Measuring Accounting Comparability *

Skylar DeTure† Elisabeth Plietzsch‡ David Windisch‡

September, 2019

Abstract

Accounting researchers and regulators are highly interested in the determinants and consequences of accounting comparability. Existing measures of comparability, however, rely on stock return data as an input, making them unsuitable for many of the questions of interest to accounting researchers. We propose that an ideal measure of comparability would satisfy three criteria. First, an ideal comparability measure would rate two firms as having more similar accounting if their reported earnings respond to "true economic" performance in the same way. Second, an ideal measure of comparability would not rely on stock return data to identify the "true economic" performance of the firms, because doing so would presuppose the capital market consequences of accounting comparability that many researchers are interested in testing. Third, an ideal measure of comparability would not rely on "input" based information, such as a checklist of the specific accounting treatments used by individual firms, because doing so would presuppose the determinants of accounting comparability. We develop and estimate a structural model to produce a firm-level measure of accounting comparability that meets the above three criteria. Our measure is distinct from the popular DeFranco et al. (2011) measure in that we do not rely on stock returns as an input.

Keywords: comparability, financial reporting, accounting, standardization, structural

Preliminary and incomplete draft: Please do not circulate.

*We appreciate helpful comments and suggestions from Ross Jennings, Christian Leuz, John McInnis, Alfred Wagenhofer, participants at the 2018 Symposium for the Economic Analysis of Firms, all DART colleagues, and colleagues from the Department of Accounting at the University of Texas at Austin. Elisabeth Plietzsch acknowledges financing from the Austrian Science Fond (FWF): W 1229-G16.

†University of Texas at Austin

‡University of Graz

1 Introduction

Accounting comparability is a key characteristic of financial reporting quality and a topic of perennial interest to investors, other stakeholders, regulators, and standard setters worldwide (e.g., FASB 2010, SEC 2010, AICPA 2011, IASB 2018). Hence, it is not surprising that accounting researchers have taken great interest in studying comparability. Recent progress in measuring comparability, most notably by De Franco et al. (2011), has paved the way for empirical research on comparability. However, despite its merits and popularity, De Franco et al.'s (2011) measurement approach, due to its strong reliance on stock returns to capture firms' true economic performance, is subject to several limitations, such as a limited scope on listed firms and rather strong assumptions about the relation between earnings and returns, making it unsuitable for many areas that are of interest to accounting researchers.

In this paper, we develop a firm-level measure of accounting comparability that relies only on time-series earnings data. Firms' accounting systems map true economic performance (which is unobservable) to accounting earnings (which are observable). We consider two accounting systems "comparable" if their mapping from true economic performance to accounting earnings is similar (De Franco et al. 2011). However, determining that similarity is not straightforward because we typically cannot observe "true" economic performance. Thus, a key challenge in measuring accounting comparability is the disentanglement of accounting differences in the reported earnings from economic differences between two firms.

Previous literature has attempted to isolate accounting comparability from economic performance by using proxies for true economic performance. For example, De Franco et al. (2011) use stock prices. The idea is that, if firm As stock price is related to its earnings per some relationship R_A, and if firm Bs stock price is related to its earnings per some relationship R_B, then firms A and B have comparable accounting if R_A and R_B are similar. However, the challenge is that prices do not necessarily capture true economic performance. Moreover, the extent to which stock prices *do* capture true economic performance likely depends on various qualities of accounting earnings–comparability in particular–and is therefore endogenous.

By using a structural model, we are able to produce a parsimonious measure of accounting comparability that does *not* rely either on stock return data or any other non-earnings data about firm performance. Thus, we consider a world where accounting earnings are the only available information about firm performance. We assume that each firm's true economic performance follows a random mean reverting process with unknown volatility, unknown mean reversion speed, and an unknown linear trend. Furthermore, we assume that each firm has a function F that maps (unobservable) economic performance to reported earnings. We can use patterns in the (observable) reported earnings of the firm, along with the assumption that true earnings follow a mean-reverting process, to back out the function F that maps economic performance to reported earnings. Then, for any two firms, we can see how similar their Fs are. If they have similar Fs, we can consider their accounting systems more comparable. We do all of this without ever knowing the firm's true economic performance; we only assume that de-trended true economic performance follows a mean reverting process.

Because our measure does not use stock returns or any other capital market data as an input, we can make three significant contributions to the accounting literature. First, we provide a measure that can be used to answer questions about the capital market consequences of accounting comparability. These might include the following: Does accounting comparability help investors incorporate earnings information into stock prices at a lower cost and in a timely, comprehensive manner? Do firms with more comparable accounting have more similar earnings response coefficients? Are a firm's earnings more value relevant if its accounting is comparable to industry peers? Do firms with more comparable earnings have more stock price co-movement? Such questions are difficult to answer using a comparability measure that assumes stock prices already capture *true* economic performance.

In addition, because our measure is not based on the earnings-return relation, it mitigates measurement error and omitted confounding influences that are related to cross-sectional differences in fundamental market variables, such as stock liquidity or price efficiency, and nonlinearities in the relationship between economic and reported earnings and returns (e.g., Freeman and Tse 1992, De George et al. 2016, Breuer and Windisch 2019) that may produce spurious results.

Finally, our measure can be used to compute comparability scores for entities that do not trade in liquid markets. For example, we could measure the accounting comparability between a private and a public firm, between two private firms, between two segments of a publicly traded firm, etc. All that is needed is for the earnings data of the two entities to be publicly available. Thus, our measure provides a tool for empirical research on the causes and effects of accounting comparability in settings where it could not previously be studied.

Our measure also contributes to the burgeoning literature on financial statement comparability. While such measures are relatively scarce, relevant research questions abound. Researchers care about how comparability affects financial statement users as well as how accounting regulation affects comparability. Regarding the former, Chen et al. (2018) study the effect of comparability on the efficiency of firms' acquisition decisions, De Franco et al. (2011) study the effect of comparability on analyst following and forecast accuracy, and Kim et al. (2013) and Kim et al. (2016) study the relationship between comparability, credit risk, and expected crash risk. Indeed, there are many more users and *uses* of financial statements that are likely to be affected by comparability, making this stream of research critical for the foreseeable future.

