Measuring Accounting Comparability

Skylar DeTure & Elisabeth Plietzsch

Discussion Paper No. 18-23
Measuring Accounting Comparability *

Skylar DeTure†  Elisabeth Plietzsch‡

September 4, 2018

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 markets 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 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

JEL Codes:

---

*We appreciate helpful comments from Ross Jennings, 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

We produce a measure of accounting comparability that relies only on time series earnings data. Firms have accounting systems that map true economic performance (which is unobservable) to accounting earnings (which is observable). We would say two firms have “comparable” accounting if the mapping from true economic performance to earnings is similar. Thus, to determine whether two firms have comparable accounting, we must determine whether they map economic performance to accounting earnings in a similar way. This is difficult to do because we typically cannot observe “true” economic performance. Thus, if we observe a difference in the accounting earnings of two firms, we cannot say whether the difference is due to differences in economic performance or differences in accounting.

Previous literature, most notably De Franco et al. (2011), have attempted to isolate accounting comparability from economic performance by using proxies for true economic performance. For example, De Franco et al. use stock prices as a proxy for true economic performance. The intuition is that if firm A’s stock price is related to its earnings according to some relationship $R_a$, and firm B’s stock price is related to its earnings according to some relationship $R_b$, then firm A and firm B have comparable accounting if $R_a$ and $R_b$ are similar. The challenge, however, is that stock 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 the firm’s accounting reports.

By using a structural model, we are able to produce a measure of accounting comparability that does not rely on stock return data or any other non-earnings data about firm performance. Thus, we imagine a world where accounting earnings is 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 unknown linear trend. We assume that each firm has a function $F$ which 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 $F$s are. If they have similar $F$s, we say they have comparable accounting. We do all of this without ever knowing the firm’s actual 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 markets data as an input, we can make two significant contributions to the accounting literature. First, we provide a measure which can be used to answer questions about the capital markets consequences of accounting comparability. For example, our measure can be used to answer questions like

- do firms with more comparable accounting have more similar earnings response coefficients?

- is 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?

These questions would be difficult to answer using a comparability measure which assumes market prices already capture true economic performance.

Second, our measure can be used to compute comparability scores for entities that don’t trade in liquid markets. For example, our technique can be used to measure the accounting comparability between a large private firm and a public firm, between two private firms, between two segments of a publicly traded firm, et cetera. All that is needed is that earnings data for the two entities be publicly available. Thus, our measure provides a tool for quantitative research on the causes and effects of accounting comparability in settings where it could not previously be studied.

Our measure contributes to the burgeoning literature on financial statement comparability. Measures for financial statement comparability are relatively scarce, and yet research
questions about financial statement comparability are abundant. Researchers care both about how comparability affects financial statement users, and about how accounting regulation affects comparability. In this first category, 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; 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, so this stream of research is unlikely to die down anytime soon.

Research in the second category, about how regulators and other gatekeepers affect accounting comparability, is also flourishing. Financial statement comparability is a topic of primary concern to regulators – 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. 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 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 Model of Accounting Comparability

We assume that each firm has true economic performance, unobservable, following a mean reverting process. Thus, firm $i$’s true economic performance at time $t$ is given by

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

(1)
with $\mu_i$ the mean economic performance for firm $i$, $\phi_i$ 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:

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

(2)

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

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

(3)

That is, accounting earnings introduce random errors $\epsilon_{i,t}^{noise} \sim N(0, \sigma_{e,i})$ to true economic performance, and those errors reverse at rate $|\rho_i|$.

