Disclosure Similarity and Future Stock Return Comovement

Travis A. Dyer
Brigham Young University

Darren T. Roulstone
Ohio State University

Andrew Van Buskirk*
Ohio State University

April 2022

Existing research often assumes that firms' financial reporting choices influence their return comovement with other firms. We examine the validity of that assumption. First, we provide initial evidence suggesting that similarity in two firms' disclosures not only predicts, but influences, future return comovement between those two firms. Second, we show that this predictive ability aggregates to the market level; disclosure similarity can be used to estimate more accurate forward-looking market betas. Taken together, these two results suggest that managers can influence their firms' betas by altering their firms' disclosures – a prominent assumption in existing research, but one with little empirical support until now.

*Corresponding author. We thank Eric Allen, Dirk Black, Jared Flake, Nick Guest, Patrick Kielty, Kyle Peterson, Amy Sheneman, Blake Steenhoven, and workshop participants at the Cornell Accounting Mini-camp, USC Accounting Mini-Conference, the BYU Accounting Research Symposium, INSEAD, the University of Colorado Boulder, the University of North Carolina-Chapel Hill, the University of Nebraska-Lincoln, the University of Rochester, the University of New South Wales, the University of Washington, Virginia Tech, and Washington University in St. Louis for helpful comments. We are grateful to the Samuel Curtis Johnson Graduate School of Management and the Fisher College of Business for financial support.

Disclosure Similarity and Future Stock Return Comovement

Abstract

Existing research often assumes that firms' financial reporting choices influence their return comovement with other firms. We examine the validity of that assumption. First, we provide initial evidence suggesting that similarity in two firms' disclosures not only predicts, but influences, future return comovement between those two firms. Second, we show that this predictive ability aggregates to the market level; disclosure similarity can be used to estimate more accurate forward-looking market betas. Taken together, these two results suggest that managers can influence their firms' betas by altering their firms' disclosures – a prominent assumption in existing research, but one with little empirical support until now.

1. Introduction

A common assumption in capital markets research is that firms' disclosure choices influence their stock's comovement with peer firms, their industry, and the market. Examples are abundant. Jin and Myers (2006) directly link opacity and market comovement (R^2), noting, "An increase in opaqueness, combined with capture by insiders, leads to lower firm-specific risk for investors and higher R^2s" (p. 258). Hutton et al. (2009) show that earnings management, as a measure of opacity, is associated with higher R^2 (i.e., a greater proportion of the firm's stock return is explained by industry and market returns). Brockman et al. (2010) adopt a similar attitude when examining the relation between information production and comovement; they conclude that comovement is lower when information production is higher.

This assumption also plays an important role in analytical papers. For example, Lambert et al. (2007) shows analytically how a firm's disclosures can affect its cost of capital via investors' perception of covariances among firms' cash flows. Crucially, this outcome relies on the assumption that a firm's financial reporting *can* influence investors' beliefs about future cash flows and the extent to which those cash flows will covary with other firms. However, while this assumption underlies numerous research studies, there is little empirical evidence to support such a causal link. Our goal is to provide initial evidence on whether that causal link exists.

We build on recent literature showing that disclosure similarity between two firms predicts their future return comovement (Box, 2018; Ibriyamova et al., 2019). Our analysis starts by examining whether the firms' disclosures themselves—beyond just predicting future return comovement—actually influence that comovement at the firm-pair level; that is, how similarities in two firms' disclosures might lead to comovement in those firms' stock returns. We then aggregate firm-pair disclosure similarity scores into a firm-level measure, and assess whether that

firm-level measure predicts a firm's comovement with the overall market (i.e., beta). If firms' disclosures influence firm-pair return comovement *and* the relation between disclosure similarity and comovement aggregates to the market level, we would conclude that the widespread assumption in the existing literature is warranted. Moreover, these results would also suggest that firms could effectively alter their industry and market betas via their reporting choices, holding their operating decisions and capital structure constant.

The obvious challenge in assessing a causal effect of disclosure on returns is that both the firms' disclosures and their returns are influenced by the economic events underlying those disclosures. This raises the possibility that, even in the absence of the firm's disclosures, other sources (e.g., media or analysts) would have provided the same information about those economic events, leading to similar return patterns. Thus, a statistical association between disclosure similarity and future return comovement would not necessarily imply a causal relation.

Our study takes three different approaches to address this empirical challenge. First, we examine the link between disclosure similarity and return comovement cross-sectionally, with the goal of isolating the effect of the firm's disclosures vs. other potential information sources.[2] In particular, we examine whether disclosure similarity between two firms exhibits a stronger relation with future return comovement between those two firms when investors are likely to place greater reliance on the firm's disclosures rather than other information sources. Our proxies for those situations are: (i) both firms are small and presumably enjoy little media coverage, (ii) both firms lack analyst coverage, and (iii) investors perceive earnings announcement periods to be relatively

[2] As we discuss in Section 2, this does not require that investors learn *directly* from the firm's disclosures. Instead, our question is whether investors' beliefs, and therefore future return patterns, would look different in a counterfactual world where the firm's disclosures did not exist. For example, if investors obtain all of their information from analysts, but those analysts generate output that is influenced by the firm's disclosures, then we would still conclude that the disclosures causally influenced investors' beliefs and future returns, with analysts playing a mediating role. The concern would be if analysts were able to generate the same information without the firm's disclosures.

more important sources of information than non-earnings announcement periods. In each case, we document a significantly stronger relation between disclosure similarity and return comovement, consistent with the firms' disclosures—as opposed to some other source—influencing that comovement.

We next study a scenario where we can more directly observe investors' reliance on the firms' disclosures – a sample of firms that issued both an original and a subsequently-amended SEC filing. We assume that the amended filing, relative to the original filing, is a more accurate reflection of the firm's financial performance and economic condition. In this setting, we calculate two measures of disclosure similarity between an amending firm *i* and a non-amending firm *j*. The first is based on the firm *i*'s original filing, while the second is based on firm *i*'s amended filing. The difference between these two similarity measures reflects the extent to which the original filing overstated or understated the similarity between firms *i* and *j*, and we characterize this misstatement as noise in the original similarity measure. By construction, the amended filing is not publicly available until after the future return comovement is measured.

If the amending firm's initial disclosure influenced future return comovement, we would expect the noise component in the original similarity measure to predict future return comovement, even though that component of similarity is subsequently shown to be unwarranted.[3] For example, if the original filing falsely characterized the two firms as being very similar and investors relied upon that disclosure in their future trading decisions, subsequent returns would covary as if the two firms truly were very similar; return comovement would be positively associated with the

[3] Our use of the term "noise" is intentionally broad. In this context, noise could represent information in the original filing that was removed or edited in the amended filing, as well as information that was absent in the original filing and added in the amended filing. The key is whether the original filing—relative to the amended filing—was *too similar* to the peer firm's disclosure (leading to higher-than-appropriate return comovement) or *insufficiently similar* to the peer firm's disclosure (leading to lower-than-appropriate return comovement).

noise in the original similarity measure. On the other hand, if the firm's performance and financial condition are revealed independently through other channels, we expect the (more accurate) amended disclosure similarity to explain future return comovement, with no relation between the noise component of similarity and comovement.[4]

Our amended filings test shows that the noise component of disclosure similarity from subsequently-amended filings predicts future return comovement, even after controlling for the similarity of the amended filing. The incremental predictive ability of similarity noise, controlling for those contemporaneous economic factors, implies that the unwarranted component of the original disclosure similarity influenced the subsequent movement of the firm's stock with the stocks of peer firms.[5] The magnitude of this effect is large – a unit of similarity noise has roughly two-thirds the explanatory power of a unit of disclosure similarity based on the firm's amended filing. In one sense, this result is not surprising; investors rely upon firms' disclosures, even if those disclosures subsequently turn out to be wrong. What is novel about this result, though, is the consequence of this reliance – more similar disclosures lead to greater return comovement, even when that similarity turns out to be misstated.

Our third approach focuses on a sample of firms entering and exiting the S&P 500. We build on Schoenfeld (2017), who argues that firms entering the S&P 500 face demand for greater disclosure from index investors, and who shows that firms respond to this demand by increasing their voluntary disclosure upon entering the S&P 500. In the same vein, we predict that if analysts and index investors have disclosure preferences for the various firms they follow/own, they will pressure the entering firms to satisfy those disclosure preferences, as well. If the entering firms

[4] Although the actual amended filing is not observable during the return measurement period, the text of the amended filing can be interpreted as a more accurate version of the original filing.

[5] A conceptually similar empirical approach—including both the original similarity measure and the similarity measured based on the amended filing in the same regression—yields the same inference.

cater to that demand, their disclosures will evolve to look more similar to peer index firms. We find evidence consistent with this prediction: Firms' disclosures indeed become more (less) similar to index peers' disclosures when those firms enter (exit) the S&P 500 index.

We then find, using an approach similar to Schoenfeld (2017), that the index-induced increases in disclosure similarity (i.e., disclosure changes induced by index membership and not by changes in other firm fundamentals) are associated with subsequent increases in return comovement with other index firms. When S&P entrance induces larger disclosure effects, we observe larger changes in return comovement, even after controlling for contemporaneous changes in institutional ownership and analyst following. Based on the same assumptions made in Schoenfeld (2017), who uses this approach to conclude that disclosure causally influences stock liquidity, we view our results as suggesting a causal relation between disclosure similarity and future returns.[6] In summary, all three of our approaches suggest that firms' disclosures influence future return comovement at the firm-pair level.

We conclude our paper by examining whether this firm-pair relation aggregates to the market level. We find that a firm-level measure of disclosure similarity, aggregated from firm pairs, is associated with a firm's future market beta. That is, a weighted average of firm *i*'s disclosure similarities with a group of peer firms predicts firm *i*'s future market beta. This significant relation is robust to controlling for the firm's recent estimated beta, its average earnings

[6] As we discuss in greater detail in Section 4.1.3, the key identifying assumption in this analysis is that, even if firms entering the index are all subject to new demand for disclosure that looks more similar to existing index firms, individual firms will be more or less sensitive to that demand (e.g., a manager may be less sensitive to this demand if that manager is more entrenched). As a result, some firms will exhibit a greater increase in disclosure similarity than other firms even though all firms were subject to the same shock in demand. Accordingly, this analysis exploits cross-sectional variation in the treatment effect on disclosure, and compares it with cross-sectional variation in the treatment effect on future return comovement. Assuming that the variation in firms' willingness to cater to demand is only associated with subsequent changes in return comovement via the disclosure channel, we interpret the relation between the firm-specific change in disclosure similarity and the residual change in return comovement as a causal one.

comovement, and the inclusion of firm fixed effects. Thus, disclosure similarity is incrementally informative about future beta and portfolio risk.

Our paper offers several contributions to both practitioners and researchers. Specifically, we provide novel evidence consistent with firms' disclosures influencing their future return comovement with other firms, as well as the overall market. This evidence provides empirical support for a longstanding assumption in capital markets research – that firms' disclosure choices affect their comovement as well as their cost of capital. In combination, these results imply that managers' disclosure choices can influence their betas, even without changes in underlying operations or capital structure. We acknowledge that establishing causality is difficult and that we cannot entirely rule out other explanations for our results; however, we believe the varied approaches we take make it unlikely that our results are driven by an obvious alternative factor. Causality aside, our results extend prior research on disclosure similarity by documenting those circumstances where disclosure similarity is more or less predictive of future return comovement. Further, our results provide a practical benefit, in that disclosure similarity based on mandatory SEC filings can be a useful tool to better estimate portfolio risk and firms' forward-looking betas.

