2024

Measuring Financial Statement Disaggregation Using XBRL

Joseph A. Johnston
Illinois State University, jajohn6@ilstu.edu

Kenneth J. Reichelt
Louisiana State University

Pradeep Sapkota
University of North Texas

This is the accepted manuscript version of the following article: Joseph A. Johnston, Kenneth J. Reichelt, Pradeep Sapkota; Measuring Financial Statement Disaggregation Using XBRL. *Journal of Information Systems* 1 March 2024; 38 (1): 119–147. https://doi.org/10.2308/ISYS-2021-004.

Measuring Financial Statement Disaggregation Using XBRL

ABSTRACT: We develop a measure of disclosure quality using the disaggregation level of XBRL filings. Our measure (*ITEMS*) extends Chen, Miao, and Shevlin's (2015) (*DQ*) measure in several ways. First, it is intuitive and does not depend on the data aggregator's collection process (e.g., S&P *Compustat*). Next, it captures the direct financial reporting practices of companies. Finally, it is readily available after the Form 10-K is filed. We validate *ITEMS* by regressing it on popular firm fundamentals and find that virtually all firm fundamentals explain *ITEMS* in the predicted direction. We further validate *ITEMS* by evaluating whether it explains the known consequences of disclosure quality: forecast errors, forecast dispersion, bid-ask spread, and cost of equity capital. In additional tests, we confirm that our results hold after controlling for accounting reporting complexity and *DQ*. We also find that *ITEMS* has explanatory power incremental to *DQ* for consequences of disclosure quality.

Keywords: Disclosure quality, Disaggregation, XBRL, SEC Financial Statement Datasets, Information Processing Cost

Data Availability: Data are available from public sources identified in the text.

1. INTRODUCTION

Our study develops an improved measure of disclosure quality using the Securities and Exchange Commission (SEC) Form 10-K XBRL filing to capture the actual disaggregated level of items disclosed. Our measure extends the disaggregation measure (*DQ*) developed by Chen, Miao, and Shevlin (2015) (CMS), which uses Standard and Poor's *Compustat* (S&P) data. A long-standing issue in the voluntary disclosure literature is deriving an appropriate empirical measure of disclosure quality (Beyer, Cohen, Lys, and Walther 2010, 311). We answer this call by developing a parsimonious measure of disaggregation that utilizes machine-readable financial disclosure data made available by the XBRL mandate.

For several reasons, our disaggregation measure (*ITEMS*) is based on the notion that disaggregation is higher disclosure quality. Disaggregation provides finer and more precise information to decision-makers (Blackwell 1951; Cheong and Thomas 2011; CMS), and separates "economically unlike items" (e.g., segmented disclosures) to improve understandability and transparency (Barth and Schipper 2008, 181). It reduces opportunism in financial reporting (Hirst, Koonce, and Venkataraman 2007), when GAAP provides considerable discretion (CMS). Consequently, we posit that *ITEMS* captures management's discretion to disaggregate line-items on the financial statements.

To capture a firm's disaggregation level, we measure *ITEMS* by the number of line-items reported on the balance sheet and income statement of the Form 10-K filing, found in the FSDS. We validate the measure by testing for the predicted direction of variable coefficients from an OLS regression of *ITEMS* on firm fundamentals and regressions of disclosure consequences on *ITEMS*. Firm fundamentals include restructuring charges, mergers and acquisitions, special items, return volatility, size, the number of segments, firm age, auditor type, industry fixed-effects (e.g., industry

practices), and year fixed-effects (e.g., macroeconomic). Disclosure consequences include analyst forecast accuracy and dispersion, bid-ask spread, and cost of equity capital. We find that the coefficient signs are in the predicted direction for virtually all variables and the results are robust to several additional tests.

Consistent with Johnston and Zhang (2021) and Boritz and No (2020), we exclude footnote tags from *ITEMS*, for several reasons. First, users will likely weight footnotes with less importance than financial statement line-items because of the higher processing costs from the intermingling of text with numbers in an ambiguous and nonuniform format (Vasarhelyi et al. 2012; Li, Ramesh, and Shen 2011; Hodge, Kennedy, and Maines 2004; Blankespoor et al. 2020, 61). Second, firms can strategically choose to obfuscate specific items by burying them in the footnotes, while emphasizing other items on the face of the financial statements (Blankespoor et al. 2020). Third, footnotes disclose more minute accounts, because the balance sheet and income statement have a practical maximum number of line-items in order to fit inside the annual report (e.g., about one page each). As such, *ITEMS* captures the level of disaggregation from the face of the financial statements.[1]

We compare our measure to those in CMS and Hoitash and Hoitash (2018) (HH), and we find several differences and that *ITEMS* has incremental explanatory power. CMS develop a disaggregation measure (*DQ*) from S&P variables, based on a count of non-missing balance sheet and income statement variables. *DQ* utilizes S&P's standard template, capturing 129 balance sheet variables and 58 income statement variables. Our paper differs from CMS because we utilize a new dataset that can calculate the actual disaggregation disclosed by firms without being constrained by the aggregator's template. In the second related study, HH create a measure of

[1] We acknowledge that not all footnote items are intended to obfuscate information and managers could disclose additional footnote items for transparency.

accounting reporting complexity (*ARC*) using all the tags reported (including footnote tags) and find that *ARC* is positively associated with misstatements, material weaknesses, audit delay, and audit fees. They argue that disclosing more items makes it more difficult for preparers of financial statements because it requires them to analyze more information and understand more accounting standards. However, *ARC* is constructed from more tags (μ=346) than *ITEMS* (μ=58), consisting of only balance sheet and income statement tags. Because HH are interested in studying the number of accounting concepts reported on the financial statements, and the implied diversity, it is reasonable for their study to include footnote tags. In additional tests, we do not find that *ARC* is associated with the external information environment (analyst forecast dispersion and accuracy and bid-ask spread), and we also find that it is positively associated with the cost of equity capital, rather than negatively as with *ITEMS*. Nevertheless, recognizing that there may be overlap with the level of disaggregation and *ARC* (HH, 277), we conduct a test that includes *DQ*, *ARC,* and *ITEMS* in the same model. We find that *ITEMS* has incremental explanatory power over *DQ* and *ARC*. In additional tests, we find that *ITEMS* does not explain reporting quality (HH's outcome variables).

Our results are significant for several reasons. First, we corroborate the findings of CMS, who find that more disaggregated disclosure is associated with a richer information environment, while identifying a better data source. Further, we extend their study and disclosure theory (Moffitt, Richardson, Snow, Weisner, and Wood 2016; Blankespoor et al. 2020). The notion that disaggregated disclosure is of higher quality becomes more important in the current era with Big Data and lower information processing costs (Moffitt et al. 2016). In the current era, XBRL data is more accessible and transparent, which permits a more accurate measure of disclosure quality compared to the more summarized version from data aggregators (Vasarhelyi et al. 2012).

Our study contributes to the disclosure quality literature by answering Beyer et al.'s (2010) call to develop a broader and more comprehensive measure of disclosure because of drawbacks in prior measures. *ITEMS* is more current than analyst ratings (e.g., Lang and Lundholm 1993) that are now dated and discontinued. *ITEMS* is available for all SEC registrants while self-constructed measures (e.g., Botosan 1997) are not feasible for large samples (Beyer et al. 2010). *ITEMS* is easier to compute than text-based measures (e.g., Li 2008) that may not be practical for many financial statement users (Beyer et al. 2010). *ITEMS* can be computed for all SEC registrants, while management earnings forecasts (e.g., Hirst et al. 2008) are prepared by a subset of firms. Utilizing management earnings forecasts to study disclosure quality could impose potential selection bias and yield less generalizable inferences.