Thus, our model reflects two critical assumptions from the accounting literature: (1) over the life of the firm, accounting earnings equal true economic performance and (2) the relationship between earnings and actual performance is stronger over longer time horizons (e.g., 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):

$$Y_t = \mu + e_1(1 - \phi)^{t-1} + \sum_{j=2}^{t-1} e_j(1 - \phi)^{t-j} + e_t + \sum_{j=0}^{t} \rho^{t-j} \epsilon_{j}^{noise}$$

(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 $\phi$ and $\sigma_e$ (the parameters governing true economic performance), but potentially different $\sigma_\epsilon$ and $\rho$ (the parameters that govern the mapping from economic performance to reported earnings). Formally stated:

\[
\text{incomparability} = E(|Y_{A,t} - Y_{B,t}|) = E\left[\left|e_{A,1}(1 - \phi)^{t-1} + \sum_{j=2}^{t-1} e_{A,j}(1 - \phi)^{t-j} + e_{A,t} + \sum_{j=0}^{t} \rho_t^{t-j} \epsilon_{A,j}^{\text{noise}}\right| - \left|e_{B,1}(1 - \phi)^{t-1} + \sum_{j=2}^{t-1} e_{B,j}(1 - \phi)^{t-j} + e_{B,t} + \sum_{j=0}^{t} \rho_t^{t-j} \epsilon_{B,j}^{\text{noise}}\right|\right]
\] (5)

We can see that $Y_{A,t} - Y_{B,t}$ is distributed normally with mean zero. Thus, $|Y_{A,t} - Y_{B,t}|$ follows a folded normal distribution with mean $\propto \text{Var}(Y_{A,t} - Y_{B,t})$. Rearranging (4) yields the variance of $Y_{A,t} - Y_{B,t}$: $\sigma_{\epsilon,A}^2 \sum_{j=0}^{t} (\rho_t^{t-j})^2 + \sigma_{\epsilon,B}^2 \sum_{j=0}^{t} (\rho_t^{t-j})^2$, which, in steady state (i.e. as $t$ goes to infinity), becomes

\[
\lim_{t \to \infty} \text{Var}(Y_{A,t} - Y_{B,t}) = \frac{(1 - \rho_B^2)\sigma_{\epsilon,A}^2 + (1 - \rho_A^2)\sigma_{\epsilon,B}^2}{(\rho_A^2 - 1)(\rho_B^2 - 1)}. \quad (6)
\]

To summarize, equation (6) gives us a formula for the incomparability of two firms $A$ and $B$ by telling us 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 (6) gives us a measure of accounting comparability:

\[
\text{comparability}(A, B) = -\frac{(1 - \rho_B^2)\sigma_{\epsilon,A}^2 + (1 - \rho_A^2)\sigma_{\epsilon,B}^2}{(\rho_A^2 - 1)(\rho_B^2 - 1)}. \quad (7)
\]

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

Essentially, we use our model from section 2 to write mathematical formulas for several moments of reported earnings. Following the economics and finance literature, 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, etcetera. For each moment of reported earnings, we can write a closed-form mathematical expression in terms of the parameters \((\rho, \sigma_e, \phi, \sigma_e)\) by plugging in our closed-form expression for reported earnings (4). For example, to derive a closed form expression of the variance of reported earnings:

\[
Var(Y_t) = Var[\mu + e_1(1 - \phi)^{t-1} + \sum_{j=2}^{t-1} e_j(1 - \phi)^{t-j} + e_t + \sum_{j=0}^{t} \rho^{t-j} e_j^{\text{noise}}]\]

(8)

We can see that the variance of the reported earnings is a function of the parameters \((\rho, \sigma_e, \phi, \sigma_e)\) and time \(t\). Assuming steady state, we can take the limit of \(f(\rho, \sigma_e, \phi, \sigma_e, t)\) 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 \((\rho, \sigma_e, \phi, \sigma_e)\).

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 \((\rho, \sigma_e, \phi, \sigma_e)\).

We can repeat this process for new moments. Each moment gives us a closed-form expression of our parameters \((\rho, \sigma_e, \phi, \sigma_e)\) 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 \((\rho, \sigma_e, \phi, \sigma_e)\). We do
this process separately for every firm, so every firm has its own values of \((\rho, \sigma, \phi, \sigma_e)\).

We rely on four moments to identify the four parameters for each firm: \(\text{Var}(Y_t), \text{Cor}(Y_t, Y_{t-1}), \text{Var}(\Delta Y_t),\) and \(
\frac{\text{Cor}(Y_t, Y_{t-5})}{\text{Var}(\Delta Y_t)}\)\). The graphs in figure… show how each of these moments changes in response to changes in the parameter values \((\rho, \sigma, \phi, \sigma_e)\). 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 whose elements are signs of the slopes in figure… to have a non-zero determinant.