Finally, our paper extends the literature on disclosure attributes and their consequences by showing that a firm's disclosure choices do not just affect that firm's stock price immediately following the disclosure; rather, a firm's *joint* disclosure choices with another firm influence the two firms' subsequent return comovement, even over relatively long periods. In particular, this paper extends the results in Koo et al. (2017), who show that qualitative attributions in firms' earnings forecasts (e.g., attributing an increased earnings forecast to positive industry trends rather than firm-specific factors) influence the immediate stock price response of peer firms. Our results

suggest that firms' joint disclosure choices also influence the way that investors interpret and respond to *future* events, leading to predictable patterns in return comovement.

2. Prior Research

2.1. Return comovement and its interpretation

Return comovement is a fundamental construct in financial economics, underlying beta in the Capital Asset Pricing Model and aiding in portfolio construction more generally. Reflecting this fundamental role, an extensive literature has investigated microeconomic and macroeconomic determinants of return covariance. Central to this research is the idea that return covariances are primarily driven by economic forces – firms' stock prices are more likely to move together if those firms are subject to common risk factors (Fama and French 1993), and comovement tends to increase in bad times as firms face similar economic pressure (Erb et al. 1994; Ding et al. 2020).

On top of those common risk factors, prior research documents other determinants of return comovement. Pindyck and Rotemberg (1993) finds greater comovement among stocks predominantly held by institutional investors, suggesting that market segmentation influences comovement. Other papers show that category investing (Boyer 2011), shared ownership (Antón and Polk 2014), investor attention (Drake et al. 2016; Huang et al. 2019), and analyst following (Muslu et al. 2014; Israelsen 2016) all play a role in return comovement. Existing literature also points to behavioral or sentiment-based factors affecting comovement: Barberis et al. (2005) shows that comovement with S&P 500 firms increases after being added to the index, while Green and Hwang (2009) finds that similarly priced stocks comove, suggesting that investors categorize stocks based on index membership and nominal per share prices.

Researchers have interpreted firms' return comovement as being informative about those firms' information environments and disclosure policies, implicitly assuming that firms' disclosures affect their observed comovement. Piotroski and Roulstone (2004) documents that information intermediaries are associated with how much a firm's returns comove with industry and market returns and interprets a lack of market-level synchronicity as an indication that there is more firm-specific information in prices. Jin and Myers (2006) and Brockman et al. (2010) apply a similar interpretation to synchronicity. Haggard et al. (2008) studies analysts' ratings of firms' disclosure policies and finds improved disclosure quality is associated with lower stock price synchronicity (i.e., correlation with market and industry returns), again implying that higher quality disclosure is associated with lower return comovement. Dang et al. (2015) provides country-level evidence that commonality in news coverage is associated with commonality in returns.

2.2. Disclosure similarity

Just as prior research has studied how firms' stock prices comove, other research investigates how firms' disclosures are similar to one another, and whether those similarities are due to economic forces, managerial discretion, or information intermediaries.[7] Drawing on measures from computational linguistics research (e.g., Salton et al. 1975), Brown and Tucker (2011) studies the similarity of firms' MD&A disclosures from one year to the next. They find that firms experiencing economic shocks are more likely to modify their MD&A language. Hoberg and Phillips (2010) uses 10-K product descriptions to measure product similarity between firms and evaluates how that product similarity influences firms' merger and acquisition decisions. Jung

[7] Here, we focus on studies specifically involving disclosure similarity and ignore other aspects of similarity like financial statement similarity (De Franco et al. 2011) and audit similarity (Knechel et al. 2015).

(2013) studies similarity in the context of commodity price risk disclosures finding that firms are more likely to follow a first mover's decision to disclose additional quantitative information when those firms have greater investor overlap.

Three recent papers explore disclosure commonality in the audit context: Drake et al. (2019) shows that auditors engage in disclosure "benchmarking," where they assess nonclient financial statement information when evaluating a client's financial statement information, resulting in greater comparability among footnotes. De Franco et al. (2020) shows that firms' MD&As are more similar when those firms share the same auditor, and McMullin (2016) provides evidence that financial statement preparers borrow footnote language from firms with whom they share an auditor.

2.3. The relation between disclosure similarity and future return comovement

Two recent papers address the relation between disclosure similarity and future return comovement directly. Box (2018) examines firm-specific newswire content and finds that the linguistic similarity between two firms' content predicts future comovement between those two firms, even after controlling for historical return comovement and other factors. Although the Box (2018) sample is based on newswire stories, some of that content represents firm-issued disclosure. Similarly, Ibriyamova et al. (2019) shows that brief company descriptions from Thomson Reuters and Yahoo! Finance can be used to predict future return comovement. In other words, both studies indicate that qualitative information about two firms is incrementally useful in predicting the future comovement between those firms.

We build on these papers by focusing on similarity in mandatory financial filings, as these filings constitute a significant source of information to investors and do not depend on media or

analyst coverage. These filings also typically include explicit discussions of forward-looking information (e.g., MD&A) that may be especially likely to predict or influence future return patterns. More importantly, while Box (2018) and Ibriyamova et al. (2019) ask whether qualitative textual information can *predict* future return comovement, our questions are in what circumstances is that predictive ability greater, and whether qualitative disclosure similarity can *affect* future return comovement. Further, we are interested in whether a relation between disclosure similarity and return comovement at the firm-pair level extends to a relation between disclosure similarity and comovement at the market level.

There are several plausible channels through which firms' disclosure choices could affect subsequent return comovement. First, firms could choose to discuss or withhold their exposure to certain risk factors. For example, suppose a firm's disclosure reveals sensitivity to a particular commodity. In that case, the firm's future returns are likely to be more sensitive to news about that commodity, and thus covary more with other firms who have disclosed a sensitivity to that commodity.[8]

Second, the firm could reveal plans to enter into a new line of business, even if that announcement is opportunistic. For example, Cooper et al. (2001) document a substantial "dotcom" effect for firms changing their corporate names to Internet-related dotcom names. More recently, several studies have documented market effects for firms making blockchain-related announcements (e.g., Jain and Jain 2019; Cheng et al. 2019; Cahill et al. 2020; Sharma et al. 2020). In both cases, investors could perceive that those firms are increasingly subject to Internet or

[8] This is true even if those same disclosure choices do not have an immediate pricing effect. Li and Ramesh (2009), for example, find that there is generally little market reaction around quarterly periodic SEC filings. However, the Li and Ramesh results do not imply that those filings will be irrelevant for future returns – disclosing a particular risk exposure might not affect the firm's *current* value, but still affect investor response to future news about that risk factor.

blockchain-related factors, leading to greater comovement with other firms exposed to those factors.

Similarly, managers have discretion in how they attribute their financial performance to market or industry factors rather than firm-specific factors. If a manager attributes much of the current period's performance to industry factors, that disclosure could lead investors to perceive an increased similarity between the firm and its industry peers, and therefore greater comovement with those peers going forward. Such an outcome would be consistent with Koo et al. (2017), who show that peer firms are more likely to experience positive information transfer from earnings forecasts when those earnings forecasts are attributed to industry trends.[9]

Finally, changes in disclosure could influence the nominal categories that investors assign to a particular stock, leading to greater comovement for firms in the same "category." Prior evidence of this phenomenon includes Barberis et al. (2005), who show increased comovement for firms entering the S&P 500; Green and Hwang (2009), who show greater comovement among similarly-priced stocks; Boyer (2011), who shows greater comovement among firms with the same S&P/Barra Value/Growth classification; Chen et al. (2016), who show that otherwise-similar firms have sharply different industry betas depending on whether they fall above or below a 50% sales threshold that determines the firm's SEC-designated industry; and Hameed and Xie (2019), who show that stocks initiating dividends comove with other dividend-paying stocks.

Overall, we build on these streams of research by asking the following questions: In what circumstances does disclosure similarity better predict future return comovement? Does disclosure similarity influence future return comovement? Does the predictive relation between disclosure

[9] There are two important distinctions between this paper and Koo et al. (2017). First, Koo et al. (2017) study the immediate short-term response to earnings forecasts, while we are interested in longer-term price behavior following the disclosures in question. Second, we focus on the consequences of two firms' joint disclosure choices, rather than the unidirectional effect of one firm's disclosure on another firm.

similarity and return comovement at the firm-pair level aggregate to a predictive relation between disclosure similarity and future market beta?

3. Sample data and the determinants of disclosure similarity

3.1. Sample and measurement

Our sample begins in 1996, when regulatory filing data for all public firms is first machine-readable via EDGAR, and ends in the third quarter of 2017. We obtain all 10-K and 10-Q filings from EDGAR, making sure to exclude amended filings from the main analysis. Since the filing data comes in a variety of formats (e.g., plain text, HTML), we standardize the documents and make them machine-readable by following the steps in Dyer et al. (2017).[10]

To form our measure of disclosure similarity, we match each firm filing (firm *i*) to a peer firm filing (firm *j*) and constrain the filings to be reporting information for the same calendar quarter.[11] In this paper, we form three distinct firm-pair samples: (1) our broad sample, used for descriptive purposes and cross-sectional analysis, (2) our amended filing sample, and (3) our S&P 500 inclusion/exclusion sample. For our broad sample, since the magnitude of possible pairings is extremely large for the entire population of firms (5,000 firms each quarter over 21 years ≈ 1 billion firm pairs) and the text processing is computationally intensive, we choose to calculate disclosure similarity for a reduced number of firm-pairs. Note that because of this choice, samples (2) and (3) are not a subset of sample (1).

[10] See the Appendix for details regarding the 10-K/Q cleaning procedures.

[11] Note that when we refer to a matched firm as a "peer" firm, we are not implying that the matched firms share industry membership or have similar fundamentals. In our paper, a "peer" firm is simply a matched firm. Further, we do not examine how a firm's returns comove with the returns of the firm's industry as is done in the stock return synchronicity literature (e.g., Jin and Myers 2006). Comovement in our paper is always between two individual firms. That being said, we show in Section 4.2 that the relation between disclosure similarity and comovement does aggregate to the market level.

To limit our broad sample while still maintaining cross-sectional variation, we calculate disclosure similarity for randomly matched firm pairs. We form these firm pairs by randomly assigning firms to one of 48 groups and only matching firms within these randomly assigned groups.[12] Having matched firm filings within these groups, we estimate the textual similarity of the two filings by using a standard natural language processing technique called cosine-similarity (e.g., Brown and Tucker 2011; Hoberg and Phillips 2010). This process constructs two vectors of words, the first being the words used in the given firm filing and the second being the words used in the matched firm filing. Both vectors of words count the frequency of word references in the respective filing. Cosine-similarity takes the cosine of the two vectors of words. The result is a measure of the similarity in word use between the two filings and is our measure of disclosure similarity (*Disclosure Similarity*).

Our sample contains one unique pairing between firm *i* and *j* in each time period *t*.[13] After constraining the sample to have all relevant data from EDGAR, Compustat, CRSP, Thomson Reuters, and I/B/E/S, we obtain a sample of 728,831 unique firm pairs and 9,889,293 observations.

