Using Machine Learning to Measure Conservatism

Jeremy Bertomeu, Edwige Cheynel, Yifei Liao, Mario Milone[*]

October 25, 2021

Abstract

Machine learning can help improve empirical proxies of conservatism by detecting patterns consistent with the data beyond the standard linear regression techniques assumed in the prior literature. Using a neural network approach, we demonstrate that the fit of differential timeliness almost doubles incorporating non-linearities and interactions absent in prior literature. The model offers the promise of reducing noise in measurements and designs more powerful tests to assess theories of conservatism. We further show that the machine-learning measures exhibit (a) fewer economically anomalous observations, (b) have economic associations consistent with the literature, and (c) less unexplained year-over-year instability. The measure reveals expected trends toward a secular decline in conservatism in the US. In simulations, we show that, while existing measures perform honorably even in the presence of a complex data-generating process that they were not designed to capture, proxies based on machine learning methods are the most robust to misspecification, feature less attenuation bias, and reduce the incidence of false negatives and positives.

Keywords: machine learning, neural network, accounting, conservatism, measure, proxy.

JEL Codes: C1, D2, M4.

Conservatism is traditionally understood as a differential response of the accounting process to losses versus gains, as manifested by a greater level of certainty required to record positive news (Watts 2003). The asymmetry is reflected empirically by a structural break in the properties of accounting numbers around favorable transactions. An extensive literature builds on this intuition to develop measures of conservatism from the differential timeliness of price and earnings, see, e.g., Basu (1997), Kothari et al. (2010) and Ball et al. (2013a,b). In this literature, the statistical models are assumed to be linear in all aspects other than differential timeliness, and fail to account for the richness of non-linearities in the identification of conservatism.

This study offers an exploratory analysis of machine learning techniques to measure conservatism when lifting functional forms restrictions plausibly unrelated to conservatism. We propose and analyze an extension of Basu (1997) and Khan and Watts (2009) in which observable characteristics related to conservatism as well as controls and their interactions need not follow a linear relation. While the current application is on the measurement of conservatism, the objective of this study is broader and suggests to incorporate algorithms from machine learning to filter noise in conventional archival studies and unlock more complex patterns across variables. Specifically, we develop measures of machine-learning conservatism (hereafter, MLC), cross-validate hyper-parameters out-of-sample for use in future research, and discuss pros and cons relative to other approaches.

To evaluate the usefulness of machine learning to improve empirical measurements, we also need to be aware of significant drawbacks. Relative to traditional statistical models, machine learning algorithms are more difficult to interpret because the model does not have a simple closed-form expression. The potential benefits of the approach must be sufficient to outweigh the costs of the limited interpretability. To this effect, we divide the objective of our study into four distinct benefits.

The first and most obvious benefit of machine learning algorithms is in terms of fit (Hastie et al. 2009), namely, whether the model captures enough of the total variation in the sample to (a) use as much as possible of the information contained in a sample about conservatism, and (b) reduce noise components, which in uses of the measure to test economic theories, could reflect unobservables endogenous to the theory being tested. A better fit does not logically imply a better measure but it alleviates potential concerns that a measure may be capturing second-order effects relative to residuals. A close cousin to a better fit is the ability of the model to explain data

with fewer variables ("data reduction"), i.e., fit the same variation with a simpler set of economic inputs. Fewer variables achieving the same fit make the interpretability of the model easier relative to estimates with separate regressions (e.g., Fama MacBeth) where the fit may be achieved by unexplained periodic variation in coefficients.

Second, machine learning does not impose any structure on the data-generating process in contrast to a traditional statistical model requiring a researcher-guided specification. Theory on conservatism does not suggest a unique fully-specified empirical specification (Bertomeu et al. 2016), leading to unresolved disagreements about functional forms that can influence the conservatism measure. Machine learning offers a perspective in which any functional form or (to an extent) choice of variables can be selected as a function of the ability of the model to reduce out-of-sample errors. The benefit of the approach is to place restrictions only for components of the model where there are clear theoretical associations but leaving the estimation non-parametric for other components for which the assumed theory is silent (Athey 2015).

Third, a linear model written as a function of a few characteristics, i.e., an index, rarely exhibits enough variation to be useable as-is in a test. Further, the index is a linear combination of other variables which, in turn, bars their inclusion as controls when using the measure as a test variable in a regression. A problem arises when the components of a measure are related to alternative explanations in tests using the measure. To address this issue, many measures are calculated by running regressions by subsamples, e.g., industry-year (Dechow et al. 2010) or, in the case of conservatism, by year (Khan and Watts 2009). This approach presents some empirical challenges because identification occurs from yearly shocks to a measure, which can make it highly unstable. Machine learning offers the advantage, even without using time as a feature, to generate a fit to the data competitive with models run by subgroups. The methods substitute for the need to use separate regressions by fiscal year in linear regressions because they do not assume any linear relations, and are well-suited for non-linear primitive statistical relations.

Lastly, machine learning algorithms can be sufficiently flexible to incorporate restrictions suggested by economic theory, and provide more economic validity for the measure. By contrast, a linear regression usually imposes restrictions on average effects and can lead to anomalous predictions. These anomalous (i.e., known to be wrong or misspecified) estimates can be problematic if they drive a result in a test using the measure. In our context, we impose that the conditional

correlation between earnings and news should be non-negative for any given set of observables. As we shall see, this intuitive restriction is violated for a significant fraction of the sample using firm or firm-year measures.[1]

Our empirical model aims to identify differential timeliness in terms of variation in observed economic proxies. The literature so far regresses earnings on returns and firm characteristics, and further include interaction terms to account for the impact on negative returns to test for the asymmetric timeliness of earnings defined as a measure of accounting conservatism. We expand the set of commonly-used variables plausibly related to conservatism and generalize the specification of these variables into the model so that the functional form need not be linear. We compare the specifications with the standard firm characteristics to the specifications augmented by additional firm characteristics, and further test the effect of relaxing the linearity on improving the model fit. Formally, the direct effect of prices on earnings ("G" function) and the interaction term conditional on negative returns ("C" function) may take any non-linear functional form. We use neural networks that combine any number of variables to encode both G and C functions. The neural networks are then combined to fit the relation between earnings and price. We impose two bounds on these functions: earnings must be non-decreasing in returns over positive and negative returns, implying that, for any observable included in the model, earnings co-move in the direction of the news on average.

Our main result is that an estimate of differential timeliness based on neural networks and eight economic characteristics, "MLC2", performs better at fitting differential timeliness than linear models estimated by year. [2] The measure requires fewer estimated coefficients and also features a lesser degree of overfitting than separate yearly regressions, as measured by comparing differences between in-sample and out-of-sample fit. The fit of an alternative machine-learning measure that includes year fixed effects, "MLC2y", is only modestly improved, which suggests that year shocks captured by firm-year measures can add noise.

---

[1]There are special cases where earnings decrease even though returns are positive: the restriction is imposed here excluding error terms so that an observation could exhibit negative earnings and positive returns (or vice-versa) due to random noise. In addition, while it is possible that, in a certain context, a firm may have systematically very negative earnings and positive returns (i.e., a firm in early growth stage with significant investments in intangibles), these are fairly unusual circumstances that do not apply to the vast majority of firms in the sample and are not captured by variables used in the analysis. In untabulated results, we also conduct the analysis without this restriction and find very similar results.

[2]In Khan and Watts (2009), the firm characteristics included are size, market-to-book and leverage. We augment by including volatility, non-operating accruals, cash flows from operating activities, investment cycle and age.

Given that our MLC measures draw from the same paradigm as prior models, our purpose is not to discredit prior measures with which, by design, we share a similar conceptual framework. But we hope that machine learning can help provide more accurate measurements and better filter noise, especially in settings with no clear consensus because prior measures are insignificant or unstable. In various tests, we show that the machine learning measures feature less unexplained instability between time periods, fewer anomalous economic properties (such as negative conditional associations between earnings and price) but otherwise share associations consistent with well-established results. The measures are therefore intended as an update of prior research useful in settings where previous approaches may not have enough power. Once the model parameters are cross-validated (the main computational step to applying the method), the measures can be easily computed by researchers with standard tools available in all statistical software.

We also offer three novel insights that can be more easily obtained with a non-linear approach. First, the measures more clearly validate a trend toward decreasing conservatism since the eighties, with a current level at its historical lowest - by contrast, this trend is only noisily observed with measures estimated by time because time and economic characteristics are confounded. Second, descriptively, we establish a few new empirical facts: the relation between conservatism and size or market-to-book is convex, while the relation between conservatism and leverage or volatility is concave. Third we revisit the relation between conservatism, investment and cost of capital in Lara et al. (2016). We replicate their main result with both the usual measures and the machine learning measures and find greater economic magnitudes with the MLC measures, consistent with less measurement noise.

We illustrate the theoretical properties of the estimation in simulations. While all measures are positively correlated, the machine learning algorithms are more robust to misspecification than other measures: if the researcher (incorrectly) uses a machine learning measure when the true model is linear or linear with time-varying coefficients, measurement errors are lower than if the researcher (incorrectly) uses a linear model when the true data-generating process is captured by the machine learning estimates. When applying the measure to simulated data in which conservatism causes a change in a variable, a researcher using machine learning is less likely to falsely accept an insignificant coefficient or falsely reject no effect.

In summary, methods from machine learning can be used to develop more accurate proxies

of conservatism based on improvements on prior frameworks. A more accurate proxy can reduce attenuation bias in coefficient estimates and lead to fewer insignificant ("false negatives") results when evaluating the effect of conservatism in small or noisy datasets. It can also reveal greater economic magnitudes when evaluating causal effects. We also heed the call by Jennings et al. (2020) to improve the measurement of accounting proxies given that, as they demonstrate in various simulations and empirical examples, noise in variables used to form controls may also bias toward finding significant coefficient estimates. The measures can be computed on the following site "https://sites.google.com/view/mlcscore/home", which calculates the measure for any user submitted data.

Literature Review. Our study links to a growing literature exploring how computational and theoretical advances in machine learning can improve accounting measurements. This literature observes that many constructs relevant to the accounting process have functional forms that are unknown a-priori to researchers. Studies such as Perols (2011) and Perols et al. (2017) are among the first to introduce machine learning methods in accounting. Recent studies have also examined the use of machine learning for textual analysis (Li 2010; Frankel et al. 2016), to analyze audit quality (Yang et al. 2020; Hu et al. 2020), to compute more accurate accruals (Ding et al. 2020), to examine the evolution of the relevance of accounting (Barth et al. 2019) or to predict SEC enforcement actions (Bao et al. 2020) and misstatements (Bertomeu et al. 2020).[3] This literature focuses on prediction models. We adopt a different perspective and employ machine learning methods to separate signal from noise, and extract an economic characteristic of interest. Our objective is to show how machine learning methods can be nested into commonly-used measures of differential timeliness, and test for the performance of machine-learning measures.[4]

There also exists an extensive literature aiming to measure conservatism which we cannot discuss in its entirety. Our baseline specification draws heavily from the differential timeliness model in Basu (1997) and Khan and Watts (2009). Taking the example of an asset impairment, Basu observes that timely reporting of bad news is reflected into earnings when a sufficient loss is expected; by contrast, gains are delayed until the earnings process is complete, implying a flatter

---

[3]There are many other studies investigating machine learning in accounting and our objective here is not to provide a comprehensive discussion; the recent surveys in Bertomeu (2020) and Minuti-Meza and Krupa (2021) provide a discussions of new perspectives in the accounting literature.

[4]These examples are not intended as a complete list of current work in this area; recent studies also include Jones (2017), Anand et al. (2019), Chen et al. (2020), Lee et al. (2020), and Hunt et al. (2020).

relation between earnings and gain expectations. Conservatism results in a broken symmetry around gains and losses, which can be tested with a linear regression of earnings on returns with an interaction in the form of an indicator variable for negative returns ("bad news"). The original Basu (1997) specification assumes that conservatism does not change over time, while Khan and Watts (2009) assume that the degree of asymmetry may vary within firm. Differential timeliness is a workhorse model to test theories of conservatism, see, e.g., Ball et al. (2000), Givoly and Hayn (2000), Ball et al. (2003), Lafond and Roychowdhury (2008). We develop a measure that retains the broken symmetry around market news but otherwise does not impose a specific parametric form on its components.

There are nevertheless continuing debates about using asymmetric timeliness in the measurement of conservatism (Beaver and Ryan 2005; Ryan 2006; Givoly et al. 2007 and a machine learning algorithm is a statistical tool that does not, on its own, provide a new econometric basis for a measure. Our machine-learning measure should likely be used in conjunction with simpler measures that capture other implications of conservatism. Roychowdhury and Watts (2007) argue market-to-book (M/B) is complementary to asymmetric timeliness in that conservatism need not be a single-dimensional property of measurement. M/B captures cumulative conservatism through systematic under-recognition of accounting assets relative to investor expectations. In addition, income statement and balance sheet conservatism may have different economic implications. For example, if a firm reduces the value of its assets, high balance sheet conservatism may be followed by low income statement conservatism if there is no further scope for impairments (Sunder et al. 2018).[5] The ability of the algorithm to allow for an asymmetric timeliness that varies with more covariates (including M/B) and in a non-linear manner, provides a more comprehensive empirical design to control for biases due to omitted variables and misspecification.

---

[5]There is no standard framework for the empirical measurement of conservatism that nests all approaches. For example, another distinction is that conservatism can be "unconditional," in the form of reduction in value of assets regardless of economic news, versus "conditional," i.e., a reduction that depends on the nature of economic events. From an information economics perspective, unconditional changes in a measure do not change the properties of the signal, so we focus here on definitions that preserve some informational role Gigler et al. (2009). While M/B is based on balance sheet numbers, asymmetric timeliness uses income numbers. The articulation of balance sheet and income statement implies that both measures would be, at least in part, affected by both types of conservatism.

1 Empirical Strategy

1.1 Model

We assume the following non-linear model of earnings and prices:

$$X_i = \underbrace{f(Y_i) + R_i g(Y_i) + R_i D_i c(Y_i) + D_i d(Y_i)}_{=m(Y_i, R_i)} + \epsilon_i, \tag{1.1}$$

where X_i is earnings, R_i is return, D_i is an indicator variable equal to one if $R_i < 0$ and 0 otherwise, Y_i is a vector of explanatory variables that does not include R_i or D_i, and ϵ_i is an i.i.d. error term.[6]

The functions $f(.)$, $g(.)$, $c(.)$ and $d(.)$ have a specified form in the regression models, whereas the machine learning algorithm does not impose any conditions on the shape of the functions. The function $c(.)$ is the measure of conservatism and will be determined in the different models.

In Basu (1997), the model is estimated with a pooled time-series regression where $f(.)$, $g(.)$, $c(.)$ and $d(.)$ are constants, and researchers commonly use firm-level regressions to estimate $c(.)$ by firm (the Basu coefficient). In Khan and Watts (2009), hereafter KW, $f(.)$, $g(.)$, $c(.)$ and $d(.)$ are linear functions of three features: size (total assets), market-to-book (M/B) and leverage (total debt to market value of equity), with the $f(.)$, $g(.)$, $c(.)$ and $d(.)$ constant between-firm within-year and estimated separately by year. These designs are special cases of this specification choosing the features Y_i as a constant, a firm fixed effect or a linear function of characteristics interacted with time fixed effects, respectively. Importantly, note that the function $c(Y_i)$ cannot be directly computed from data because we do not observe it, so a statistical method and assumptions about Y_i are required to estimate it.

Our objective is to estimate equation (1.1) for any potential choice of Y_i and without assuming a specific function form for $g(.)$ and $c(.)$. Given that the model features four separate objects that are combined to obtain the estimate of the function $m(.)$, we use a neural network approach where the functions are created separately using networks and combined into a prediction. While other methods can be used as well, a benefit of neural networks is that they allow to easily encode

---

[6]That errors are i.i.d. is not a realistic assumption, due to persistence in economic shocks or reversals of news that was not incorporated by the conservative system in prior periods. Ball et al. (2013a) offer a discussion of the validity of asymmetric timeliness in the presence of reversals; however, we recognize that their conclusions hold within the linear specification. For this reason, we do not claim that zero asymmetric timeliness should reflect no conservatism but a higher measure should be associated to more conservatism.

some structure into the analysis by building the network structure as a function of the problem examined.[7] We develop next the implementation of the neural network approach and illustrate what we mean by structuring the problem.

1.2 Artificial Neural Networks

Because a researcher-chosen specification of the algorithm is used to encode the economic restrictions from (1.1), it is helpful to obtain a general understanding of how a neural network fits data. Artificial neural networks are general function approximators inspired by biological structures. Figure 1 displays how a neural network method connects the inputs to the output. A network is composed of perceptrons (also called nodes or neurons) connected to each other. Each perceptron is a mathematical function that receives a vector as input and returns a scalar. This scalar is a function of a linear combination of the input vector. If we denote $\boldsymbol{x}$ the input vector, a perceptron returns a scalar

$$z = \Phi(\boldsymbol{w}\boldsymbol{x} + b),$$

where $\boldsymbol{w}$ is a vector of weights, b is a bias term and $\Phi(.)$ is called the activation function. In its simplest form, the activation function is linear and the perceptron returns $z = \boldsymbol{w}\boldsymbol{x} + b$. When connected in a network, the output value of a perceptron becomes a component of the input vector of a different perceptron. A single perceptron with a linear activation is thus equivalent to a linear model.

We estimate equation (1.1) with a custom designed neural network whose structure is illustrated in Figure 2. The input layer of the network is a vector of dimension $n+2$ where the first n inputs correspond to the variables Y_i, and the last two inputs are the return R_i and a dummy variable D_i equal to one when the return $R_i < 0$ is negative. The square nodes correspond to embedded neural networks whose outputs are the values of the functions $f(.)$, $g(.)$, $g(.)+c(.)$, and $d(.)$. Their inputs consist only of the variables Y_i, and exclude the return R_i and dummy D_i. The output layer consists of a single perceptron whose inputs are the values of the functions as well as R_i and D_i. This perceptron does not have any weights or bias that are modified during training. Its only

---

[7]Other algorithms, such as regression trees, are very effective at simpler tasks that involve the prediction of a single object without any assumed structure and can do better than neural networks in an unstructured problem. As we report in further analyses, these tools can be adapted for the estimation of an equation such as (1.1) but, given the assumed structure, lose some of their benefits and do not do better than neural networks.