We lean on the accounting and information systems literature to suggest that *ITEMS* has several incremental advantages over extant measures. First, by accessing XBRL data from the SEC's Financial Statement Data Sets (FSDS), *ITEMS* employs substantially more datapoints (15,000 or more standard tags) compared to datasets provided by data aggregators (e.g., S&P has approximately 1,000 variables) (Debreceny, Farewell, Piechocki, Felden, Gräning, and d'Eri 2011).[2] Further, the XBRL taxonomy permits firms to create "extended tags" when a standard tag does not exist for the reporting item. A case in point is the revenue reporting by Ryder System, Inc. (gvkey 009299, cik 85961) in their 2015 fiscal-year Form 10-K filing (Appendix 1, Panel A). They report three revenue categories along with total revenues, which XBRL reports individually. They seemingly report two discretionary categories (services revenue and fuel services) following

[2] Hoitash, Hoitash, and Morris (2020) provide an excellent review of XBRL data and the extant research that utilizes this data, as well as potential avenues for future research. They suggest that the XBRL data is rich, unique and broadly available.

Regulation S-X, which are both extended tags.[3, 4] On the other hand, S&P effectively reports one category, total revenues (REVT: $6,571.89M), because the only other S&P revenue category, sundry revenues (RIS), is zero for the current and prior years; thus, XBRL reports three more revenue categories. The FASB XBRL taxonomy provides firms the flexibility to report multiple revenue categories from the available general ledger tags (XBRL-GL), while S&P only provides a standard template. Hence, we construct *ITEMS* to capture the number of tags reported in the firm's Form 10-K financial statements proper.

Second, *ITEMS* uses XBRL, which is arguably more accurate. Because S&P normalizes the data to fit its template, some variables are incomplete and inaccurate (Sanmiguel 1977; Kinney and Swanson 1993; Chychyla and Kogan 2015; Boritz and No 2020). Third, because firms furnish XBRL data with the Form 10-K filing, *ITEMS* does not have the time lag associated with data aggregators, particularly for some non-S&P 1500 companies (Howard and Zhou 2021; Blankespoor et al. 2020; Boritz and No 2020). Finally, XBRL is publicly available for all SEC registrants and does not require a paid subscription with a data aggregator (e.g., S&P, Moody's, and Bloomberg).[5] These arguments are corroborated by a testimonial from the then SEC Assistant Director of the Office of Structured Disclosure, Mike Willis, who publicly stated that the FSDS is timely, complete, granular, freely available, and transparent to the public (Willis 2019).

Next, our results indicate that *ITEMS* is associated with a richer information environment, in contrast to the *ARC* measure. Our results reaffirm that *ARC* is capturing complex accounting

[3] SEC Regulation S-X (17 CFR § 210.5) requires that companies report revenues from different activities (products, rentals, services, and other) that exceed 10 percent of total revenues. FASB ASC topic 606 is similar in that it requires reporting by major source, but it provides more discretion.

[4] Another example relates to accounts receivable. Some companies report accounts receivable as a net number while others break it down into gross accounts receivable and allowance for doubtful accounts (Vasarhelyi, Chan, and Krahel 2012). XBRL permits companies to disaggregate gross accounts receivable and allowance for doubtful accounts, allowing users to identify companies with a higher level of disaggregation.

[5] We acknowledge that data aggregator subscription fee is nominal compared to the information processing cost (Blankespoor et al. 2020) but it could be substantial for non-institutional investors (Vasarhelyi et al. 2012).

reporting (e.g., more diverse operations and a more complex capital structure) by counting the number of footnote tags rather than the number of financial statement tags. In the bigger picture, *ITEMS* measures a positive antecedent of optimal disclosure quality, while *ARC* measures a negative antecedent. Our findings support the FASB's current push to improve the effectiveness of footnote disclosure (Tysiac 2018). The findings of our paper may interest regulators because they infer that the additional disclosure inherent in the XBRL taxonomy is more informative to market participants than the data provided by third-party vendors. Finally, our use of XBRL data corroborates the optimism for its usefulness in that it "will likely drive future research" (Kothari 2019).

The remainder of our paper is as follows. The next section presents the background and literature. We then discuss the measure of *ITEMS* and the empirical design including validation of *ITEMS*. Next, we present the results including additional analyses and robustness tests. Finally, we conclude the study with a discussion of caveats.

2. BACKGROUND AND LITERATURE

Disaggregated Disclosures

Our theoretical argument stems from the notion that more disaggregated information is higher quality because it is more reliable, it reduces information asymmetry with finer information (CMS) and it deters managerial opportunism (Blackwell 1951; Williamson 1985; Hirst et al. 2007; CMS). Within mandatory filings, managers have the discretion to disaggregate disclosures to provide finer information (CMS; Bonner, Clor-Proell, and Koonce 2014). For example, more transparent managers can choose to disaggregate net gains and losses from asset sales and early debt extinguishment, (Bonner et al. 2014) revealing further details of prior decisions that require verification. On the other hand, more opaque managers draw attention to summary numbers when

smoothing income (D'Souza, Ramesh, and Shen 2010), and will aggregate individual segments to hide poor performance (Hayes and Lundholm 1996). Finer information can also deter managerial opportunism by giving managers "fewer degrees of freedom to manage the reported numbers" (CMS; Hirst et al. 2007; D'Souza et al. 2010). Consequently, disaggregated information is more transparent (Barth and Schipper 2008), and has greater information content (e.g., Lansford, Lev, and Tucker 2013; Lim 2014).[6]

SEC Regulation S-X (17 CFR § 210.5-02,03) (Reg. S-X) and the FASB codification mandate specific line-items on the balance sheet and the income statement. Reg. S-X mandates based on the relative magnitude, providing discretion for reporting smaller amounts. For instance, Reg. S-X requires that notes receivable be reported separately from accounts receivable if it is more than 10 percent of the aggregate, providing discretion for reporting lesser amounts.[7] Further, revenue classes (products, rentals, services, and other) should be reported separately if each is more than 10 percent of total revenue. For instance, Ryder System, Inc. (Appendix 1 – Panel A) disaggregates fuel services revenue even though it could have aggregated it with the broader category of "Services revenue."[8] In addition, FASB's topic, *Revenue from Contracts with Customers* (ASC 606-10-55-89), provides broader discretion than Reg. S-X for reporting revenue categories. The FASB topic requires consideration of categories reported to the public outside the financial statements and the company's organization of operating segments.

XBRL and the SEC Financial Statement Data Sets

To permit users to process the content of the Form 10-K and 10-Q filings more efficiently, SEC Rule 33-9002 requires registrants to furnish the filings with eXtensible Business Reporting

[6] A counter argument is that users incur marginal processing costs for additional disaggregation, which may not exceed the marginal benefit. However, as we argue later, processing costs are higher for footnotes than for financial statement line-items.
[7] Other examples include separately reporting prepaid expenses more than five percent of current assets and intangible asset classes more than five percent of total assets.
[8] See footnote 33 for a comparison to other companies in the same industry and to other industries.