Research regarding how regulators, standard setters, and other gatekeepers affect accounting comparability is also flourishing. Financial statement comparability is a topic of primary concern to standard setters; it is mentioned numerous times in both the FASB and IASB conceptual frameworks and is frequently mentioned as a primary benefit of financial statement regulation and harmonization. In addition, Brochet et al. (2013), Wang (2014), and Neel (2016) study the effect of accounting standard harmonization on comparability. As accounting standards and regulations continue to evolve, we can expect a continuing need to measure the effect of regulation on financial statement comparability.

The remainder of the paper proceeds as follows. Section II outlines our structural model of accounting comparability. Section III describes our procedure for estimating the measure of accounting comparability. Section IV presents empirical tests of our measure, and comparisons to the DeFranco et al. *CompAcct* measure of comparability. Section V discusses future research and concludes.

2 Analytical Framework

In this section, we propose a simple model of a firm's earnings process to derive a theory-based measure of accounting comparability. Therefore, we assume that each firm has true economic performance, unobservable, following a mean reverting process (e.g., Bhattacharya 1978, Freeman et al. 1982, Fama and French 2000). Thus, firm i's true economic performance at time t is given by

$$X_{i,t} = X_{i,t-1} + \phi_i(\mu_i - X_{i,t-1}) + e_{i,t} \tag{1}$$

where μ_i is the mean economic performance for firm i, ϕ_i is the mean reversion constant of economic performance for firm i, and $e_{i,t} \sim N(0, \sigma_{e,i})$ is the shock to firm i's economic performance in period t.

We assume that firms convert unobservable economic performance (X_t) to observable reported earnings (Y_t) as follows:

$$Y_{i,t} = X_{i,t} + \epsilon_{i,t} \tag{2}$$

where ϵ represents any factors causing reported earnings to be different from true economic performance. In particular, we assume

$$\epsilon_{i,t} = \epsilon_{i,t}^{error} + \rho_i \epsilon_{i,t-1} \text{ with } \rho_i \in (-1, 0) \tag{3}$$

That is, reported accounting earnings Y_t measure true economic performance X_t with error $\epsilon_{i,t}^{error} \sim N(0, \sigma_{\epsilon,i})$, and this error reverses at rate $|\rho_i|$.

Thus, our model is consistent with the general assumption in the accounting literature that in steady state and over a sufficiently long time period, accounting earnings are expected to converge to the true economic performance, and that the relationship between economic performance and earnings is stronger over longer periods of aggregation (e.g., Easton et al. 1992, Dechow 1994).

Plugging equations (1) and (3) into (2), and solving the resulting recurrence equation gives us a closed form solution for the reported earnings of firm i in period t (suppressing the i subscript for ease of notation):

$$Y_t = \mu + \sum_{j=1}^{t} e_j (1-\phi)^{t-j} + \sum_{j=0}^{t} \rho^{t-j} \epsilon_j^{error} \tag{4}$$

Following DeFranco et al. (2011), we define two firms as comparable to the extent that they report similar earnings when they experience similar economic performance. Conversely, two firms are incomparable to the extent that they report dissimilar earnings after experiencing similar economic performance. Thus, we can calculate the incomparability of two firms as the expected difference in their reported earnings assuming they had identical underlying economics. In particular, we calculate the expected absolute value of $Y_A - Y_B$ assuming the firms have identical ϕ and σ_e (the parameters governing true economic performance), but potentially different σ_ϵ and ρ (the parameters that govern the mapping from economic performance to reported earnings). More formally stated:

$$\begin{aligned} incomparability &= E(|Y_{A,t} - Y_{B,t}|) \\ &= E[|(e_{A,1}(1-\phi)^{t-1} + \sum_{j=2}^{t-1} e_{A,j}(1-\phi)^{t-j} + e_{A,t} + \sum_{j=0}^{t} \rho_A^{t-j} \epsilon_{A,j}^{error}) \\ &\quad - (e_{B,1}(1-\phi)^{t-1} + \sum_{j=2}^{t-1} e_{B,j}(1-\phi)^{t-j} + e_{B,t} + \sum_{j=0}^{t} \rho_B^{t-j} \epsilon_{B,j}^{error})|] \end{aligned} \tag{5}$$

We can see that $Y_{A,t} - Y_{B,t}$ is normally distributed with mean zero. Thus, $|Y_{A,t} - Y_{B,t}|$ follows a half-normal distribution with mean $= \sqrt{\frac{2 * Var(Y_{A,t} - Y_{B,t})}{\pi}}$. Therefore, we first derive the variance of $Y_{A,t} - Y_{B,t}$. Rearranging (4) yields the variance of $Y_{A,t} - Y_{B,t}$: $\sigma^2_{\epsilon,A} \sum_{j=0}^{t} (\rho_A^{t-j})^2 + \sigma^2_{\epsilon,B} \sum_{j=0}^{t} (\rho_B^{t-j})^2$, which, in steady state (i.e., as t goes to infinity), becomes

$$\lim_{t \to \infty} Var(Y_{A,t} - Y_{B,t}) = \frac{(1-\rho_B^2)\sigma^2_{\epsilon,A} + (1-\rho_A^2)\sigma^2_{\epsilon,B}}{(\rho_A^2 - 1)(\rho_B^2 - 1)}. \tag{6}$$

Then, $E(|Y_{A,t} - Y_{B,t}|)$ is given by

$$incomparability = \sqrt{\frac{2((1-\rho_B^2)\sigma^2_{\epsilon,A} + (1-\rho_A^2)\sigma^2_{\epsilon,B})}{(\rho_A^2 - 1)(\rho_B^2 - 1)\pi}}. \tag{7}$$

To summarize, equation (7) provides us with a formula that describes the incomparability of two firms A and B as the expected absolute difference in their reported earnings, assuming they both have true economic performance governed by the same data generating process. Taking the negative of (7) gives us a measure of accounting comparability:

$$comparability(A,B) = -\sqrt{\frac{2((1-\rho_B^2)\sigma^2_{\epsilon,A} + (1-\rho_A^2)\sigma^2_{\epsilon,B})}{(\rho_A^2 - 1)(\rho_B^2 - 1)\pi}} \tag{8}$$

Note that we could calculate the numerical value of this measure if we had estimates of $\rho_A, \rho_B, \sigma_{\epsilon,A}$ and $\sigma_{\epsilon,B}$.