Figure 1: Structure of a feed-forward neural network.

Circles represent perceptrons. Thick arrows are vectors and thin arrows are scalars.
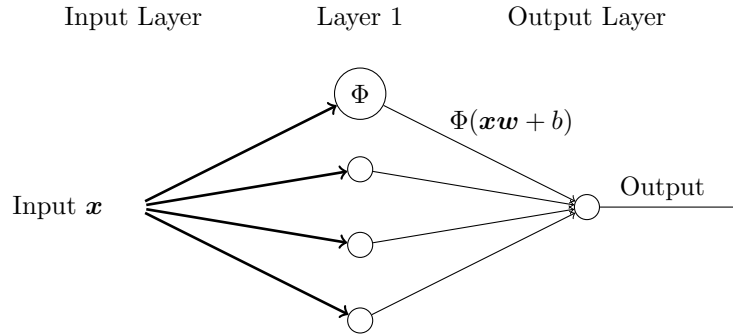
purpose is to compute the function $m(.)$ in Equation (1.1). The trained network is used to provide a measure of machine learning conservatism (the function $c(.)$, hereafter, denoted MLC) that only depends on Y_i.

An advantage of neural networks is that it is easy to impose a model structure on the network by choosing an appropriate network and activation functions. As shown in Figure 2, we write the network to extract the $g(.)$ and $c(.)$ semi-parametrically by imposing the break at $R_i = 0$ in (1.1) that is core to the theory of differential timeliness, without requiring other parametric assumptions. The input layer of the networks encoding functions $f(.)$ and $d(.)$ is a vector of dimension n corresponding to the variables Y_i, the intermediate layer uses hyberbolic tangent activation functions, and the output layer consists of a single perceptron whose output is the value of the respective functions as exhibited in Figure 3a. As we do not impose any restriction on the values of $f(.)$ and $d(.)$, the activation function of their output perceptron is chosen to be linear. However, we impose the correlation between earnings and prices to be positive: in order to impose this constraint, we use an exponential activation function for the output perceptron of functions $g(.)$ and $g(.) + c(.)$ as shown in Figure 3b.[8]

1.3 Training and Validation

The training of the neural network, which describes the process of adjusting the weights to learn the mapping between input and output, is performed by an algorithm known as back-propagation.

---

[8]We do not restrict conservatism $c(.)$ to be positive and some firms do exhibit negative conservatism. For purposes of interpretation, it has been observed in the literature that "zero" asymmetric timeliness need not mean neutral accounting (Beaver and Ryan 2005; Ryan 2006; Givoly et al. 2007; Breuer and Windisch 2019) - rather, $c(.)$ is intended as a proxy such that higher $c(.)$ means more conservatism.

Figure 2: Structure of the Neural Network.

Circles represent perceptrons and rectangles represent embedded neural networks. Thick arrows are vectors and thin arrows are scalars.
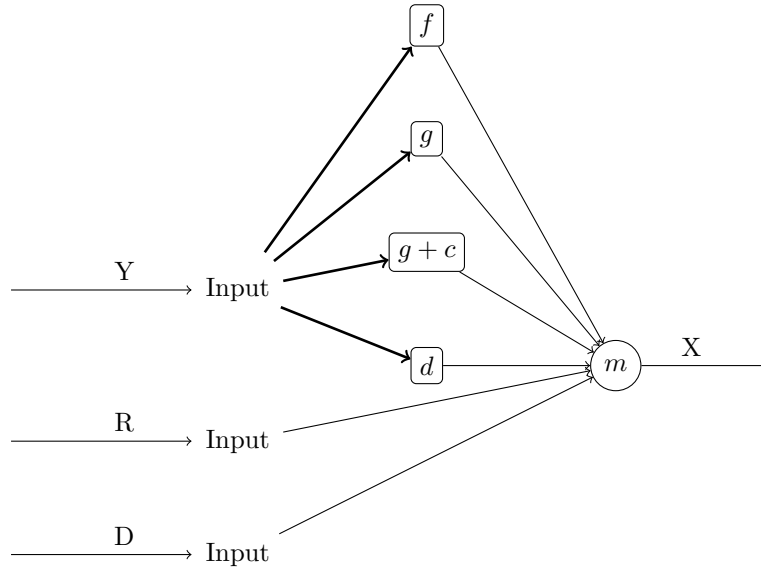
Figure 3: Structure of the embedded Neural Networks.

For functions $f(.)$ and $d(.)$, the activation function of the output layer is linear (not constrained). For functions $g(.)$ and $g(.)+c(.)$, the positivity constraint is imposed by using an exponential activation function. All intermediate layer perceptrons use the hyperbolic tangent activation function. Thick arrows are vectors and thin arrows are scalars.

(a) $f(.)$ and $d(.)$

(b) $g(.)$ and $g(.)+c(.)$

10
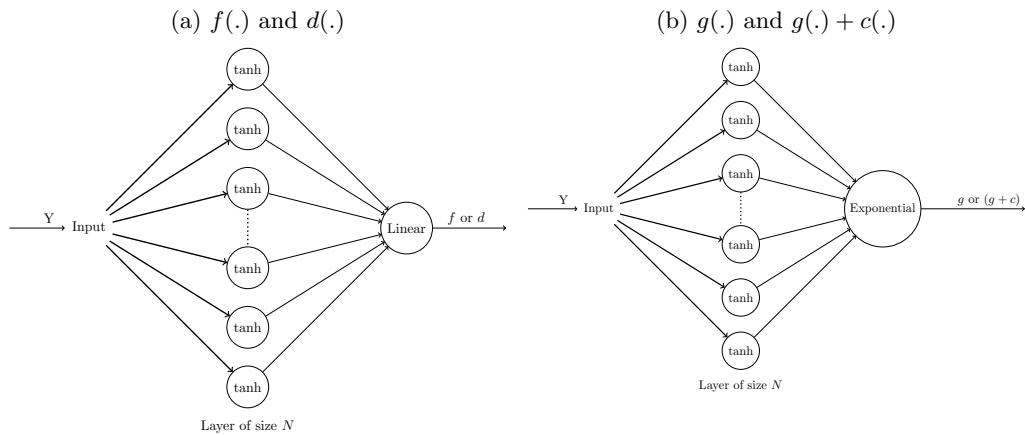
Back-propagation consists of updating the weights of all perceptrons in order to minimize a loss function. This is usually performed by computing the gradient of the loss function with respect to each weight of the network.[9]

We split the full sample into two subsamples, one for training and one for testing. The training sample is used to estimate the model and the testing sample is used to evaluate out-of-sample performance. The neural network is a function of several model parameters (sometimes called hyperparameters) that control the fit of the model to each individual observation. Machine learning algorithms do not have a closed-form expression for the optimal parameters and, instead, the tuning is done with computational methods. The optimal parameters are found by a process of cross-validation where the training sample is further split into a training - we call it the cross-validation (CV) training sample for the training subsample used in the cross-validation - and a validation sample. Given a set of model parameters, we train the network on the CV training sample and assess its fit on the validation sample. Using the optimal model parameters found by cross-validation, we estimate the model on the training sample and assess its performances on the test sample. To ensure the robustness and replicability of the results, we use an ensemble model (average of one hundred models estimated on the training sample) instead of a single estimation.[10] This ensemble model allows us to estimate how well our neural network approach fits the data.

An ensemble model that fits earnings is however impractical for use by researchers whose main objective is to compute a proxy for conservatism. Furthermore, that proxy ought to use as much information as possible – not only the one contained in the training sample. Once our approach is validated using the training and test samples, we train a separate neural network on the average C score from an ensemble model trained on the full sample. Apart from the fit of the models, the results presented in this paper use this final neural network.[11]

The usual approach when machine learning is used for prediction (as is the case with most current applications of the method) is to divide the dataset by year, with earlier years assigned to

---

[9]In order to reduce sensitivity to outliers, we use the Huber loss (with $\delta = 1$) when training the neural network as opposed to the more traditional mean square error. We conduct analyses using other loss functions in untabulated results and find similar results.

[10]Because neural networks typically use stochastic optimizers, models trained with the same hyperparameters exhibit slightly different convergence. We use the average of an ensemble of one hundred models trained with different seeds to reduce noise and ensure replicability.

[11]We conduct analysis in untabulated results to ensure that the final C score model gives similar results than the average C score from the ensemble model.

the CV training, later years assigned to validation, and final years in the sample assigned to the test sample (Bao et al. 2020; Bertomeu et al. 2020). This research design avoids using information to build a prediction that may not be known yet by the time the prediction is made. However, our use of machine learning is different in that we are not predicting but fitting a non-linear model whose form is unknown. Furthermore, comparability with existing measures estimated by time or with time fixed effects requires cross-validation, validation and test samples that have *all* years.[12] For this reason, we build each subsample by sampling by firm (rather than by year). The size of the training sample (CV training and validation) is set at 80% of the full sample. This sample is further split into the CV training and validation such that the CV training represents approximately 50% of the full sample. The result is three non-overlapping sets where one firm can only belong to one subsample but all years are present in each sample.

The cross-validation focuses on the two fitting parameters that have the greatest impact in the neural network: the number of iterations and the intermediate layer size. The number of iterations is the number of times the gradient of the loss function is computed and the weights are updated. The layer size refers to the number of perceptrons used in the intermediate layer of the network. More perceptrons increases the ability of the network to capture non-linearities. To reduce the complexity of the search, we constrain all networks ($f, g, g + c$ and d) to have the same number of perceptrons in the intermediate layer.[13]

1.4 Variable Selection

As benchmarks, we compute several standard measures of time-varying differential timeliness. The models L1 and L1y are ordinary least squares estimate of (1.1): in model L1, we define the variables Y_i as size, M/B and leverage, and restrict $g(.)$ and $c(.)$ to be linear to create an index variable; in model L1y, following KW, $g(.)$ and $c(.)$ are assumed to have coefficients that depend on

---

[12]Another limitation of choosing a validation and test subsample by time period is that, while this corrects for within-year correlation between the subsamples, it does not correct for within-cluster correlations, e.g., within firm or industry Bertomeu (2020). Unfortunately, there is still limited theoretical guidance on choosing an out-of-sample in panels with two-way correlation structures.

[13]We focus on networks with only one intermediate layer. We did not find significant benefits to using more than one layer and, in theory, a single layer can fit any functional form (in complex problems with unstructured data, e.g., image processing, it has been nevertheless shown that multiple layers outperform a single layer). We also cross-validated on the learning rate and found similar results. We use the Adam optimizer with the default parameters of the tensorflow library (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-7}$). The CV indicates the following hyper parameters (iterations; layer size): MLC1 (12,000; 25), MLC1y (4,000; 40), MLC2 (10,000; 40), MLC2y (4,500; 50).

calendar time and the model is estimated with separate yearly regressions. Model L1y is the most common empirical specification of the measure.[14] Models L2 and L2y add 5 additional variables identified to be potentially related to conservatism in KW: volatility, net operating accruals, cash flow from operations, investment cycle, and firm age, as defined in Table 1. Similar to models L1 and L1y, L2 includes only these variables while L2y is estimated by year. Together, the four models L1, L1y, L2 and L2y are benchmarks achieved with a standard linear regression.

Next, we develop four models estimated using the neural network approach by choosing the set of variables Y_i and derive four measures of machine-learning conservatism (MLC). The models MLC1 and MLC1y use the same three variables as KW and are the most parsimonious models in terms of assumed variables. In MLC1, we only include the three firm characteristics and assume that the relation does not change over time. In MLC1y, we follow the assumption in KW that the relation is different between years (i.e., their model is estimated with separate yearly regressions). Within a neural network approach, year-level variation can be incorporated by adding time fixed effects as additional variables in Y_i. Unlike fixed effects in a pooled time-series linear regression, the algorithm interacts these fixed effects to capture any time-variation in coefficients of the model. Models MLC2 and MLC2y are computed similarly to models L2 and L2y but using the neural network approach.[15]

Overall, we derive eight different models, which can be compared pairwise over any of the following dimensions: (i) linear vs. machine learning, (ii) fewer vs. more variables, and (iii) only economic variables vs. time effects. Regarding the first dimension (i), on one hand, linear models may imply a lower fit, and in models L1y and L2y, Fama MacBeth averaged coefficients need not summarize well the data-generating process (for example, in untabulated results, the out-of-sample adjusted R2 of the averaged coefficients in L1y is negative); on the other hand, machine learning methods may overfit the dataset. On the second dimension (ii), using more variables makes interpretation more difficult, especially when the variables are collinear and their causal links are not fully known. Finally on the third dimension (iii), many economic phenomena may change over time. Including time variation in the statistical model does not explain the causes for the changes

---

[14]See, e.g., Goh and Li (2011), Ettredge et al. (2012), Jayaraman (2012), Tan (2013), Ahmed and Duellman (2013), Martin and Roychowdhury (2015), Kim and Zhang (2016), Lara et al. (2016), DeFond et al. (2016) and Huang and Hilary (2018).

[15]Although we used the variables included in Khan and Watts (2009) as baseline, the ability of the model to incorporate many more non-linear factors will allow researchers to use additional terms in the measure.

and due to the increase in the number of parameters, it can lead to overfitting even in a linear regression.

2 Estimation

2.1 Data and Preliminaries

Table 2 reports the different steps resulting in our final data sample. We start with firm-year observations from both Compustat and CRSP database with fiscal year ends from 1963 to December 31st, 2019, for a total of 229,387 firm-year observations and 21,392 unique firms. For the full sample, we require non-missing net income, share price, common outstanding shares, book equity, total debt, assets and, as in KW, require a match to the CRSP database and non-missing compounded returns over the past eight months (to exclude the earnings announcement from the prior year) and drop all firm-year observations with share price below $1. For measures using additional variables, we require non-missing leverage, net operating accruals, cash flows from operations, investment cycle, volatility and age, which shrinks the sample further to 185,581 firm-year observations and 17,645 unique firms as reported in Table 3. To improve comparability across years in pooled regressions, we deflate size by the consumer price index from the St Louis Federal Reserve.

The descriptive statistics in Table 4 map to known characteristics of the Compustat universe. A median firm has a price to earnings of about of about 17 (1/.057), achieved a yearly return of 4.1%, has a size of 70 billion USD (log 5.094), a market-to-book of 1.58 and a enterprise value that is about .23 in debt (.3/(1 + 0.3)). We also collect a few additional variables following KW such as the (estimated) probability of litigation from Shu (2000), the probability of informed trade (Easley and O'Hara 2004; Brown and Hillegeist 2007) and the closing bid-ask spread; note that these are more sparsely populated than the variables used for the construction of the measures.[16] We report the correlations in Table 5. We find, as expected, that earnings are positively associated to returns and accruals, and negatively associated to leverage and stock volatility. Further, firms with higher earnings have lower probability of litigation and informed trade, and lower bid-ask spreads.

In Table 6, we further establish comparability to prior studies by comparing the coefficient estimates in our full sample, model L1y, relative to the coefficients in Table 3 of KW. For complete-

---

[16] We thank Stephen Brown for sharing the data on PIN.

ness, we also replicate their research design and recover very similar coefficients. The coefficients are similar in magnitude to the original estimates, with the recent periods featuring a $c(.)$ function with attenuation in market to book (coefficient $D \times Ret \times M/B$) and greater sensitivity to size and leverage (coefficients $D \times Ret \times Size$ and $D \times Ret \times Lev$).

To further motivate the benefits of a machine learning based measure, we analyze some of the properties, and potential drawbacks, of the benchmark model L1y. Given that the model is estimated with separate yearly regressions, the coefficients on size, M/B and leverage are expected to vary from year to year. As shown in Figure 4, the coefficients exhibit large spikes with only limited interpretable persistence that would map to trends in economic transactions or macroeconomic changes. In fact, all coefficients change sign over the sample period: the M/B coefficient oscillates around zero and the coefficients on size and leverage can change three-fold year-over-year and occasionally become zero or slightly negative.

Figure 4: Plots of the C score coefficients over time using yearly regressions.

Subfigure (a) uses sample restrictions used in Khan and Watts (2009) and subfigure (b) uses our full sample.