Language (XBRL) tags (SEC 2009). XBRL is a data standard that better permits the movement of digital information in an accurate and timely manner (XBRL International 2016). The standard is independent of any particular platform, which simplifies its functionality. It is free of charge to preparers and users of SEC filings and is managed by a global not-for-profit consortium, XBRL International.

The FASB developed the U.S. GAAP XBRL tags that form part of the SEC's EDGAR system (XBRL U.S. 2016).[9] XBRL tags identify data elements in an SEC filing, such as account descriptions, amounts, and textual note disclosures. The tags enable corporate transparency by enclosing metadata that describes the content of the Form 10-K and 10-Q filings (Debreceny et al. 2011, 632). The XBRL tags are organized into flat files that are machine-readable with all the metadata (SEC 2016b).

Constructing a disclosure quality measure from XBRL files entails two challenges: a longer learning curve and accuracy concerns. A longer learning curve discourages users to adopt the technology (Hodge et al. 2004), so the SEC Office of Structured Disclosure (OSD) created the FSDS. The FSDS provides the public with a website having a series of flat text files that are machine readable into relational databases (SEC 2016b). We use the website to collect the XBRL tags from the Form 10-K filings and construct *ITEMS*.

The second challenge is that researchers have raised concerns about the accuracy of the XBRL data because the FSDS are not independently verified against the Form 10-K and 10-Q filings (Debreceny et al. 2011; Boritz and No 2016; Alles and Gray 2012; AICPA 2017; Basoglu and White Jr 2015). Nevertheless, filers have an incentive to ensure the data are of high quality

[9] The EDGAR database uses *standard* tags and *extended* tags. Standard tags are from the U.S. GAAP XBRL taxonomy (http://bit.ly/2hEH0FI) and cannot be modified. Extended tags permit customized reporting when standard tags are insufficient to comply with either SEC filing requirements or business reporting (XBRL U.S. 2016; Debreceny et al. 2011). Standard tags account for 79 percent of all tags (HH).

because analysts, regulators, and others use it extensively (Alles and Gray 2012; Boritz and No 2020). Filers can indicate the level of auditor assurance (SEC 2011), and the AICPA XBRL Assurance Task Force provides guidance and procedures (AICPA 2013).[10] Further, the accuracy of *ITEMS* is improving as companies file more often, consistent with learning curve theory (Du, Vasarhelyi, and Zheng 2013). Also, many of the errors do not affect the count of *ITEMS* because they are more often related to identifying the correct tag or the correct value (Debreceny et al. 2011, Du et al. 2013). The recent move towards inline XBRL could address some of the data accuracy concerns because filers have an added incentive with users having the ability to examine further details of a datapoint by simply clicking on it (SEC 2020).

Prior Disclosure Quality Measures

Given that XBRL data is readily available and is reasonably reliable, our measure is more parsimonious than extant measures from prior studies. Extant measures capture specific aspects but have several shortcomings. AIMR scores capture analyst ratings of annual, quarterly and other published information and investor relations (e.g., Lang and Lundholm 1993) but they have become dated (1985-1996) and discontinued. Further, it is unclear whether the analysts on the AIMR panels take the ratings seriously, whether the firms that are selected are representative (biased to larger firms) (Healy and Palepu 2001), and whether the analysts are unbiased. Self-constructed measures capture the presence of specific disclosures (e.g., summary financial ratios, nonfinancial measures, and forecasts) (Botosan 1997; Francis et al. 2008) but they are not feasible for large samples (Beyer et al. 2010). Text-based measures capture readability and other textual qualities (e.g., Li 2008) but may not be practical for many financial statement users and may not always be appropriate in the context of business reports (Loughran and McDonald 2014).

[10] Extant research suggests that regulators may soon mandate assurance (Boritz and No 2016; Farewell and Pinsker 2015; Alles and Gray 2012).

Management earnings forecasts are prepared by a subset of companies, suggesting potential selection bias and less generalizability because their absence does not distinguish between a concealed forecast and none at all. The shortcomings of these measures necessitate the call for a broader and more comprehensive framework of voluntary disclosure determinants and consequences (Beyer et al. 2010).

A recent study by CMS begins to address these shortcomings by developing a measure based on the disaggregation of financial statement accounts. They base their measure (*DQ*) on a count of non-missing S&P balance sheet (129) and income statement (58) variables. While the use of the S&P datasets has its own merits (e.g., normalized data for cross-sectional tests), it likely understates the true level of disaggregation because it is constrained by a fixed number of variables. Further, prior research has raised concerns about the accuracy of S&P (Chychyla and Kogan 2015; Kinney and Swanson 1993; Boritz and No 2020).[11] Given that S&P differs from the Form 10-K filing, FSDS could be more accurate, and more importantly, studies that utilize S&P data (e.g., *DQ*) likely understate the true level of disaggregation. Finally, *ITEMS* could be available in a timelier basis, particularly for smaller companies.[12] XBRL data is filed concurrently with the Form 10-K filing in nearly all cases (99%) (Howard and Zhou 2021) and the FSDS datasets are available within one week of the quarter-end when most SEC registrants file (SEC 2016).

Recently, HH show that the count of all XBRL tags measures accounting reporting complexity (*ARC*). They posit that firms that disclose more tags burden the preparer and harm

[11] For example, Chychyla and Kogan (2015) find that 17 out of 30 S&P variables differ significantly from the amounts reported in the XBRL tags for the Form 10-K filing. However, for the purpose of calculating *DQ* and *ITEMS*, if the amounts differ but the line-items are the same, then the accuracy of *DQ* is not affected. Further, Boritz and No (2020) suggest that between 48 and 63 percent of financial statement items available in XBRL are not available from aggregators, suggesting that XBRL has more line-items and is a more accurate measure of disaggregation.

[12] S&P informs us that the number of days to update the Compustat annual fundamentals file depends on company size (S&P 1500 status) and whether the company subscribes to S&P's quality control program. S&P 1500 (non-S&P) companies that are not part of the quality control program are updated within three (14) days. S&P (non-S&P) companies that are part of the quality program are updated within 14 (25) days. In contrast, the FSDS takes approximately one week for the quarterly update.

financial reporting quality, evident from more misstatements, a longer audit delay, and higher audit fees. Our study differs from HH because *ITEMS* is a positive antecedent of disclosure quality, capturing disaggregation and measured by the count of tags from the face of the balance sheet and income statement (including the statement of comprehensive income).[13] The positivity partly arises from the constraint imposed by the maximum number of line-items on each of the two financial statements printed in the annual report (e.g., one to two pages each – see Ryder System, Inc. 2015 FY Form 10-K filing (p. 63-65)). In contrast, *ARC* is potentially a negative antecedent capturing reporting complexity, measured by the count of the footnote tags and the financial statement tags. Its negativity partly arises from the variable cost for users to locate, read and understand the footnotes (Blankespoor et al. 2019), and a higher maximum and ambiguous practical number of footnotes (e.g., tens of pages – Ryder System, Inc. 2015 FY 10-K has 52 pages [p. 68 to 119]).