Table 1 presents the descriptive statistics for the broad sample. For our measures of disclosure similarity and earnings and return comovement, we report descriptive statistics for the raw variables.[14] In our statistical tests, we convert continuous variables to a standard normal for interpretational convenience. The data in Table 1 shows that firm pairs in the sample share a Fama-French 48 industry classification 6% of the time. That is, although we matched firms based on

[12] We chose to use 48 groups simply because that number of groups is commonly used in empirical research to identify industry membership. However, as our groups are assigned randomly, the assigned groups have no inherent meaning in our study.
[13] The two firms in the matched pair remain matched until one of the firms disappears from EDGAR. As new firms appear in EDGAR, they are assigned randomly to one of the 48 groups and matched to each firm in the group.
[14] Using mandated regulatory filings to calculate disclosure similarity results in high cosine similarity scores. The median firm-pair in our sample has a cosine similarity score of 0.94. For comparison, Brown and Tucker (2011) document a median similarity score of 0.89 using only the MD&A section.

randomly-assigned groups, firm pairs shared the same actual industry in about 6% of the observations. Additionally, sharing a common dedicated investor (Bushee, 2001), analyst, and auditor occurs in 17%, 1%, and 15% of the sample, respectively.

3.2. Determinants of disclosure similarity

To provide some context for *Disclosure Similarity* in our broad sample, we first perform an analysis of its determinants using the following regression specification:

$$Disclosure\ Similarity_{i,j,t} = \beta \cdot Determinants_{i,j,t} + \alpha_{i,t} + \alpha_{j,t} + \varepsilon_{i,j,t} \quad (1)$$

Determinants is a vector of variables we expect to be related to similarity in disclosure, where each variable measures the relation between firm *i* and firm *j* in time *t*. We include fixed effects associated with firm *i* (firm *j*) in time *t,* which controls for time-specific shocks to disclosure similarity that uniformly affect all the pairings of firm *i* (firm *j*) in time *t*.[15]

We have several predictions regarding the determinants of disclosure similarity, and we broadly classify these determinants into two categories: fundamental and clientele effects. We expect firms with similar economic fundamentals to use similar language to describe those fundamentals in their periodic filings. For example, we expect firms that share a common industry classification or have a similar reliance on capital markets to disclose similar content. Additionally, extant research suggests that firms that share similar market-to-book, market value of equity, performance, and stock price momentum experience similar return patterns (Fama and French, 1993; Carhart, 1997). As a consequence, we expect the similarities underlying those correlated returns to translate into disclosure similarity.

[15] As is pointed out in prior work (e.g., Brown and Tucker 2011), measures using cosine similarity are mechanically related to the length of the two input documents. Additionally, we know that other features of corporate disclosures have systematically changed over time, including the extent of boilerplate language (Dyer et al. 2017). The employed fixed effects remove such mechanical effects from contributing to the inferences drawn in this paper.

To capture each of these dimensions of fundamental similarity, we identify instances where firm *i* and firm *j* share the same Fama-French 48 industry classification (*Same Industry*). Additionally, we create indicator variables *Similar MTB*, *Similar SIZE*, *Similar ROA*, and *Similar MOM,* which take the value of one if firm *i* and firm *j* share the same decile of market-to-book, market value of equity, return on assets, and stock return, respectively, during the reporting period. We expect each of these determinants to be positively related to *Disclosure Similarity*.

The second group of predictions centers on clientele effects.[16] As discussed earlier, prior research suggests that firm clientele can influence firms' disclosure choices (Boone and White 2015; Schoenfeld 2017; Chapman and Green 2018; De Franco et al. 2020; Park et al. 2019; Abramova et al. 2020). We study three dimensions of firm clientele: analysts, investors, and auditors. We expect that for each dimension, shared clientele will increase the similarity of the two firms' disclosures. *Same Analyst* is an indicator variable taking the value of one if firm *i* and firm *j* received a forecast from the same analyst during the fiscal period, and zero otherwise. *Same Investor* is an indicator variable taking the value of one if firm *i* and firm *j* share the same dedicated investor (Bushee, 2001) during the fiscal period, and zero otherwise.[17] *Same Auditor* is an indicator variable taking the value of one if firm *i* and firm *j* share the same auditor during the fiscal year, and zero otherwise.

Table 2 presents the determinants of disclosure similarity.[18] The first key takeaway from this table is that dimensions of fundamental similarity are related to disclosure similarity. Specifically, sharing the same industry is associated with a 0.244 standard deviation increase in

[16] We acknowledge that our two groups of determinants are not mutually exclusive. Industry-focused mutual funds or size-based index funds, for example, will result in some firm pairs having both fundamental and clientele similarity.
[17] We use dedicated investors, rather than the entire population of institutional investors because some diversified investors tend to hold positions in most firms simultaneously (e.g., Total Stock Market funds), resulting in almost all firm pairs sharing at least one common institutional investor.
[18] With the exception of the Schoenfeld-style regressions in Tables 6 and 7, our regressions cluster standard errors by firm *i* and year-quarter.

disclosure similarity, the largest effect in the table. In addition, measures of similarity in common risks are also positively associated with disclosure similarity – two firms are more likely to share common disclosure language if they are in the same decile of *Size*, market-to-book (*MTB*), profitability (*ROA*), and momentum (*MOM*).

The second takeaway is that there seem to be economically and statistically significant clientele effects. Sharing an analyst (or a dedicated institutional investor) with a peer firm is associated with a 0.141 standard deviation increase (0.046 standard deviation increase) in disclosure similarity. These associations broadly support the notion that common analysts and investors encourage commonality in disclosure. That being said, the associations do not necessarily imply causality; disclosure similarity (or fundamental similarity underlying that disclosure similarity) is likely to attract common analyst and investor coverage, which would lead to the same association without clientele influencing disclosure. We also find that sharing an auditor is positively related to disclosure similarity, as documented by De Franco et al. (2020).

3.3. Baseline relation between disclosure similarity and return comovement

We next evaluate the association between *Disclosure Similarity* and future return comovement within our broad sample, where *Future Return Comovement*$_{i,j,t}$ is the Pearson correlation of the stock returns of firm *i* and *j* over the 90 days following the later date of firm *i* and firm *j*'s disclosure. Figure 1 Panel A graphically illustrates the measurement timing for the main variables of interest in the broad sample. We begin by showing the univariate relation between return comovement and three intuitive determinants. Specifically, Figure 2 shows *Future Return Comovement* by quartile of disclosure similarity, earnings comovement, and lagged return comovement. Unsurprisingly, past return comovement shows the sharpest relation with future return comovement. Earnings comovement shows a positive, but small, association with future

return comovement. Finally, disclosure similarity shows a clear positive association with future return comovement.

We next measure the determinants of return comovement in a multivariate setting using the following OLS regression:

$$Future\ Return\ Comovement_{i,j,t} = \beta_1 \cdot Disclosure\ Similarity_{i,j,t} + Controls \qquad (2)$$

Disclosure Similarity is as previously defined and represents the degree of qualitative similarity between two firms. Higher degrees of correlation (β_1) suggest that *Disclosure Similarity* can be used to incrementally predict *Future Return Comovement* between two firms. *Controls* represents a vector of control variables that includes previous return comovement during the fiscal quarter (*Return Comovement*$_{i,j,t}$) and quarterly earnings comovement over the prior three years (*Earnings Comovement*$_{i,j,t}$), as well as fixed effects for each firm *i* calendar quarter and each firm *j* calendar quarter.

Table 3 presents the results of this multivariate regression. Column (1) controls for historical stock return comovement, historical earnings comovement, and both firm *i*'s and firm *j*'s average disclosure similarity during that period. As expected, in Column (1) we see that historical return comovement (*Return Comovement*$_{i,j,t}$) is an economically significant predictor of future return comovement, with a correlation of 0.135 (i.e., a one-standard deviation increase in past return comovement is associated with a 0.135 standard deviation increase in future return comovement).[19] Further, we find that *Disclosure Similarity* is incrementally predictive of subsequent return comovement, controlling for the previous period's return comovement as well as earnings comovement. When we control for commonality in firm characteristics (Column (2)),

[19] This auto-correlation in return comovement becomes larger when fixed effects are omitted. Disclosure similarity continues to be a significant predictor of future return comovement with fixed effects omitted.

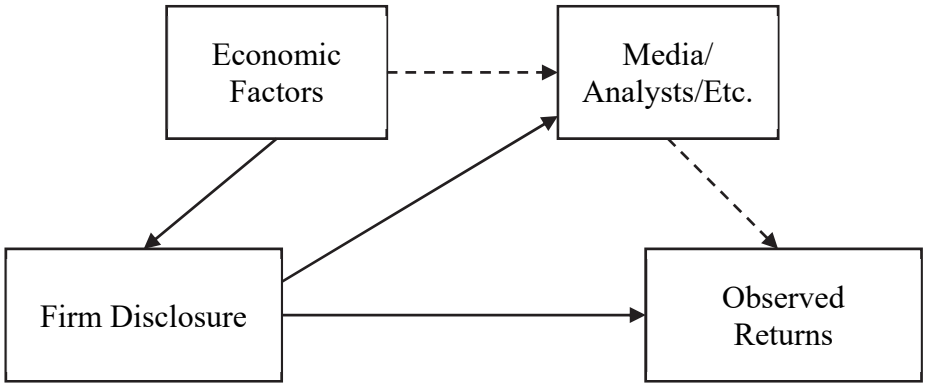
we continue to observe that *Disclosure Similarity* is positively associated with *Future Return Comovement*.

The results in Table 3 are broadly in line with the results in Box (2018) and Ibriyamova et al. (2019), who show that similarity in newswire text and company descriptions can be used to predict future return comovement. While an important extension of prior work is that we relate return comovement to disclosure similarity derived from firm-provided, mandatory filings, our main research questions are addressed in the following sections.

4. Empirical Analysis – disclosure similarity and future return comovement

4.1 Does disclosure influence return comovement?

Our first question is whether two firms' joint disclosure similarity influences their future return comovement. What makes this a difficult question to address is that a firm's disclosures (and therefore its disclosure similarity with other firms) are driven by underlying economic events that could be communicated by other information channels. Thus, even if we observe a statistical association between disclosure and future returns, that relation could be due to the firm's disclosures themselves, as we hypothesize, or via other information sources. In graphical terms, we are interested in the following scenario:



Here, a firm's disclosure can have both a direct effect on post-disclosure returns and an indirect effect on returns (mediated by other information sources who rely upon the firm's disclosures); either effect would be consistent with our predictions. However, those other information sources present a problem to the extent that they could also generate information *without* relying upon the firm's disclosures (i.e., a situation where information flows only along the dashed path in the graphic). Because we cannot completely measure and control for those other information sources, they represent a potential correlated omitted variable. The following sections discuss the three different empirical approaches we take to address this challenge.

4.1.1 Cross-sectional variation in comovement predictability

Our first approach starts with the basic statistical relation between disclosure similarity and subsequent return comovement, and then investigates how that relation varies as a function of investor reliance on the firms' financial reporting output. We expect that if disclosure similarity itself drives return comovement, that relation should be stronger in circumstances where investors place greater reliance on those disclosures. Said differently, if other information channels drive the relation between disclosure similarity and comovement, we expect the relation to be weaker in the absence of those other information channels.

We proxy for those circumstances in three ways. First, we identify firm pairs where each firm is relatively small (i.e., below the median when firms are ranked on market value of equity), based on the assumption that investors in small firms have fewer sources of external information such as sell-side analysts or media coverage. Second, we identify firm pairs where neither of the two firms has sell-side analyst coverage, again based on the assumption that investors will place greater reliance on firms' own disclosures when they have fewer external sources of information.