(a) Replication

(b) Full Sample
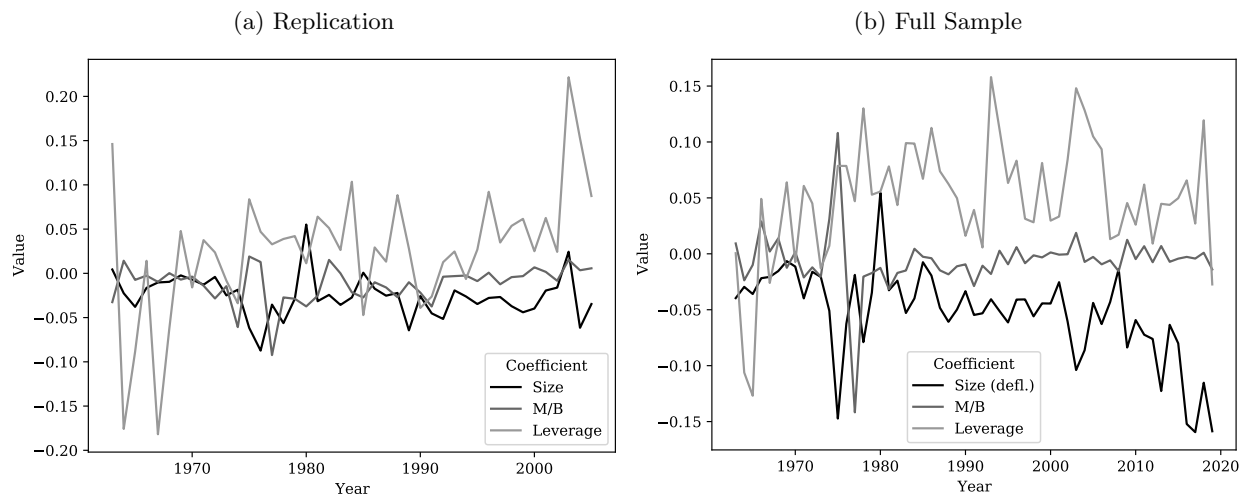
Another symptom that may indicate overfitting is that the relation between earnings and returns is often negative, which can only occur under unusual circumstances. We plot in Figure 5 the density of the relation conditional on positive news $g(.)$ and conditional on negative news $g(.)+c(.)$. In both cases, we expect the average effect to be such that earnings reflect information in the direction of the news, consistent with most prior research (see, e.g., Ohlson (2001) and Penman and Zhang (2002)). However, according to the measure, $g(.)$ is centered around zero, indicating that about half of the

firms with positive returns feature lower earnings conditional on higher returns. The magnitudes of the coefficient (between $-.2$ and $.2$) suggest that this feature is not due solely to the limited association between earnings and returns (Ball and Brown 2014). Conditional on negative returns, observations are more likely to feature a positive association $g(.) + c(.) > 0$ but about a quarter of all observations feature a negative association. Given that the variables used in L1y are unlikely to capture the special circumstances leading to a negative association (early stage growth firms with unusually high levels of investments in intangibles), this distribution of parameters suggests that the linear model is misspecified and the functions $g(.)$ and $g(.) + c(.)$ should be restricted to remain positive.

Figure 5: Plots of the $g(.)$ and $g(.) + c(.)$ scores densities estimated using yearly regressions on the full sample with Size, Market-to-book and Leverage (measure L1y).

(a) G Score

(b) G+C Score
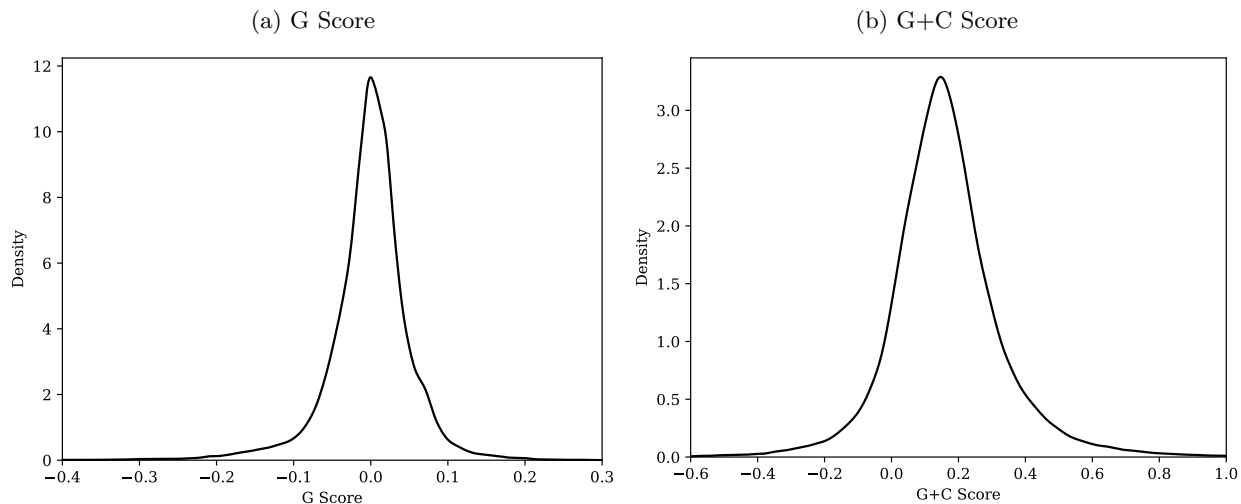
2.2 Fit and Time Trends

We proceed next to estimate the eight models: the linear models L1, L1y, L2, and L2y, where models "2" refers to the inclusion of the additional features and models ending with "y" are estimated with separate yearly regressions. The models MLC1, MLC1y, MLC2, and ML2y are similarly defined but estimated with the neural network approach. We use the same training and validation procedure for all models to ensure comparability. We always use the largest possible sample for

which variables are non-missing. [17]. We use here a long time-series feasible on Compustat: the long time horizon tends to benefit "y" models because more structural breaks are likely to occur. Plausibly, models without time effects and interaction should be more accurate using more recent data. The fit is reported in Table 7.

In models 1, we find that L1, estimated with a pooled regression features a very similar fit in-sample (training) and out-of-sample (test): with four variables, the model is estimated with many degrees of freedom and features very little overfitting. The R^2 of the model is at 10.7% in the test sample. The fit of the model more than doubles when estimating the model by year in L1y, at 22.9% in the test sample. While not based on a machine learning algorithm, there is evidence that the time regressions slightly overfit the data, with the fit decreasing by about 2.7% relative to the training sample. In summary, L1y is not as parsimonious as L1 (or even a regression with time fixed effects): the separate regressions imply 4 variables $\times$ 57 years, for a total of 228 estimated coefficients. We observe throughout all models that the MLC version (without year) use fewer fitted coefficients than the linear model with year, which intuitively suggests that the degree of overfitting in MLC is not necessarily greater than that in linear model estimated by year. Of course, models MLC1y and MLC2y are the models with the most estimated coefficients.

While MLC1 is about double the fit of the simple linear model L1, at 20.4%, with only three economic factors and a constant, it (honorably) fails to match the fit of L1y; hence, non-linearities in three factors cannot fully explain changes due to yearly shocks. Yet, the decrease in fit is surprisingly small for MLC1 relatively to L1y (2.5% lower), and furthermore, MLC1 does not seem to exhibit as much overfitting, with a modest 1.4% decrease in fit relative to the training sample. The importance of time effects is much reduced when moving from MLC1 to MLC1y. The fit increases by 8.6% (versus 12.2% for L1y), or, in relative terms, an increase in R^2 of $(29 - 20.4)/20.4 = 42.2\%$ for the machine learning model versus $(22.9 - 10.7)/10.7 = 114\%$ for the linear model. In L1y, the coefficients are fixed within each period, and the fit due to changes over time may be capturing an underlying non-linear relationship. In contrast, with machine learning, if a true non-linear relationship exists, it would be established using the changes in levels across period.

We examine the measures in models 2 incorporating additional accounting variables. Interest-

---

[17]In untabulated analyses, we show that model 2 performs significantly better than models 1 holding samples constant.

ingly, L2 fits considerably better with the additional variables and the differences with L2y are smaller with a relative change in R^2 of $(40.5 - 31.5)/31.5 = 28.6\%$. Thus, these variables appear to be important components to capture conservatism even in a regression framework. The main result from Table 7 is that MLC2, a model with 8 variables and a constant but allowing for non-linearities, fits slightly better than L2y, which uses 9 variables (including the intercept) $\times$ 57 years, for a total of 513 estimated coefficients. L2y achieves an R^2 of 40.5% in the test sample versus 45.5% for MLC2. The reduction in performance between training and test sample is due to the fact that L2y fits a large number of coefficients. The degree of overfitting is less pronounced for MLC2, with an absolute reduction in fit of 2.7% for the machine learning algorithm MLC2 versus 4.4% for L2y when comparing training and test samples. These results are unexpected given the long time-series used but indicate that non-linearities play an important role in explaining differential timeliness, at least as important as changes over time. The fit of MLC2 is further penalized relative to L2y because the positivity restrictions on $g(.)$ and $g(.) + c(.)$ reduce the fit in the training sample.

For completeness, we also report the fit of MLC2y, which we expect to be higher than MLC2. The results confirm this conjecture: in the training sample, the fit of MLC2y is 7.1% higher than MLC2. However, the difference shrinks in the test sample to about 4.5% given that MLC2y is less parsimonious. Based on these results, MLC2 achieves the best balance between parsimony and fitting the sample, explaining almost half of the variation in differential timeliness.

We compare these measures in the correlation matrix in Table 8. As expected, all measures are positively correlated, suggesting that they directionally capture a common construct. Machine learning measures tend to feature a correlation of .4 or above with their linear counterparts, providing a sanity check that the first-order factors recovered in linear models are still present with the machine learning measures. These correlations are the lowest when comparing models with time versus models without time effects, indicating that time changes are difficult to estimate. Models 1 and 2 with time effects feature the lowest correlation, implying that models with time effects are also more sensitive to the inclusion of additional variables. We also find that the probability of litigation is positively associated to conservatism for all measures. PIN and bid-ask spreads have a positive association with conservatism in all measures as documented in LaFond and Watts (2008). If information asymmetry stems from withholding bad news, firms exposed to more information asymmetries may be conservative by recognizing in a more timely fashion bad news (Wittenberg-

Moerman (2008)). All in all, these associations are consistent with the evidence documented in the literature.

Figure 6: Plot over time of the C scores measures L1, L1y, MLC1, MLC1y.

(a) L1

(b) L1y

(c) MLC1
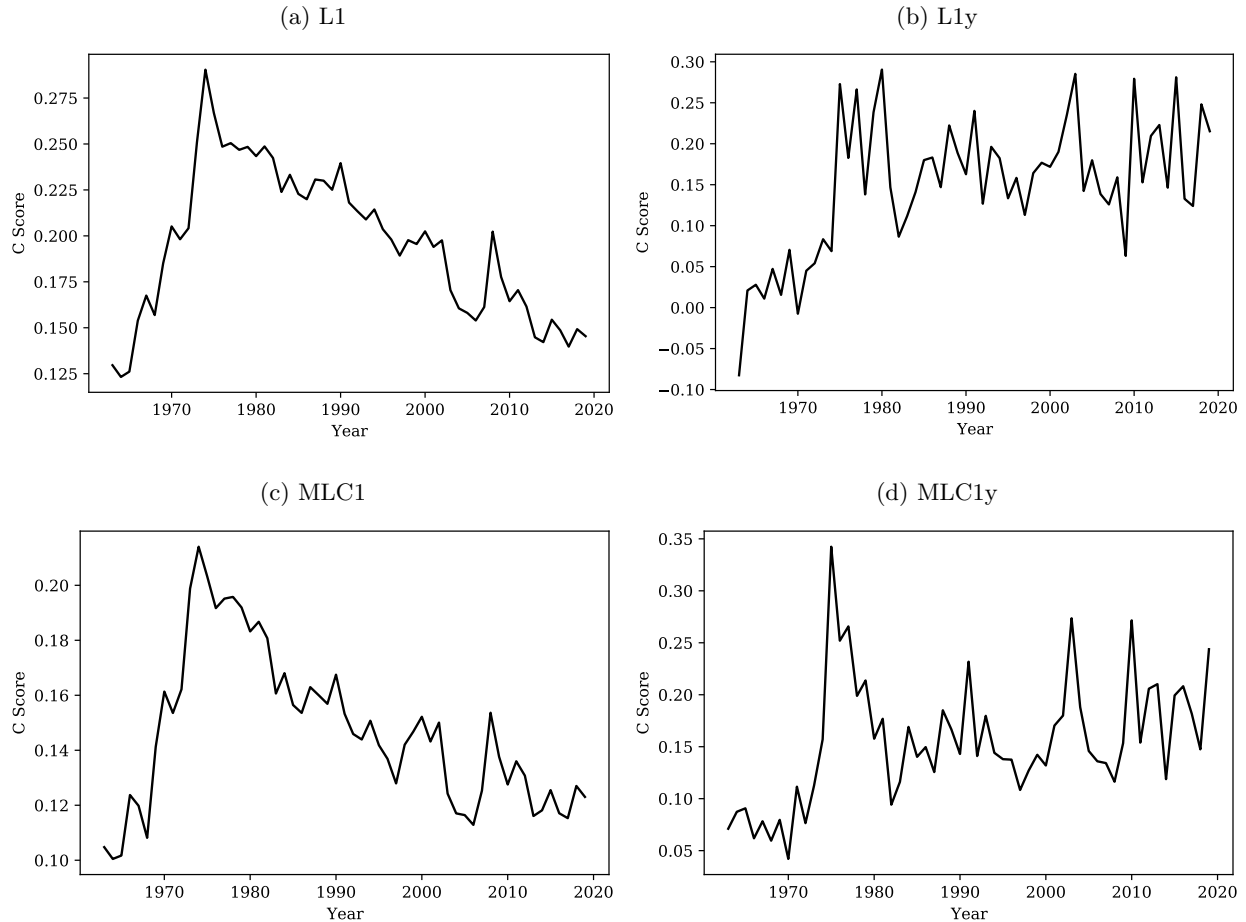
(d) MLC1y

Figures 6 and 7 explore time-series changes in the construct of interest, along the lines of using machine learning to examine long-term trends in accounting (Barth et al. 2019). Models L1, MLC1, L2 and MLC2 exhibit an increase until the mid-seventies, followed by a clear long-term trend of declining conservatism, starting in the decade of the 70s and show that conservatism is near its lowest ever in 2020. To our knowledge, this has not been documented with archival data in prior literature but is highly consistent with a shift in conceptual statements towards relevance, initiated with the first concepts statements of the FASB in 1978 and, later, the convergence with international accounting standards and the change in focus toward neutrality and prudence (Zeff 2013). The models with yearly effects do not clearly show this trend expected from institutional

Figure 7: Plot over time of the C scores measures L2, L2y, MLC2, MLC2y.

(a) L2

(b) L2y

(c) MLC2
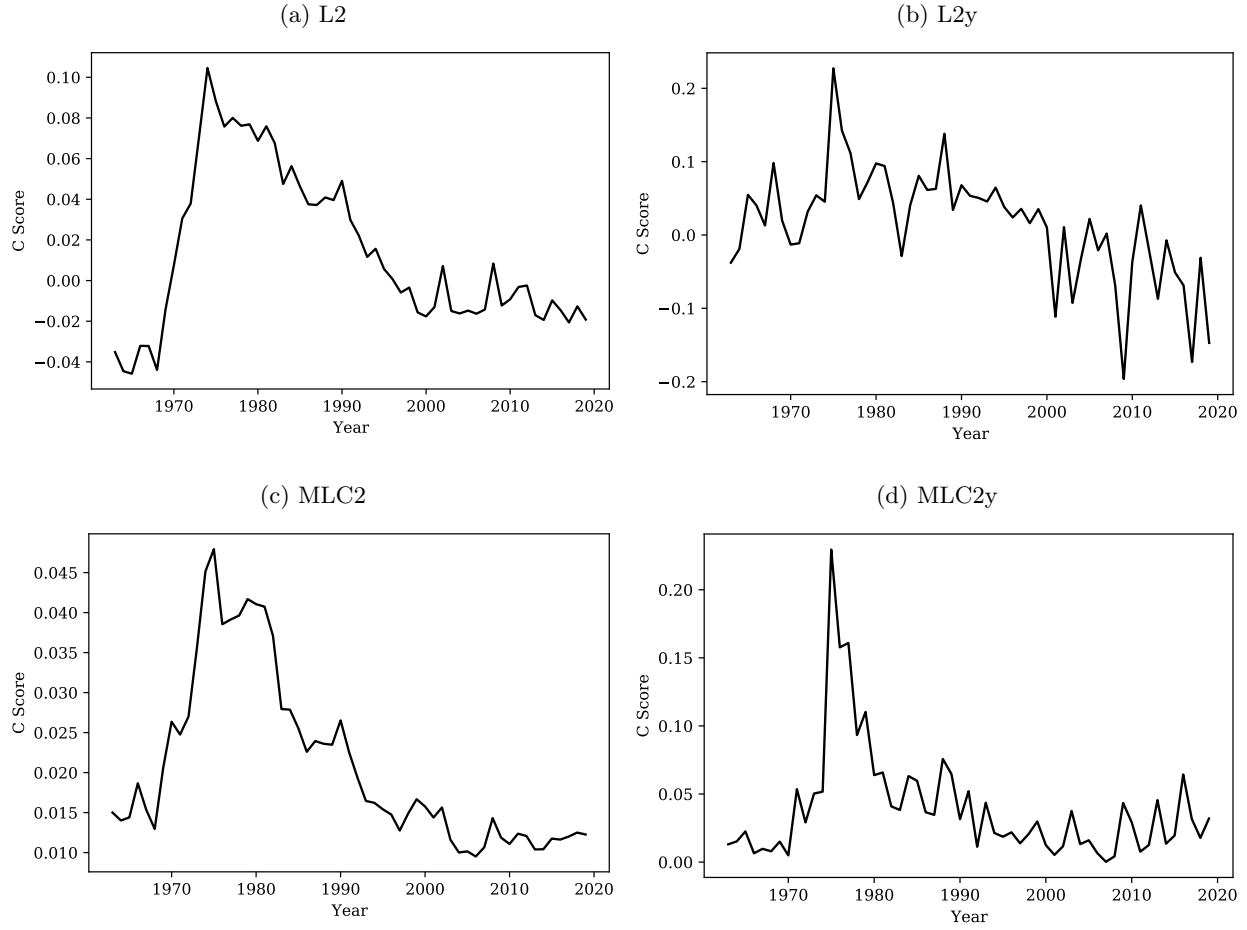
(d) MLC2y

facts, plausibly because the measures are too volatile. Overall, the time-series behavior of L2 and MLC2 seems to capture the institutional changes more than L2y and MLC2y.

We plot in Figure 8 the distribution of measures across observations: these plots reveal some of the greatest differences between measures. The benchmark model L1y features extreme variation in $c(.)$, with magnitudes close to one. L1, L1y, MLC1, MLC1y, and MLC2y display a vast majority of positive $c(.)$; by contrast, L2 and L2y are centered with their mode close to zero. The distributions also tend to have positive skewness. Machine learning models with more variables (MLC2 and MLC2y) feature lower dispersion in conservatism, which is driven in large part by the positivity restrictions on $c(.)$ and $g(.) + c(.)$.