Because footnotes impose a variable cost on the user to process each, particularly less-sophisticated users who may find the cost prohibitive (Blankespoor et al. 2019), it also provides management with the opportunity to bias reporting. Like classification shifting (McVay 2006), managers can strategically blur the importance of specific accounts by burying them in a footnote rather than recognizing them upfront on the face of the financial statements. Less sophisticated users may only read a part or none of the footnote to avoid processing the nonuniform format and intermingled text and numbers (Blankespoor et al. 2020; Vasarhelyi et al. 2012; Li et al. 2011; Hodge et al. 2004).[14]

[13] If disclosure quality is a humped shared function of disclosure quantity, *ITEMS* represents the positive part and *ARC* represents the negative part.

[14] Anecdotal evidence also suggests that footnote items overload the user's cognition. For example, Ghai and Rapp (2016) suggests "most analysts are unwilling and/or unskilled enough to brave the ocean of text in individual, 40,000-word (yes!) SEC filings and painstakingly trace back numbers buried in footnotes to relevant line-items elsewhere on the financial statement."

While footnotes are costlier for users to process, it is not necessarily poorer disclosure quality, because they provide more minute accounts (e.g., fixed asset cost by class) and narratives (e.g., depreciation rates and policies). As mentioned, the balance sheet and the income statement have fewer tags than the footnotes because they are constrained by page-length. On the other hand, the footnote section can be tens of pages long to allow for a more detailed breakdown of line-items and providing narrative descriptions of economic events. However, the higher cost for users to process the footnotes implies that they should weight them less than the items reported on the face of the two financial statements.

In summary, we create an improved measure of disaggregation that extends CMS's *DQ* measure by utilizing XBRL data and isolating the measure to items recognized on the face of the balance sheet and income statement. We do, however, agree with HH's assertion that even if *DQ* and *ARC* (and by construction *ITEMS*) are distinct and have their own merits, they "capture some aspect of disclosure detail" (HH, pg., 277).

3. DISCLOSURE QUALITY MEASURE AND EMPIRICAL DESIGN

Disaggregation Measure

We construct our disaggregation measure from the SEC's FSDS published by the Office of Structured Disclosure from the filers' XBRL annual financial statement (SEC 2016b). The SEC provides the data in a flat-file with a relational database structure that can be read as a table by many software packages (e.g., MS SQL Server, SAS, and MS Excel) (SEC 2016b). The FSDS contains several tables (SUB, PRE, NUM, and TAG), which are distributed in archive files and compiled quarterly.[15] One table that is particularly useful is the presentation table (PRE) because it allows us to systematically count the number of line-items on the face of the balance sheet and

[15] For further details of the FSDS, see https://www.sec.gov/data/financial-statements/aqfs.pdf.

the income statement and construct *ITEMS*.[16] PRE identifies the financial statement type, the item number, whether the item is parenthetical (e.g., allowance for doubtful accounts), and the item's XBRL tag name. Financial statement types include the balance sheet, income statement, statement of comprehensive income, cash flow statement, and statement of equity. The number of line-items in *ITEMS* intuitively speaks to the degree of granularity (Blackwell 1951).

Appendix 1, Panel B, provides an example of the PRE file from the 2015 income statement of Ryder System, Inc. For comparison purposes, Panel A provides the income statement from the Form 10-K filing. The PRE file reports the number of line-items (28). It also distinguishes standard tags (adsh "us-gaap/2014") from extended tags (adsh "0000085961-16-000079"), which we use in our robustness tests.

To construct our disaggregation measure, we count the number of line-items on the balance sheet and the income statement, as follows. We select all rows in the PRE file and tags that are not an abstract type. We consider only the items reported in U.S. dollars that only have a monetary data type (we exclude per-share items).[17] We require annual amounts rather than quarterly amounts. For Ryder System, Inc., the count of the income statement items is 20 (after excluding two abstract items and six per-share items).[18] When a firm also reports a statement of comprehensive income, we count both financial statements. In the case of Ryder System, Inc., the count of the statement of comprehensive income items is 11 (after excluding one abstract tag). Thus, the total income statement disclosure score is 31. For the balance sheet (Appendix 1 – Panel

[16] Other files are as follows. The "SUB" file contains information about the submission, including CIK, SIC, filing date, fiscal period, and form type. The "NUM" file contains the numerical values from the financial statements, and includes the date of the item, the number of quarters the item covers, the unit of measure, and whether the value is for the consolidated entity. The "TAG" file has the specifics of the XBRL tag used, which include the type of tag - a taxonomy tag, an extended tag, or an abstract tag (a row section heading), the datatype (monetary vs. per-share), the sign of the account (Dr. or Cr.), and the duration (quarter or year).
[17] We exclude per-share items because they potentially measure organization and operating choices (dilutive securities, discontinued operations, and extraordinary items). In untabulated results, we include items of *any* datatype and find qualitatively similar results.
[18] We include subtotals in our count as they allow users to more quickly process and verify the financial statements.

C), the disclosure score is 26 (excluding five abstract tags). We include zero amounts for the current year when the comparative figure is non-zero (e.g., pension lump sum settlement expense).[19]

Consistent with CMS, we only include items on the balance sheet and income statement/statement of comprehensive income. We exclude the statement of equity because it is unclear whether the number of items reflects the firm's disaggregation choices or its organization (e.g., multiple share classes and stock option use). We exclude items on the cash flow statement because it is unclear whether the direct method of presentation is more informative than the indirect method (CMS).

Disaggregation and Firm Fundamentals

We validate our disaggregation measure (*ITEMS*) by testing the predicted association with firm fundamentals. We use both *ITEMS* (raw) and the log of *ITEMS* as the dependent variable. We follow CMS to identify firm fundamentals that influence the disclosure policies of the firm. CMS identifies asset restructuring (*Restructure*), mergers and acquisitions (*M&A*), and special items (*SPI*) as disclosure quality determinants.[20] Firms with more mergers and acquisitions, restructurings, and special items are more likely to have greater economic activities outside normal operations and additional line-items on the financial statements. We predict a positive coefficient on the three variables. We also include return volatility (*Std*(*RET*)), firm size (*Log*(*AT*)), the number of business segments (*Log*(*NSEG*)), and firm age (*Log*(*AGE*)). *Std*(*RET*) captures the proprietary information associated with growth options and financial leverage (Christie 1982), and

[19] We include the current period zero amounts with the non-zero comparative amount, for two reasons. First, they inform the user of the absence of an item in the current period compared to the prior (e.g., an asset, liability, revenue, or expense). Second, a financial item that is reported as having a zero amount in the current period means that that item is not aggregated with a summary item (e.g., "Other Expenses"). Ryder System, Inc. reported zero for "Pension lump sum settlement expense" in 2015, but reported $97,231 in 2014, which confirms that the item was not part of a summary item in 2015.