Third, we identify firm pairs where both firms' earnings announcement periods (i.e., when the information in firms' 10-Ks and 10-Qs are first revealed) are perceived to be relatively more informative than non-earnings announcement periods. We use option-based implied volatility to measure the extent to which investors view earnings announcement periods as relatively more informative than non-earnings announcement periods. This measure of relative importance follows Iselin and Van Buskirk (2020) and is based on the ratio of the earnings announcement implied volatility to a baseline level of implied volatility; the greater that ratio, the greater the perceived importance of the firm's earnings announcement. For each of these proxies, we expect a stronger relation between disclosure similarity and future return comovement if the firms' disclosures influence that comovement. On the other hand, if disclosure similarity predicts future return comovement because of some other correlated information source, we would expect the relation between disclosure similarity and comovement to be weaker in the absence of those other information sources.

Table 4 shows the results of this analysis, where each of the three columns features a different proxy for whether the firm-pair is likely to have less external information and therefore greater investor reliance on those firms' financial reporting output: both firms having below-median size (Column 1), neither firm having analyst coverage (Column 2), or both firms having above-median levels of anticipated earnings announcement importance, using the measure in Iselin and Van Buskirk (2020) (Column 3).[20] In each case, we include both the investor reliance measure and the interaction of the reliance measure and disclosure similarity.

[20] Specifically, this proxy is measured immediately prior to a firm's earnings announcement using option-based implied volatilities. The measure is equal to the ratio of excess earnings variance to baseline variance, where the variance measures are calculated as follows: Excess earnings variance is $\frac{\sigma^2_{t,30} - \sigma^2_{t,60}}{\frac{252}{21} - \frac{252}{42}}$, and baseline variance is $\frac{\sigma^2_{t,30}}{252} - \frac{\sigma^2_{EA\ Excess}}{21}$, where $\sigma^2_{t,30}$ and $\sigma^2_{t,60}$ refer to implied volatiles from 30-day and 60-day standardized options, respectively.

For each proxy, the regression results in Table 4 show that return comovement is not only increasing with firms' disclosure similarity, but that the relation is significantly stronger when investors are likely to place greater reliance on the firms' disclosures. The magnitude of this effect is large in comparison to the baseline relation between similarity and comovement. For example, when we use size as a proxy for investor reliance, the incremental predictive effect of disclosure similarity is more than three times larger when both firms are small than for the remaining firm pairs (0.012 as a baseline effect compared to 0.012 plus 0.032 for small firm-pairs). Even the smallest effect, based on options-based implied volatility in Column 3, implies a 10 percent larger relation in situations where both firms' earnings announcements are perceived as being more influential than the median.

In summary, disclosure similarity between two firms predicts future return comovement between those firms. Moreover, that effect is larger in situations where investors place greater reliance on firms' disclosures, suggesting that the disclosures themselves affect that comovement. The following two sections further explore this causal relation.

4.1.2 Amended filings

As described above, a challenge in establishing a causal link between disclosure similarity and future return comovement is the possibility that relevant economic fundamentals are disclosed through other sources independent of corporate disclosures. Our second approach addresses this challenge by focusing on a situation where the firm's disclosures were inaccurate in a way that was unlikely to be shared by independent outside sources. Specifically, we examine situations where firms filed an initial SEC filing, and then subsequently amended that filing because of

Iselin and Van Buskirk (2020) show that this measure reflects an *ex ante* firm-level measure of the degree to which earnings announcements are, in expectation, important sources of information.

perceived inaccuracies. To the extent that investors incorporate those disclosure inaccuracies in their pricing decisions, this would be consistent with disclosure (rather than independent channels for economic information) influencing future return comovement.

For each amending firm *i* and a matched non-amending firm *j*, we create two measures of disclosure similarity between those two firms – one based on firm *i*'s original filing and a second based on firm *i*'s amended filing. We assume that firm *i*'s original filing captures the firm's fundamentals with some amount of noise, and that the amended filing better reflects firm *i*'s economic fundamentals. Therefore, disclosure similarity based on the original filing (*Disclosure Similarity – Original*) should be a noisier measure of the economic similarity between firm *i* and firm *j* than disclosure similarity based on the amended filing (*Disclosure Similarity – Amended*).

The difference between those two measures (*Disclosure Similarity – Error*) reflects the *unwarranted* disclosure similarity between the two firms in the original filing, and could be positive or negative. When that measure is positive, it indicates that the original filing *overstated* the extent of similarity with the peer firm's filing. When that measure is negative, it indicates that the original filing *understated* the extent of similarity with the peer firm's filing.

If investors assess firms' comparability based on accurate firm characteristics, as reflected in the amended filing, we would expect *Disclosure Similarity – Amended* to predict future return comovement. Furthermore, if investors derive their assessed comparability from independent sources, that assessment will be unrelated to the unwarranted disclosure similarity implied by the original filing.[21] In this scenario, observed return comovement would be associated with the warranted similarity reflected in *Disclosure Similarity – Amended*, and unrelated to *Disclosure Similarity – Error*.

[21] The assumption here is that if non-firm information sources generate their information without relying on the firm's disclosures, that information will not include the noise that was in the firm's original disclosure.

On the other hand, if investors do rely upon the firm's disclosure to assess comparability, and then use that assessment in their subsequent pricing, we would expect the same positive relation between future return comovement and *Disclosure Similarity – Amended*, but also a positive relation between comovement and *Disclosure Similarity – Error*. That is, if investors are influenced by an original filing that overstates the similarity between firms *i* and *j (*i.e., *Disclosure Similarity Error* >0*)*, the subsequent return comovement between those two firms will be higher than the comovement implied by (the more accurate) *Disclosure Similarity – Amended*.

To test these predictions, we estimate the following specification:

$$\begin{aligned} Future\ Return\ Comovement_{i,j,t} = \ & \alpha_1 \cdot Disclosure\ Similarity - Amended_{i,j,t} \\ & + \alpha_2 \cdot Disclosure\ Similarity - Error_{i,j,t} \\ & + Controls \end{aligned} \tag{3}$$

A positive and significant α_2 estimate would indicate that return comovement is associated with the noise component of disclosure similarity and would imply that disclosure similarity influences comovement in future returns.

We obtain a sample of amended 10-K/Q filings that occur 90 days after the initial filing date. This design choice means that the amended filings themselves could not have influenced the observed return comovement, as they are filed after the future return comovement is measured.[22] Figure 1 Panel B depicts the measurement timing for this test. We constrain amended filings to be at least 90 percent of the length of the original disclosure to remove cases where the amended disclosure is not likely to be a complete 10-K/Q disclosure.[23] The resulting sample of original and amended filings is paired with all possible peer firms' filings that occurred during the same time period. As a result, the sample for these tests is different from the sample used in Tables 1-4.

[22] To be more precise, the *new information* in the amended filings could not influence the observed return comovement. The information that was disclosed in both the original and the amended filing could.

[23] Results are identical if we constrain the length to be larger than the original filing; however, we include smaller amended filings in our sample to allow for cases where information was removed from the original 10-K.

Table 5 presents the association between future return comovement and disclosure similarity using the amended filing sample. Both columns present the association between *Future Return Comovement* and the two measures of disclosure similarity (*Disclosure Similarity – Error* and *Disclosure Similarity – Amended)*, with Column (2) featuring a rich set of control variables. We find that both components of disclosure similarity (i.e., the warranted similarity reflected in the amended measure and the unwarranted noise component) are related to future return comovement patterns, with the estimated coefficient for the noise component equal to about 65% of the estimated coefficient for the warranted component (0.020 compared to 0.031 in Column 2).[24]

One interpretation of these results—namely, the fact that even subsequently-amended information is associated with returns—is that they simply indicate that investors find firms' original disclosures credible and revise their beliefs about firm value accordingly. That interpretation is not only intuitive, but implicit in prior studies showing a significant investor reaction to firms' announcing restatements of their results (e.g., Palmrose et al. 2004). However, our results go beyond that conclusion. These results imply that when firms' disclosures are similar, that similarity leads to investors pricing the firms' stocks more similarly in the future, even when that disclosure similarity wasn't warranted by economic circumstances. Thus, we document that disclosure has an effect on future returns even when it is not simply proxying for the underlying economics.

4.1.3 Disclosure and comovement changes around S&P 500 entrance/exit

[24] An alternative design choice is to simply include both the Original similarity measure and the Amended similarity measure, rather than focusing on the "error" portion of the original disclosure. Untabulated results indicate that the two approaches yield the same inferences.

Our third approach exploits a setting where firms experience a shock in corporate disclosure preferences (by investors), and assesses whether that induced change in disclosure leads to changes in subsequent return comovement. In particular, this test is based on a shock to the number of shared index investors, driven by firms' inclusion in (exclusion from) the S&P 500. This approach has two aspects. First, we predict that index inclusion will generate demand on the entering firm for disclosure that looks similar to disclosures provided by firms already in the index. For example, new investors may demand specific forward-looking information like capital expenditure guidance, or more industry- or market-level information than the entering firm had previously provided to their former investors. Second, we predict that the resulting change in disclosure similarity will be associated with subsequent changes in return comovement.

The first prediction is based on existing evidence that investor demand influences firms' disclosure choices. For example, Boone and White (2015) shows that institutional ownership (related to index membership) influences firms' disclosure decisions, leading to a higher level of disclosure. Schoenfeld (2017) shows a similar effect for firms entering the S&P 500. More recently, Park et al. (2019) shows that common ownership increases disclosure levels because it decreases proprietary costs, and Abramova et al. (2020) shows that short-term changes in attention from passive institutional owners influence firms' disclosure quantity. While these papers document firms responding to investor demand by providing *more* information, we expect that the nature of that information will be similar to what those investors already demand from the other firms they own. This prediction is in line with the results in Jung (2013), who shows that firms are more likely to emulate peer firms' commodity price risk disclosures when they share investors

with the peer firms. If so, the entering firms' disclosures will evolve to look more similar to those of the existing index members.[25]

The second prediction is that the induced changes in disclosure will be associated with changes in return comovement. We recognize that changes in shared ownership almost certainly have a direct (i.e., unrelated to disclosure) effect on return comovement (e.g., Barberis et al., 2005; Antón and Polk, 2014); our interest is in the indirect effect: Changes in shared ownership influence disclosure similarity, which then influences return comovement. The benefit of this setting is that we can observe and empirically measure both the direct and indirect effects on return comovement.

The key identifying assumption is that, even if firms entering the S&P 500 index are subject to a common shock in demand for new disclosure, the response of those firms will be heterogeneous. For example, some managers may be more entrenched than others, and therefore be less sensitive to investor demand for disclosure.[26] As a consequence, entering firms will exhibit changes in disclosure similarity that are induced by index inclusion, but that are not perfectly correlated with changes in other factors like common analyst coverage or institutional ownership that are likely to affect return comovement on their own.

To perform this analysis, we follow the empirical approach in Schoenfeld (2017). This entails estimating the treatment effect of S&P 500 index inclusion on all our independent and dependent variables, for each inclusion/exclusion event. In doing so, we evaluate whether larger disclosure effects from index inclusion are related to larger changes in return comovement, even

[25] The results in Pawliczek et al. (2021) are also consistent with this prediction. They show that BlackRock's annual Dear CEO letters, asking/pressuring firms to provide greater disclosure along environmental and regulatory lines, result in those portfolio firms providing more disclosure on those topics.