Figure 8: Densities of the C scores measures.
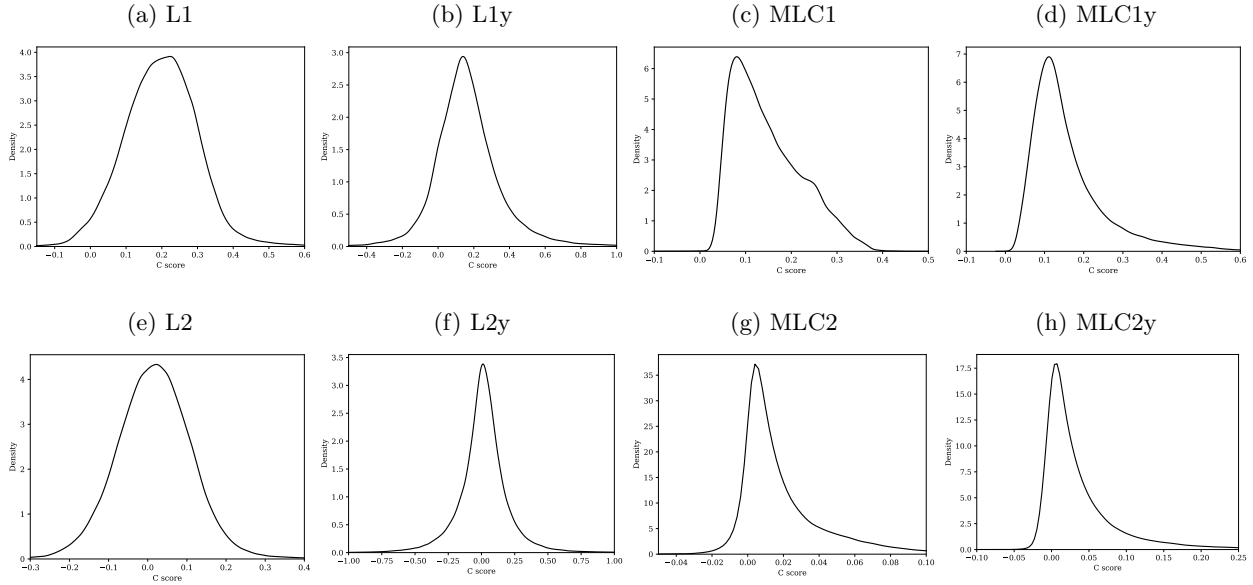
2.3 Economic Factors associated to Conservatism

Figures 9 to 12 explore in greater detail the nature of the relation between conservatism and the economic factors identified in each model, by measuring average conservatism for each input variable: size, M/B, leverage, volatility, firm age, as well as probability of litigation, probability of informed trading (PIN) and closing bid-ask spread. In these Figures, we average the conservatism conditional on key economic characteristics.

Figure 9: Plots of the C score measures as a function of Size with 95% confidence intervals in gray.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1 (b) L1y (c) MLC1 (d) MLC1y

(e) L2 (f) L2y (g) MLC2 (h) MLC2y

Figure 10: Plots of the C score measures as a function of market-to-book ratio.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1 (b) L1y (c) MLC1 (d) MLC1y

(e) L2 (f) L2y (g) MLC2 (h) MLC2y

22
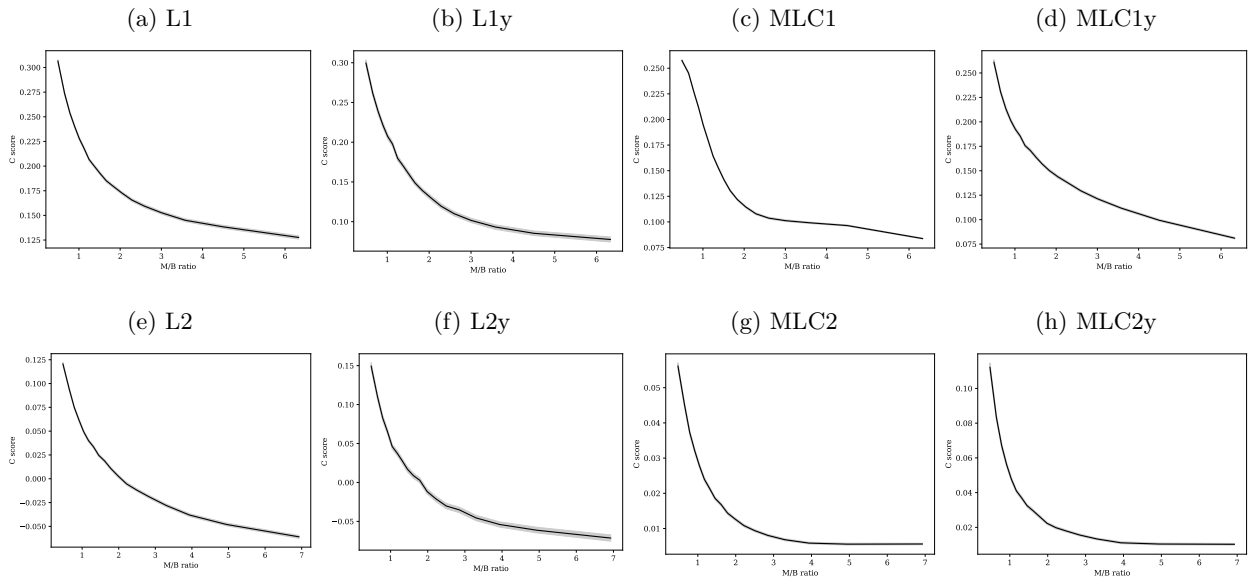
The negative relationship between conservatism and size is consistent across all measures in Figure 9. The neural network reveals a convex relation between conservatism and size. The steepest decrease in conservatism occurs across smaller sizes but the effect flattens over larger firms. We conjecture that smaller firms tend to have less access to fluid capital markets and face greater informational frictions, requiring increasingly conservative measurements. In addition, larger firms are more likely to average out many transactions, which may smooth asymmetries between returns and earnings.

We also find similar patterns in the case of M/B. Note that because M/B is correlated to size and leverage, the relation need not be linear even with a linear specification (as we average over the entire sample). In fact, Figure 10 shows a convex relationship between conservatism and M/B. Firms with greater M/B feature lower levels of conservatism but the relationship flattens over high M/B levels. Presumably, firms with very high M/B exhibit a looser connection between prices and accounting numbers regardless of the content of the news (Rajgopal et al. 2020). The linear models fit this relation indirectly, by averaging over other correlated characteristics.

Figure 11: Plots of the C score measures as a function of leverage.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1   (b) L1y   (c) MLC1   (d) MLC1y

(e) L2   (f) L2y   (g) MLC2   (h) MLC2y
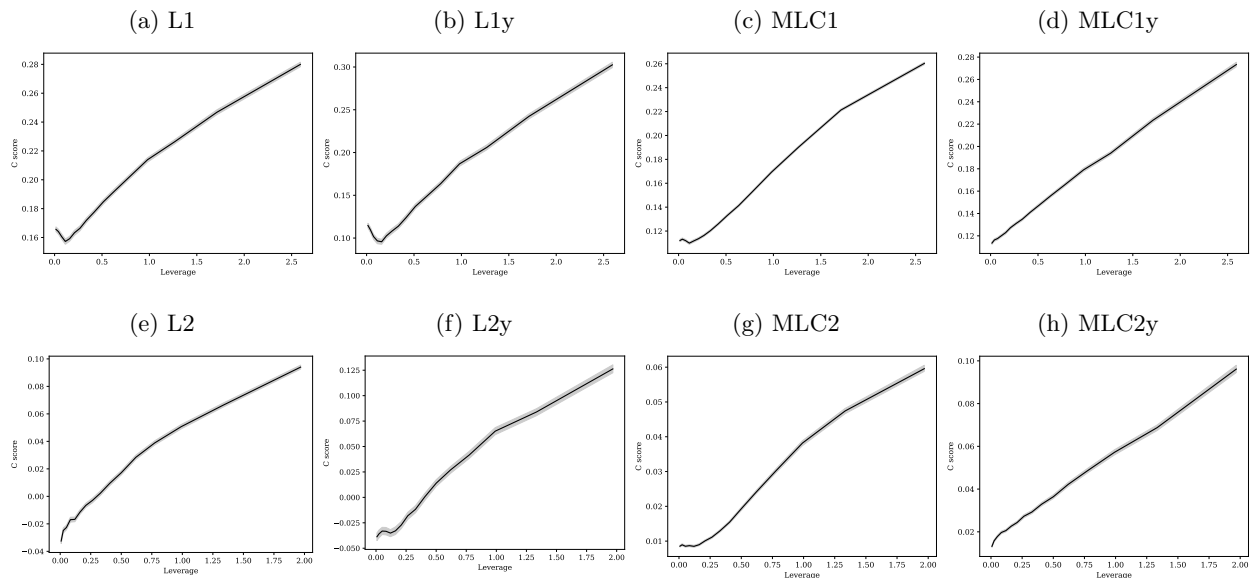
While conservatism is increasing in leverage in all specifications in Figure 11, there are some differences across measures. For MLC2, and to a lesser degree L1, L1y, MLC1 and L2y, conser-

Figure 12: Plots of the C score measures as a function of age.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1   (b) L1y   (c) MLC1   (d) MLC1y

(e) L2   (f) L2y   (g) MLC2   (h) MLC2y
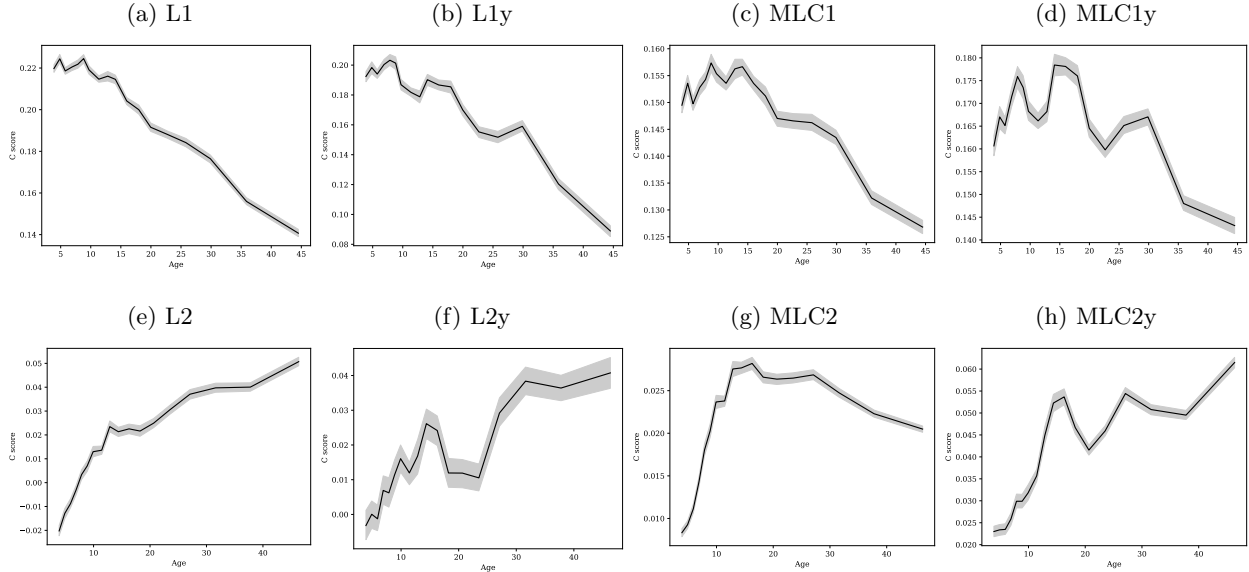
vatism is initially convex (slightly decreases or flat for low levels of leverage) and then becomes slightly concave. Conservatism is commonly thought to offer contracting benefits in the presence of financial risk. When leverage and financial risk are very low, the benefits may be other or muted, implying a weaker relationship. Vice-versa, if leverage is very high, conservatism may not fully resolve the causes of the high leverage (e.g., low past performance) and, therefore, the relationship may become less steep. As is intuitive, we find in all measures that firms with high levels of leverage also tend to implement more conservatism. However, no attempt is made here to attribute a causal effect between leverage and conservatism, which would require an exogenous shock to leverage or conservatism. An extensive literature studies the causal effect, see, e.g., Tan (2013) for a recent example. Our main objective is to provide a more accurate measure of conservatism when identifying causal effects.

A factor that appear to be ambiguously related to conservatism is firm age in Figure 12. Measures L1, L1y, MLC1 and MLC1y suggest a decreasing relation; however, when firm age is included in the model, the relationship is different: increasing in the linear models L2, L2y and MLC2y, and inverse U-shaped for MLC2. While there is (to our knowledge) no complete theory explaining the link between life cycle and conservatism, the fact that firm age is correlated with time effects is
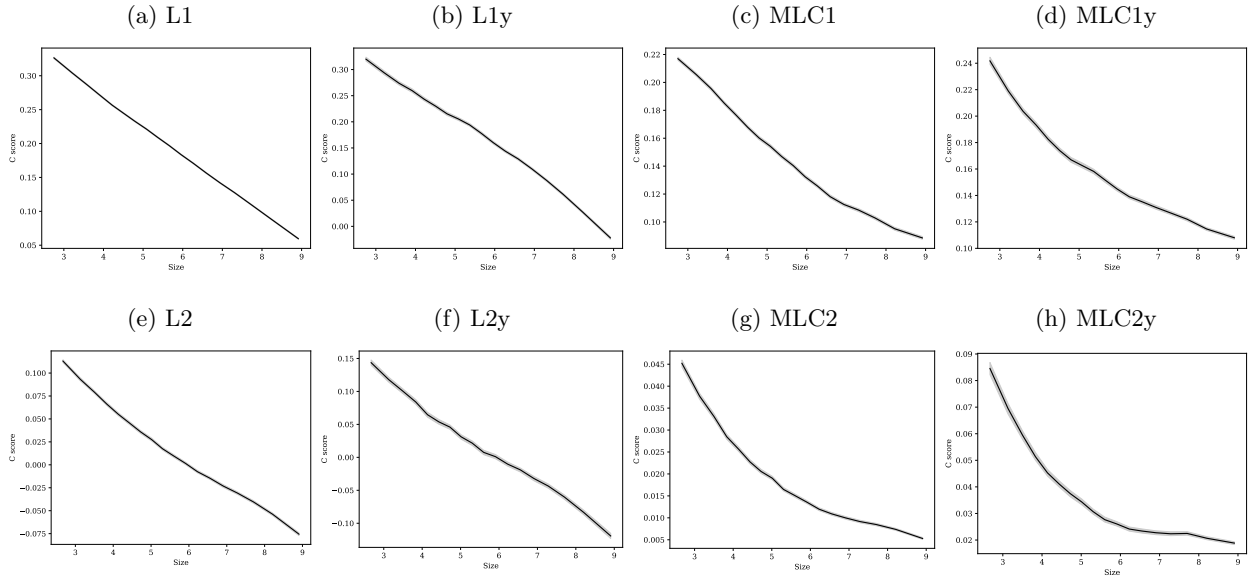
24

Figure 13: Plots of the C score measures as a function of Volatility.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1   (b) L1y   (c) MLC1   (d) MLC1y

(e) L2   (f) L2y   (g) MLC2   (h) MLC2y
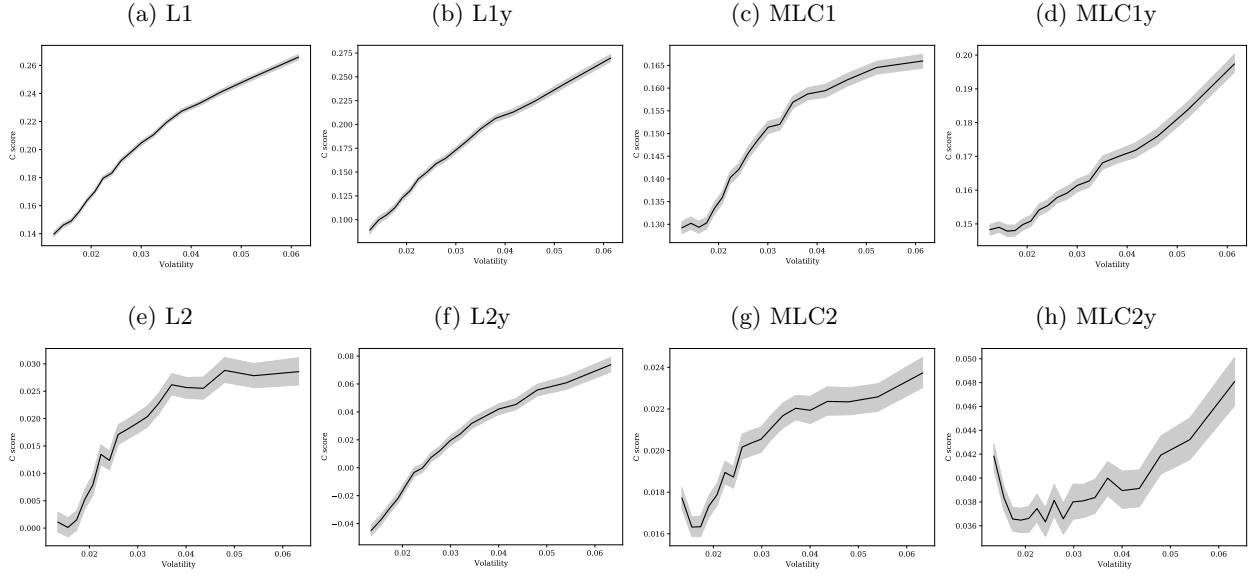
making its identification particularly difficult for models that vary with time effects. Our preferred measures on the basis of fit without time effects, MLC2, suggest that firms with moderate age have the highest level of conservatism: new cohorts featuring less conservatism is intuitive due to changes in the nature of transactions Srivastava (2014). However, the usefulness of conservatism may decline for very old firms that need less access to capital markets and have already established durable financing relationships.