[20] Appendix 2 defines all variables.

we predict a positive coefficient. Size is a proxy for the firm's information environment and should have a positive association with *ITEMS*. However, CMS find the opposite association with *DQ*; thus, we make no prediction on the coefficient sign.[21] The number of business segments captures operating diversity, and may also capture firm size; thus, we make no prediction on the coefficient sign. *Log*(*AGE*) controls for the effects of capital structure over time (e.g., Myers 1984; Myers and Majluf 1984), and we predict a positive coefficient.[22] We also include audit firm size (big 4 and second tier) in the model because larger audit firms tend to have more structured audits that influence their clients' disclosure practices (Kothari, Ramanna, and Skinner 2010; Francis, Pinnuck, and Watanabe 2014; De Franco, Fogel-Yaari, and Li 2020; Johnston and Zhang 2021). Namely, structured audits standardize and reduce the number of line-items. However, since larger auditors are associated with higher financial reporting quality, they could also be associated with more disaggregation; thus, we make no prediction on the coefficient sign on *BIG4* and *Tier2*. Industry fixed-effects control for disclosure related to the firm's operating environment (e.g., number of inventory items),[23] and industry-wide disclosure practices. Year fixed-effects control for firm-invariant temporal effects. All continuous variables in this and subsequent regressions are winsorized at the 1st and 99th percentile. We estimate equation (1) to test the association between *ITEMS* and firm fundamentals, as follows:

$$\begin{aligned} ITEMS_{i,t} = {} & \alpha + \beta_1 Restructure_{i,t} + \beta_2 M\&A_{i,t} + \beta_3 SPI_{i,t} + \beta_4 Std\left(RET_{i,t}\right) + \\ & \beta_5 Log\left(AT_{i,t}\right) + \beta_6 Log\left(NSEG_{i,t}\right) + \\ & \beta_7 Log\left(AGE_{i,t}\right) + \beta_8 BIG4 + \beta_9 Tier2 + YearFE + IndFE + \varepsilon_{i,t} \ . \end{aligned} \tag{1}$$

[21] CMS finds a negative association between firm size and *DQ* using *Compustat* data. They suggest that this relationship is driven by "constraints imposed by U.S. GAAP, which impose an upper bound on the number of items they can report, while inherently having more items available to be reported" (CMS, 1036). If the constraints are imposed by the smaller number of S&P variables, rather than by U.S. GAAP, when compared to the number of XBRL tags, then we should observe a positive association.

[22] Including the number of financing items on the balance sheet measures not only disclosure quality but also firm age and growth options. Capital structure theory (e.g., Myers 1984; Myers and Majluf 1984) suggests that older firms are more likely to have a legacy of diverse debt and equity classes, while contract cost theory (e.g., Myers 1977) suggests that investments with higher growth options are more often financed with shorter-term debt and preferred stock.

[23] For example, firms in the manufacturing industry have multiple inventory items, and those in the business service industry often have none.

Analyst Forecast Properties

We further validate *ITEMS* by testing the predicted association with disclosure consequences: analysts' information, liquidity (bid-ask spread) and cost of equity capital. Starting with analysts' information, greater disaggregation is arguably more informative to analysts by it reducing forecast error (*FE*) and dispersion (*DISP*) (Lang and Lundholm 1996), and we predict a negative coefficient on *ITEMS*. We use analyst earnings forecasts for the year t+1, issued or revised in the 90 days following the Form 10-K filing for year t. *FE* is the average of the absolute value of the earnings per share forecast error, and *DISP* is the standard deviation of the forecasts. Both are scaled by stock price per share at the fiscal year-end.[24] We include several control variables (*Std*(*E*), *Growth*, *ROA*, and *Log*(*COV*)), following CMS. The standard deviation of earnings (*Std*(*E*)) and the five-year average sales growth (*Growth*) control for earnings forecast accuracy (Dichev and Tang 2009), and we predict a positive coefficient. Return on assets (*ROA*) controls for extreme performance, and we predict a negative coefficient. The number of analysts (*Log*(*COV*)) controls for the firm's information environment, and we predict a negative coefficient. Finally, we include the log of the average price per share over the 90-day period following the filing (*Log*(*PRICE*)) to control for scale factors that the denominator of the independent variable may induce, and we predict a negative coefficient. We include firm fundamentals, year fixed-effects, and industry fixed-effects from equation (1). We estimate equation (2) to test the association between analysts' information and *ITEMS*, as follows:

$$\begin{aligned} FE_{i,t}\ (DISP_{i,t}) = {} & \alpha + \beta_1 ITEMS_{i,t} + \beta_2 Std(E)_{i,t} + \beta_3 Growth_{i,t} + \beta_4 ROA_{i,t} + \\ & \beta_5 Log\left(COV_{i,t}\right) + \beta_6 Log\left(PRICE_{i,t}\right) + \beta_7 BIG4 + \beta_8 Tier2 + \\ & \beta_9 Restructure_{i,t} + \beta_{10} M\&A_{i,t} + \beta_{11} SPI_{i,t} + \beta_{12} Log\left(AT_{i,t}\right) + \end{aligned} \quad (2)$$

[24] In unreported results, we find qualitatively similar results if we scale forecast error and dispersion by the magnitude of the average forecast and the magnitude of reported earnings, or if we use the unscaled forecast error and dispersion as suggested by Cheong and Thomas (2011).

$$\beta_{13}Std(RET_{i,t}) + \beta_{14}Log(NSEG_{i,t}) + \beta_{15}Log(AGE_{i,t}) + YearFE + IndFE + \varepsilon_{i,t}\ .$$

Bid-Ask Spread Analysis

We also validate *ITEMS* by testing the predicted association with illiquidity. Greater disaggregation should be negatively associated with illiquidity because disaggregation reveals additional information investors use to trade, and we predict a negative coefficient on *ITEMS*. We measure illiquidity by the bid-ask spread, using both the quoted bid-ask spread (*QBAS*) and the effective bid-ask spread (*EBAS*), following CMS. *QBAS* is equal to (*ASK - BID)/MIDQUOTE*, and *EBAS* is the absolute value of 2*(*PRICE - MIDQUOTE*)/*MIDQUOTE*, where *MIDQUOTE* is (*BID + ASK*)/2. *ASK*, *BID* and *PRICE* are the average of the respective closing daily values for the 90 days after the filing of the Form 10-K. They are from the CRSP daily stock file. We control for factors affecting the bid-ask spread (*Log*(*PRICE*) and *Log*(*VOL*)). *Log*(*PRICE*) controls for market-maker processing costs, and *Log*(*VOL*) controls for inventory holding costs (CMS), and we predict a negative coefficient. We also control for growth potential by including the firm's Book-to-Market ratio (*BTM*), and we predict a positive coefficient. We include firm fundamentals, year fixed-effects, and industry fixed-effects. We estimate equation (3) to test the predicted association between the bid-ask spread and *ITEMS*, as follows:

$$\begin{aligned} QBAS_{i,t}\ (EBAS_{i,t}) \ = \ & \alpha + \beta_1 ITEMS_{i,t} + \beta_2 Log(PRICE)_{i,t} + \beta_3 Log(VOL)_{i,t} + \\ & \beta_4 BTM_{i,t} + \beta_5 Log(AT_{i,t}) + \beta_6 BIG4 + \beta_7 Tier2 + \\ & \beta_8 Restructure_{i,t} + \beta_9 M\&A_{i,t} + \beta_{10} SPI_{i,t} + \beta_{11} Std(RET_{i,t}) + \\ & \beta_{12} Log(NSEG_{i,t}) + \beta_{13} Log(AGE_{i,t}) + YearFE + IndFE + \varepsilon_{i,t} \end{aligned} \quad (3)$$

