(0.293)*	*(0.594)*	*(0.413)*	*(0.999)*	*(0.910)*	*(0.429)*	
	0.115	0.048	0.058	0.067	0.011	0.067	0.037	0.054	**0.119**	**0.110**	0.037	**2**
	(0.110)	*(0.507)*	*(0.341)*	*(0.278)*	*(0.857)*	*(0.236)*	*(0.486)*	*(0.310)*	***(0.023)***	***(0.032)***	*(0.470)*	
BM	**-0.153**	-0.111	-0.026	0.018	-0.010	0.011	-0.052	0.031	**-0.087**	**-0.111**	0.033	3
	(0.034)	*(0.124)*	*(0.671)*	*(0.765)*	*(0.862)*	*(0.851)*	*(0.327)*	*(0.556)*	***(0.097)***	***(0.031)***	*(0.525)*	
	-0.182	**-0.148**	**-0.123**	-0.060	0.014	-0.022	-0.060	0.061	-0.037	-0.072	-0.056	**3**
	(0.011)	***(0.041)***	***(0.043)***	*(0.324)*	*(0.816)*	*(0.698)*	*(0.262)*	*(0.249)*	*(0.476)*	*(0.162)*	*(0.276)*	
KW	-0.044	-0.055	**-0.106**	**-0.111**	0.075	-0.061	0.068	0.061	**-0.158**	**-0.116**	**-0.116**	5
	(0.546)	*(0.450)*	***(0.082)***	***(0.071)***	*(0.204)*	*(0.281)*	*(0.200)*	*(0.248)*	***(0.003)***	***(0.023)***	***(0.025)***	
	-0.075	**-0.137**	-0.087	**-0.162**	0.081	**-0.152**	0.026	**-0.193**	**-0.213**	**-0.166**	**-0.188**	**7**
	(0.302)	***(0.057)***	*(0.154)*	***(0.008***	*(0.168)*	***(0.007)***	*(0.631)*	***(<0.001)***	***(<0.001)***	***(0.001)***	***(<0.001)***	

Table 4 shows the correlation between the conservatism proxies and the sales-to-employee ratio (calculated as total sales divided by number of employees) by year. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure. Spearman's correlation coefficients are shaded. Pearson's correlation coefficients are unshaded. *p*-values are provided in parenthesis. Correlations significant at the 10% level are in bold. The rightmost column gives the number of years in which correlations are significant at the 10% level.

Table 5: Results for the Tests of Convergent Validity Correlations by Year

1997

Variable	NA	HR	BM	KW
NA		-0.093	-0.104	**-0.189**
		(0.197)	*(0.149)*	***(0.008)***
HR	0.053		-0.020	0.068
	(0.466)		*(0.787)*	*(0.349)*
BM	**-0.191**	0.078		0.118
	(0.008)	*(0.283)*		*(0.102)*
KW	**-0.132**	0.078	**0.168**	
	(0.068)	*(0.280)*	***(0.019)***	

1998

Variable	NA	HR	BM	KW
NA		-0.068	-0.075	-0.014
		(0.349)	*(0.301)*	*(0.846)*
HR	0.026		0.054	0.225
	(0.716)		*(0.459)*	*(0.002)*
BM	**-0.203**	0.049		-0.033
	(0.005)	*(0.497)*		*(0.648)*
KW	-0.091	0.021	0.088	
	(0.210)	*(0.772)*	*(0.223)*	

1999

	NA	HR	BM	KW
NA		-0.058	-0.094	-0.037
		(0.346)	*(0.123)*	*(0.547)*
HR	**0.178**		**0.157**	**0.262**
	(0.003)		***(0.010)***	***(<0.001)***
BM	**-0.164**	**0.119**		-0.038
	(0.007)	***(0.050)***		*(0.529)*
KW	0.038	**0.223**	0.041	
	(0.537)	***(<0.001)***	*(0.503)*	