To calculate the observed values of our moments, we use quarterly Compustat income before extraordinary items. We use quarterly data because we need as much time series variation as possible within each firm to calculate firm-specific parameters \((\rho, \sigma, \phi, \sigma_e)\). For this reason, we also restrict our sample to firms with at least 40 quarters of available Compustat data. We remove any linear trend from each firm’s earnings to ensure that our data reflect the model assumption that true 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.

After calculating the observed values of each moment for each firm, we use a two-step approach to estimate the parameters \((\rho, \sigma, \phi, \sigma_e)\) 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 \((\rho, \sigma, \phi, \sigma_e)\), i.e., 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.

\(^1\)Closed form expressions of the moment formulas are in the Appendix, alongside their derivations.
and standard errors for each firm. For any pair of firms, we can use their parameter estimates from the GMM to produce an estimate of their accounting comparability by plugging their parameter estimates into equation (7). 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 \((\rho, \sigma_e, \phi, \sigma_e)\) at random, and simulating 40 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% of the time. Panel A of figure 2 shows the conditional mean of our estimated comparability score, given the true comparability score, overall 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. \(\text{CompAcct}\) measure using our simulated data. To benchmark against \(\text{CompAcct}\), 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 \(\text{CompAcct}\) the largest possible advantage over our structural estimate. \(\text{CompAcct}\), 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 DeFranco et al. \(\text{CompAcct}\) 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 Results

In this section, we provide initial evidence on how our structural comparability measure performs compared to the DeFranco et al. measure when using real (as opposed to simulated) data. Therefore, we start with two very “quick and dirty” ways of validating our measure: 1) we check whether firms in the same industry are more comparable and 2) we analyze the correlation between our structural measure and the DeFranco et al. measure.

Further, we replicate selected results of two previous comparability papers. First, we replicate DeFranco et al. (2011), regarding the relationship between comparability and correlated analyst forecast errors. Second, we replicate Francis et al. (2014), regarding the influence of auditor style on comparability.

4.1 Very ”Quick and Dirty” Validation

We start our very “quick and dirty” validation by examining the effect of same industry membership on our structural comparability measure. Which accounting standards are mainly relevant depends hugely on the industry. For example, the rules of ASC 905 – Agriculture will be mainly applied by companies that operate primary in the agriculture industry, while companies in the retail industry should be unaffected by ASC 905. In order to test this relationship, we regress our structural comparability measure on a dummy variable that indicates whether both firms belong to the same industry.\(^2\) In untabulated results, we find that the coefficient is positive and highly significant (at the 1 % level).\(^3\) Thus, our structural measure classifies indeed two firms as more comparable when these are in the same industry. However, it should be also noted that the adjusted \(R^2\) is close to zero, indicating that same

\(^2\)Note, that this test is not able to indicate whether a comparability measure successfully separates the accounting comparability from the underlying economic comparability. The prediction with respect to same industry membership on economic comparability is ambiguous. On the one hand, firms in an industry underlie similar economic shocks and thereby economic comparability should be higher. On the other hand, an increase in one firm’s earnings might be also on the cost of a competitor, and therefore economic performance could be negatively correlated.

\(^3\)We find also a positive and significant coefficient when we use a log or rank transformation on our structural comparability measure. The comparability measure is in all regression winsorized at the 2.5 and 97.5 percentile.
industry membership is not able to explain a large portion of variation in comparability. The results are robust (but with higher p-values) to the inclusion of industry fixed effects for firm $i$ and firm $j$, when we use the log or rank transformation on our structural comparability measure.

Next, we calculate the spearman correlation between our comparability estimates and the corresponding DeFranco et al. estimates. The correlation is 0.22 with a p-value of $< 2 \times 10^{-16}$ using only data since 2005, and .41 with a p-value of $< 2 \times 10^{-16}$ using all data since 1962. 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 are high enough to indicate that both measures relate to a similar underlying construct but low enough that we can say that both measures are distinct.