The relation between conservatism and volatility is consistently positive across measures, except for a subset of values in MLC2 and MLC2y, where it is convex and decreasing for very low levels of volatility. For all other measures, we tend to observe a concave relation between conservatism and volatility - indicating that models with time effects indirectly capture non-linearities. As for the rest of these associations, these are not intended as causal and the firm may be responding to higher levels of volatility by changing its disclosure environment. Reporting choices are less likely to be effective in environments with high levels of volatility related to shocks to the business model (Billings et al. 2015).

The probability of litigation is found in prior studies to be positively associated to conservatism. We confirm this finding in Figure 14 for most levels of litigation risk and, in particular, if litigation

Figure 14: Plots of the C score measures as a function of Probability of Litigation.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1   (b) L1y   (c) MLC1   (d) MLC1y

(e) L2   (f) L2y   (g) MLC2   (h) MLC2y
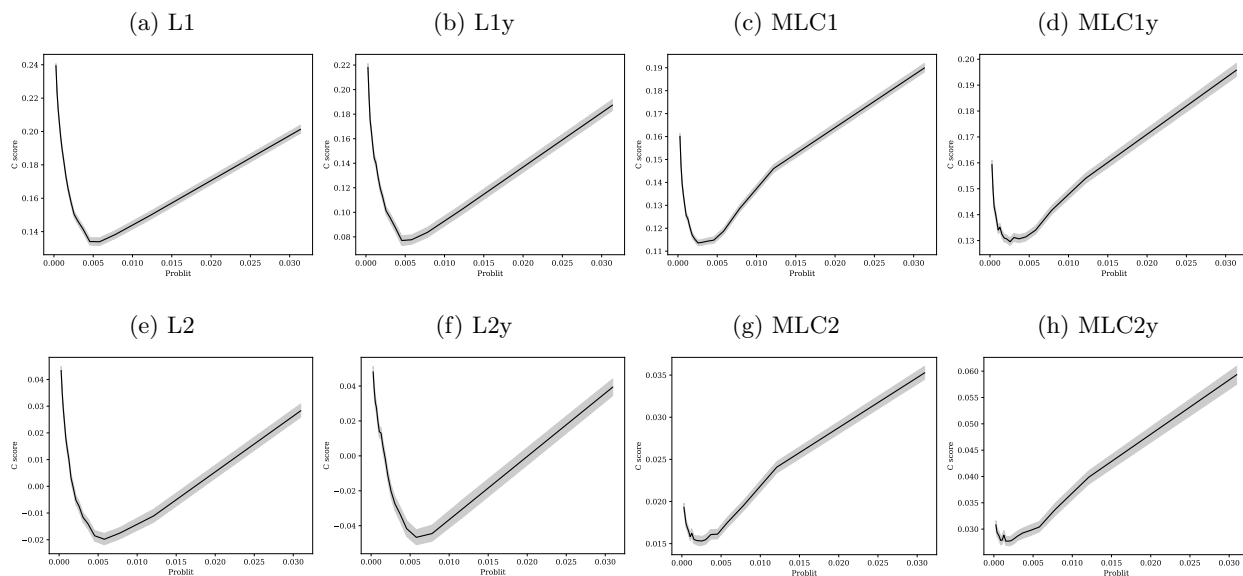
risk is large. However, we also find that for low levels of litigation, conservatism can be decreasing in litigation. A possible reason for non-monotonicity (found even with the measure L1y) is that litigation is also endogenous to conservatism and may imply a very low level of litigation when the firm chooses to be very conservative, see also Kronenberger and Laux (2021) for a theoretical model explaining this relation. Interestingly, the decreasing relation occurs for a very small subset of values in MLC2 and MLC2y, suggesting that these models tend to be more in line with the intuitive positive relation (Watts 2003).

The last two plots consider the association between market characteristics indicating information asymmetry: the probability of informed trading (PIN) in Figure 15 and bid-ask spread in Figure 16. In all specifications, that conservatism is increasing in PIN, consistent with prior literature (LaFond and Watts 2008). Firms with greater levels of conservatism tend to feature more informed trading. This fact is consistent with two possible explanations: that firms with greater information asymmetry address it with greater levels of conservatism or that the omission of information in earnings creates profit opportunities from informed trading. Unfortunately, an observational documentation such as ours cannot distinguish between the two explanations. We document consistent results with bid-ask spreads, as bid-ask spreads are increasing in conservatism

Figure 15: Plots of the C score measures as a function of Probability of Informed Trading (PIN).

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1　　　(b) L1y　　　(c) MLC1　　　(d) MLC1y

(e) L2　　　(f) L2y　　　(g) MLC2　　　(h) MLC2y
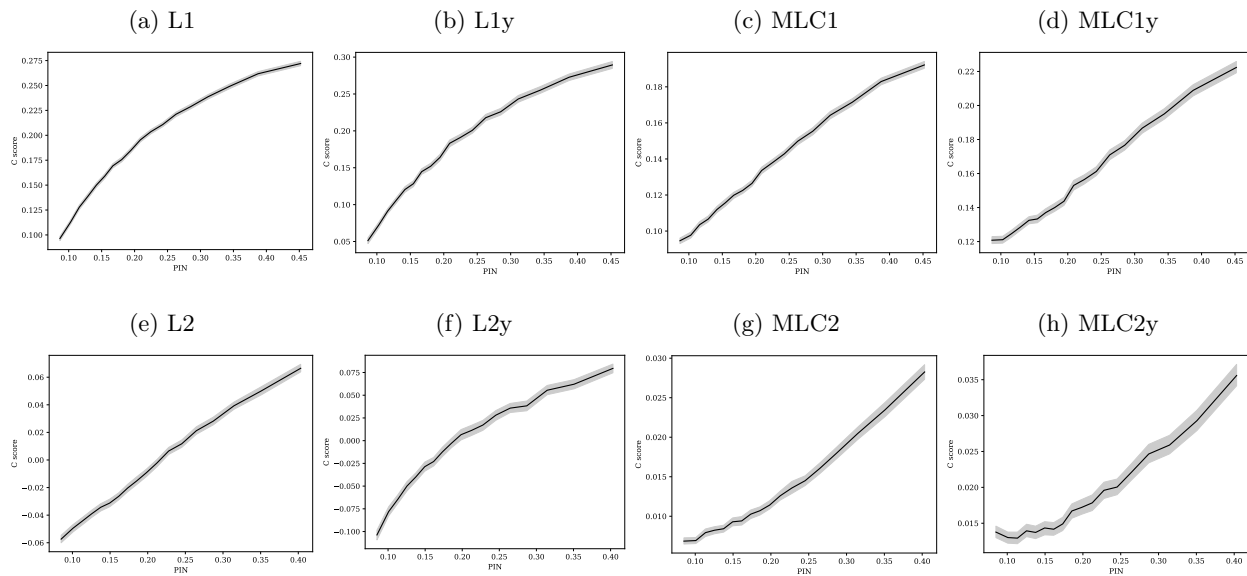
across all measures. The models disagree in terms of the convexity of the relation between PIN and conservatism. Our expectation is that for firms with very low PIN and almost no informed trading, there should be no relation with conservatism because investors can assume that all trading occurs for liquidity reasons - this is more consistent with the convex relationship in MLC2 and MLC2y.

All models show an increasing relationship between bid-ask spreads and conservatism which, as is intuitive, suggests more reliance on conservatism when there are more informational frictions. Interestingly, all models seem to suggest a special circumstance for firms with near-zero bid-ask spreads, which feature much lower conservatism as they may indicate firms that are almost perfectly liquid. This jump is, however, much more muted in MLC2 and MLC2y where, in terms of economic magnitudes, the region occurs for bid-ask spreads below .5% which may suggest very small transaction prices. For MLC2 and MLC2y, the relation is close to linear for most bid-ask spreads, while it appears concave for all other measures.

We observe similar results in multivariate analyses in Tables 9 and 10, as we run a regression of each measure on (standardized) explanatory variables. We include two additional columns for each measure, to better understand how each measure is constructed and decompose the separate time effects or non-linearities or interactions from the neural network into additive terms. For L1 and L2,

27

Figure 16: Plots of the C score measures as a function of bid-ask Spread.

Average C score for 20 quantiles. First and last quantiles are removed.

(a) L1   (b) L1y   (c) MLC1   (d) MLC1y

(e) L2   (f) L2y   (g) MLC2   (h) MLC2y
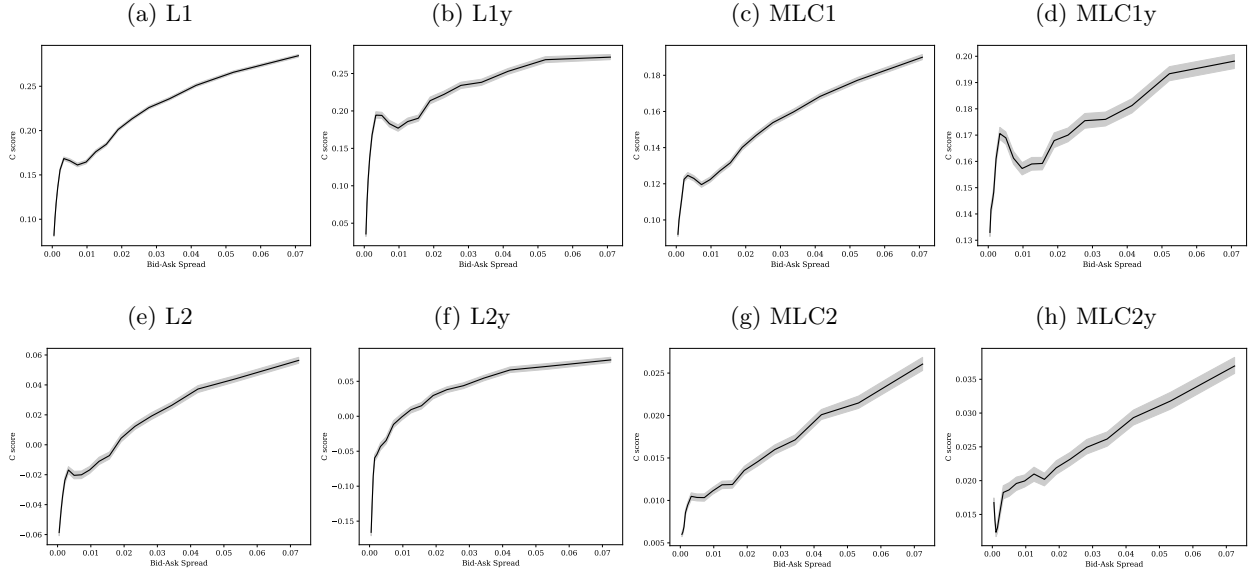
we skip these decompositions given that the measures are a linear combination of its components, so column (1) in Table 9 and 10 achieve an R^2 equal to one. We focus on the interactions of size, M/B, leverage and volatility.[18]

We find that all linear models with time effects, L1y and L2y, exhibit hidden non-linearities and interactions that would not be immediately apparent in the specification. For example, in the multivariate model, L1y and L2y are concave in M/B and leverage, and most of the interactions are significant. The significance of the coefficients of the quadratic terms stems from the model in L1y, which is restricted to be linear in each year, to fit non-linear correlation structures with other variables given this linear constraint. Across both L1y and L2y in column (4) of Tables 9 and 10, we provide support for a negative interaction between size and leverage. Firms with higher size tend to respond less to leverage. We also find that volatility increases the effect of M/B, amplifying the importance of future prospects contained in M/B.

Turning to measures based on the neural networks, we find that interactions and non-linearities increase the adjusted R^2 of the models more than in the linear models, suggesting that, by and large, the MLC models are better designed to assign more of the variation to these components.

---

[18]We can expand the set of interactions by including more variables, but report here the most common firm characteristics that are documented to correlate the most with conservatism measures.

Columns (5) and (8) of Tables 9 and 10 include no interaction or non-linearities and confirm results in Figures 9-12. Columns 6 and 9 of Tables 9 and 10 convey the same non-linearities as with univariates. Several new insights are obtained by considering columns 7 and 10 in Tables 9 and 10, which speak to the interactions captured by the neural network approach. Across all specifications, there is a strong positive interaction between M/B and size, which was insignificant or very small in linear models. MLC2 also features a negative interaction between leverage and the variables size, M/B and volatility consistent with L1y and L2y, and positive interactions between M/B and volatility.

Given that we standardize all explanatory variables, columns (1), (2), (5) and (8) in each panel also allows us to evaluate the importance of a variable to explain $c(.)$.[19] For all "1" measures, as expected, the three constructs used for the measure are the most important explanatory variables; however, in these models, other variables that were not used to construct the measure often carry importance that is comparable to the original components. This provides more evidence that models "2," containing more variables, help fit differential timeliness. Examining models "2" in Table 10, size seems to be the most important factor across models but, interestingly, the importance is evenly spread across all other variables and leverage or M/B do not always carry, on their own, substantially greater explanatory power than many other variables.

3 Robustness Analyses

3.1 Measurement Errors

As our baseline analysis, we examine economically appealing features of a measure (e.g., whether it has characteristics consistent with firmly established priors in the literature) and whether it fits the data well (e.g., whether it can extract sufficient information on a construct of interest). However, while necessary conditions, these are not sufficient to guarantee the validity of the measure. Unfortunately, the researcher does not know the true data-generating process (hereafter, DGP) for conservatism which creates complications given that all measures are misspecifying the true process.

---

[19]Unfortunately, neural networks do not have a standard tool to recover the importance of a variable within a perceptron, so we develop intuition with the help the average effect in a regression.

A common approach is to conduct simulations where the researcher defines a representative DGP and examines the behavior of measures when applied to a problem with finite samples and/or if the measure itself may be misspecified to some degree either because of its assumptions or computational methodology. Any of these reasons will lead to statistical error in our setting, given that even the most general algorithms presume some regularities in the data, we use a finite number of observations, and the computational methods are imperfect. Evaluating the properties of an estimator by simulation enables us to assess qualities that cannot be analytically derived and is sometimes used in applied work to evaluate complex models.

As in Breuer and Schütt (2019) and Jennings et al. (2020), we use simulations to assess errors in the empirical analysis. Clearly, each model is expected to work best if it is the true DGP, so we repeat the analysis 8 times by rotating the true DGP and then running each measure against this DGP. To be specific, in Table 11, we block bootstrap the variables (Y_i) of 100 new datasets by firm, construct earnings using the functions $g(.)$ and $c(.)$ estimated in the earlier section and applying (1.1), and let a researcher unaware of these functions run their own model in each row. To keep the analysis as transparent as possible, we do not cross-validate the machine learning models for each DGP but use the hyperparemeters of our previous cross-validation.[20] The mean absolute error (MAE) is the average absolute error between the true c and the estimate of the researcher. Then, we average across all simulations and compute a standard error across the 100 point estimates of MAEs.

For example, in column L1y, we reconstruct earnings X_i for each observation in the sample by using estimation L1y. Then, in row MLC2, the researcher runs the analysis and obtains a new MLC2-based C score for this particular sample - of course, the researcher is potentially misspecified and uses a model that is too complex for the DGP. We find that a researcher using MLC2 (by mistake) when the true DGP is L1y would do an error equal to 52.4% (half of a standard deviation of the true measure), versus 22.3% if (correctly) using L1y.[21]

As intuitive, estimation procedures do well in the diagonal, since they are correctly specified

---

[20]For us to do this exercise by cross-validating each simulated sample would suppose that we intend each researcher using these measures to cross-validate again; however, this is the main time-consuming part of the analysis and the main objective is of our study is to offer reasonable hyperparameters which can be easily re-used to recover measures of conservatism.

[21]For simplicity, we calibrate the DGP as if the estimates of the model were the actual DGP. Of course, these estimates contain error and need not be an exact match of the DGP - so these results only speak to the desirability of a method if an underlying process is either more or less complex than the method assumes.

for the process of interest. Errors are, unsurprisingly, larger when the process is more complex and includes non-linearities or time effects. As to robustness, a better model can be read in rows and must be such that its errors are smaller than other rows (i.e., comparing row vectors). All "1" models tend to perform poorly if the true DGP has more variables (columns for "2" models), and are often multiple standard errors away from the true value. MLC2 and MLC2y are the most robust approaches across models with similar errors across all specifications (hovering at 50% of a standard error). Also, MLC2y appears to be less robust than MLC2: MLC2's error against MLC2y remains at 40.8% while MLC2y's error against MLC2 is 91.3%, likely because MLC2y has many variables and incorrectly assigns the fit to time effects and their interactions. This table provides additional support for MLC2 to be the preferred measure for differential timeliness.

3.2 Testing Errors

Given that most researchers are not interested in using a model to make precise measurements of conservatism, we simulate a generic statistical model in which a researcher uses the measure to test an economic theory, but is not interested in the measure's accuracy as an end goal. We specify the test as follows:

$$Z_i = \alpha C_i + \beta Y_i' + \epsilon_i, \tag{3.1}$$

where Z_i is an object of interest, C_i is the true level of conservatism (unobserved to the researcher) and Y_i' is a subset of controls. We proceed as in Table 11 using the simulations and measures, but now additionally computing the variable Z_i and letting the researcher employ their model estimate of C_i to recover the parameter α. To keep the model as transparent as possible, we set the true $\alpha = 3$ and (for now) do not add additional error terms ($\epsilon_i = 0$). Vector Y_i' includes size, M/B and leverage, with respective β coefficients of 4, -2, and 1.[22]

The main results are reported in Table 12. As expected, measures still perform best in the diagonal. Throughout, all measures tend to feature attenuation bias given that they are measured with error and, consistent with conventional uses, we do not attempt to correct for this bias. The attenuation bias is notable even for machine learning models which suggests that MLC measures do

---

[22]Note that in untabulated analyses, these are not critical assumptions as long as one only focuses on relative comparisons between models.

not fully solve the problem of not having enough assumed economic structure to estimate measures precisely. The attenuation bias is more severe with more complex DGPs and works against the researcher rejecting no effect. Note that L1 is only stated as a benchmark: because L1 is an index and its components are controlled for, the model is not identified in the structural equations. Hence, as observed in the first column, α cannot be recovered regardless of the model used.