2000

	NA	HR	BM	KW
NA		-0.079	-0.014	-0.101
		(0.198)	*(0.816)*	*(0.100)*
HR	0.096		**-0.242**	0.028
	(0.117)		***(<0.001)***	*(0.645)*
BM	-0.058	**0.182**		**0.326**
	(0.343)	***(0.003)***		***(<0.001)***
KW	-0.065	0.079	**-0.331**	
	(0.285)	*(0.195)*	***(<0.001)***	

2001

	NA	HR	BM	KW
NA		-0.051	-0.044	-0.060
		(0.388)	*(0.452)*	*(0.307)*
HR	0.007		**0.154**	**0.216**
	(0.900)		***(0.009)***	***(<0.001)***
BM	-0.043	**0.108**		**0.363**
	(0.463)	***(0.067)***		***(<0.001)***
KW	-0.077	0.050	**0.273**	
	(0.190)	*(0.392)*	***(<0.001)***	

2002

	NA	HR	BM	KW
NA		**0.098**	-0.027	0.039
		(0.083)	*(0.638)*	*(0.488)*
HR	0.052		**0.182**	-0.102
	(0.360)		***(0.001)***	*0.069*
BM	-0.027	**0.116**		**-0.516**
	(0.636)	***(0.040)***		***(<0.001)***
KW	0.068	-0.048	**-0.454**	
	(0.225)	*(0.400)*	***(<0.001)***	

2003

	NA	HR	BM	KW
NA		0.035	-0.017	**0.125**
		(0.513)	*(0.753)*	***(0.019)***
HR	-0.031		**0.154**	**-0.190**
	(0.564)		***(0.004)***	***(<0.001)***
BM	-0.016	**0.129**		-0.021
	(0.759)	***(0.015)***		*(0.693)*
KW	**0.093**	-0.074	-0.029	
	(0.080)	*(0.164)*	*(0.589)*	

2004

	NA	HR	BM	KW
NA		0.034	**0.099**	**0.130**
		(0.519)	***(0.059)***	***(0.013)***
HR	0.050		0.056	**-0.121**
	(0.348)		*(0.287)*	***(0.022)***
BM	0.042	0.030		0.002
	(0.424)	*(0.574)*		*(0.973)*
KW	**0.157**	0.076	0.028	
	(0.003)	*(0.151)*	*(0.599)*	

2005

	NA	HR	BM	KW
NA		**0.096**	**-0.192**	**0.152**
		(0.068)	***(<0.001)***	***(0.004)***
HR	0.057		**0.100**	-0.082
	(0.278)		***(0.056)***	*(0.119)*
BM	**-0.125**	-0.014		**-0.306**
	(0.017)	*(0.794)*		***(<0.001)***
KW	**0.168**	0.029	**-0.243**	
	(0.001)	*(0.579)*	***(<0.001)***	

2006

	NA	HR	BM	KW
NA		0.006	**-0.199**	0.057
		(0.914)	***(<0.001)***	*(0.269)*
HR	-0.022		0.040	**0.150**
	(0.670)		*(0.437)*	***(0.003)***
BM	**-0.138**	**-0.132**		**0.110**
	(0.007)	***(0.010)***		***(0.032)***
KW	0.101	0.022	**0.133**	
	(0.049)	*(0.668)*	***(0.010)***	

2007

	NA	HR	BM	KW
NA		0.070	**0.095**	0.045
		(0.177)	***(0.067)***	*(0.382)*
HR	-0.023		**0.165**	**0.102**
	(0.653)		***(0.001)***	***(0.049)***
BM	0.020	-0.066		-0.042
	(0.694)	*(0.200)*		*(0.423)*
KW	**0.148**	**-0.137**	-0.006	
	(0.004)	***(0.008)***	*(0.911)*	

Table 5 provides the annual correlation matrices in full. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure. Spearman (Pearsons) correlation coefficients are above (below) the diagonal. *p*-values are provided in parenthesis. Significant at the 10% level are in bold.