### 4.2 Correlated Analyst Forecast Errors

*Note: the results of this subsection are based on an initial smaller sample and are not updated yet.*

We replicate DeFranco et al.’s (2011) test regarding the association of correlated analyst forecast errors and financial statement comparability. Firms that have more comparable accounting systems are expected to have more correlated forecast errors. Comparable accounting leads to similar deficiencies in the firm’s financial reporting, and therefore, analysts are more like to make similar errors when forecasting the firm’s earnings.

Table 1 presents the results of our replication. All variables are defined as in DeFranco et al. (2011). We include Industry fixed effects 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 importantly, all measures for accounting comparability are significantly positive as predicted. Our results differ concerning
the magnitudes. While DeFranco et al. report that a one-standard-deviation increase in CompAcct is associated with a 56% increase in CorrFestError, we find that a one-standard-deviation increase in CompAcct is associated with only a 21% (17%) increase in CorrFestError in our full (overlapping) sample. This difference appears to be driven by differences in sample selection between our paper and DeFranco et al.: compare to DeFranco et al., our sample exhibits lower means and standard deviations for both CompAcct and CorrFestError.

Next, we test the relationship between correlated analyst forecast error and comparability using our structural 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 CorrFestError (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, providing validation for both measures. However, the results are different enough to suggest that the measures do not capture the same construct. Importantly, the difference in results across the measures is consistent with the theoretical differences between the two measures: namely, 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 structural comparability measure only captures the first component. Since the second component mechanically correlates with analyst forecast error, we should expect the DeFranco et al. measure to give larger coefficient estimates than our structural measure, in this particular test.

An alternative explanation for why our structural comparability measure yields a smaller coefficient estimate than the DeFranco et al. measure is that our structural measure is too
noisy. We do not think the structural comparability measure is noisy enough to create these differences. First, the simulation tests in section 3 suggest that our estimation procedure is reasonably precise. Second, we test whether our measure gives similar results when we repeat each test several times on random subsamples. We find that the results using our measure are robust across subsamples, which we would not expect if our measure was pure noise.

4.3 Auditor Style

In this section, we replicate Francis et al.’s (2014) test on 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. Thus, Francis et al. (2014) expect that firms with the same Big 4 audit firm have more comparable accounting than firms with different auditors.

We choose to 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). Auditors are unlikely able to affect a firm’s underlying economics. Therefore, if any effect of having a same auditor exists, this effect should relate to accounting comparability.

We extend Francis et al.’s argumentation also to non-Big 4 auditors and expect that having the same auditor in general should increase accounting comparability compared to Big 4 and non-Big 4 firms that do not share the same auditor. We replicate Francis et al.’s test for three different accounting comparability measures: (1) the ECOMP measure from Francis et al. (2014), (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. (2014) except for Same_Auditor which indicates whether the
firm pair has the same Big 4 or non Big 4 auditor. We include 2-digit SIC industry fixed
effects for firm $i$ and $j$ and cluster standard errors at the firm $i$ level. We winsorize all
continuous variables at the 2.5 % level.

Table 2 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 structural measure as dependent variable. The significance is the highest in case of
our structural measure. The Same_Auditor coefficient is insignificant for the DeFranco et
al. measure indicating that this measure is either too noisy or captures mainly economic
comparability.

Next, we analyze how the comparability measures perform with smaller sample sizes.
Therefore, we assume that the sample used for table 2 represents the whole population and
that there should exists a positive relation 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 each 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 3 reports the
corresponding results. The columns $+*$, $-*$ and 0 state the proportion of samples with
a significant positive, significant negative or non-significant coefficient of Same_Auditor,
respectively. Table 2 shows that our structural measure still performs really good in smaller
sample sizes. Even in samples that consists only of 5 % of the original sample size, our
measure finds nearly always the expected positive association between Same_Auditor and
accounting comparability. For comparison, the ECOMP_COV measure finds only in 15.8 %
of the samples a significant positive effect. In summary, the results in table 2 support that
our measure seems to capture accounting comparability more precise.