Overall, MLC2 is still remarkably able to recover the coefficient across DGPs and better than in any other procedure, including MLC2y. Even in the case of L1y, where MLC2 performs the worst, the coefficient is significantly different from zero. By contrast, all "y" models can have insignificant coefficients or the wrong sign if the true model is different. MLC1 and MLC2y reasonably perform at this exercise but tend to feature more bias; especially, the attenuation is particularly large when MLC2y is used against a true DGP based on MLC2. These findings are consistent with the main insight from Table 11.

In Table 13, we also simulate the model with non-zero errors ϵ_i, sampling with a normal distribution with a standard-error equal to five (which roughly matches cross-sectional variation due to errors and due to observables in a linear model). We then examine the fraction of draws such that the estimated parameter is different from zero. On this metric, MLC2 and MLC2y appear the most robust at correctly signing the parameter. MLC2y is better at recovering a positive coefficient against L2y while MLC2 is better at recovering positive coefficients against MLC2y than vice-versa. L2y comes as a distant third, and performs poorly when the DGP is MLC1, MLC2 and MLC2y.

The above analysis focuses on type II errors (false acceptance of no effect). We repeat this analysis to assess the probability that the models incorrectly rejects $\alpha = 0$, or type I errors, by simulating equation (3.1) with α set to zero and measuring the effect (Table 14) as well as the frequency that the model incorrectly rejects a true $\alpha = 0$ (Table 15). All models perform extremely well at this task, and cases in which an effect is found despite no effect in the true DGP are the exception rather than the norm (except for the non-identified DGP in L1). Interestingly, for this type of error, the linear model can occasionally outperform machine learning because the attenuation bias caused by misspecification tends to produce estimates that are close to zero.

3.3 An application to the cost of capital

We develop next a case study showing how a machine learning model can be used to refine our understanding of the economic effects of conservatism documented in a prior literature. Conservatism is often a mean to mitigate information asymmetries and facilitate debt contracting. Firms experiencing difficulties in levying debt tend to be more conservative. Lara et al. (2016) hypothesize that conservatism helps to alleviate underinvestment because it enables firms endowed with profitable investment opportunities to finance them. Conversely, conservatism can also discipline firms that tend to overinvest by singling out in a more timely fashion unprofitable projects. To test their main hypothesis, Lara et al. (2016) construct a measure of conservatism based on the total timeliness of bad news recognition in the lines of KW. They define a firm-year measure CONS, the conservatism proxy, as the annual decile ranks of the three-year average (years t, $t-1$, and $t-2$) of the total timeliness of loss recognition (G-Score+C-Score).

They adapt the method of Biddle et al. (2009) to test the effect of conservatism on investment as follows:

$$Investment_{t+1} = \beta_t + \delta_1 CONS_t + \delta_2 CONS_t \times UnderInvest_t + \delta_3 UnderInvest_t$$

$$+ \delta_4 FRQ_t + \delta_5 FRQ_t \times UnderInvest_t$$

$$+ \delta_6 GOV_t + \delta_7 GOV_t \times UnderInvest_t + \gamma Controls_t + \epsilon_{t+1} \qquad (3.2)$$

In Equation (3.2), Investment is a measure of total future investment in both capital and research and development.[23] UnderInvest is an industry-year ranked variable capturing settings where under- or overinvestment is more likely. FRQ is a measure of financial reporting quality. GOV is a set of governance variables, and Controls is a vector of control variables that affect investment and financing. Further, Lara et al. (2016) predict and find supporting evidence that the effect of conservatism on investment should be the strongest in environments featuring higher information asymmetries. The additional regression includes different proxies for information asymmetries and

---

[23]To be more precise, Investment is a measure of total investment calculated as capital expenditures plus research and development plus acquisition expenditures less cash receipts from sales of PPE, multiplied by 100, and scaled by lagged sales.

their interaction with the measure CONS:

$$
\begin{aligned}
Investment_{t+1} = {} & \beta_t + \delta_1 CONS_t + \delta_2 CONS_t \times InfoAsym_t \\
& + \delta_3 CONS_t \times UnderInvest_t \\
& + \delta_4 CONS_t \times UnderInvest_t \times InfoAsym_t + \gamma Controls_t + \epsilon_{t+1} \qquad (3.3)
\end{aligned}
$$

Information asymmetry is proxied with three different measures. First, a market-based measure is derived as a continuous variable computed as the average of the standardized values of bid-ask spread, volatility, and idiosyncratic risk. Second, information asymmetry is measured with a dummy variable that takes the value of 1 if the firm has no credit rating (NCR). Finally, the variable young is the negative of the difference between the first year when the firm appears in CRSP and the current year.

We replicate the main regression in Lara et al. (2016) Table 2 (equation 3.2) and the secondary regression Table 3 to further test the incremental role of information asymmetry on investment and conservatism (equation 3.3), but in place of their CONS measure, we use the different conservative measures derived and tested in our study.

Table 16 reports the results of the regression in 3.2. Across all measures with the exception of L2y, similar to Lara et al. (2016), the coefficients δ_1 and δ_2 are respectively negative and positive significant at the 1% level. The coefficients in L2y model have the same signs as in Lara et al. (2016) but they are insignificant. This lack of power can be attributed to the volatility in the measure. All machine learning measures of conservatism have a larger impact on investment relatively to the linear measures, potentially because they are better apt at extracting the noise in the data. They also seem to capture similar underlying factors to differential timeliness: firms facing potential underinvestment issues tend to be more conservative. Conservatism also reduces overinvestment as implied by $\delta_1 > 0$.

Tables 17-19 evaluate the incremental effect of information asymmetry. Lara et al. (2016) document that the coefficient δ_4 in Equation (3.3) on the three-way interaction among their conservative measure, UnderInvest, and InfoAsym is positive and significant. This finding supports that the effect of conservatism is stronger when underinvestment is more likely to occur and information asymmetry is large. The machine learning measures return significant positive coefficients but with

levels of significance that vary across the information asymmetry proxies. These lower levels of significance compared to L1y (systematically with a highly significant positive coefficient) suggests that the machine learning measures take into account more features than a standard linear measure.

4   Concluding Remarks

Measuring characteristics of accounting information is notoriously difficult because, as researcher, one only observes the joint result of inherent economics and the choices made in the accounting system (Dechow et al. 2010). Conservatism reliably presents all the features of this empirical challenge: new transactions are continuously being recorded and affected by the economic environment. Isolating the effect of conservatism requires to gauge whether these transactions are within the scope of conservative rules defined by GAAP, the level of verifiability and, more broadly, the level of certainty about future cash flows. The discretion that accounting policies offer in choices of conservatism can only be extracted after these economic factors have been satisfactorily included in the model. Machine learning is a statistical tool that, on its own, is not a complete toolbox to solve this non-trivial identification challenge - presumably, the identification could only be resolved in principle via better consensus over the data-generating process of economic transactions in a structural model. However, machine learning provides new benefits in using observable economic variables more thoroughly given its ability to fit a flexible model that does not presume particular interactions or functional forms.

Our main result is that a parsimonious neural network with relatively few variables provides a proxy for differential timeliness with a fit equivalent to standard models estimated with separate yearly regressions. We cross-validate the hyperparameters of the network and provide routines that allow for a simple recovery of the measure with limited computing needs. The measure provides three primary benefits. First, the measure is in terms of interpretable economic factors widely-used across studies, and avoids the difficulties in interpreting average coefficients in separate regressions. Second, the measure offers a new look at long-term trends in conservatism, revealing a secular decrease in conservatism over most of the time periods following the FASB conceptual statements. Third, the proxy exhibits new interactions and convexities that were previously unknown to researchers, and may be used in extended tests of economic theories that may have been insignificant

due to lack of power.

The analysis is part of a broader research agenda, showing that machine learning tools can be used beyond prediction tasks. Refined algorithms provide new methods to filter noise when the relation between signal and noise is not fully captured by a closed-form statistical model. This is particularly important when, as is common in accounting research, most theories do not fully specify the correct empirical research design and researchers may argue about functional forms, interactions or variable inclusions. We illustrate this objective by improving a standard approach to differential timeliness: the same approach can nevertheless be used to improve any noisy proxy, such as accounting quality (Barth et al. 2019), cost of capital (Lambert et al. 2007, 2012; Bertomeu and Cheynel 2016) or accruals (Dechow et al. 1995; Armstrong et al. 2016). Given the relative young age of these methods in social sciences, there does not yet exist a consensus on methodology and whether the tools will dramatically improve empirical inference; yet, their growing use nevertheless presents an exciting domain of applications in accounting.

Appendix

Name	Definition	Formula
Notes: Subscript '-n' refers to the value of the variable n years in the past. The formulas use the names of the original variables from the WRDS data sources.		
COMPUSTAT		
Earning	Net income before extraordinary items deflated by lagged market value of equity.	$\frac{ib}{(prcc\_f*csho)_{-1}}$
Size	Natural logarithm of the market value of equity.	$log(prcc\_f * csho)$
Size (defl.)	Natural logarithm of the market value of equity corrected for inflation.	$log\left(\frac{prcc\_f*csho}{gdpdef}\right)$ where $gdpdef$ is the GDP price deflator series.[a]
M/B Ratio	Market to book ratio. Ratio of market value of equity to book value of equity.	$\frac{prcc\_f*csho}{ceq}$
Leverage	Financial leverage. Long-term debt plus short term debt deflated by market value of equity.	$\frac{dltt+dlc}{prcc\_f*csho}$
NOACC	Non-Operating Accruals scaled by lagged assets.	$\frac{(ib+dp)-cfo-oacc}{at_{-1}}$
CFOA	Cash Flow from operating activities scaled by lagged assets.	$\frac{oancf}{at_{-1}}$ if $fyear \geq 1987$, $\frac{fopt-oacc}{at_{-1}}$ if $fyear < 1987$
InvCycle	Investment Cycle. Depreciation expenses deflated by lagged assets.	Standard deviation of $\frac{capx}{at}$ over the past 5 years (3 years minimum)
CRSP		
Return	Compounded Monthly Returns over the past eight months (beginning four months after the previous fiscal year end).	$exp(\sum_{m=-8}^{0} ln(1+ret_m))-1$ where ret_{-n} is the monthly return of the nth previous month.
Volatility	Standard deviation of daily returns over the past year.	$std(ret_d)$ where ret_d is the series of daily returns over the past 253 trading days.
Age	Age of the firm in years. Number of years with return history on CRSP or number of years in Compustat if not available.	
Bid-Ask Spread	Average of the daily bid-ask spread scaled by the midpoint of the spread over the past year.	$\frac{1}{N}\sum_{d=-253}^{0} \frac{ask_d-bid_d}{\frac{ask_d+bid_d}{2}}$ were ask_{-n} is the ask price of the dth previous day.
OTHER		
Problit	Probability of Litigation.	As computed in Shu (2000)
PIN	Probability of Informed Trade.	From Easley et al model. Obtained from Stephen Brown website.[b] Data available annually from 1993 to 2010.

Table 1: Description of the variables.

[a]The GDPDEF series is compiled by the U.S. Bureau of Economic Analysis and available from the FRED, Federal Reserve Bank of St. Louis under the series name 'GDPDEF' (https://fred.stlouisfed.org/series/GDPDEF).

[b]The PIN data can be obtained here: https://scholar.rhsmith.umd.edu/sbrown/probability-informed-trade-easley-et-al-model.

	Firm-Year obs.	Unique Firms
COMPUSTAT data	537,971	40,951
Merge with CRSP data	331,097	29,054
Non-Null Earnings, Size, M/B and Leverage	253,314	22,809
Non-Null Returns	246,761	22,606
Data between fiscal year 1963 and December 2019	242,603	22,367
Price per share ≥ $1	229,387	21,392

Table 2: Data selection.

	Timespan	Models	Firm-year obs.	Unique firms	Features
Full Sample	1963 - 2019	L1, L1y MLC1, MLC1y	229,387	21,392	Size, M/B ratio, Leverage
Restricted Sample	1963 - 2019	L2, L2y MLC2, MLC2y	185,581	17,645	Size, M/B ratio, Leverage Volatility, NOAcc, CFOA, Inv. Cycle, Age

Table 3: Data selection for models 1 and 2.

	count	mean	std	min	25%	50%	75%	max
Earnings	229,387	0.022	0.212	-2.797	0.002	0.057	0.101	1.150
Return	229,387	0.181	0.828	-0.962	-0.258	0.041	0.385	12.905
Size	229,387	5.246	2.238	-0.098	3.571	5.094	6.780	11.533
M/B ratio	229,387	2.503	4.947	-60.220	0.953	1.582	2.780	92.052
Leverage	229,387	0.831	1.623	0.000	0.056	0.299	0.870	22.362
Volatility	227,101	0.032	0.018	0.002	0.018	0.027	0.040	0.144
NOAcc	187,623	-0.019	0.114	-3.571	-0.032	-0.008	0.006	0.708
CFOA	213,144	0.044	0.204	-4.809	0.006	0.069	0.132	0.612
Inv. Cycle	218,528	0.044	0.037	0.000	0.021	0.037	0.057	0.634
Age	229,387	17.953	15.124	0.762	7.005	13.011	24.016	94.063
Problit	157,367	0.031	0.145	0.000	0.001	0.002	0.005	1.000
PIN	99,173	0.232	0.134	0.000	0.135	0.201	0.297	1.000
Bid-Ask Spread	146,025	0.023	0.032	0.000	0.002	0.011	0.031	0.281

Table 4: Descriptive Statistics of the full sample.

	Earnings	Return	Size	M/B ratio	Leverage	Volatility	NOAcc	CFOA	Inv. Cycle	Age	Problit	PIN	Bid-Ask Spread
Earnings		0.080*	0.112*	-0.033*	-0.166*	-0.387*	0.287*	0.337*	-0.090*	0.085*	-0.223*	-0.031*	-0.120*
Return	0.304*		0.072*	0.160*	-0.126*	0.079*	-0.001	0.049*	0.031*	-0.032*	-0.090*	-0.049*	0.008*
Size	0.048*	0.172*		0.128*	-0.142*	-0.416*	-0.005*	0.192*	0.008*	0.387*	-0.085*	-0.574*	-0.641*
M/B ratio	-0.137*	0.274*	0.369*		-0.137*	0.054*	-0.049*	-0.070*	0.046*	-0.035*	-0.038*	-0.130*	-0.056*
Leverage	0.124*	-0.139*	-0.100*	-0.476*		0.078*	-0.002	-0.021*	-0.081*	-0.014*	0.721*	0.117*	0.071*
Volatility	-0.377*	-0.177*	-0.447*	-0.014*	-0.112*		-0.159*	-0.290*	0.148*	-0.301*	0.160*	0.084*	0.569*
NOAcc	0.195*	0.025*	-0.071*	-0.093*	0.057*	-0.119*		0.127*	-0.089*	0.076*	-0.048*	0.035*	0.009*
CFOA	0.386*	0.180*	0.261*	0.129*	-0.125*	-0.243*	-0.149*		0.089*	0.140*	-0.082*	-0.069*	-0.108*
Inv. Cycle	-0.075*	-0.011*	0.029*	0.114*	-0.047*	0.126*	-0.056*	0.291*		-0.048*	0.021*	-0.125*	0.059*
Age	0.151*	0.059*	0.340*	-0.068*	0.102*	-0.362*	0.082*	0.149*	0.006*		-0.040*	-0.208*	-0.231*
Problit	0.021*	-0.184*	0.364*	-0.127*	0.557*	-0.092*	-0.042*	0.010*	0.017*	0.106*		0.017*	0.031*
PIN	0.006*	-0.032*	-0.645*	-0.304*	0.123*	0.109*	0.075*	-0.187*	-0.143*	-0.221*	-0.378*		0.507*
Bid-Ask Spread	-0.074*	-0.066*	-0.801*	-0.215*	0.023*	0.516*	0.069*	-0.211*	0.058*	-0.352*	-0.408*	0.651*	

Table 5: Pearson (top) and Spearman (bottom) correlation matrix.

	Full Sample		Replication		Khan and Watts (2009)	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Intercept	0.046	3.797	0.067	4.554	0.083	7.530
D	-0.010	-1.583	0.004	0.692	-0.024	-3.560
Ret	-0.035	-2.994	0.024	3.775	0.031	1.840
Ret x Size	0.007	5.749	0.000	0.333	0.005	2.250
Ret x M/B	-0.001	-0.556	0.000	0.211	-0.006	-2.000
Ret x Lev	-0.003	-0.851	-0.005	-1.304	0.005	0.770
D x Ret	0.417	10.731	0.234	12.397	0.237	10.780
D x Ret x Size	-0.053	-10.052	-0.026	-7.052	-0.033	-7.420
D x Ret x M/B	-0.006	-1.722	-0.012	-3.774	-0.007	-0.930
D x Ret x Lev	0.050	7.122	0.026	2.354	0.033	1.860
Size	0.007	7.648	0.007	6.241	0.005	4.830
M/B	-0.010	-4.788	-0.016	-5.321	-0.017	-7.930
Lev	-0.006	-3.253	-0.011	-4.736	-0.008	-3.610
D x Size	0.001	0.522	-0.001	-0.975	0.003	3.450
D x M/B	0.000	0.214	0.001	0.752	-0.001	-0.420
D x Lev	-0.002	-1.096	-0.007	-2.847	-0.002	-0.880
R2	0.231		0.220		0.240	

Table 6: Replication of Table 3 of Khan and Watts (2009).