[26] Indeed, Schoenfeld (2017) shows that when firms enter the S&P 500, the observed increase in disclosure quantity is greater for firms whose managers are less entrenched.

after controlling for contemporaneous changes in direct effects (i.e., shared index ownership) and potential coincidental changes in firm fundamentals (e.g., convergence in firm size).

If the induced increases in disclosure similarity around S&P 500 entrance are associated with increased return comovement, after controlling for other contemporaneous changes in the firm's environment, we would conclude that the changes in disclosure similarity influenced that comovement. However, if S&P 500-driven increases in disclosure similarity are not accompanied by an increase in future return comovement, it would suggest that disclosure similarity on its own does not influence return comovement.

We start by identifying 377 firms that either entered or exited the S&P 500 between 1996 and 2017 and also have requisite data for control variables (firm *i*). We then calculate firm *i*'s *Future Return Comovement* and *Disclosure Similarity* with a peer firm (firm *j*) in each period *t,* where peer firms are a combination of (1) every firm in the S&P 500 in time *t* and (2) 1,000 randomly selected firms.[27] When the paired firm is an S&P firm, the firm-pair observation is identified as *Treat* = 1; when the paired firm is a randomly selected firm, the firm-pair observation is identified as *Treat* = 0. For those time periods when the entering firm is in the S&P 500, we set *Inclusion* equal to 1; *Inclusion* is equal to 0 for the time periods when the entering/exiting firm is not in the index. The benefit of benchmarking against a set of randomly selected firms is that it allows us to rule out the concern that the new S&P 500 firms were changing their disclosures in such a way as to become more similar to *all* firms around the time they were entering the index (e.g., adding more common or boilerplate language to their disclosures).

[27] We require the set of firms available for random selection to survive more than half the sample period. This constraint increases the number of matches with data in each period and allows for time-series comparisons. We find similar results when drawing from the entire population of firms and randomly selecting new firms each period.

For each entering/exiting firm, we estimate a difference-in-differences effect for disclosure similarity and future return comovement by regressing those variables on *Treat x Inclusion*, while controlling for firm and time-period fixed effects. We do the same for all other control variables. The estimated coefficient on *Treat x Inclusion* reflects the extent to which *Disclosure Similarity (*or *Return Comovement)* increases between entering firms and S&P 500 peer firms for the periods when the entering firm is in the index, relative to disclosure similarity between entering firms and random firms for the same periods.[28] Table 6 shows the results for *Disclosure Similarity*, and illustrates that our first prediction appears to be true. That is, firms entering the S&P 500 exhibit a statistically significant increase in disclosure similarity with their new index, consistent with the entering firms responding to a demand for information from common owners and analysts.

Table 7 Panel A shows the firm-level estimates for the remaining variables, with each estimated firm-level effect labeled with the prefix *TxI*. Here, as expected, we observe positive changes in *Return Comovement*, as well as other intuitive effects like significant increases in common analyst coverage and common institutional ownership. In essence, the treatment effects summarized in Panel A are analogous to the difference-in-difference estimates used in the analysis in Schoenfeld (2017).

We then use these estimated treatment effects to assess whether the estimated treatment effect on disclosure similarity (*TxI_Disclosure Similarity*) is correlated with the estimated treatment effects for other firm-pair attributes (e.g., *TxI_Future Return Comovement*). The results of this analysis are shown in Panel B. In Column (1), the most important result is that *TxI_Same Quasi-Indexer Holdings* is positively correlated with *TxI_Disclosure Similarity*. This correlation

[28] The phrase "estimated effect" does not necessarily imply that index inclusion *caused* the observed change in each variable. Instead, this process estimates, for each firm, the contemporaneous change in each variable around the index entrance/exit. In cases where a single firm has both an entrance and an exit, we estimate a single treatment effect for that firm being included in the index.

is important because index investors' choice of *whether* to hold a pair of stocks is determined primarily by index membership, rather than firm-pair fundamentals. Thus, the documented correlation is unlikely to be driven by a correlated omitted variable.

We next evaluate the relation between future return comovement and disclosure similarity based on their respective treatment effects (Column (2)). We find a significantly positive association between the intensity of change in disclosure similarity and the intensity of the change in subsequent return comovement, even after controlling for the various effects that index inclusion has on our control variables. These results suggest that holdings by common investors affect return comovement through changes in disclosure similarity.[29] This indirect effect of quasi-index holdings on return comovement through disclosure similarity is approximately 9% of the magnitude of the direct effect ((0.232 · 0.172) / 0.437).[30] Overall, the results in Tables 6 and 7 suggest that S&P 500 index inclusion induces disclosure changes, and that those induced changes causally influence future return comovement.

[29] As in Schoenfeld (2017), this inference requires a set of assumptions to provide causal inference: (1) the change in shared investors must directly affect *Disclosure Similarity* and *Future Return Comovement*, and not vice versa; (2) the change in shared investors must not affect changes in *Disclosure Similarity* through *Future Return Comovement*; and (3) the error terms in the empirical models for *Disclosure Similarity* and *Future Return Comovement* must be uncorrelated. As in Schoenfeld (2017), these assumptions appear reasonable in our case. Index investors determine *whether* to hold a stock based on index membership rather than firm-pair features such as disclosure similarity or return comovement. Additionally, it is unlikely that induced changes in *future* return comovement, coming from shared investor holdings, will affect previously-made disclosure choices. Lastly, we are unaware of a factor that affects both *Disclosure Similarity* and *Future Return Comovement* that is not accounted for by the time-varying firm fixed effects, time-varying control-firm fixed effects, difference-in-difference adjustment, and difference-in-difference control variables. Moreover, we estimate (using the konfound function in Stata) that in order for a correlated omitted variable to invalidate our inference, that omitted variable would have to have a higher correlation with return comovement and disclosure similarity than any of the independent variables in our regression.

[30] An alternative approach to evaluating this research question is to start with a regression (similar to the main regression in Barberis et al. 2005) that explains comovement as a function of index inclusion, and then add disclosure similarity to that regression. If the disclosure similarity dampens the relation between index inclusion and comovement, it would imply that a portion of that relation is due to the disclosure effect. We perform this analysis (untabulated) and find similar inferences; disclosure similarity accounts for approximately 5% of the increase in comovement documented by Barberis et al. (2005) for firms entering the S&P 500.

4.2 Does the relation between disclosure similarity and return comovement aggregate to the market level?

Having shown evidence that disclosures influence return comovement at the firm-pair level, we now investigate whether these relations aggregate to the market level. Such an aggregation, if present, would suggest that a firm can, in fact, influence its beta via disclosure choices. To do so, we calculate each firm's *weighted average* disclosure similarity (*Ave. Disclosure Similarity*) across all other matched firms in the same period, where the firm pairings are the same random pairings as those used in Tables 2 and 3. The weighting is based on each matched firm's market value of equity. We then estimate a forward-looking beta (*Future Beta*) as the estimated coefficient on market return in a regression explaining firm *i*'s future return as a function of the future market return where the market return is the value-weighted average of the matched firms' returns.

The results of this analysis are shown in Table 8, and they indicate that the firm-pair levels can aggregate into an informative firm-level measure. Specifically, Table 8 shows a positive and significant relation between *Ave. Disclosure Similarity* and *Future Beta*. This relation appears to be economically meaningful, as well, considering that the relation between disclosure similarity and future beta is approximately twice as large as the relation between earnings comovement and future beta. This relation is consistent with our earlier evidence that disclosure similarity predicts comovement at the firm-pair level, and is robust to controlling for a firm's historical beta, its average earnings comovement, and the inclusion of firm and year fixed effects. Our results, in combination, support the longstanding assumption in capital markets research – that firms' disclosure choices affect their comovement as well as their cost of capital. Furthermore, this

evidence suggests that disclosure similarity, based on mandated filings, offers a practical tool to better estimate market- or portfolio-level covariances.

5. Conclusion

We use textual analysis to examine disclosure similarity among firms, and how that similarity not only predicts, but influences future return comovement. Across three different empirical settings, we find evidence that disclosure similarity (and changes in disclosure similarity) between two firms causally affects the future comovement between those two firms' stock prices. Furthermore, we show that this relation aggregates to the market level. A firm-level measure of disclosure similarity, aggregated from firm pairs, is predictive of future market beta.

Our study provides novel empirical evidence suggesting that firms' disclosures causally affect future return comovement. As a result, we provide validation for a common assumption in prior research (e.g., Jin and Myers 2006; Hutton et al. 2009; Brockman et al. 2010) and provide novel empirical support for the proposition in Lambert et al. (2007) that firms' disclosure choices can influence their cost of capital by affecting investors' assessed covariances. Our paper contributes to the broad literature studying return comovement, and shows that disclosure similarity (like common ownership, analysts, and auditors) both predicts and influences future stock return comovement. Finally, our paper extends the literature on disclosure attributes and their consequences by showing that a firm's disclosure choices do not just affect that firm's stock price; rather, a firm's *joint* disclosure choices with another firm influence the two firms' subsequent return comovement.

References

Abramova, I., J. E. Core, and A. Sutherland. 2020. Institutional Investor Attention and Firm Disclosure. *The Accounting Review* 95 (6): 1-21.

Antón, M., Polk, C., 2014. Connected Stocks. *Journal of Finance* 69 (3): 1099–1127.

Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. *Journal of Financial Economics* 75 (2): 283–317.

Boone, A. L., White, J. T., 2015. The Effect of Institutional Ownership on Firm Transparency and Information Production. *Journal of Financial Economics* 117 (3): 508–533.

Box, T., 2018. Qualitative Similarity and Stock Price Comovement. *Journal of Banking & Finance* 91: 49–69.

Boyer, B. H., 2011. Style-related comovement: Fundamentals or labels? *The Journal of Finance* 66 (1): 307-332.

Brockman, P., Liebenberg, I., Schutte, M., 2010. Comovement, Information Production, and the Business Cycle. *Journal of Financial Economics* 97 (1): 107–129.

Brown, S. V., Tucker, J. W., 2011. Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research* 49 (2): 309–346.

Bushee, B.J., 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207-246.

Cahill, D., Baur, D. G., Liu, Z., Yang, J. W., 2020. I Am a Blockchain Too: How Does the Market Respond to Companies' Interest in Blockchain? *Journal of Banking & Finance* 113: 105740.

Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52 (1): 57-82.

Chapman, K., Green, J. R., 2018. Analysts' influence on managers' guidance. *The Accounting Review* 93 (1): 45-69.

Chen, H., Cohen L., Lou, D. 2016. Industry Window Dressing. *Review of Financial Studies* 29 (12): 3354–3393.

Cheng, S. F., De Franco, G., Jiang, H., Lin, P., 2019. Riding the Blockchain Mania: Public Firms' Speculative 8-K Disclosures. *Management Science* 65 (12): 5901–5913.

Cooper, M., Dimitrov, O., Rau, P., 2001. A Rose.com by any Other Name. *Journal of Finance* 56 (6): 2371-2388.

Dang, T.L., Moshirian, F., Zhang, B., 2015. Commonality in News around the World. *Journal of Financial Economics* 116: 82-110.

De Franco, G., Fogel-Yaari, H., Li, H., 2020. MD&A Textual Similarity and Auditors. *Auditing: A Journal of Practice & Theory* 39 (3): 105-131.

De Franco, G., Kothari, S. P., Verdi, R. S., 2011. The Benefits of Financial Statement Comparability. *Journal of Accounting Research* 49 (4): 895–931.