3 Data and Estimation Procedure

We can calculate the accounting comparability, as defined in section 2, for any two firms if we know the values of ρ and σ_ϵ for those firms. To estimate ρ and σ_ϵ (the parameters describing how economic performance maps to accounting earnings) we also need to estimate ϕ and σ_e (the parameters that govern economic performance). We estimate the four parameters, $\rho, \sigma_\epsilon, \phi, \sigma_e$, for each firm using the Generalized Method of Moments (GMM).

Essentially, we use our model from section 2 to obtain formulas for several moments of reported accounting earnings. Following the economics and finance literatures, we use the word "moment" loosely to refer to any function of a random variable, such as the variance of that random variable, the auto-correlation of that random variable, et cetera. For each moment of reported earnings, we can write a closed-form mathematical expression in terms of the parameters $\rho, \sigma_\epsilon, \phi$, and σ_e by plugging in our closed-form expression for reported earnings in equation (4). For example, to derive a closed-form expression of the variance of reported earnings:

$$\begin{aligned} Var(Y_t) &= Var\Big[\mu + e_1(1-\phi)^{t-1} + \sum_{j=2}^{t-1} e_j(1-\phi)^{t-j} + e_t + \sum_{j=0}^{t} \rho^{t-j}\epsilon_j^{error})\Big] \\ &= f(\rho, \sigma_\epsilon, \phi, \sigma_e, t) \end{aligned} \tag{9}$$

We can see that the variance of the reported earnings is a function of the parameters $\rho, \sigma_\epsilon, \phi, \sigma_e$, and time t. Assuming steady state, we can take the limit of $f(\rho, \sigma_\epsilon, \phi, \sigma_e, t)$ as $t \to \infty$, which removes t from our expression for the variance. Thus, we end up with a formula for the variance of reported earnings that depends only on the parameters $\rho, \sigma_\epsilon, \phi$, and σ_e. If we set our formula for the variance of reported earnings equal to the observed variance of reported earnings from Compustat, we have an equation of four variables $(\rho, \sigma_\epsilon, \phi, \sigma_e)$.

We can repeat this process for other moments . Each moment gives us a closed-form expression of our parameters which we can set equal to the empirical value for that moment. Thus, if we have at least four moments, then we also have at least four equations of our four variables, allowing us to solve for the parameters $\rho, \sigma_\epsilon, \phi$, and σ_e. We repeat this process separately for every firm, so every firm has its own values of $\rho, \sigma_\epsilon, \phi$, and σ_e.

We rely on four moments to identify the four parameters for each firm: $\mathrm{Var}(Y_t)$, $\mathrm{Cor}(Y_t, Y_{t-1})$, $\mathrm{Var}(\Delta Y_t)$, and $\frac{\mathrm{Cor}(Y_t, Y_{t-5})}{\mathrm{Var}(\Delta Y_t)}$.[1] The graphs in figure 1 show how each of these moments changes in response to changes in the parameter values $(\rho, \sigma_\epsilon, \phi, \sigma_e)$. It is important to choose moments which respond differently to each parameter. This way, if the moments change, we can determine which parameters are driving that change. More formally, we want the matrix

[1] Closed form expressions of the moment formulas are in the Appendix, alongside their derivations.

whose elements are signs of the slopes in figure 1 to have a non-zero determinant.

To compute the corresponding empirical values for the moments, we use quarterly Compustat income before extraordinary items (Compustat item *ibq*). We use quarterly data because we need as much time series variation as possible within each firm to calculate firm-specific parameters. Our main comparability measure is based on 40 quarters of data, using all firms that have at least 20 quarters of data available in Compustat in the period from 2007 to 2016. We remove any linear trend from each firm's earnings to ensure that our data reflect the model assumption that true economic performance follows a mean reverting process. We also normalize each firm's mean earnings to 1, ensuring that differences in firm size do not drive our results. Overall, this results in a sample of 8,261 firms for which we can calculate our comparability measure.

After calculating the empirical values of each moment for each firm, we use a two-step approach to estimate the parameters $\rho, \sigma_\epsilon, \phi$, and σ_e for each firm. First, we use General Simulated Annealing to find an approximate solution to the moment conditions. General Simulated Annealing is a numerical optimization approach for finding the minimum of a function over a large search space. The function we minimize is a weighted sum of squared distances between theoretical moment values and observed moment values. The goal is to minimize this function by choice of the parameters, that is, to choose parameters such that the moment conditions are as close to being satisfied as possible.

We use the parameter estimates from the General Simulated Annealing as an initial guess for our GMM procedure. We use the iterative GMM proposed by Hansen et al. (1996) with an optimal weighting matrix. The output from GMM is a list of parameter estimates and standard errors for each firm. For any pair of firms, we can produce an estimate of their accounting comparability by plugging their GMM parameter estimates into equation (8). We can also produce standard errors for these comparability estimates using the delta method.

To validate our estimation procedure, we benchmark its performance on simulated data where we know the actual parameter values (and hence the true comparability scores). We simulate 200 firms by drawing 200 parameter vectors $(\rho, \sigma_\epsilon, \phi, \sigma_e)$ at random, and simulating 40 (20) quarters worth of economic performance and reported earnings data for each one. Then, we use our estimation procedure to estimate the parameter values for each simulated firm. We then calculate the "true" and "estimated" comparability for each firm-pair, using the true and estimated parameter values respectively. Lastly, we randomly select 2,000 pairs of firm-pair observations and compare whether our estimated comparability measure correctly identifies which pair is more comparable according to the true comparability score. Our estimation procedure correctly identifies the more comparable firm-pair over 85% (75%) of the time. Panel A of figure 2 shows the conditional mean of our estimated comparability score, given the true comparability score, over all 198,000 firm-pairs in our simulated data set. Note the positive monotonic relationship between our estimates and the true values, as well as the narrow margin of error for the majority of possible comparability values.