Cost of Capital Analysis

Finally, we validate *ITEMS* by testing the predicted association with the cost of equity capital (*COEC*). Greater disaggregation should be associated with a lower cost of equity capital,

and we predict a negative coefficient on *ITEMS*. Disaggregation provides investors with more detailed and credible information for valuation and mitigation of mispricing, and it reduces information asymmetry (CMS). We measure the cost of equity capital as the average of three methods in the three months following the Form 10-K filing: *r_MPAG*, *r_GLS*, and *r_CT* following Li and Mohanram (2014), and Hou, van Dijk, and Zhang (2012). We use the most recent earnings forecast and the month-end stock price. We control for several factors affecting the cost of equity capital (*Beta*, *BTM*, and *Log*(*MV*)). *Beta* controls for systematic risk, and we predict a positive coefficient. *Log*(*MV*) and *BTM* control for size and growth potential, respectively. Larger firms tend to be more stable, and we predict a negative coefficient on *Log*(*MV*). Higher *BTM* firms are riskier, and we predict a positive coefficient. We include firm fundamentals,[25] year fixed-effects, and industry fixed-effects. We estimate equation (4) to test the predicted association between the cost of equity capital and *ITEMS*, as follows:

$$\begin{aligned} COEC_{i,t} = {} & \alpha + \beta_1 ITEMS_{i,t} + \beta_2 Log(MV)_{i,t} + \beta_3 BTM_{i,t} + \beta_4 Beta_{i,t} \\ & + \beta_5 BIG4 + \beta_6 Tier2 + \beta_7 Restructure_{i,t} \\ & + \beta_8 M\&A_{i,t} + \beta_9 SPI_{i,t} + \beta_{10} Std(RET_{i,t}) \\ & + \beta_{11} Log(NSEG_{i,t}) + \beta_{12} Log(AGE_{i,t}) \\ & + YearFE + IndFE + \varepsilon_{i,t} \end{aligned} \quad (4)$$

4. RESULTS

Descriptive Statistics

Table 1 outlines our sample selection for the four samples: firm fundamentals (equation 1), analyst (equation 2), bid-ask spread (equation 3), and cost of capital (equation 4). Our sample period covers fiscal years 2009 to 2017. We start by obtaining 47,509 Form 10-K filings from the SEC FSDS. For the firm fundamentals sample, we exclude 512 firm-year observations that are

[25] We exclude *Log*(*AT*) from firm fundamentals due to high multicollinearity. The VIF for *Log*(*AT*) and *Log*(*MV*) are 16.90 and 16.09, respectively. When *Log*(*AT*) is excluded, the VIF for *Log*(*MV*) is 2.69. However, the inferences are unchanged if we include *Log*(*AT*) in equation (4).

either reported in a foreign currency or do not identify the financial statement type. We exclude 9,787 firm-year observations that are not in *Compustat* and 7,376 firm-year observations that are not in CRSP. We exclude 282 firm-year unusual observations where the firm reports less than three items or more than 100 items (multiple entities) on both the income statement and the balance sheet. We exclude 1,874 firm-year observations, where we cannot calculate the firm fundamental variables. A base sample of 27,678 firm-year observations is the starting point for the other three samples. To mitigate the influence of outliers, we winsorize all continuous variables at the 1st and 99th percentiles and exclude 47 observations where the external studentized residual magnitude exceeds five. The firm fundamentals sample is 27,631 firm-year observations.

For the analyst sample, we require that a firm have at least three revisions in the 90 days following the Form 10-K filing. This requirement and the data needed for additional variables in equation (2) exclude 13,286 firm-year observations. After dropping observations with a studentized residual magnitude exceeding five (382), the analyst sample is 14,010 firm-year observations.

For the bid-ask spread sample, we require each firm-year observation to have at least 30 bid and ask prices in the 90 days following the Form 10-K filing. This requirement and the data needed for additional variables in equation (3) exclude 150 firm-year observations. After dropping observations with a studentized residual magnitude exceeding five (133), the bid-ask sample size is 27,395 firm-year observations.

For the cost of capital sample, after dropping missing data to compute the cost of equity capital (6,600) and observations that exceed the studentized residual magnitude of five (296), the sample size is 20,782 firm-year observations.

[INSERT TABLE 1 HERE]

Table 2 presents summary statistics for the sample. Panel A, B, C, and D report for the firm fundamentals, analyst, bid-ask spread, and cost of capital samples, respectively.[26] Panel A reports that the mean (median) firm-year observation in our sample has 57.509 (56) line-items on the balance sheet and income statement. The mean of *Restructure* (0.283), *M&A* (0.170), and *SPI* (0.018) are consistent with HH. The standard deviation of daily returns has a mean (median) of 0.109 (0.089), which is consistent with CMS. The mean of *Log*(*NSEG*) (0.278), *Log*(*AT*) (6.943), *Log*(*AGE*) (2.884), *BIG4* (0.703) and *Tier2* (0.097) are consistent with HH.

Panel B presents the descriptive statistics for the analyst sample. Mean (median) forecast error per share is 0.034 (0.028), forecast dispersion is 0.003 (0.105) and raw *STD*(*E*) is 5.624 (6.000), which are consistent with prior studies (e.g., Wu and Wilson 2016). Mean and median sales growth, *ROA*, and analyst coverage are comparable to prior studies (e.g., Feng, Li, McVay, and Skaife 2015). We note that the firms in our analyst sample are slightly larger, with a mean total assets of 2.581 billion (mean *Log*(*AT*) of 7.856).

Panel C presents the descriptive statistics for the bid-ask spread sample. The mean effective (quoted) bid-ask spread is 48.412 (54.476) basis points, and the median effective (quoted) spread is 12.182 (9.944) percent of the mid-quote. The implied mean daily volume is 280,688 shares, while the implied median average daily volume is 338,405 shares. The mean (median) book-to-market ratio is 0.625 (0.494). Panel D presents the descriptive statistics for the cost of capital sample. The mean cost of equity capital is 11.695, and the firm-specific beta is 1.020. All the variables presented are consistent with prior studies (e.g., HH; Li and Mohanram 2014).

[INSERT TABLE 2 HERE]

[26] Correlations among the control variables are not high. Most correlations are below 0.40 in magnitude. The maximum correlation in the firm fundamentals sample is -0.50, between *BIG4* and *Tier2.* In the analyst sample, it is -0.66, between *BIG4* and *Tier2*. In the bid-ask spread sample, it is -0.65, between *EBAS* (%) and *Log*(*PRICE*). In the cost of capital sample, it is 0.58, between *BIG4* and *SIZE*.

Predictors of Disaggregation

Table 3 presents the results from estimating equation (1) with OLS. The first and second columns report the results from the dependent variable, *ITEMS*, in the raw and log form, respectively. The predictors of disclosure quality have coefficients in the expected direction and are economically significant. *Restructure* has a positive coefficient (p<0.01), which indicates that a restructuring charge has 1.450 (3 percent) more line-items.[27, 28] *M&A* has a positive coefficient (p<0.01), which indicates that M&A firms have 1.006 (2.4 percent) more line-items. *SPI* has a positive coefficient (p<0.01), which indicates that special items have 5.342 (10.5 percent) more line-items. *Log*(*NSEG*) and *Log*(*AGE*) have positive coefficients (p<0.01), suggesting that a one percent increase in the number of segments and age yields 0.02 more line-items.[29]

Log(*AT*) has a positive coefficient (p<0.01),[30] and with similar economic significance to *Log*(*NSEG*) and *Log*(*Age*). However, CMS report a negative coefficient on *Log*(*AT*). CMS argue that the constraints imposed by U.S. GAAP drive the negative relation. Our results suggest that the *Compustat* database imposes constraints rather than U.S. GAAP. *BIG4* has a negative coefficient which could be due to the greater similarity of financial statement tags between Big 4 clients engaging the same audit firm and thus requiring fewer (and more similar) line-items (Johnston and Zhang 2021). We leave it to future research to further explore the role of auditor quality on disaggregation quality. In short, these results for the most part suggest that firm fundamentals explain our measure of disaggregation (*ITEMS*) in the predicted direction.