Ding, R., Zhou, H., Li, Y., 2020. Social Media, Financial Reporting Opacity, and Return Comovement: Evidence from Seeking Alpha. *Journal of Financial Markets* 50: 100511.

Drake, M. S., Jennings, J., Roulstone, D. T., and Thornock, J. R., 2016. The Comovement of Investor Attention. *Management Science* 63 (9): 2847–2867.

Drake, M. S., Lamoreaux, P. T., Quinn, P. J., and Thornock, J. R., 2019. Auditor Benchmarking of Client Disclosures. *Review of Accounting Studies* 24 (2): 393-425.

Dyer, T., Lang, M., and Stice-Lawrence, L., 2017. The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting & Economics* 64 (2/3): 221–245.

Erb, C. B., Harvey, C. R., Viskanta, T. E., 1994. Forecasting International Equity Correlations. *Financial Analysts Journal* 50 (6): 32–45.

Fama, E. F., French, K. R., 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33 (1): 3–56.

Fama, E. F., French, K. R., 1997. Industry Costs of Equity. *Journal of Financial Economics* 43 (2): 153–193.

Green, T. C., Hwang, B., 2009. Price-based return comovement. *Journal of Financial Economics* 93 (1): 37-50.

Haggard, K. S., Martin, X., Pereira, R., 2008. Does voluntary disclosure improve stock price informativeness? *Financial Management* 37 (4): 747-768.

Hameed, A., Xie, J., 2019. Preference for dividends and return comovement. *Journal of Financial Economics* 132 (1): 103-125.

Hoberg, G., Phillips, G. M., 2010. Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *Review of Financial Studies* 23 (10): 3773–3811.

Huang, S., Huang, Y., Lin, T., 2019. Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics* 132 (2): 369-383.

Hutton, A. P., Marcus, A. J., Tehranian, H., 2009. Opaque financial reports, R2, and crash risk. *Journal of Financial Economics* 94 (1): 67-86.

Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., Stolin, D., 2019. Predicting Stock Return Correlations with Brief Company Descriptions. *Applied Economics* 51 (1): 88–102.

Iselin, M., and A. Van Buskirk. 2020. Event-Specific Uncertainty and its Expected Resolution. *SSRN eLibrary*.

Israelsen, R. D., 2016. Does common analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis* 51 (4): 1193-1229.

Jain, A., Jain, C., 2019. Blockchain Hysteria: Adding "Blockchain" to Company's Name. *Economics Letters* 181: 178–181.

Jin, L., Myers, S. C., 2006. R2 around the world: New theory and new tests. *Journal of Financial Economics* 79 (2): 257-292.

Jung, M. J., 2013. Investor Similarity and Diffusion of Disclosure Practices. *Review of Accounting Studies* 18 (1): 167–206.

Knechel, W. R., Vanstraelen, A., Zerni, M., 2015. Does the Identity of Engagement Partners Matter? An Analysis of Audit Partner Reporting Decisions. *Contemporary Accounting Research* 32 (4): 1443–1478.

Koo, D. S., J. Julie Wu, and P. E. Yeung. 2017. Earnings Attribution and Information Transfers. *Contemporary Accounting Research* 34 (3): 1547–1579.

Li, E., Ramesh, K., 2009. Market reaction surrounding the filing of periodic SEC reports. *The Accounting Review* 84 (4): 1171-1208.

Lambert, R., Leuz, C., Verrecchia, R. E., 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* 45 (2): 385-420.

McMullin, J. L., 2016. Can I Borrow Your Footnotes? Footnote Boilerplate's Learning Externality. *SSRN eLibrary*.

Muslu, V., Rebello, M., Xu, Y., 2014. Sell-Side Analyst Research and Stock Comovement. *Journal of Accounting Research* 52 (4): 911–954.

Palmrose, Z.-V., V. J. Richardson, and S. Scholz. 2004. Determinants of Market Reactions to Restatement Announcements. *Journal of Accounting and Economics* 37 (1): 59–89.

Park, J., Sani, J., Shroff, N., White, H., 2019 Disclosure incentives when competing firms have common ownership. *Journal of Accounting and Economics* 67 (2-3): 387-415.

Pawliczek, A., A. N. Skinner, and L. A. Wellman. 2021. A New Take on Voice: The Influence of BlackRock's 'Dear CEO' Letters. *Review of Accounting Studies* 26 (3): 1088–1136.

Pindyck, R. S., Rotemberg, J. J., 1993. The Comovement of Stock Prices. *Quarterly Journal of Economics* 108 (4): 1073–1104.

Piotroski, J. D., Roulstone, D. T., 2004. The Influence of Analysts, Institutional Investors, and Insiders on the Incorporation of Market, Industry, and Firm-Specific Information into Stock Prices. *The Accounting Review* 79 (4): 1119–1151.

Salton, G., Wong, A., Yang, C. S., 1975. A Vector Space Model for Automatic Indexing. *Communications of the ACM* 18 (11): 613–620.

Schoenfeld, J., 2017. The Effect of Voluntary Disclosure on Stock Liquidity: New Evidence from Index Funds. *Journal of Accounting and Economics* 63 (1): 51–74.

Sharma, P., Paul, S., Sharma, S., 2020. What's in a Name? A Lot If It Has "Blockchain." *Economics Letters* 186: 108818.

Appendix

Variable Definitions:

Variable Name	Variable Definitions
Disclosure Similarity	The cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter.
Future Return Comovement	The Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure.
Return Comovement	The Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the calendar quarter prior to 10-K/Q disclosure.
Earnings Comovement	The Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure.
Same Industry	An indicator variable taking the value of one if firm *i* and firm *j* share the same Fama-French 48 industry classification during the fiscal quarter, and zero otherwise (Fama and French, 1997).
Same Investor	An indicator variable taking the value of one if firm *i* and firm *j* share the same dedicated investor (Bushee, 2001) during the fiscal period, and zero otherwise.
Same Analyst	An indicator variable taking the value of one if firm *i* and firm *j* received a forecast from the same analyst during the fiscal period, and zero otherwise.
Same Auditor	An indicator variable taking the value of one if firm *i* and firm *j* share the same auditor during the fiscal year, and zero otherwise.
Similar MTB	An indicator variable taking the value of one if firm *i* and firm *j* share the same decile of Market-to-Book, and zero otherwise. Market-to-Book is measured as the firm's Market Value of Equity scaled by Book Value of Equity as of the end of the fiscal quarter.
Similar SIZE	An indicator variable taking the value of one if firm *i* and firm *j* share the same decile of Market Value of Equity, and zero otherwise. Market Value of Equity is measured as of the end of the fiscal quarter.
Similar ROA	An indicator variable taking the value of one if firm *i* and firm *j* share the same decile of Return on Assets, and zero otherwise. Return on Assets is measured as net income scaled by total assets as of the end of the fiscal quarter.
Similar MOM	An indicator variable taking the value of one if firm *i* and firm *j* share the same decile of Stock Return, and zero otherwise. Stock Return is measured as the percentage change in stock price over the fiscal quarter.
Both Small Firms	An indicator if both firm *i* and firm *j* have below-median size.
No Analyst Coverage	An indicator if neither firm *i* nor firm *j* has analyst coverage.
Both High EA Importance	An indicator if both firm *i* and firm *j* have above-median levels of anticipated earnings announcement importance, using the measure in Iselin and Van Buskirk (2020).

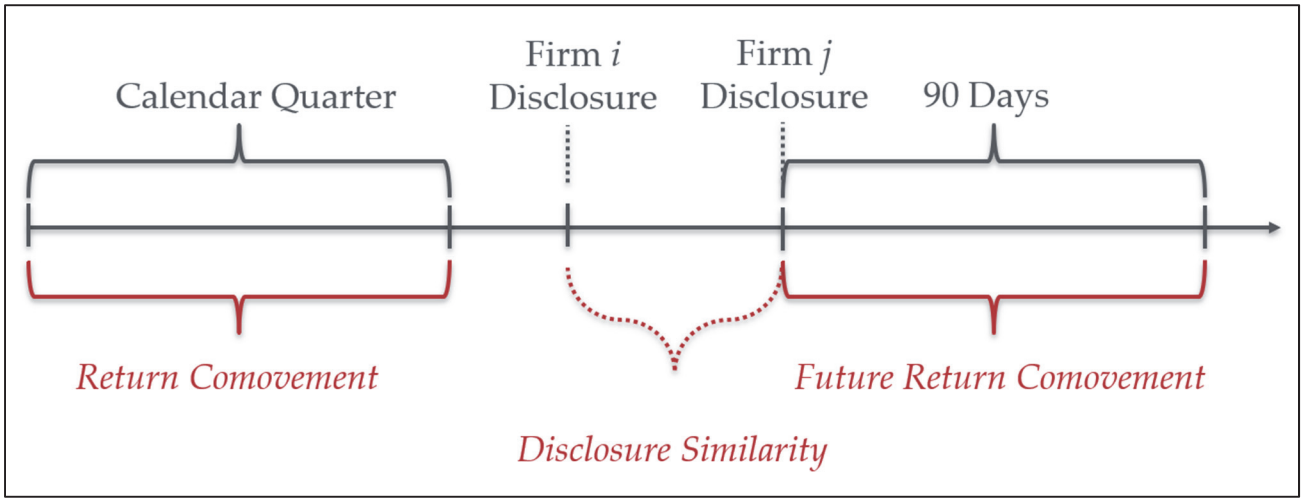
Same Quasi-Indexer Holdings	The number of institutional investors classified as Quasi-indexer that hold both firm *i* and firm *j*.
Same Dedicated Investor Holdings	The number of institutional investors classified as Dedicated Investors that hold both firm *i* and firm *j*.
Same Transient Investor Holdings	The number of institutional investors classified as Transient Investors that hold both firm *i* and firm *j*.
Future Beta	The estimated regression coefficient on the market return in a regression explaining firm *i*'s stock return, adjusted for the risk-free rate. The market return is defined as the weighted average return for the randomly selected groupings in the broad sample. The beta coefficient is estimated using daily data over the 90 days following the last disclosure for the calendar quarter in the randomized grouping.
Beta	The estimated regression coefficient on the market return in a regression explaining firm *i*'s stock return, adjusted for the risk-free rate. The market return is defined as the weighted average return for the randomly selected groupings in the broad sample. The beta coefficient is estimated using daily data over the calendar quarter.
Ave. Disclosure Similarity	The weighted average cosine-similarity of firm *i*'s 10-K/Q report with that of each firm j, weighted by the Market Value of Equity of firm *j* in the randomly selected grouping.
Ave. Earnings Comovement	The weighted average Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure, weighted by the Market Value of Equity of firm *j* in the randomly selected grouping.

10-K/Q Cleaning Procedure

We follow the steps from Dyer et al. (2017) in cleaning 10-K/Q documents. First, we remove all header and appendix information. This content includes the SEC header section at the start of all 10-K documents, graphics, zip files, XML files, excel files, 101 exhibits, 100 exhibits, pdf files, and XBRL. Next, we remove HTML tagging from the text using the HTML::Parser Perl module. We further remove remaining tags (e.g., <TEXT>, <PAGE>, <DOCUMENT>, <TYPE>) and lines with certain tags (e.g., <S>, <C>) following Miller (2010). Next, we exclude lines with fewer than 20 characters or 15 alphanumeric characters. Lastly, we exclude paragraphs with (1) more than 50% non-alphabetic characters and (2) fewer than 80 characters.