We also benchmark our estimation procedure against the DeFranco et al. (2011) *CompAcct* measure using our simulated data. To benchmark against *CompAcct*, we assume that stock price at time t is the discounted sum of future actual economic performance. We make this strong efficient market assumption to give *CompAcct* the largest possible advantage over our structural estimate. *CompAcct*, even using the highly informative stock price data, correctly identifies the more comparable firm-pairs only 53% of the time. Panel B of figure 2 shows the conditional mean of the *CompAcct* measure as a function of the true simulated comparability over our full sample of 198,000 simulated firm pairs. The relationship is non-monotonic with a relatively large margin of error.

4 Validation and Empirical Tests

In this section, we provide some initial evidence on how our comparability measure performs in real (as opposed to simulated) data. Therefore, we start with some descriptive analyses to explore the validity our measure: 1) we examine whether firms in the same industry are more comparable and 2) we examine the correlation between our structural measure and the DeFranco et al. (2011) measure.

In further analyses, we replicate the empirical tests of two previous empirical studies on accounting comparability. Specifically, we replicate DeFranco et al.'s (2011) test of the relation between comparability and correlated analyst forecast errors and Francis et al.'s (2014) test of the influence of auditor style on accounting comparability.

4.1 Descriptive Analysis

We start with a descriptive analysis to evaluate our comparability measure. The mean (median) of our comparability measure based on all firm A–firm B pairs of Compustat firms with available data in the period from 2007 to 2016 is -3.905 (-1.544), with a standard deviation of 6.620. The values can be interpreted as the average (median) absolute accounting difference in the reported quarterly earnings of two firms in the dataset.[2]

To explore our measure's validity, we examine the relationship between industry membership and our comparability measure. Which accounting standards are relevant for a firm depends, among other things, on the industry in which it operates. For example, the rules of ASC 905 – Agriculture will be applied mainly by firms that operate in the agricultural industry, whereas firms in the retail industry should be largely unaffected by ASC 905. To test this relationship, we regress our comparability measure on a dummy variable that takes the value of one if two firms belong to the same SIC2 industry, and zero if they belong to different SIC2 industries.[3] In untabulated results, we find that the coefficient of the dummy variable is positive and highly significant (at the 1 % level).[4] Thus, our measure effectively classifies the accounting of two firms as more comparable when these firms belong to the

[2] The comparability measure is in winsorized at the 2.5 and 97.5 percentile in all analyses.

[3] Note that this test is not able to determine whether our measure successfully separates accounting comparability from economic comparability. The effect of industry membership on economic comparability is ambiguous. Firms in the same industry are exposed to similar economic shocks and therefore economic comparability should be high. However, an increase in a firm's earnings might harm a direct competitor, leading to a negative correlation in the economic performance of the two firms.

[4] We also find a positive and significant coefficient when we use a log or rank transformation on our comparability measure.

same industry. However, it should be noted that the adjusted R^2 is close to zero, indicating that same industry membership is not able to explain a large portion of the variation in comparability. The results are robust (but with higher p-values) to the inclusion of industry fixed effects for firm A and firm B, and when we use a log or rank transformed version of our measure.

Next, we calculate the Spearman correlation between our comparability estimates and the corresponding DeFranco et al. (2011) estimates. The correlation is, depending on the sample period, between 0.15 and 0.22, with a p-value less than 0.001.[5] For a random selection of 2,000 pairs of firm-pairs, our measure agrees with the DeFranco et al. measure on which firm-pair is more comparable about 60% of the time. These correlations indicate that the two are empirically related but distinct measures of comparability.

4.2 Correlated Analyst Forecast Errors[6]

In this analysis, we replicate DeFranco et al.'s (2011) test of the association between financial statement comparability and correlated analyst forecast errors. DeFranco et al. argue that firms with more comparable accounting systems exhibit more correlated forecast errors. The intuition is that comparable accounting leads to similar deficiencies in firms' financial reporting, and therefore, analysts are more likely to make similar errors when forecasting the earnings of two firms. Table 2 presents the results of our replication. All variables are defined as in DeFranco et al. We include industry fixed effects based on the 2-digit SIC industry classification and cluster standard errors at the firm i and analyst k level. We winsorize all continuous variables at the 1% level. The full sample results serve as a benchmark for the overlapping sample to mitigate concerns regarding the smaller sample size of the overlapping sample. Most of the control variables are significant with the predicted sign. More impor-

[5] Note that De Franco et al. (2011) use 16 quarters of earnings data, whereas in our main sample, we use 40 quarters. However, we also compute our measure for smaller subsamples based on 20 quarters and modify the De Franco et al. (2011) measure accordingly. The correlation coefficient is signficantly positive for all subsamples.

[6] Note: The results of this subsection are based on an initial smaller sample and are not updated yet.

tantly, the coefficients of the accounting comparability measures are significantly positive as predicted. Our results differ from DeFranco et al. only in terms of magnitude. While DeFranco et al. report that a one-standard-deviation increase in $CompAcct$ is associated with a 56% increase in $CorrFcstError$, we find that a one-standard-deviation increase in $CompAcct$ is associated with only a 21% (17%) increase in $CorrFcstError$ in our full (overlapping) sample. This difference appears to be driven by differences in the sample selection; compared to DeFranco et al., our sample exhibits lower means and standard deviations for both $CompAcct$ and $CorrFcstError$.

Next, we test the relationship between accounting comparability and correlated analyst forecast errors using our comparability measure. We use the overlapping sample for these tests to facilitate comparison between our measure and the DeFranco et al. measure. For our structural measure, we find that a one-standard-deviation increase is associated with a 10% increase in $CorrFcstError$ (compared to 17% for the DeFranco et al. measure). The results are similar enough to suggest that our measure and the DeFranco et al. measure capture related constructs. However, the results are different enough to suggest that the measures do not capture the same construct. Importantly, the difference in the results is consistent with the conceptual differences between the two measures: the DeFranco et al. measure relies on stock price data to isolate economic performance, whereas our measure uses only information from earnings to isolate economic performance. Since analyst forecasts influence stock price data (and vice versa), the DeFranco et al. measure picks up two components of comparability: (1) the similarity in how two firms' earnings respond to an economic event and (2) the similarity in how two firms' stock prices respond to an economic event. Notably, our comparability measure only captures the first component. Since the second component likely correlates with analyst forecast errors, we expect to see larger coefficient estimates for the DeFranco et al. measure in this particular test.