[INSERT TABLE 3 HERE]

[27] Consistent with CMS, we multiply all coefficients in the first column by 100 for ease of exposition.
28 Semi-elasticities from the logged form of *ITEMS* are presented in parenthesis.
[29] The economic significance is based on the raw form of *ITEMS*. A one percent increase in the variable of interest is equal to the coefficient divided by 100 (Wooldridge, 2016, 639).
[30] The relation between *SIZE* and *ITEMS* is consistent if we use the log of market value instead of *Log*(*AT*).

Validation of ITEMS as a measure of disaggregation

ITEMS and Analysts' Information

Table 4 presents the results of estimating equation (2). The first and second column present the results, where the dependent variable is *FE* and *DISP*, respectively. The coefficient on *ITEMS* is negative (p<0.01) for both dependent variables, as predicted. Further, the control variable coefficients are in the predicted direction. These results are consistent with the notion that *ITEMS* is positively associated with analyst information.

[INSERT TABLE 4 HERE]

ITEMS and Illiquidity

Table 5 presents the results of estimating equation (3). The first and second column present the results, where the dependent variable is the quoted bid-ask spread (*QBAS*) and effective bid-ask spread (*EBAS*), respectively. The coefficient on *ITEMS* is negative (p<0.01) for both dependent variables, as predicted. Further, the control variable coefficients are in the predicted direction. These results are consistent with the notion that *ITEMS* is negatively associated with illiquidity.

[INSERT TABLE 5 HERE]

ITEMS and Cost of Equity Capital

Table 6 presents the results of estimating equation (4), where the dependent variable is the cost of equity capital (*COEC*). The coefficient on *ITEMS* is negative (p<0.01), as predicted. Further, the control variable coefficients are in the predicted direction. These results are consistent with the notion that *ITEMS* is negatively associated with the cost of equity capital.

[INSERT TABLE 6 HERE]

Additional Analyses

Decomposing by Financial Statement

Following CMS, we redefine *ITEMS* by the line-item count of the balance sheet ($ITEMS_{BS}$) and the income statement ($ITEMS_{IS}$). CMS find that income statement line-items primarily drive the association between analyst information and *DQ*. In contrast, they find that both balance sheet and income statement line-items affect the association between the bid-ask spread and *DQ*.

Table 7 presents the results of estimating modified equations (2), (3) and (4) by replacing *ITEMS* with $ITEMS_{BS}$ and $ITEMS_{IS}$. In the first two columns, we report equation (2) results. The coefficient on $ITEMS_{BS}$ is not statistically significant, whereas the coefficient on $ITEMS_{IS}$ is negative for *FE* ($p<0.01$). The result is consistent with CMS. It also supports the conventional view that analysts use the income statement more than the balance sheet. The third and fourth columns report the modified equation (3) results. The coefficient on $ITEMS_{BS}$ is negative ($p<0.01$), while the coefficient on $ITEMS_{IS}$ is not statistically significant. These results suggest that the disclosure of balance sheet line-items reveals more about firm risk (e.g., financial leverage and liquidity). The last column presents the modified equation (4) results. The coefficient on $ITEMS_{BS}$ is negative ($p<0.01$), but the coefficient on $ITEMS_{IS}$ is not statistically significant. These results also suggest that the disclosure of balance sheet items reveals more about firm risk than income statement items.

[INSERT TABLE 7 HERE]

Disaggregation Quality (*DQ*) and Accounting Reporting Complexity (*ARC*)

ITEMS has similarities to and differences from CMS's *DQ* disaggregation measure and HH's *ARC* complexity measure.[31] *ITEMS* is similar to *DQ* in that it measures disaggregation using income statement and balance sheet data. However, they differ in that *ITEMS* uses XBRL data from a taxonomy of more than 15,000 tags, whereas *DQ* uses *Compustat's* balancing model with around 145 variables, and *ITEMS* excludes footnote tags. *ITEMS* is similar to *ARC* in that it does not use *Compustat* variables, and its construction includes the count of balance sheet and income statement line-items. Consequently, *ITEMS* and *ARC* are correlated.[32]

However, *ITEMS* differs from *ARC* in several respects. First, *ITEMS* only counts the line-items on the face of the balance sheet and the income statement and not the footnote tags, whereas *ARC* counts all the financial statement and footnote tags. Further, *ITEMS* is a different construct from *ARC* because it measures disaggregation of financial statement line-items, rather than accounting reporting complexity, which have different consequences.[33] As mentioned, *ITEMS* is more constrained by the item-count than *ARC* because the income statement and balance sheet have a practical maximum number of line-items. In contrast, *ARC* includes the count of footnotes of detailed accounts (e.g., fixed asset cost by class) that have a larger (and more ambiguous) practical maximum (several pages).

[31] HH compares *DQ* to *ARC* and concludes they are constructively different because "*DQ* is based on around 145 *Compustat* variables, while *ARC* is based on all monetary items that companies disclose in their financial statements and notes. Thus, *ARC* relies on a broader set of disclosures" (HH, 227). HH further suggest that "*ARC* is more directly linked to accounting, is based on *more detailed disclosures*, and exhibits greater variation than operating and linguistic complexity measures" (HH, 264, italics added).

[32] *ARC* and *ITEMS* are correlated because more detailed disclosures report more financial datapoints, which increases accounting reporting complexity (HH). The Pearson correlation between *ARC* and *ITEMS* is approximately 0.60. We recommend size-industry adjusting *ITEMS* or controlling for firm size and industry in a multivariate regression. When we size-industry adjust *ITEMS*, the correlation with *ARC* is 0.28 and the correlation with *DQ* is 0.14.

[33] We examine the positive consequences of disaggregation such as analyst forecast properties, firm liquidity, and cost of capital. HH examine the negative consequences of accounting reporting complexity such as material weaknesses, audit fees, misstatements, audit delay, and accruals quality.

Consequently, we posit that financial statement users weight footnotes differently (and possibly less) than the line-items on the face of the financial statements, following views by practitioners (e.g., Taub 2012; Ghai 2016; Tysiac 2018) and academic studies (e.g., Hirshleifer and Teoh 2003). For example, in an ongoing effort by the FASB to improve the effectiveness of footnote disclosure, comments from an exposure draft suggest that footnote disclosure grew 28 percent over the six years to 2012 (Tysiac 2018). The exposure draft comments suggest that footnotes for defined benefit plans, deferred taxes, stock options, and fair value measurements are too detailed for analysts and investors (sophisticated and naïve) to comprehend (Taub 2012; Ghai and Rapp 2016). Because investors have a limited attention span and limited processing power, they fixate on line-items on the face of the financial statements and discount items in the footnotes (Hirshleifer and Teoh 2003). Further, as mentioned, footnotes are potentially biased because management can obfuscate accounts by placing them in the footnotes rather than on the financial statements proper.