Figure 1

Panel A. Main Sample Measurement



Panel B. Amended Filing Sample Measurement

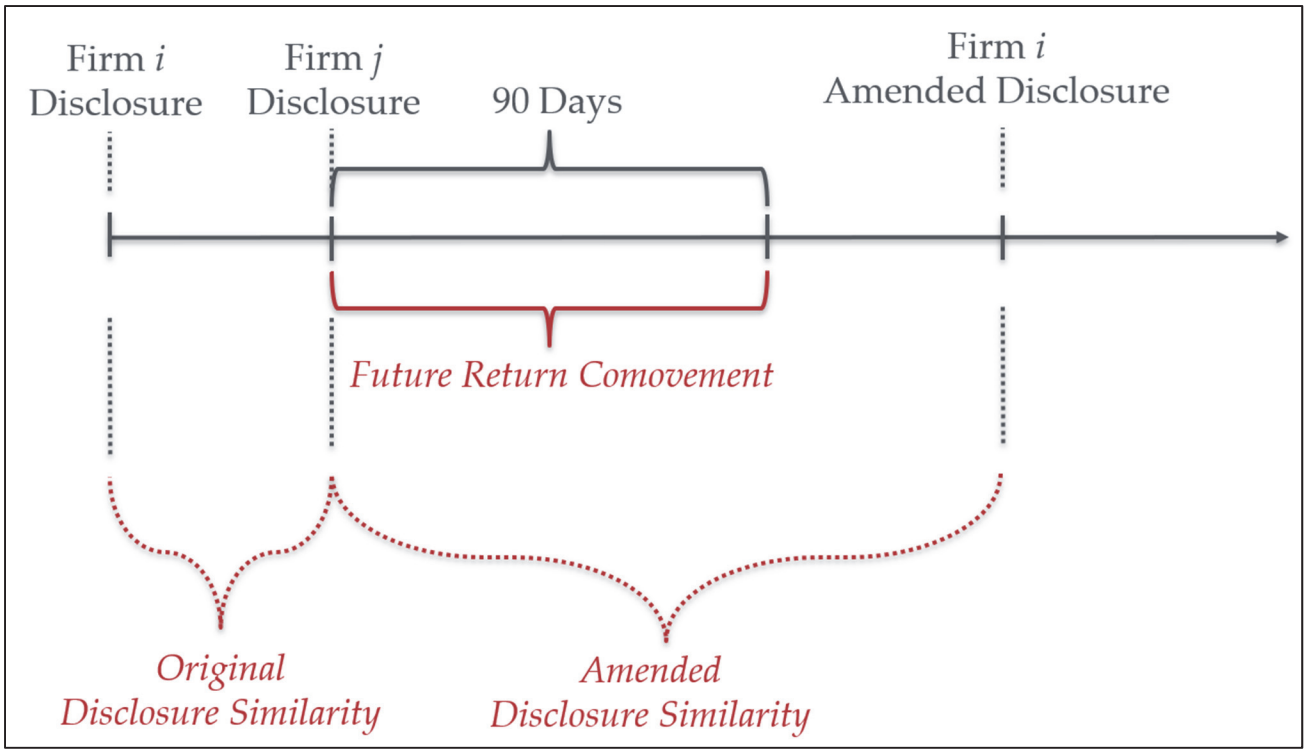
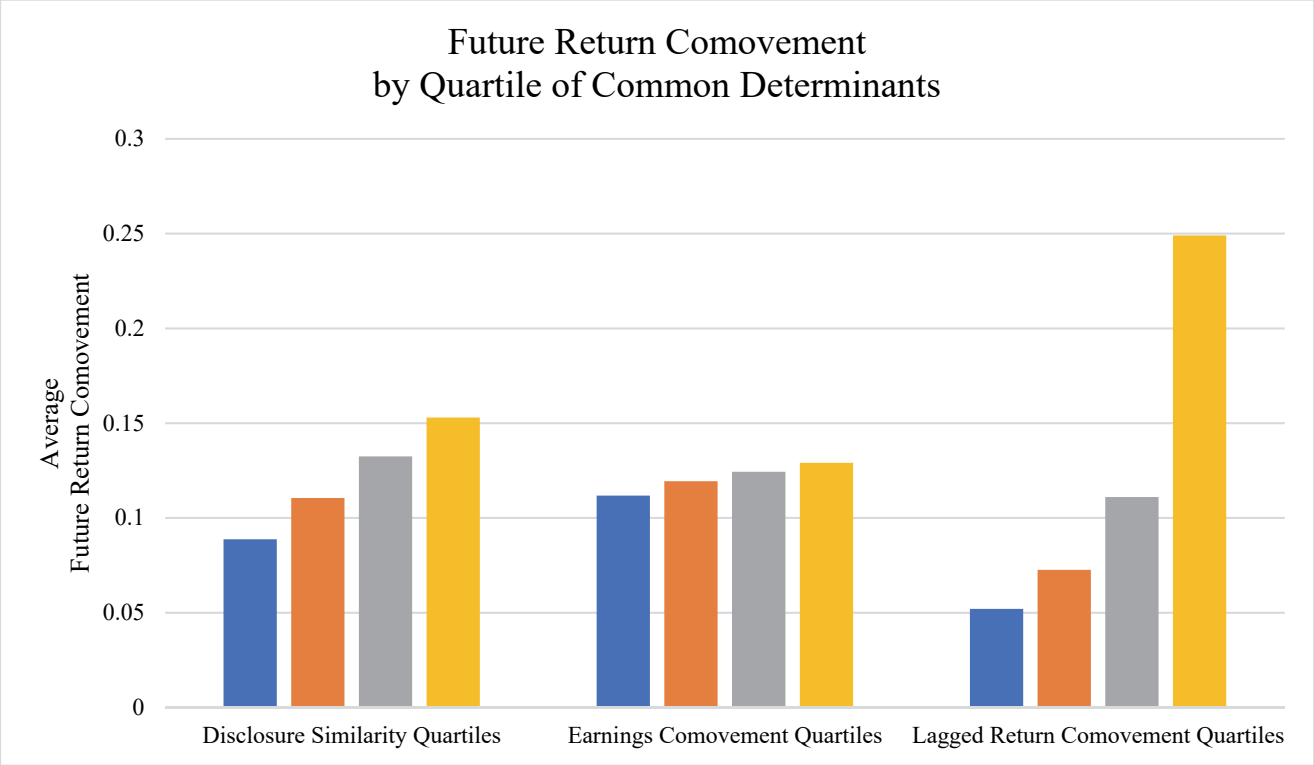


Figure 2



Notes: This figure presents the average value of future return comovement by quartile of disclosure similarity, earnings comovement, and the preceding return comovement. Each observation in the data represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). Firms are randomly assigned to fixed groupings across time and are only paired to firms within this grouping in order to reduce the magnitude of possible pairings. *Future Return Comovement* is measured as the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. *Earnings Comovement* is the Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure. *Return Comovement* is the Pearson correlation coefficient between a firm's stock return (firm *i*) and a peer firm's stock return (firm *j*) over the calendar quarter prior to 10-K/Q disclosure.

Table 1 - Descriptive Statistics for the Broad Sample

	N	Mean	Std Dev	Q1	Median	Q3
*Disclosure Similarity**	9,889,293	0.93	0.04	0.91	0.94	0.95
*Future Return Comovement**	9,889,293	0.12	0.19	-0.01	0.10	0.23
*Return Comovement**	9,889,293	0.12	0.18	-0.01	0.10	0.23
*Earnings Comovement**	9,889,293	0.03	0.39	-0.25	0.02	0.30
Same Industry	9,889,293	0.06	0.23	0	0	0
Same Investor	9,889,293	0.17	0.38	0	0	0
Same Analyst	9,889,293	0.01	0.08	0	0	0
Same Auditor	9,889,293	0.15	0.36	0	0	0
Similar MTB	9,889,293	0.11	0.31	0	0	0
Similar SIZE	9,889,293	0.11	0.32	0	0	0
Similar ROA	9,889,293	0.11	0.31	0	0	0
Similar MOM	9,889,293	0.10	0.30	0	0	0

Notes: This table presents the descriptive statistics for our main sample. Each observation in the data represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). Firms are randomly assigned to fixed groupings across time and are only paired to firms within this grouping in order to reduce the magnitude of possible pairings. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. *Future Return Comovement* is measured as the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure. *Return Comovement* is the Pearson correlation coefficient between a firm's stock return (firm *i*) and a peer firm's stock return (firm *j*) over the calendar quarter prior to 10-K/Q disclosure. *Earnings Comovement* is the Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure. *Same Industry* is an indicator variable if firm *i* and firm *j* share the same Fama-French 48 classification. *Same Analyst* is an indicator variable taking the value of one if firm *i* and firm *j* both received a forecast from the same analyst during the fiscal period, and zero otherwise. *Same Investor* is an indicator variable taking the value of one if firm *i* and firm *j* both share the same dedicated investor (Bushee, 2001) during the fiscal period, and zero otherwise. *Same Auditor* is an indicator if firm *i* and firm *j* share the same auditor during the period. *Similar MTB*, *Similar SIZE*, *Similar ROA*, *Similar MOM* are indicator variables taking the value of one if firm *i* and firm *j* share the same decile of Market-to-Book, Market Value of Equity, Return on Assets, and Stock Return, respectively, during the prior fiscal quarter, and zero otherwise. *Continuous variables are not converted into standard normal in this table.

Table 2 - Determinants of Disclosure Similarity

	Disclosure Similarity		
	Coef.	*t-Stat*	
Same Industry	0.244	22.82	***
Same Investor	0.046	20.34	***
Same Analyst	0.141	18.19	***
Same Auditor	0.027	13.97	***
Similar MTB	0.023	12.31	***
Similar SIZE	0.024	15.38	***
Similar ROA	0.053	22.06	***
Similar MOM	0.015	9.17	***
Firm *i* x Yearqtr FE	Yes		
Firm *j* x Yearqtr FE	Yes		
R-squared	0.75		
Obs	9,889,293		

Notes: This table presents the correlation between disclosure similarity and measures of firm commonality. Each observation in the data represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). Firms are randomly assigned to fixed groupings across time and are only paired to firms within this grouping in order to reduce the magnitude of possible pairings. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. It is further standardized by subtracting off the mean and dividing by the standard deviation. *Same Industry* is an indicator variable if firm *i* and firm *j* share the same Fama-French 48 classification. *Same Analyst* is an indicator variable taking the value of one if firm *i* and firm *j* both received a forecast from the same analyst during the fiscal period, and zero otherwise. *Same Investor* is an indicator variable taking the value of one if firm *i* and firm *j* both share the same dedicated investor (Bushee, 2001) during the fiscal period, and zero otherwise. *Same Auditor* is an indicator if firm *i* and firm *j* share the same auditor during the period. *Similar MTB, Similar SIZE, Similar ROA, Similar MOM* are indicator variables taking the value of one if firm *i* and firm *j* share the same decile of Market-to-Book, Market Value of Equity, Return on Assets, and Stock Return, respectively, during the prior fiscal quarter, and zero otherwise. Singleton observations are dropped. Standard errors are clustered by firm *i* (11,696 clusters) and year-quarter (85 clusters).