	L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
R squared								
Training Sample	0.115	0.256	0.218	0.321	0.315	0.449	0.482	0.554
Test Sample	0.107	0.229	0.204	0.290	0.315	0.405	0.455	0.500
Full Sample	0.114	0.252	0.215	0.314	0.316	0.445	0.477	0.544
Number of coefficients								
Full Model	16	912	504	9924	36	2052	1604	13404
C score	4	228	126	2481	9	513	401	3351

Table 7: Fits.

	L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y	Size	M/B ratio	Leverage	Volatility	Problit	PIN	Bid-Ask Spread
L1		0.718*	0.668*	0.620*	0.805*	0.522*	0.508*	0.475*	-0.860*	-0.426*	0.524*	0.376*	0.298*	0.536*	0.534*
L1y	0.735*		0.502*	0.782*	0.578*	0.599*	0.340*	0.446*	-0.539*	-0.223*	0.491*	0.287*	0.306*	0.422*	0.286*
MLC1	0.776*	0.588*		0.578*	0.642*	0.414*	0.585*	0.388*	-0.561*	-0.269*	0.420*	0.151*	0.184*	0.438*	0.399*
MLC1y	0.628*	0.762*	0.668*		0.587*	0.459*	0.427*	0.656*	-0.363*	-0.238*	0.624*	0.149*	0.367*	0.331*	0.194*
L2	0.785*	0.573*	0.713*	0.603*		0.549*	0.599*	0.601*	-0.654*	-0.420*	0.521*	0.051*	0.277*	0.416*	0.336*
L2y	0.562*	0.535*	0.502*	0.416*	0.604*		0.346*	0.499*	-0.471*	-0.163*	0.348*	0.162*	0.193*	0.307*	0.313*
MLC2	0.587*	0.422*	0.651*	0.481*	0.760*	0.473*		0.525*	-0.480*	-0.140*	0.328*	0.048*	0.078*	0.317*	0.321*
MLC2y	0.426*	0.330*	0.422*	0.494*	0.672*	0.460*	0.624*		-0.352*	-0.155*	0.582*	0.023*	0.356*	0.204*	0.222*
Size	-0.901*	-0.579*	-0.634*	-0.403*	-0.678*	-0.514*	-0.537*	-0.364*		0.128*	-0.142*	-0.416*	-0.085*	-0.574*	-0.641*
M/B ratio	-0.571*	-0.411*	-0.679*	-0.588*	-0.606*	-0.387*	-0.524*	-0.451*	0.369*		-0.137*	0.054*	-0.038*	-0.130*	-0.056*
Leverage	0.331*	0.283*	0.468*	0.478*	0.398*	0.302*	0.459*	0.450*	-0.100*	-0.476*		0.078*	0.721*	0.117*	0.071*
Volatility	0.405*	0.323*	0.182*	0.113*	0.074*	0.192*	0.044*	-0.074*	-0.447*	-0.014*	-0.112*		0.160*	0.084*	0.569*
Problit	-0.187*	-0.114*	-0.009*	0.086*	-0.081*	-0.069*	0.034*	0.084*	0.364*	-0.127*	0.557*	-0.092*		0.017*	0.031*
PIN	0.617*	0.531*	0.491*	0.407*	0.454*	0.412*	0.340*	0.173*	-0.645*	-0.304*	0.123*	0.109*	-0.378*		0.507*
Bid-Ask Spread	0.690*	0.441*	0.490*	0.234*	0.424*	0.501*	0.324*	0.192*	-0.801*	-0.215*	0.023*	0.516*	-0.408*	0.651*	

Table 8: Correlation matrix of the measures (Pearson top and Spearman bottom)

					C Score					
	L1	L1y			MLC1			MLC1y		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Size	-0.083***	-0.161***	-0.174***	-0.162***	-0.039***	-0.111***	-0.044***	-0.050***	-0.183***	-0.036***
	(0.000)	(0.000)	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
M/B ratio	-0.036***	-0.019***	-0.017***	-0.026***	-0.007***	-0.004***	-0.040***	-0.012***	-0.010***	-0.052***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Leverage	0.027***	0.063***	0.082***	0.085***	0.022***	0.046***	-0.005***	0.047***	0.066***	0.052***
	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Volatility	-0.000***	0.001*	0.011***	0.000	-0.001***	0.009***	0.015***	-0.003***	-0.005***	0.033***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
NOAcc	0.000***	-0.002***	-0.001	-0.002***	0.002***	0.003***	0.002***	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFOA	0.000***	-0.005***	-0.006***	-0.005***	0.005***	0.005***	0.006***	0.000**	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Inv. Cycle	-0.000***	-0.002***	-0.002***	-0.002***	-0.003***	-0.006***	-0.003***	-0.003***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	-0.000***	0.008***	0.026***	0.008***	0.004***	0.003***	0.002***	0.010***	0.017***	0.007***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Problit	0.000***	-0.000	-0.020***	0.000	-0.006***	0.022***	-0.004***	-0.004***	-0.007***	-0.003***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
Bid-Ask Spread	-0.000***	-0.046***	-0.095***	-0.045***	0.007***	-0.003***	0.005***	-0.010***	-0.055***	-0.019***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
$Size^2$			0.004*			0.069***			0.120***	
			(0.002)			(0.001)			(0.001)	
$M/B\ ratio^2$			-0.001*			-0.004***			-0.002***	
			(0.000)			(0.000)			(0.000)	
$Leverage^2$			-0.023***			-0.032***			-0.024***	
			(0.001)			(0.000)			(0.000)	
$Volatility^2$			-0.010***			-0.011***			-0.000	
			(0.001)			(0.000)			(0.001)	
$NOAcc^2$			0.001			0.001***			0.001***	
			(0.000)			(0.000)			(0.000)	
$CFOA^2$			0.001			-0.001***			-0.001***	
			(0.000)			(0.000)			(0.000)	
$Inv.\ Cycle^2$			0.002**			0.003***			0.002***	
			(0.001)			(0.000)			(0.000)	
Age^2			-0.020***			-0.002***			-0.013***	
			(0.001)			(0.000)			(0.001)	
$Problit^2$			0.021***			-0.025***			0.005***	
			(0.002)			(0.001)			(0.001)	
$Bid\text{-}Ask\ Spread^2$			0.048***			0.002***			0.030***	
			(0.001)			(0.000)			(0.000)	
Size X M/B				0.004***			0.032***			0.034***
				(0.001)			(0.000)			(0.001)
Size X Leverage				-0.007***			0.045***			0.003***
				(0.001)			(0.000)			(0.001)
M/B X Leverage				-0.001*			0.005***			0.004***
				(0.000)			(0.000)			(0.000)
Size X Volatility				0.002***			-0.010***			-0.031***
				(0.001)			(0.000)			(0.000)
M/B X Volatility				0.005***			0.003***			0.009***
				(0.001)			(0.000)			(0.000)
Leverage X Volatility				-0.018***			-0.019***			-0.011***
				(0.001)			(0.000)			(0.001)
Intercept	0.173***	0.160***	0.160***	0.160***	0.125***	0.125***	0.125***	0.153***	0.153***	0.153***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967
R^2	1.000	0.677	0.692	0.678	0.556	0.674	0.659	0.497	0.574	0.523

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Regression of the measures L1, L1y, MLC1, MLC1y on firms characteristics. Explanatory variables are standardized.

	C Score									
	L2	L2y			MLC2			MLC2y		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Size	-0.084***	-0.125***	-0.055***	-0.129***	-0.010***	-0.026***	-0.010***	-0.012***	-0.032***	-0.010***
	(0.000)	(0.001)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
M/B ratio	-0.025***	-0.002***	0.002***	-0.010***	0.001***	0.000***	-0.007***	-0.001***	-0.004***	-0.008***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Leverage	0.030***	0.038***	0.065***	0.081***	0.008***	0.017***	0.017***	0.020***	0.024***	0.041***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Volatility	-0.017***	-0.026***	-0.005***	-0.036***	-0.001***	0.003***	0.001***	-0.001***	-0.001**	-0.001***
	(0.000)	(0.001)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NOAcc	-0.018***	-0.029***	-0.037***	-0.029***	0.001***	0.003***	0.001***	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFOA	0.021***	0.017***	0.018***	0.018***	0.002***	0.004***	0.002***	0.001***	0.001***	0.000***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Inv. Cycle	-0.021***	0.020***	0.034***	0.019***	-0.004***	-0.006***	-0.004***	-0.001***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.045***	0.032***	0.047***	0.034***	0.006***	0.014***	0.006***	0.017***	0.022***	0.017***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Problit	0.000***	0.000	-0.019***	0.001	-0.004***	-0.003***	-0.004***	0.001***	-0.006***	0.001***
	(0.000)	(0.001)	(0.003)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Bid-Ask Spread	0.000***	0.007***	0.018***	0.010***	0.003***	0.000	0.002***	0.005***	0.001***	0.004***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$Size^2$			-0.062***			0.016***			0.019***	
			(0.003)			(0.000)			(0.000)	
$M/B\ ratio^2$			-0.006***			0.002***			0.006***	
			(0.001)			(0.000)			(0.000)	
$Leverage^2$			-0.032***			-0.012***			-0.005***	
			(0.001)			(0.000)			(0.000)	
$Volatility^2$			-0.017***			-0.003***			-0.000	
			(0.002)			(0.000)			(0.000)	
$NOAcc^2$			-0.015***			0.003***			-0.001***	
			(0.001)			(0.000)			(0.000)	
$CFOA^2$			0.009***			0.002***			-0.000	
			(0.001)			(0.000)			(0.000)	
$Inv.\ Cycle^2$			-0.019***			0.002***			0.001***	
			(0.001)			(0.000)			(0.000)	
Age^2			-0.011***			-0.009***			-0.006***	
			(0.001)			(0.000)			(0.000)	
$Problit^2$			0.022***			0.000			0.007***	
			(0.003)			(0.000)			(0.000)	
$Bid\text{-}Ask\ Spread^2$			-0.004***			0.000***			0.001***	
			(0.001)			(0.000)			(0.000)	
Size X M/B				0.001			0.007***			0.006***
				(0.002)			(0.000)			(0.000)
Size X Leverage				-0.016***			-0.002***			-0.018***
				(0.002)			(0.000)			(0.000)
M/B X Leverage				-0.004***			0.000***			-0.003***
				(0.001)			(0.000)			(0.000)
Size X Volatility				0.012***			0.000			0.000
				(0.001)			(0.000)			(0.000)
M/B X Volatility				0.010***			0.001***			0.002***
				(0.001)			(0.000)			(0.000)
Leverage X Volatility				-0.030***			-0.008***			-0.004***
				(0.001)			(0.000)			(0.000)
Intercept	-0.003***	-0.016***	-0.016***	-0.016***	0.014***	0.014***	0.014***	0.024***	0.024***	0.024***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967	104,967
R^2	1.000	0.389	0.408	0.394	0.405	0.549	0.432	0.533	0.562	0.561

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Regression of the measures L2, L2y, MLC2, MLC2y on firms characteristics. Explanatory variables are standardized.

		DGP							
		L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
Estimation	L1y	0.422***	0.223***	1.044***	0.785***	1.448***	0.771***	4.949***	2.106***
		[163.536]	[179.337]	[299.0]	[264.396]	[400.423]	[609.142]	[518.484]	[550.636]
	MLC1	0.287***	0.528***	0.16***	0.469***	1.458***	0.873***	4.454***	1.734***
		[126.712]	[876.061]	[53.258]	[429.501]	[436.168]	[873.679]	[430.505]	[403.679]
	MLC1y	0.472***	0.356***	0.529***	0.33***	1.26***	0.771***	3.521***	1.479***
		[119.118]	[313.932]	[87.668]	[79.777]	[306.028]	[596.826]	[303.83]	[278.008]
	L2	0.061***	0.474***	0.561***	0.594***	0.062***	0.555***	2.025***	0.793***
		[38.566]	[921.03]	[234.252]	[349.548]	[30.052]	[2231.862]	[311.836]	[357.783]
	L2y	0.689***	0.325***	1.199***	0.847***	0.648***	0.304***	3.299***	1.432***
		[235.515]	[254.508]	[274.604]	[294.085]	[261.234]	[239.56]	[363.851]	[406.629]
	MLC2	0.333***	0.524***	0.273***	0.462***	0.441***	0.584***	0.505***	0.408***
		[111.492]	[641.096]	[68.725]	[243.505]	[151.695]	[1017.629]	[48.984]	[192.465]
	MLC2y	0.571***	0.411***	0.644***	0.402***	0.562***	0.551***	0.931***	0.376***
		[73.765]	[226.23]	[73.954]	[64.707]	[148.565]	[638.046]	[43.564]	[57.685]

Estimated using the average over 100 draws. All errors are normalized by the standard deviation of the DGP measure. t-statistics in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 11: Mean Absolute Errors (MAE) between the true C score (DGP) and the estimated one.

		DGP							
		L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
Estimation	L1y	0.0**	2.484***	0.276***	1.082***	-0.02***	1.633***	-0.053***	0.153***
		[2.018]	[230.96]	[60.476]	[160.818]	[-7.79]	[181.13]	[-58.125]	[46.148]
	MLC1	-0.0	0.107***	3.01***	1.531***	0.288***	0.549***	0.274***	0.268***
		[-0.297]	[8.102]	[206.36]	[132.747]	[15.103]	[37.712]	[47.134]	[38.091]
	MLC1y	-0.0	2.413***	1.39***	2.007***	0.1***	1.193***	-0.008***	0.386***
		[-0.807]	[125.625]	[58.224]	[107.935]	[16.528]	[67.092]	[-3.712]	[62.002]
	L2	0.0	0.956***	1.775***	0.709***	3.015***	2.299***	0.332***	1.264***
		[0.446]	[44.757]	[49.845]	[93.809]	[213.681]	[217.018]	[131.687]	[188.003]
	L2y	0.0	1.933***	0.082***	0.636***	0.586***	2.452***	0.026***	0.367***
		[0.629]	[197.445]	[39.644]	[111.926]	[103.649]	[232.309]	[30.412]	[113.72]
	MLC2	-0.0	0.387***	2.663***	1.479***	2.641***	2.562***	1.638***	1.916***
		[-0.941]	[25.188]	[114.061]	[60.376]	[129.764]	[57.879]	[30.438]	[37.26]
	MLC2y	-0.0	1.925***	0.909***	1.634***	1.222***	2.298***	0.571***	1.809***
		[-0.67]	[70.23]	[35.034]	[59.57]	[33.008]	[85.861]	[22.549]	[59.236]

Estimated using the average over 100 draws. The C score coefficient used to generate the data equals 3. t-statistics in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 12: Average C score coefficients found using an OLS regression.

		DGP							
		L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
Estimation	L1y	0.020	1.000	0.410	1.000	0.010	1.000	0.030	0.190
	MLC1	0.010	0.030	1.000	1.000	0.010	0.160	0.050	0.070
	MLC1y	0.030	1.000	1.000	1.000	0.010	1.000	0.020	0.790
	L2	0.010	0.370	0.420	0.170	1.000	1.000	0.090	0.980
	L2y	0.010	1.000	0.050	1.000	1.000	1.000	0.020	0.960
	MLC2	0.000	0.160	1.000	0.960	1.000	1.000	0.710	0.950
	MLC2y	0.000	1.000	0.990	1.000	1.000	1.000	0.380	1.000

Estimated using the average over 100 draws. The C score coefficient used to generate the data equals 3. t-statistics in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 13: Fraction of significant coefficients found using an OLS regression.

		DGP							
		L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
Estimation	L1y	0.0**	0.0***	0.0	0.0	-0.0*	-0.0	-0.0**	-0.0
		[1.917]	[2.898]	[0.592]	[0.154]	[-1.402]	[-0.22]	[-2.008]	[-0.35]
	MLC1	-0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0**
		[-0.447]	[-0.694]	[0.434]	[0.072]	[1.172]	[0.314]	[0.036]	[1.731]
	MLC1y	-0.0	0.0*	-0.0	0.0	0.0	-0.0	0.0	0.0*
		[-0.772]	[1.514]	[-0.859]	[0.547]	[0.424]	[-0.964]	[0.372]	[1.555]
	L2	0.0	0.0**	-0.0	-0.0*	0.0	-0.0	0.0	-0.0
		[0.925]	[1.75]	[-1.018]	[-1.476]	[0.155]	[-0.973]	[0.644]	[-0.712]
	L2y	0.0	0.0*	-0.0	0.0	-0.0**	-0.0	0.0	0.0
		[0.46]	[1.585]	[-0.152]	[0.975]	[-1.821]	[-0.979]	[0.881]	[0.176]
	MLC2	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0
		[-0.996]	[-0.959]	[-1.133]	[-0.573]	[-0.258]	[-0.097]	[-0.078]	[-1.227]
	MLC2y	-0.0	-0.0***	-0.0	-0.0*	-0.0	-0.0**	0.0	0.0
		[-0.733]	[-2.675]	[-0.558]	[-1.504]	[-0.682]	[-1.842]	[0.247]	[0.139]

Estimated using the average over 100 draws. The C score coefficient used to generate the data equals 0. t-statistics in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 14: Average C score coefficients found using an OLS regression.

		DGP							
		L1	L1y	MLC1	MLC1y	L2	L2y	MLC2	MLC2y
Estimation	L1y	0.020	0.010	0.020	0.000	0.020	0.010	0.000	0.020
	MLC1	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.010
	MLC1y	0.000	0.010	0.020	0.000	0.030	0.000	0.010	0.010
	L2	0.010	0.000	0.030	0.000	0.000	0.020	0.000	0.000
	L2y	0.000	0.020	0.020	0.020	0.010	0.000	0.010	0.010
	MLC2	0.000	0.000	0.010	0.010	0.010	0.030	0.000	0.000
	MLC2y	0.010	0.000	0.020	0.020	0.020	0.000	0.030	0.010

Estimated using the average over 100 draws. The C score coefficient used to generate the data equals 0. t-statistics in parentheses. $^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 15: Fraction of significant coefficients found using an OLS regression.