An alternative explanation for why our comparability measure yields a smaller coefficient estimate than the DeFranco et al. measure is that our measure is more noisy. However, we do not believe that noise explains these differences. First, the simulation tests in section 3 suggest that our estimation procedure is reasonably precise. Second, we repeat our tests several times using random subsamples. We find that the results are robust across these subsamples, which should not be the case if our measure is pure noise.

4.3 Auditor Style[7]

In this section, we replicate Francis et al.'s (2014) test of how auditor style affects accounting comparability. Financial reports are the result of negotiations between a firm and its auditor. Big 4 audit firms have their own unique audit test approaches and in-house working rules for interpreting and applying GAAP. Therefore, different Big 4 audit firms should affect firms' reported earnings differently. In other words, Francis et al. expect that firms with the same Big 4 auditor have more comparable accounting earnings than firms with different auditors.

We replicate Francis et al.'s auditor style test because it provides a setting to reasonably assess whether the comparability measures capture accounting comparability or economic comparability (or just noise). It is unlikely that auditors are able to affect a firm's underlying economics. Therefore, if any comparability effect of having the same auditor exists, this effect is more likely to be driven by accounting rather than economic comparability.

In our analysis, we extend Francis et al.'s argument to non-Big 4 auditors and expect that having the same auditor should, in general, increase accounting comparability compared to firm-pairs that do not share the same auditor. We replicate Francis et al.'s test for three different accounting comparability measures: (1) their $ECOMP\_COV$ measure, (2) our structural measure, and (3) the DeFranco et al. measure. We use a log-transformation of these measures to deal with their skewness. All variables are defined as in Francis et al. except for $Same\_Auditor$ which indicates whether the firm-pair has the same Big 4 or non Big 4 auditor.[8] We include 2-digit SIC industry fixed effects for firm A and B. We

[7] Note: the results of this subsection are based on an initial smaller sample and are not updated yet.
[8] We do not limit the $Same\_Auditor$ indicator to firm-pairs that share the same auditor over the whole construction period of the comparability measure. However, this should just introduce noise and therefore

winsorize all continuous variables at the 2.5 % level.

Table 3 reports our results. All coefficients are standardized. We find the expected positive and significant coefficient of *Same_Auditor* only when using *ECOMP_COV* or our comparability measure as the dependent variable. Interestingly, the *Same_Auditor* coefficient is insignificant for the DeFranco et al. measure.

Finally, we analyze how the comparability measures perform with smaller sample sizes. Therefore, we assume that the sample used for table 3 represents the whole population and that there should exist a positive relationship between having the same auditor and accounting comparability. For sample sizes of 1, 5, 10, 30, 50, 70 and 90 % of the original sample size, we draw 1,000 samples and check how often we find a positive and significant coefficient. We count coefficients with p-values of less than 0.1 as significant. Table 4 reports the corresponding results. The columns +*, −*, and 0 state the proportion of samples with a significant positive, a significant negative, and a non-significant coefficient of *Same_Auditor*, respectively. Table 3 shows that our structural measure performs well in smaller samples. Even in samples that consist only of 5 % of the original sample size, our measure almost always provides the expected positive association between *Same_Auditor* and accounting comparability. For comparison, the *ECOMP_COV* measure finds a significant positive effect in only 15.8 % of the samples.

5 Conclusion

We developed and validated a new firm-level measure of accounting comparability that uses only reported earnings as an input. We define the accounting comparability of two firms as the similarity with which these firms' accounting systems map (unobservable) economic performance into (observable) reported earnings. Thereby, our measure is designed to distinguish between economic and accounting comparability, as firms can have very comparable accounting systems even if they have a starkly different economic performance.

work against finding significant results.

Our measure is conceptually appealing because it is based on earnings, which should help users of financial statements to assess firm performance. Thus, our measure is consistent with the idea that financial reporting provides information to the capital markets, not the other way around. By contrast, previous measures of accounting comparability have often assumed that the information content of stock prices is not affected by qualitative characteristics of accounting earnings, such as comparability.

Our measure is practically appealing because (1) it can be used to test the capital market consequences of accounting comparability and (2) it can be used to study accounting comparability in contexts in which not all entities have prices in a liquid market. Measures of comparability that rely on stock returns are not suitable for testing capital market consequences of comparability, because the stock return information embedded in the measure can mechanically induce the observed capital market results.

We show that our measure is moderately, but not perfectly, correlated with De Franco et al.'s (2011) comparability measure, suggesting that these two measures capture related but distinct constructs. Furthermore, we replicated selected results from the comparability literature using our measure. The results using our measure differ from the results using previous measures in predictable ways, which is consistent with the theoretical differences between our measure and previous measures. For example, in tests of capital market consequences of comparability, our measure produces attenuated results compared to previous measures. We would expect this to be the case, because previous measures, which rely on capital market data as inputs, should have mechanically stronger relationships with capital market outputs.

Similarly, in tests of non-capital market determinants of comparability, our measure produces more pronounced results than previous measures. We would expect this to be the case, because previous measures rely on capital market outcomes, which are beyond the auditor's control, whereas our measure depends only on reported earnings, which is within the auditor's control.

We contribute to the existing literature by developing a new accounting comparability measure that relies only on time-series earnings data. This measure is a parsimonious measure that can be calculated for public and private firms at the firm-level and does not require any assumptions about the relation between earnings and returns. Moving forward, we believe that our measure can be used by accounting researchers to address new research questions for a broader set of firms in the economy and to validate existing findings on the determinants and consequences of accounting comparability.

References

1. American Institute of Certified Public Accountants 2011. International Financial Reporting Standards (IFRS)An AICPA Backgrounder. New York, NY.