Table 3 - Future Return Comovement and Disclosure Similarity

	Future Return Comovement					
	(1)			(3)		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*	
Disclosure Similarity	0.034	15.39	***	0.020	15.14	***
Return Comovement	0.135	13.58	***	0.119	12.98	***
Earnings Comovement	0.009	17.96	***	0.006	14.16	***
Same Industry				0.107	13.97	***
Same Investor				0.206	11.71	***
Same Analyst				0.384	25.86	***
Same Auditor				0.037	9.38	***
Similar MTB				0.027	15.30	***
Similar SIZE				0.145	24.18	***
Similar ROA				0.030	17.16	***
Similar MOM				0.023	12.00	***
Firm *i* x Yearqtr FE	Yes			Yes		
Firm *j* x Yearqtr FE	Yes			Yes		
R-squared	0.54			0.55		
Obs	9,889,293			9,889,293		

Notes: This table presents the predictive ability of disclosure similarity in forecasting future return comovement. Each observation represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). Firms are randomly assigned to fixed groupings across time and are only paired to firms within this grouping in order to reduce the magnitude of possible pairings. *Future Return Comovement* is measured as the Pearson correlation coefficient between a firm's stock return (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. *Return Comovement* is the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the calendar quarter prior to 10-K/Q disclosure. *Earnings Comovement* is the Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure. All continuous variables are standardized by subtracting off the mean and dividing by the standard deviation. All other variables are as previously defined. Singleton observations are dropped. Standard errors are clustered by firm *i* (11,696 clusters) and year-quarter (85 clusters).

Table 4 - Future Return Comovement and Disclosure Similarity: Greater Reliance on Accounting Disclosures

	Future Return Comovement								
	(1)			(2)			(3)		
Greater Reliance =	*Both Small Firms*			*No Analyst Coverage*			*Both High EA Importance*		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*		*Coef.*	*t-Stat*	
Disclosure Similarity	0.012	5.99	***	0.017	12.49	***	0.033	15.11	***
Greater Reliance	0.382	15.89	***	0.152	12.71	***	0.026	9.09	***
Disclosure Similarity x Greater Reliance	0.032	4.96	***	0.023	4.96	***	0.003	2.91	***
Controls	Yes			Yes			Yes		
Firm *i* x Yearqtr FE	Yes			Yes			Yes		
Firm *j* x Yearqtr FE	Yes			Yes			Yes		
R-squared	0.56			0.55			0.66		
Obs	9,889,293			9,889,293			2,429,389		

Notes: This table presents cross-sectional variation in the predictive ability of disclosure similarity in forecasting future return comovement. Each observation represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). Firms are randomly assigned to fixed groupings across time and are only paired to firms within this grouping in order to reduce the magnitude of possible pairings. *Future Return Comovement* is measured as the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. *Return Comovement* is the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the calendar quarter prior to 10-K/Q disclosure. *Both Small Firms* is an indicator if both firms have below-median size. *No Analyst Coverage* is an indicator if neither firm has analyst coverage. *Both High EA Importance* is an indicator if both firms have above-median levels of anticipated earnings announcement importance, using the measure in Iselin and Van Buskirk 2020. All continuous variables are standardized by subtracting off the mean and dividing by the standard deviation. All other variables are as previously defined. Singleton observations are dropped. Standard errors are clustered by firm *i* and year-quarter. Columns (1) - (2) have 11,696 firm clusters and 85 year-quarter clusters. Column (3) has 5,828 firm clusters and 85 year-quarter clusters.

Table 5 - Future Return Comovement and Disclosure Similarity: Amended Filings Sample

	Future Return Comovement					
	(1)			(2)		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*	
Disclosure Similarity - Amended	0.041	15.42	***	0.031	13.39	***
Disclosure Similarity - Error	0.026	8.50	***	0.020	7.04	***
Return Comovement	0.104	11.36	***	0.098	11.31	***
Earnings Comovement	0.008	8.69	***	0.007	7.73	***
Same Industry				0.098	12.32	***
Same Investor				0.066	10.03	***
Same Analyst				0.366	7.10	***
Same Auditor				0.049	7.80	***
Similar MTB				0.024	6.89	***
Similar SIZE				0.160	19.13	***
Similar ROA				0.028	11.24	***
Similar MOM				0.024	8.41	***
Firm *i* x Yearqtr FE	Yes			Yes		
Firm *j* x Yearqtr FE	Yes			Yes		
R-squared	0.45			0.46		
Obs	11,752,771			11,752,771		

Notes: This table presents the association between disclosure similarity and future return comovement. Each observation represents a pairing of firms (firms *i* and *j*) during a calendar quarter (time *t*). In this sample, each firm *i* amended its original filing more than 90 days after the original release. *Future Return Comovement* is measured as the Pearson correlation coefficient between a firm's stock returns (firm *i*) and a peer firm's stock return (firm *j*) over the 90 days following the latter of firm *i* and firm *j*'s disclosure. *Disclosure Similarity* (*Disclosure Similarity - Amended*) is measured as the cosine-similarity of a firm's original (amended) 10-K/Q report with that of a peer firm's original 10-K/Q report, where peer reports come from the same calendar quarter. *Disclosure Similarity - Error* is the difference between *Disclosure Similarity* and *Disclosure Similarity - Amended*. All continuous variables are standardized by subtracting off the mean and dividing by the standard deviation. All other variables are as previously defined. Standard errors are clustered by firm *i* (1,864 clusters) and year-quarter (84 clusters).

Table 6 - Disclosure Similarity and S&P 500 Inclusion/Exclusion

	Disclosure Similarity					
	(1)			(2)		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*	
Treat x Inclusion	0.019	4.28	***	0.016	3.74	***
Earnings Comovement				0.005	5.98	***
Same Industry				0.143	14.32	***
Same Analyst				0.132	18.31	***
Same Auditor				0.011	4.55	***
Similar MTB				0.012	6.04	***
Similar SIZE				0.006	1.96	*
Similar ROA				0.023	9.72	***
Similar MOM				0.010	6.12	***
Firm *i* x Yearqtr FE	Yes			Yes		
Firm *j* x Yearqtr FE	Yes			Yes		
R-Squared	0.70			0.70		
Observations	6,503,038			6,503,038		

Notes: This table evaluates the change in disclosure similarity for firms that enter/exit the S&P 500 index. *Disclosure Similarity* is measured as the cosine-similarity of a firm's 10-K/Q report with that of a peer firm's 10-K/Q report, where peer reports come from the same calendar quarter. *Treat* is an indicator variable taking the value of one if firm *i* is paired with an S&P 500 firm *j* and zero otherwise. *Inclusion* is an indicator if firm *i* is in the S&P 500. All other variables are as previously defined. Standard errors are clustered by firm *i* (377 clusters) and year-quarter (85 clusters).

Table 7 - Return Comovement, Disclosure Similarity, and S&P 500 Inclusion/Exclusion: Schoenfeld Approach

Panel A. Schoenfeld Approach - Descriptives

	N	Mean	
TxI_Future Return Comovement	377	0.10	***
TxI_Disclosure Similarity	377	0.02	***
TxI_Earnings Comovement	377	-0.01	
TxI_Same Quasi-Indexer Holdings	377	0.65	***
TxI_Same Dedicated Investor Holdings	377	0.31	***
TxI_Same Transient Investor Holdings	377	0.49	***
TxI_Same Industry	377	0.01	***
TxI_Same Analyst	377	0.01	***
TxI_Same Auditor	377	0.00	
TxI_Similar MTB	377	0.00	
TxI_Similar SIZE	377	0.21	***
TxI_Similar ROA	377	0.00	
TxI_Similar MOM	377	0.00	***

Panel B. Schoenfeld Approach - Intensity of Treatment

	TxI_Disclosure Similarity			*TxI_Future Return Comovement*		
	(1)			(2)		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*	
TxI_Disclosure Similarity				0.172	3.42	***
TxI_Same Quasi-Indexer Holdings	0.232	2.19	**	0.437	4.18	***
TxI_Same Dedicated Investor Holdings	0.039	0.71		-0.017	-0.31	
TxI_Same Transient Investor Holdings	-0.227	-2.07	**	-0.515	-4.91	***
TxI_Earnings Comovement	0.025	0.32		0.068	1.17	
TxI_Same Industry	0.028	0.42		0.057	0.88	
TxI_Same Analyst	0.238	2.72	***	0.220	2.93	***
TxI_Same Auditor	0.004	0.07		0.044	1.08	
TxI_Similar MTB	-0.251	-3.37	***	-0.018	-0.23	
TxI_Similar SIZE	0.005	0.10		0.092	1.99	**
TxI_Similar ROA	0.066	0.95		0.113	1.78	*
TxI_Similar MOM	0.079	1.14		0.183	2.82	***
R-Squared	0.145			0.341		
Observations	377			377		

Notes: This table evaluates if S&P 500 inclusion events that elicit larger (smaller) changes in disclosure similarity also exhibit larger (smaller) changes in future return comovement. For each firm that entered or exited the S&P 500 during our sample period (firm *i*), we estimate a difference-in-differences analysis for future return comovement, controlling for firm and time-period fixed effects. We store the resulting *Treatment x Inclusion* indicator for each event and label it *TxI_Future Return Comovement*. As such, each observation in this analysis represents one firm that entered/exited the S&P 500 during our sample period. We perform similar difference-in-difference analyses for each variable of interest and add the prefix *TxI* to indicate the treatment effect (e.g., *TxI_Disclosure Similarity*). Panel A presents descriptive statistics for the treatment effects. Panel B presents the result of regressing *TxI_Disclosure Similarity* and *TxI_Future Return Comovement* against other treatment effects. Coefficient estimates are beta coefficients. All variables are as previously defined. Standard errors are adjusted using Huber-White.

Table 8 - Future Beta and Disclosure Similarity

	Future Beta					
	(1)			(2)		
	Coef.	*t-Stat*		*Coef.*	*t-Stat*	
Ave. Disclosure Similarity	0.030	4.27	***	0.021	2.85	***
Beta	0.397	24.16	***	0.150	12.63	***
Ave. Earnings Comovement	0.016	3.35	***	0.009	2.62	***
Firm FE	No			Yes		
Yearqtr FE	No			Yes		
R-squared	0.16			0.35		
Obs	303,891			303,398		

Notes: This table presents the predictive ability of a firm's average disclosure similarity in forecasting future Beta for firm *i* in time *t*. *Future Beta* is the estimated regression coefficient on the market return in a regression explaining firm *i*'s stock return, adjusted for the risk-free rate. In this analysis, the market return is defined as the weighted average return for the randomly selected groupings in the broad sample. *Future Beta* is estimated using daily data over the 90 days following the last disclosure for the calendar quarter in the randomized grouping. *Beta* is estimated similarly over the calendar quarter. *Ave. Disclosure Similarity* is the weighted average cosine-similarity of firm *i*'s 10-K/Q report with that of each firm *j*, weighted by the Market Value of Equity of firm *j* in the randomly selected grouping. *Ave. Earnings Comovement* is the weighted average Pearson correlation coefficient between a firm's quarterly earnings and a peer firm's quarterly earnings over the three years prior to 10-K/Q disclosure, weighted by the Market Value of Equity of firm *j* in the randomly selected grouping. All continuous variables are standardized by subtracting off the mean and dividing by the standard deviation. Standard errors are clustered by firm *i* and year-quarter. Column (1) has 12,042 firm clusters and 85 year-quarter clusters. Column (2) has 11,549 firm clusters and 85 year-quarter clusters.