	Investment$_{t+1}$					
	L1y	MLC1	MLC1y	L2y	MLC2	MLC2y
	(1)	(2)	(3)	(4)	(5)	(6)
CONS	−5.413***	−11.556***	−12.326***	−4.411	−7.490***	−5.265**
	(−3.227)	(−8.213)	(−2.715)	(−1.046)	(−8.274)	(−2.539)
CONS*UnderInvest	12.334***	15.190***	17.087***	9.285	13.752***	15.297***
	(5.463)	(7.448)	(4.112)	(1.572)	(8.271)	(6.227)
CONS + CONS*UnderInvest	6.921***	3.634***	4.761***	4.874**	6.262***	10.032***
	(4.181)	(3.333)	(3.479)	(2.189)	(6.861)	(5.438)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,586	41,586	41,586	41,584	41,584	41,584
R^2	0.338	0.341	0.340	0.338	0.339	0.338

Note: *p<0.1; **p<0.05; ***p<0.01

CONS is the conservatism measure of our different models (L1y to MLC2y). t-statistics are in parenthesis.

Table 16: Association between future investment and accounting conservatism.

	Investment$_{t+1}$					
	L1y	MLC1	MLC1y	L2y	MLC2	MLC2y
	(1)	(2)	(3)	(4)	(5)	(6)
CONS	−6.108**	−10.709***	−10.173**	−2.325	−6.876***	−4.414**
	(−2.640)	(−7.939)	(−2.453)	(−0.595)	(−6.338)	(−2.036)
CONS*IA	−7.702***	−4.864***	−5.540***	−4.756**	−1.902*	−0.713
	(−4.326)	(−4.580)	(−3.434)	(−2.385)	(−1.917)	(−0.594)
CONS + CONS*IA	−13.809***	−15.574***	−15.713***	−7.081	−8.777***	−5.127*
	(3.668)	(7.693)	(3.349)	(1.355)	(7.056)	(1.995)
CONS*UnderInvest	15.844***	15.054***	15.157***	6.312	12.923***	13.595***
	(5.795)	(7.093)	(4.092)	(1.176)	(7.008)	(5.062)
CONS*UnderInvest*IA	8.788***	7.528***	8.253***	7.408**	3.392**	1.134
	(3.031)	(4.742)	(3.507)	(2.608)	(2.464)	(0.536)
CONS + CONS*IA + CONS*UnderInvest + CONS*UnderInvest*IA	10.822***	7.008***	7.698***	6.639**	7.538***	9.601***
	(4.518)	(4.037)	(4.518)	(2.521)	(6.143)	(3.810)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,586	41,586	41,586	41,584	41,584	41,584
R^2	0.342	0.344	0.343	0.340	0.340	0.339

Note: *p<0.1; **p<0.05; ***p<0.01

CONS is the conservatism measure of our different models (L1y to MLC2y). t-statistics are in parenthesis.

Table 17: Effect of information asymmetry. Proxy: IA.

	Investment$_{t+1}$					
	L1y	MLC1	MLC1y	L2y	MLC2	MLC2y
	(1)	(2)	(3)	(4)	(5)	(6)
CONS	−2.373	−8.744***	−11.155**	−0.040	−9.107***	−3.792
	(−1.363)	(−7.925)	(−2.327)	(−0.010)	(−5.006)	(−1.406)
CONS*NCR	−5.709**	−4.913**	−1.757	−8.247***	2.123	−2.245
	(−2.356)	(−2.522)	(−0.888)	(−3.934)	(1.251)	(−1.230)
CONS + CONS*NCR	−8.082***	−13.657***	−12.912***	−8.287*	−6.984***	−6.037***
	(3.564)	(6.711)	(2.856)	(1.767)	(8.216)	(3.072)
CONS*UnderInvest	7.894***	10.617***	14.512***	3.284	15.678***	12.909***
	(3.262)	(6.594)	(3.515)	(0.640)	(5.889)	(4.516)
CONS*UnderInvest*NCR	8.187**	7.820***	3.888	11.440***	−2.663	3.835
	(2.125)	(2.856)	(1.424)	(3.201)	(−1.141)	(1.287)
CONS + CONS*NCR + CONS*UnderInvest + CONS*UnderInvest*NCR	7.999***	4.780***	5.488***	6.437**	6.030***	10.707***
	(4.278)	(3.647)	(3.488)	(2.524)	(6.782)	(5.334)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,586	41,586	41,586	41,584	41,584	41,584
R^2	0.338	0.341	0.341	0.339	0.339	0.338

Note: *p<0.1; **p<0.05; ***p<0.01

CONS is the conservatism measure of our different models (L1y to MLC2y). t-statistics are in parenthesis.

Table 18: Effect of information asymmetry. Proxy: NCR.

	Investment$_{t+1}$					
	L1y	MLC1	MLC1y	L2y	MLC2	MLC2y
	(1)	(2)	(3)	(4)	(5)	(6)
CONS	−8.042***	−14.042***	−17.207***	−4.949	−8.169***	−4.754*
	(−4.590)	(−7.665)	(−3.401)	(−0.866)	(−6.000)	(−1.824)
CONS*Young	−0.144***	−0.156***	−0.292***	−0.056	−0.071	−0.024
	(−3.752)	(−3.169)	(−4.283)	(−0.598)	(−1.256)	(−0.311)
CONS + CONS*Young	−8.186***	−14.198***	−17.499***	−5.005	−8.240***	−4.779*
	(4.639)	(7.610)	(3.418)	(0.862)	(5.869)	(1.801)
CONS*UnderInvest	13.672***	18.012***	23.040***	9.888	14.704***	14.456***
	(5.072)	(6.192)	(4.497)	(1.158)	(6.099)	(3.952)
CONS*UnderInvest*Young	0.084	0.191**	0.375***	0.099	0.113	0.079
	(1.229)	(2.309)	(3.579)	(0.641)	(1.389)	(0.723)
CONS + CONS*Young + CONS*UnderInvest + CONS*UnderInvest*Young	5.570***	4.005***	5.915***	4.982	6.576***	9.756***
	(2.860)	(2.724)	(3.963)	(1.444)	(5.271)	(3.819)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	41,586	41,586	41,586	41,584	41,584	41,584
R^2	0.339	0.342	0.342	0.338	0.339	0.338

Note: *p<0.1; **p<0.05; ***p<0.01

CONS is the conservatism measure of our different models (L1y to MLC2y). t-statistics are in parenthesis.

Table 19: Effect of information asymmetry. Proxy: Young.

Bibliography

Ahmed, Anwer S, and Scott Duellman (2013) 'Managerial overconfidence and accounting conservatism.' *Journal of Accounting Research* 51(1), 1–30

Anand, Vic, Robert Brunner, Kelechi Ikegwu, and Theodore Sougiannis (2019) 'Predicting profitability using machine learning.' *Available at SSRN 3466478*

Armstrong, Christopher, George Foster, and Daniel Taylor (2016) 'Abnormal accruals in newly public companies: opportunistic misreporting or economic activity?' *Management Science* 62(5), 1316–1338

Athey, Susan (2015) 'Machine learning and causal inference for policy evaluation.' In 'Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining' pp. 5–6

Ball, Ray, and Philip R Brown (2014) 'Ball and Brown (1968): A retrospective.' *Accounting Review* 89(1), 1–26

Ball, Ray, Ashok Robin, and Joanna Shuang Wu (2003) 'Incentives versus standards: properties of accounting income in four east asian countries.' *Journal of Accounting and Economics* 36(1-3), 235–270

Ball, Ray, SP Kothari, and Ashok Robin (2000) 'The effect of international institutional factors on properties of accounting earnings.' *Journal of Accounting and Economics* 29(1), 1–51

Ball, Ray, SP Kothari, and Valeri V Nikolaev (2013a) 'Econometrics of the basu asymmetric timeliness coefficient and accounting conservatism.' *Journal of Accounting Research* 51(5), 1071–1097

_ (2013b) 'On estimating conditional conservatism.' *Accounting Review* 88(3), 755–787

Bao, Yang, Bin Ke, Bin Li, Y Julia Yu, and Jie Zhang (2020) 'Detecting accounting fraud in publicly traded us firms using a machine learning approach.' *Journal of Accounting Research* 58(1), 199–235

Barth, Mary E, Ken Li, and Charles McClure (2019) 'Evolution in value relevance of accounting information'

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

Beaver, William H, and Stephen G Ryan (2005) 'Conditional and unconditional conservatism: Concepts and modeling.' *Review of accounting studies* 10(2), 269–309

Bertomeu, Jeremy (2020) 'Machine learning improves accounting: discussion, implementation and research opportunities.' *Review of Accounting Studies* 25(3), 1135–1155

Bertomeu, Jeremy, and Edwige Cheynel (2016) 'Disclosure and the cost of capital: a survey of the theoretical literature.' *Abacus* 52(2), 221–258

Bertomeu, Jeremy, Anne Beyer, and Daniel J. Taylor (2016) 'From casual to causal inference in accounting research: The need for theoretical foundations.' *Foundations and Trends in Accounting* 10(2-4), 262–313

Bertomeu, Jeremy, Edwige Cheynel, Eric Floyd, and Wenqiang Pan (2020) 'Using machine learning to detect misstatements.' *Review of Accounting Studies* pp. 1–52

Biddle, Gary C, Gilles Hilary, and Rodrigo S Verdi (2009) 'How does financial reporting quality relate to investment efficiency?' *Journal of Accounting and Economics* 48(2-3), 112–131

Billings, Mary Brooke, Robert Jennings, and Baruch Lev (2015) 'On guidance and volatility.' *Journal of Accounting and Economics* 60(2-3), 161–180

Breuer, Matthias, and David Windisch (2019) 'Investment dynamics and earnings-return properties: A structural approach.' *Journal of Accounting Research* 57(3), 639–674

Breuer, Matthias, and Harm H Schütt (2019) 'Accounting for uncertainty: An application of bayesian methods to accruals models.' *Available at SSRN 3417406*

Brown, Stephen, and Stephen A Hillegeist (2007) 'How disclosure quality affects the level of information asymmetry.' *Review of accounting studies* 12(2), 443–477

Chen, Xi, Yang Ha Cho, Yiwei Dou, and Baruch Lev (2020) 'Fundamental analysis of xbrl data: A machine learning approach.' *Available at SSRN*

Dechow, Patricia M, Richard G Sloan, and Amy P Sweeney (1995) 'Detecting earnings management.' *Accounting Review* pp. 193–225

Dechow, Patricia, Weili Ge, and Catherine Schrand (2010) 'Understanding earnings quality: A review of the proxies, their determinants and their consequences.' *Journal of Accounting and Economics* 50(2-3), 344–401

DeFond, Mark L, Chee Yeow Lim, and Yoonseok Zang (2016) 'Client conservatism and auditor-client contracting.' *Accounting Review* 91(1), 69–98

Ding, Kexing, Baruch Lev, Xuan Peng, Ting Sun, and Miklos A Vasarhelyi (2020) 'Machine learning improves accounting estimates: evidence from insurance payments.' *Review of Accounting Studies* 25(3), 1098–1134

Easley, David, and Maureen O'Hara (2004) 'Information and the cost of capital.' *Journal of Finance* 59(4), 1553–1583

Ettredge, Michael, Ying Huang, and Weining Zhang (2012) 'Earnings restatements and differential timeliness of accounting conservatism.' *Journal of Accounting and Economics* 53(3), 489–503

Frankel, Richard, Jared Jennings, and Joshua Lee (2016) 'Using unstructured and qualitative disclosures to explain accruals.' *Journal of Accounting and Economics* 62(2-3), 209–227

Gigler, Frank, Chandra Kanodia, Haresh Sapra, and Raghu Venugopalan (2009) 'Accounting conservatism and the efficiency of debt contracts.' *Journal of accounting research* 47(3), 767–797

Givoly, Dan, and Carla Hayn (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, Dan, Carla K Hayn, and Ashok Natarajan (2007) 'Measuring reporting conservatism.' *Accounting Review* 82(1), 65–106

Goh, Beng Wee, and Dan Li (2011) 'Internal controls and conditional conservatism.' *Accounting Review* 86(3), 975–1005

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009) 'An introduction to statistical learning'

Hu, Hanxin, Ting Sun, Miklos A Vasarhelyi, and Min Zhang (2020) 'A machine learning approach of measuring audit quality: Evidence from china.' *Available at SSRN 3732563*

Huang, Sterling, and Gilles Hilary (2018) 'Zombie board: Board tenure and firm performance.' *Journal of Accounting Research* 56(4), 1285–1329

Hunt, Joshua OS, David M Rosser, and Stephen P Rowe (2020) 'Using machine learning to predict auditor switches: How the likelihood of switching affects audit quality among non-switching clients.' *Journal of Accounting and Public Policy* p. 106785

Jayaraman, Sudarshan (2012) 'The effect of enforcement on timely loss recognition: Evidence from insider trading laws.' *Journal of Accounting and Economics* 53(1-2), 77–97

Jennings, Jared N, Jung Min Kim, Joshua A Lee, and Daniel J Taylor (2020) 'Measurement error and bias in causal models in accounting research.' *Available at SSRN 3731197*

Jones, Stewart (2017) 'Corporate bankruptcy prediction: a high dimensional analysis.' *Review of Accounting Studies* 22(3), 1366–1422

Khan, Mozaffar, and Ross L Watts (2009) 'Estimation and empirical properties of a firm-year measure of accounting conservatism.' *Journal of Accounting and Economics* 48(2-3), 132–150

Kim, Jeong-Bon, and Liandong Zhang (2016) 'Accounting conservatism and stock price crash risk: Firm-level evidence.' *Contemporary accounting research* 33(1), 412–441

Kothari, S.P., Karthik Ramanna, and Douglas J. Skinner (2010) 'Implications for gaap from an analysis of positive research in accounting.' *Journal of Accounting and Economics*

Kronenberger, Sebastian, and Volker Laux (2021) 'Conservative accounting, audit quality, and litigation.' *Management Science*

LaFond, Ryan, and Ross L Watts (2008) 'The information role of conservatism.' *Accounting Review* 83(2), 447–478

Lafond, Ryan, and Sugata Roychowdhury (2008) 'Managerial ownership and accounting conservatism.' *Journal of accounting research* 46(1), 101–135

Lambert, Richard, Christian Leuz, and Robert E. Verrecchia (2007) 'Accounting information, disclosure, and the cost of capital.' *Journal of Accounting Research* 36(2), 385–420

__ (2012) 'Information asymmetry, information precision, and the cost of capital.' *Review of Finance* 16(1), 1–29

Lara, Juan Manuel García, Beatriz García Osma, and Fernando Penalva (2016) 'Accounting conservatism and firm investment efficiency.' *Journal of Accounting and Economics* 61(1), 221–238

Lee, Gene Moo, James P Naughton, Xin Zheng, and Dexin Zhou (2020) 'Predicting litigation risk via machine learning.' *Available at SSRN 3740954*

Li, Feng (2010) 'The information content of forward-looking statements in corporate filings – a naïve bayesian machine learning approach.' *Journal of Accounting Research* 48(5), 1049–1102

Martin, Xiumin, and Sugata Roychowdhury (2015) 'Do financial market developments influence accounting practices? credit default swaps and borrowers? reporting conservatism.' *Journal of Accounting and Economics* 59(1), 80–104

Minuti-Meza, Miguel, and Jake Krupa (2021) 'Regression and machine learning methods to predict discrete outcomes in accounting research'

Ohlson, James A (2001) 'Earnings, book values, and dividends in equity valuation: An empirical perspective.' *Contemporary accounting research* 18(1), 107–120

Penman, Stephen H, and Xiao-Jun Zhang (2002) 'Accounting conservatism, the quality of earnings, and stock returns.' *Accounting Review* 77(2), 237–264

Perols, Johan (2011) 'Financial statement fraud detection: An analysis of statistical and machine learning algorithms.' *Auditing: A Journal of Practice & Theory* 30(2), 19–50

Perols, Johan L, Robert M Bowen, Carsten Zimmermann, and Basamba Samba (2017) 'Finding needles in a haystack: Using data analytics to improve fraud prediction.' *Accounting Review* 92(2), 221–245

Rajgopal, Shivaram, Anup Srivastava, and Rong Zhao (2020) 'Do digital technology firms earn excess profits? an alternative perspective.' *An Alternative Perspective (November 25, 2020)*

Roychowdhury, Sugata, and Ross L Watts (2007) 'Asymmetric timeliness of earnings, market-to-book and conservatism in financial reporting.' *Journal of Accounting and Economics* 44(1-2), 2–31

Ryan, Stephen G (2006) 'Identifying conditional conservatism.' *European Accounting Review* 15(4), 511–525

Shu, Susan Zhan (2000) 'Auditor resignations: Clientele effects and legal liability.' *Journal of Accounting and Economics* 29(2), 173–205

Srivastava, Anup (2014) 'Why have measures of earnings quality changed over time?' *Journal of Accounting and Economics* 57(2-3), 196–217

Sunder, Jayanthi, Shyam V Sunder, and Jingjing Zhang (2018) 'Balance sheet conservatism and debt contracting.' *Contemporary Accounting Research* 35(1), 494–524

Tan, Liang (2013) 'Creditor control rights, state of nature verification, and financial reporting conservatism.' *Journal of Accounting and Economics* 55(1), 1–22

Watts, Ross L (2003) 'Conservatism in accounting part i: Explanations and implications.' *Accounting horizons* 17(3), 207–221

Wittenberg-Moerman, Regina (2008) 'The role of information asymmetry and financial reporting quality in debt trading: Evidence from the secondary loan market.' *Journal of Accounting and Economics* 46(2-3), 240–260

Yang, Jui-Chung, Hui-Ching Chuang, and Chung-Ming Kuan (2020) 'Double machine learning with gradient boosting and its application to the big n audit quality effect.' *Journal of Econometrics* 216(1), 268–283

Zeff, Stephen A (2013) 'The objectives of financial reporting: a historical survey and analysis.' *Accounting and Business Research* 43(4), 262–327