2. Bhattacharya, S. (1978). Project Valuation with Mean-Reverting Cash Flow Streams. The Journal of Finance 33: 1317-1331.

3. Breuer, M., and D. Windisch (2019). Investment Dynamics and Earnings-Return Properties: A Structural Approach. Journal of Accounting Research, forthcoming.

4. Brochet, F., A.D. Jagolinzer, and E.J. Riedl (2013). Mandatory IFRS adoption and financial statement comparability. Contemporary Accounting Research 30: 1373-1400.

5. Chen, C.W., D.W. Collins, T.D. Kravet, and R.D. Mergenthaler (2018). Financial statement comparability and the efficiency of acquisition decisions. Contemporary Accounting Research 35: 164-202.

6. Dechow, P.M. (1994) Accounting Earnings and Cash Flows as Measures of Firm Performance: The Role of Accounting Accruals. Journal of Accounting and Economics 18: 342.

7. De Franco, G., S.P. Kothari, and R.S. Verdi (2011). The benefits of financial statement comparability. Journal of Accounting Research 49: 895-931.

8. De George, E.T, X. Li, and L. Shivakumar (2016). A review of the IFRS adoption literature. Review of Accounting Studies 21: 898-1004.

9. Easton, P.D., T.S. Harris, and J.A. Ohlson (1992). Aggregate accounting earnings can explain most of security returns. Journal of Accounting and Economics 15: 119-142.

10. Fama, E.F., and K.R. French (2000). Forecasting Profitability and Earnings. Journal of Business 73: 161-175.

11. Financial Accounting Standards Board (2010). Statement of Financial Accounting Concepts No. 8, Conceptual Framework for Financial Reporting. Norwalk, CT.

12. Francis, J.R., M.L. Pinnuck, and O. Watanabe (2013). Auditor style and financial statement comparability. Journal of Accounting Research 89: 605-633.

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

14. Freeman, R.N., and S.Y. Tse (1992). A Nonlinear Model of Security Price Responses to Unexpected Earnings. Journal of Accounting Research 30: 185209.

15. Hansen, L.P., J. Heaton, and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators. Journal of Business and Economic Statistics 14: 262-80.

16. International Accounting Standards Board (2018). The Conceptual Framework for Financial Reporting 2018. IASB, London.

17. Kim, J.-B., L. Li, L.Y. Lu, and Y. Yu (2016). Financial statement comparability and expected crash risk. Journal of Accounting and Economics 61: 294-312.

18. Kim, S., P. Kraft and S.G. Ryan (2013). Financial statement comparability and credit risk. Review of Accounting Studies 18: 783-823.

19. Neel, M. (2017). Accounting comparability and economic outcomes of mandatory IFRS adoption. Contemporary Accounting Research 34: 658-690.

20. SEC (2010). Commission statement in support of convergence and global accounting standards. Washington, DC SEC.

21. Wang, C. (2014). Accounting standards harmonization and financial statement comparability: Evidence from transnational information transfer. Journal of Accounting Research 52: 955-992.

Tables

Table 1: Parameter Estimates

Parameter	Estimation	Standard Error
ϕ	0.639	0.004
σ_e	2.554	0.185
σ_ϵ	2.276	0.117
ρ	-0.358	0.004

This table reports the average firm-specific parameter estimates obtained via the Generalized Method of Moments (GMM) for a sample of Compustat firms in the period from 2007 to 2016.

Table 2: Correlated Forecast Errors of Comparable Firms

	Prediction	Full Sample DeFranco et al. Measure	Overlap Sample	Overlap Sample
$CompAcct_{ijt}$	+	0.003***	0.002***	
$Comparability_{ijt}$	+			0.0001***
$Size_{jt}$	+	0.002***	0.001***	-0.002***
$Book\text{-}Market_{jt}$	?	-0.002***	-0.009***	-0.008***
$Volume_{jt}$	?	-0.001***	-0.001***	-0.001***
ROA_{jt}	?	0.045***	0.072***	0.082***
$Predictability_{jt}$	+	0.001**	0.001	-0.000
$Volatility\ Earn_{jt}$	−	0.016	-0.008	-0.045***
$Volatility\ Ret_{jt}$	−	-0.088***	-0.141***	-0.081***
$Size\ Difference_{ijt}$	−	-0.001***	0.000	-0.000
$Book\text{-}Market\ Diff_{ijt}$	−	-0.013***	-0.020***	-0.202***
$Pseudo\ R^2$		27.2 %	30.9 %	29.8 %
No. of Obs.		106,397	16,918	16,918

This table reports an analysis of the relation between the pairwise accounting comparability measures (i.e., at the firm i – firm j level) and analyst forecast errors of firm j in the same industry as the sample firm i. We estimate various specifications of the following probit model

$$CorrFcstError_{ijt} = \alpha + \beta_1 CompAcct_{ijt} + \gamma Controls_{jt} + \epsilon_{ijt}$$

The dependent variable is $CorrFctEror$, which proxies for the correlation in forecast errors between firm i and j. *CompAcct* denotes DeFranco et al.'s (2011) comparability measure. *Comparability* denotes our structural comparability measure. Industry fixed effects on the 2-digit SIC industry classification are included but not reported. Standard errors are clustered at the firm i and analyst k level. All continuous variables are winsorized at the 1% level. Column 1 presents the results for the full sample, consisting of all firm-pairs in 2014 for which we have the DeFranco et al. measure and sufficient data is available. Column 2 and 3 present results for the overlap sample, consisting of all firm-pairs in 2014 for which both the DeFranco et al. and our structural measure is so far available. ***, ** and * denote significance at the 1%, 5% and 10% (two-sided) levels, respectively. Variables are defined as in DeFranco et al. (2011).