\footnote{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
work against finding significant results.}
5 Conclusion

We develop and validate a 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 those firms map (unobservable) economic performance to reported earnings. Firms can have very comparable accounting even if they have starkly different economic performance.

Our measure is conceptually appealing because it assumes that earnings data is the only information available to users of financial statements. Thus, our measure is consistent with the common belief among accounting researchers that financial reporting provides information to the capital markets, not the other way around. In contrast with previous measures of accounting comparability have assumed that the information content of stock prices is uncorrelated with the information content of accounting earnings.

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

We find that our measure is moderately but not perfectly correlated with the De Franco et al. comparability measure, suggesting that these two measures capture related but distinct constructs. We replicate selected results from the comparability literature using our measure. Results using our measure differ from results using previous measures in predictable ways, consistent with the theoretical differences between our measure and previous measures. For example, in tests of capital markets 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 markets data as inputs, should have mechanically stronger relationships with capital markets 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 document that firms with a higher comparability score have more similar earnings response coefficients than firms with lower comparability scores. The similarity of earnings response coefficients is a critical capital markets consequence of comparability which previous measures have not been able to test. Moving forward, we plan to replicate more existing studies of comparability using our structural measure. Further, we plan to re-run our current replication tests on larger samples which covering more time periods. We would encourage other researchers interested in accounting comparability to use our measure, which we plan to make publicly available as the paper progresses.


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

$$\text{CorrFctError}_{ijt} = \alpha + \beta_1 \text{CompAcct}_{ijt} + \gamma \text{Controls}_{jt} + \epsilon_{ijt}$$

The dependent variable is $\text{CorrFctError}$, which proxies for the correlation in forecast errors between firm $i$ and $j$. $\text{CompAcct}$ denotes DeFranco et al.’s (2011) comparability measure. $\text{Comparability}$ 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 2: Effect of Big 4 Auditor Style on Comparability

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Structural Measure</th>
<th>DeFranco et al. Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ECOMP_COV</td>
<td></td>
</tr>
<tr>
<td>Same_Auditor</td>
<td>0.004**</td>
<td>0.025***</td>
</tr>
<tr>
<td>Size_Diff</td>
<td>-0.003</td>
<td>-0.015</td>
</tr>
<tr>
<td>Size_Min</td>
<td>-0.006</td>
<td>-0.068**</td>
</tr>
<tr>
<td>LEV_Diff</td>
<td>-0.017***</td>
<td>-0.022</td>
</tr>
<tr>
<td>LEV_Min</td>
<td>-0.025***</td>
<td>-0.056***</td>
</tr>
<tr>
<td>MB_Diff</td>
<td>0.025***</td>
<td>0.046***</td>
</tr>
<tr>
<td>MB_Min</td>
<td>0.030***</td>
<td>0.006</td>
</tr>
<tr>
<td>Loss_Prob_Diff</td>
<td>-0.020***</td>
<td>-0.002</td>
</tr>
<tr>
<td>Loss_Prob_Min</td>
<td>-0.029***</td>
<td>-0.113***</td>
</tr>
<tr>
<td>STD_Sales_Diff</td>
<td>0.016</td>
<td>0.029</td>
</tr>
<tr>
<td>STD_Sales_Min</td>
<td>0.013*</td>
<td>0.008</td>
</tr>
<tr>
<td>STD_CFO_Diff</td>
<td>-0.011</td>
<td>0.036</td>
</tr>
<tr>
<td>STD_CFO_Min</td>
<td>-0.012*</td>
<td>0.029*</td>
</tr>
<tr>
<td>STD_Sales_Grch_Diff</td>
<td>-0.007</td>
<td>0.039**</td>
</tr>
<tr>
<td>STD_Sales_Grch_Min</td>
<td>0.012**</td>
<td>0.013</td>
</tr>
<tr>
<td>CFO_Comp_Cov</td>
<td>0.015***</td>
<td>0.130***</td>
</tr>
<tr>
<td>RET_Cov</td>
<td>0.019***</td>
<td>0.054***</td>
</tr>
<tr>
<td>industry fixed effects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.9 %</td>
<td>11.0 %</td>
</tr>
<tr>
<td>No. of Obs.</td>
<td>415,380</td>
<td>415,380</td>
</tr>
</tbody>
</table>