Table 6: Results for the Tests of Predictive Validity

Panel A: Fama-Macbeth Results					Panel B: Pooled Regression Results				
	Size(+)	**LEV (+)**	**ROA (+)**	**AUD (+)**		**Size(+)**	**LEV (+)**	**ROA (+)**	**AUD (+)**
NA					**NA**				
Mean Coefficient	-0.002***	0.040***	-0.094***	-0.005	Coefficient	-0.003***	-0.004	-0.065***	-0.003***
t-statistic	*-3.350*	*5.030*	*-6.460*	*-1.670*	*p*-value	*(<0.001)*	*(0.583)*	*(<0.001)*	*(0.378)*
HR					**HR**				
Mean Coefficient	0.019	-0.353	-0.820	-0.046	Coefficient	0.007	-0.421*	-0.773***	-0.032
t-statistic	*1.090*	*-1.000*	*-0.110*	*-0.460*	*p*-value	*(0.758)*	*(0.059)*	*(0.002)*	*(0.733)*
BM					**BM**				
Mean Coefficient	0.083***	-0.170	0.318	-0.056	Coefficient	0.083***	0.598**	0.384	-0.157
t-statistic	*5.430*	*-1.370*	*0.970*	*1.010*	*p*-value	*(<0.001)*	*(0.014)*	*(0.153)*	*(0.121)*
KW					**KW**				
Mean Coefficient	-0.037	0.169	-0.121	-0.021*	Coefficient	0.009	0.072	-0.137**	-0.001
t-statistic	*-1.620*	*0.370*	*-1.460*	*-2.090*	*p*-value	*(0.144)*	*(0.253)*	*(0.048)*	*(0.974)*

Table 6 panel A provides the average coefficients and Fama and Macbeth (1973) *t*-statistics obtained using the annual cross-sectional regression $CON_{it} = \alpha_0 + \alpha_1 SIZE_{it} + \alpha_2 LEV_{it} + \alpha_3 ROA_{it} + \alpha_4 AUD_{it} + \varepsilon_{it}$. Size is the natural log of the market value of common equity. LEV is leverage measured as long-term debt divided by total assets. ROA is return on assets measured as net income before extraordinary items divided by total assets. AUD is a dichotomous variable which equals 1 if the firm uses a first-tier auditor, 0 otherwise. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure. ***, ** and * indicates significance at the 1%, 5% and 10% levels respectively. Panel B provides the pooled regression. *p* -values are provided in parenthesis. Intercept, firm and year fixed effects included but not reported.

Table 7: Summary of the Results of the Tests of Construct Validity

Criteria Assessed	Description	Empirical Test	NA	HR	BM	KW
Concurrent Validity	The proxy is able to distinguish between 'high' and 'low' uncertainty firms.	*t*-test for the difference in mean conservatism score for firms with 'high' and 'low' uncertainty.	2	2	4	1
Discriminant Validity	The proxy is not correlated with efficiency (an unrelated construct).	Correlations between the proxies and the sales-to-employee ratio.	1	2	3	4
Average Rank			1.5	2	3.5	2.5
Predictive Validity	The proxy is related to firm characteristics which are known to be associated with conservatism.	Annual cross-sectional regressions with Fama-Macbeth (1973) *t*-statistics and a pooled regression.	1	3	1	3
Average Rank			1.33	2.33	2.67	2.67
Convergent Validity	The proxies give results which are correlated with each other.	Correlation between the conservatism score indicated by each of the proxies.	NA shows the lowest level of correlation with the other proxies.	HR and BM show the strongest correlation.		KW appears correlated with the other proxies equally.
Trend	What trend in conservatism do the proxies detect?	Time regressed on the mean values of each conservatism proxy.	Increasing and Significant.	Decreasing but not significant.	Increasing but not significant.	Increasing but not significant.

Table 7 provides the summary of the results of construct validity detailed in section 6Five. NA is the Givoly and Hayn (2000) negative accruals conservatism proxy. HR is the Penman and Zhang (2002) Hidden-Reserves measure. BM is the bias component of the BTM ratio estimated in accordance with Beaver and Ryan (2000). KW is the Khan and Watts (2009) firm-year AT measure.