Table 3: Effect of Big 4 Auditor Style on Comparability

	Dependent Variable		
	$ECOMP\_COV$	Structural Measure	DeFranco et al. Measure
$Same\_Auditor$	0.004**	0.025***	0.007
$Size\_Diff$	-0.003	-0.015	0.074***
$Size\_Min$	-0.006	-0.068**	0.144***
LEV_Diff	-0.017***	-0.022	-0.079***
LEV_Min	-0.025***	-0.056***	-0.134***
MB_Diff	0.025***	0.046***	0.073***
MB_Min	0.030***	0.006	0.067***
$Loss\_Prob\_Diff$	-0.020***	-0.002	-0.149***
$Loss\_Prob\_Min$	-0.029***	-0.113***	-0.210***
STD_Sales_Diff	0.016	0.029	-0.036
STD_Sales_Min	0.013*	0.008	-0.029
STD_CFO_Diff	-0.011	0.036	0.002
STD_CFO_Min	-0.012*	0.029*	0.010
$STD\_Sales\_Grth\_Diff$	-0.007	0.039**	-0.020
$STD\_Sales\_Grth\_Min$	0.012**	0.013	0.007
CFO_COMP_COV	0.015***	0.130***	0.069***
RET_COV	0.019***	0.054***	0.131***
industry fixed effects	yes	yes	yes
$Pseudo\ R^2$	0.9 %	11.0 %	28.7 %
No. of Obs.	415,380	415,380	415,380

This table reports an analysis of the relation between auditor style and the pairwise accounting comparability measures (i.e., at the firm i – firm j level). We estimate the following OLS regression model

$$ln(ComparabilityMeasure) = \alpha + \beta_1 Same\_Big4 + \gamma Controls_{jt} + \epsilon_{ijt}$$

The dependent variable is $ln(ComparabilityMeasure)$ and denotes the natural logarithm of three comparability proxies: $ECOMP\_COV$, our structural comparability measure, and DeFranco et al.'s (2011) *CompAcct*. Industry fixed effects for firm i and firm j on the 2-digit SIC industry classification are included but not reported. Standard errors are clustered at the firm i level. All continuous control variables are winsorized at the 2.5% level. Column 1 presents the results when $ECOMP\_COV$ is the dependent variable. Column 2 presents the result when our structural measure is the dependent variable. Column 3 presents the results when *CompAcct* is the dependent variable. The estimation is based on a sample consisting of all firm-pairs in 2014 for which we have all three comparability measures and sufficient data is available. In contrast to Francis et al. (2014), firm-pairs are also included when either one or both firms are audited by a non-Big 4 audit firm. ***, ** and * denote significance at the 1%, 5% and 10% (two-sided) levels, respectively. Variables are defined as in Francis et al. (2014).

Table 4: Smaller Sample Performance Based on the Effect of Auditor Style on Comparability

	Dependent Variable								
	ECOMP_COV			Structural measure			DeFranco measure		
sample size	+*	−*	0	+*	−*	0	+*	−*	0
1%	0.099	0.021	0.88	0.551	0.001	0.448	0.113	0.009	0.878
5%	0.158	0.009	0.833	0.990	0	0.010	0.135	0	0.865
10%	0.231	0.004	0.765	1	0	0	0.126	0	0.874
30%	0.432	0	0.568	1	0	0	0.046	0	0.954
50%	0.635	0	0.365	1	0	0	0.013	0	0.987
70%	0.843	0	0.157	1	0	0	0	0	1
90%	0.994	0	0.006	1	0	0	0	0	1

This table reports the proportion of samples with a significant positive (+* column), significant negative (−* column) or insignificant (0 column) coefficient of *Same_Auditor*. Coefficients with a p-values of less than 0.1 are count as significant. We use the sample of 415,380 firm-pairs from the analysis in table 3 as original sample. For different samples sizes (1, 5, 10, 30, 50, 70 and 90 %) we draw each 1,000 random samples and estimate for each sample the following OLS regression model (the same as in table 3)

$$ln(ComparabilityMeasure) = \alpha + \beta_1 Same\_Auditor + \gamma Controls_{jt} + \epsilon_{ijt}$$

The dependent variable is $ln(ComparabilityMeasure)$ and stands for the natural logarithm of three comparability proxies: *ECOMP_COV*, our structural measure and DeFranco et al.'s (2011) *CompAcct* measure. Industry fixed effects for firm i and firm j on the 2-digit SIC industry classification are included Standard errors are clustered at the firm i level.

Figures

The plots below show the relationship between the model parameters and the moments. Because the slopes go in different directions [9], we can identify how the parameters of two firms must differ to explain how the observed moments of those two firms differ.