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:

$$\ln(\text{Comparability Measure}) = \alpha + \beta_1 \text{Same}_4 + \gamma \text{Controls}_jt + \epsilon_{ijt}$$

The dependent variable is $\ln(\text{Comparability Measure})$ and stands for 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 also 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), with exception of Same_Auditor, which indicates whether both firms are audited by the same audit firm in general (not only Big 4 audit firms).
Table 3: Smaller Sample Performance Based on the Effect of Auditor Style on Comparability

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>ECOMP COV</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural measure</td>
<td>DeFranco measure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+*</td>
<td>-*</td>
<td>0</td>
</tr>
<tr>
<td>1%</td>
<td>0.099</td>
<td>0.021</td>
<td>0.88</td>
</tr>
<tr>
<td>5%</td>
<td>0.158</td>
<td>0.009</td>
<td>0.833</td>
</tr>
<tr>
<td>10%</td>
<td>0.231</td>
<td>0.004</td>
<td>0.765</td>
</tr>
<tr>
<td>30%</td>
<td>0.432</td>
<td>0</td>
<td>0.568</td>
</tr>
<tr>
<td>50%</td>
<td>0.635</td>
<td>0</td>
<td>0.365</td>
</tr>
<tr>
<td>70%</td>
<td>0.843</td>
<td>0</td>
<td>0.157</td>
</tr>
<tr>
<td>90%</td>
<td>0.994</td>
<td>0</td>
<td>0.006</td>
</tr>
</tbody>
</table>

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 2 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 2):

\[
\ln(\text{Comparability Measure}) = \alpha + \beta_1 \text{Same_Auditor} + \gamma \text{Controls}_{ijt} + \epsilon_{ijt}
\]

The dependent variable is \(\ln(\text{Comparability Measure})\) 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 \(^5\), we can identify how the parameters of two firms must differ to explain how the observed moments of those two firms differ.

\[
\begin{align*}
\text{Cov}(Y_t, Y_{t+1}) & & \text{Var}(Y_t) & & \text{Var}(\Delta Y) & & \frac{\text{Cov}(Y_t, Y_{t+5})}{\text{Var}(\Delta Y_t)} \\
\rho & & & & & \\
\phi & & & & & \\
\sigma_e & & & & & \\
\sigma_e & & & & & 
\end{align*}
\]

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

\(^5\)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 traditional AcctComp score correctly identifies the more comparable firm pair 53% of the time.

**Structural Estimates vs True Simulated Values**

**DeFranco 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 \((\rho, \sigma_e, \phi, \sigma_e)\) drawn from uniform distributions. For the purpose of calculating the DeFranco 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.

\[
\lim_{t \to \infty} \text{Var}(Y_t) = -\frac{\sigma^2_e}{\rho^2 - 1} - \frac{\sigma^2_e}{(\phi - 2)\phi}
\]  \hspace{1cm} (9)

\[
\lim_{t \to \infty} \text{Var}(\Delta Y_t) = \frac{2\sigma^2_e}{\rho + 1} - \frac{2\sigma^2_e}{\phi - 2}
\]  \hspace{1cm} (10)

\[
\lim_{t \to \infty} \text{Cov}(Y_t, Y_{t+1}) = \frac{(\rho^2 - 1)\sigma^2_e(\phi - 1) + 2\rho\sigma^2_e\phi}{2(\rho - 1)^2\phi}
\]  \hspace{1cm} (11)

\[
\lim_{t \to \infty} \frac{\text{Cov}(Y_t, Y_{t+5})}{\text{Var}(\Delta Y_t)} = \frac{\sigma^2_e(\phi - 1)^5}{(\phi - 2)\phi} - \frac{\rho^5\sigma^2_e}{\rho^2 - 1}
\]  \hspace{1cm} (12)