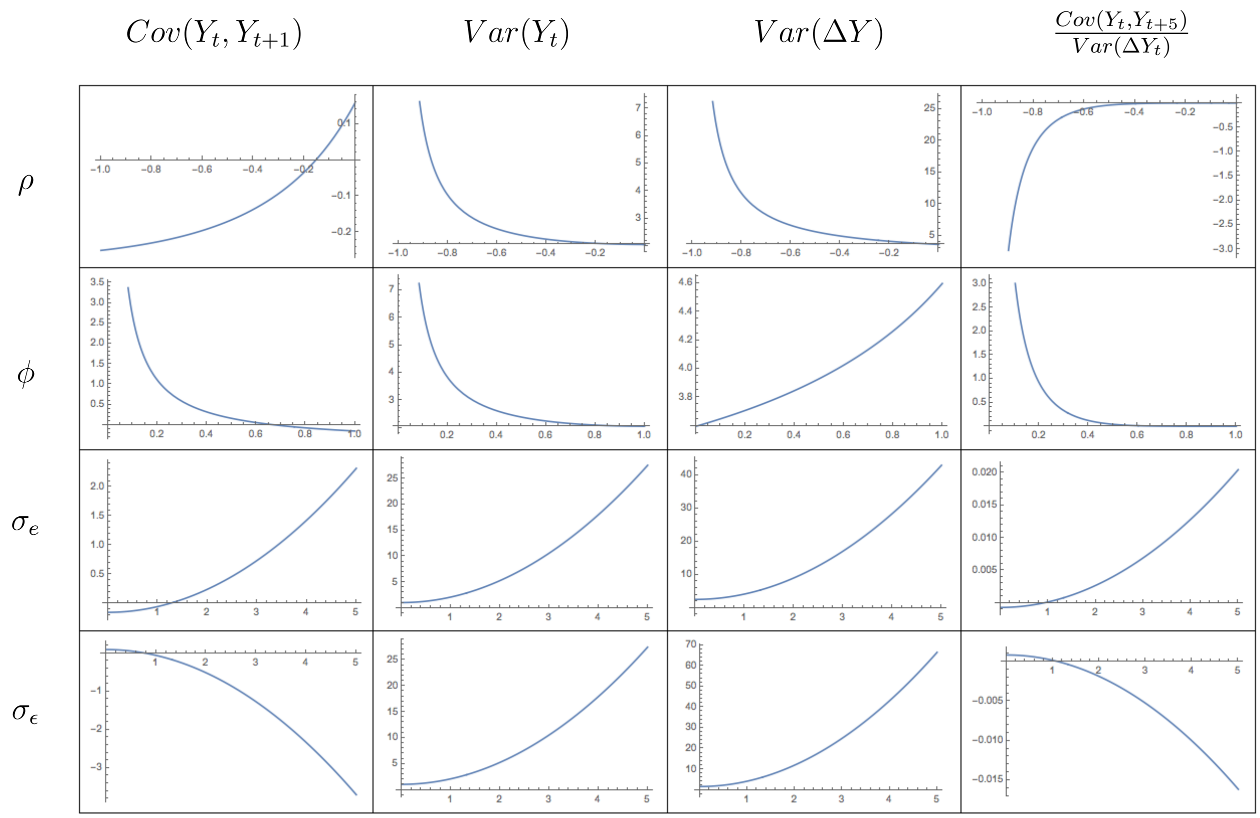
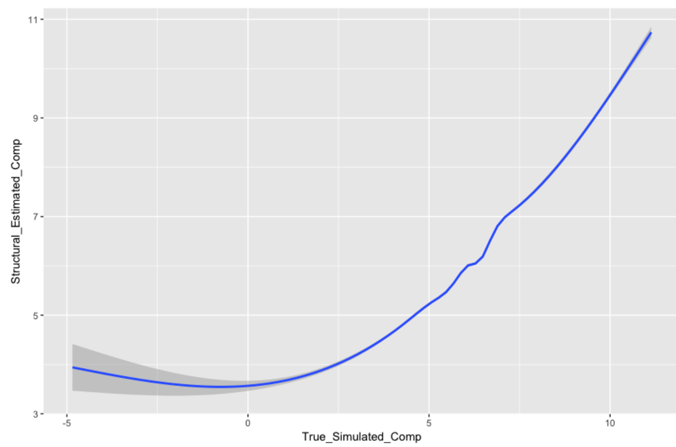


Figure 1: Identification of the parameters (x-axis, rows) based on the moments (y-axis, columns). Median values are used for the respective parameters ($\rho = -0.23$, $\phi = 0.76$, $\sigma_e = 1$, $\sigma_\epsilon = 1$).

[9] In particular, the matrix of slopes has non-zero determinant. Equivalently, each row of slopes is linearly independent of the others

The plots below show the relationship between estimated comparability and true comparability on a simulated data set. For any two pairs of simulated firms, our structural estimates can determine which pair is more comparable with 85% accuracy. In contrast, the *AcctComp* score correctly identifies the more comparable firm pair 53% of the time.

Structural Estimates vs True Simulated Values



DeFranco et al. AcctComp vs True Simulated Values

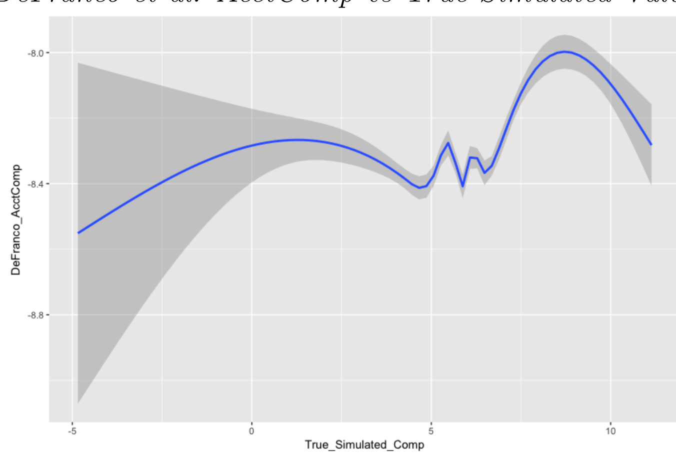


Figure 2: Graphs are based on simulated data which consist of 200 firms (forming 198,000 unique firm pairs) each with 40 quarters of "true economic performance", 40 quarters of earnings data, and 40 quarters of market return data. Each firm's data is generated according to the mean reverting model described in section (2), with parameters $(\rho, \sigma_\epsilon, \phi, \sigma_e)$ drawn from uniform distributions. For the purpose of calculating the DeFranco et al.'s AcctComp measure, we also generate simulated stock returns under the assumption that market prices are a noisy measure of the discounted value of true future economic performance.

Appendix A

Below are the closed form expressions for each of our moments. These are the values we would expect to observe for each moment, in steady state, given the true parameters.

$$\lim_{t\to\infty} Var(Y_t) = -\frac{\sigma_\epsilon^2}{\rho^2-1} - \frac{\sigma_e^2}{(\phi-2)\phi} \tag{10}$$

$$\lim_{t\to\infty} Var(\Delta Y_t) = \frac{2\sigma_\epsilon^2}{\rho+1} - \frac{2\sigma_e^2}{\phi-2} \tag{11}$$

$$\lim_{t\to\infty} Cov(Y_t, Y_{t-1}) = \frac{(\rho^2-1)\,\sigma_e^2(\phi-1) + 2\rho\sigma\epsilon^2\phi}{2(\rho-1)^2\phi} \tag{12}$$

$$\lim_{t\to\infty} \frac{Cov(Y_t, Y_{t-5})}{Var(\Delta Y_t)} = \frac{\sigma_e^2(\phi-1)^5}{(\phi-2)\phi} - \frac{\rho^5\sigma_\epsilon^2}{\rho^2-1} \tag{13}$$