To evaluate whether *DQ* and *ARC* are different from *ITEMS*, we conduct three empirical tests. In the first test, Table 8, Panel A, reports the results from including *ARC*, *DQ*, and two other complexity variables in equations (2), (3) and (4), while controlling for firm fundamentals. We find that the coefficient on *ITEMS* is negative and statistically significant ($p<0.05$, except *FE* - $p<0.10$) and it has incremental explanatory power over the coefficients on *ARC* and *DQ*. Note that *ARC* is positively associated with *COEC*, but *ITEMS* is negatively associated, suggesting that *ARC* is capturing a construct distinct from *ITEMS*.

[INSERT TABLE 8 PANEL A HERE]

In the second test, we examine the informational role of the line-items recognized on the balance sheet and income statement and the footnote items. Table 8, Panel B, reports the results of

separating *ARC* into the line-items recognized on the balance sheet and income statement (*ITEMS* (*FS*)) and the footnote items (*ITEMS* (*NOTES*)). The results suggest that most of the associations we document between *ITEMS* and the external information environment are not affected by footnote items. However, *ITEMS* (*NOTES*) is positively associated with cost of capital, suggesting that investors discount the value of footnotes possibly because of the higher information processing cost.

[INSERT TABLE 8 PANEL B HERE]

In the final test, we explore the association of the dependent variables examined in HH (misstatements, materials weaknesses, accrual quality, audit delay, and audit fees) with *ITEMS* (*FS*) and *ITEMS* (*NOTES*). Table 8, Panel C, presents the results that suggest that most of the results documented in HH are driven by footnote item tags.[34] This is plausible because items recognized on the balance sheet and the income statement only make up a small portion of the tags.[35] The coefficient on *ITEMS* (*NOTES*) is also comparable to those documented by HH on *ARC*. Consistent with HH, we fail to document an association between either *ITEMS* (*FS*) or *ITEMS* (*NOTES*) and accrual quality (*AQ*). Further, we document a positive association between *ITEMS* (*FS*) and audit delay, suggesting that firms with more financial reporting complexity and more financial statement line-items require more time to prepare and audit the financial statements.

[INSERT TABLE 8 PANEL C HERE]

[34] Following HH, we control for *BUSINESS_SEG*, *GEO_SEG*, *FOREIGN*, *10-K File Size*, *SIZE, LOSS*, *GC*, *DISTRESS*, *EXT_GROWTH*, *STD_CFO*, *STD_SALES*, *LEVERAGE*, *INV_REC*, *RESTRUCTURE*, *ACQUISITION*, *SPECIAL_ITEMS*, *FIRM_AGE*, *LIT_IND*, *BIG4*, *MAT_WEAK*, and *RESTATE*. When misstatements (*RESTATE*) is the dependent variable, material weakness (*MAT_WEAK*) is a control variable. Similarly, when the dependent variable is *MAT_WEAK*, we control for *RESTATE*. In all other regressions, both *MAT_WEAK* and *RESTATE* are control variables. When accrual quality (*AQ*) is the dependent variable, we add three additional control variables (*LOSS_PROP, STD_ROA, MEAN_CYCLE*), following HH. Two-digit SIC industry and year fixed-effects are added, and results are robust to using alternative industry classifications. Please refer to HH's Appendix A for the variable definitions.

[35] The mean of *ITEMS* is 58 line-item tags (Table 2, Panel A). HH report the mean of *ARC* is 346 line-item and footnote tags.

Disaggregation by Type

Next, we focus on whether certain the item-type (standard, extended and parenthetical) drives our results. We first investigate the power of extended tags on the financial statements, which firms create when the line-item is not defined by the FASB. If extended tags are idiosyncratic, *ITEMS* is driven by greater disclosure of firm-specific information rather than by more line-items per se. In an untabulated analysis, we create a variable, $ITEMS_{Extended}$, which is the count of extended tags on either the balance sheet or the income statement. Across all our models, we find that standard tags, rather than extended tags, drive our results.

We next test whether the placement of the item matters. Within the FSDS, we can identify if an item is presented as a line-item or as a parenthetical item. Parenthetical items include allowance for doubtful accounts, accumulated depreciation/amortization, and tax effects for discontinued items. While they likely are relevant, they are not reported as a separate line-item. To test, we create a variable, $ITEMS_{Parenthetical}$, which is the count of parenthetical items. In an untabulated analysis, we find that the coefficient on $ITEMS_{Parenthetical}$ is statistically insignificant while the coefficient on *ITEMS* remains significant, suggesting that parenthetical items do not affect our results.

Robustness Tests

For robustness, we test various specifications. First, we check whether our results are robust to removing financial (SIC 6000 – 6999) and utility firms (SIC 4900 – 4999) because these industries generally report more tags. When we remove them, we find qualitatively similar results. Next, we test whether our results are sensitive to the year of XBRL implementation. XBRL implementation was phased-in by filer size, and our sample excludes smaller firms before June 30, 2010. When we exclude *all* filings before June 30, 2010, we find qualitatively similar results.

Finally, we alter our clustering of standard errors by firm only, and by industry and year, and we find similar results.

In our main tables, we winsorize continuous variables at 1st and 99th to mitigate undue outlier influence and we delete observations with a studentized residual greater than five in magnitude to remove the influence. Leone, Minutti-Meza, and Wasley (2019) note that some empirical results may be the result of undue influence by outliers, and they recommend using robust regression to reduce the influence of outliers. We repeat our main tests using the MM-estimator suggested by Leone, et al. (2019) as an alternative to winsorization and deletion based on studentized residuals. Our results are largely consistent with our main results, with the only exception that the association between *ITEMS* and *DISP* is insignificant. We also consider cut-offs of two, three and four for studentized residuals and we find similar results to our main analysis. We also remove observations if Cook's D is greater than 4/N, rather than by studentized residuals, where N is the number of observations in the respective sample for the test, and our results hold. Thus, our results are largely robust to controlling for the influence of outliers.

Finally, we test for whether there is a learning effect during the early years of XBRL implementation, and whether our results are robust to excluding the early years (Du et al. 2013). We find some evidence of a learning effect, though we also find that our results are robust to excluding the first three years when the learning effect is more pronounced. In our first set of tests, we regress each of disclosure consequence variables (*FE*, *DISP*, *EBAS*, *QBAS*, and *COEC*) on the interaction of *ITEMS* and the year fixed-effect variable. We find that there is a statistically significant learning effect (i.e., a more negative coefficient over time) with *COEC*, evident from a negative trend of the coefficient on the interaction term over time ($p<0.01$), and a marginally negative trend with *QBAS* ($p<0.1$). Next, we conduct three tests using the same dependent variables

and interacting *ITEMS* with 1) a pre-June 30, 2011 fiscal-year indicator variable (*Pre-2011*), 2) an indicator variable of the first three annual XBRL filings (*NumFiling3*), and 3) the cumulative number of annual XBRL filings (*NumFilings*). We find evidence of a significant learning effect with the *Pre-2011* interaction term and *EBAS*, *QBAS* and *COEC* (p<0.05), and the *NumFilings* interaction term and *FE* (p<0.05). We find marginal significance with the *NumFiling3* interaction term and *FE* (p<0.1). However, we find a positive association between the *NumFilings* interaction term and *EBAS* and *QBAS*. Last, we construct two sub-samples that exclude the first three years. In the first sub-sample, we drop fiscal years 2009-2011, and in the second sub-sample we drop the first three years of annual XBRL filings. We estimate equations (2) to (4), and we find that in eight of ten tests the coefficient on *ITEMS* is negative and statistically significant (p<0.01) and marginally significant in the two remaining tests (*DISP*) (p<0.1).

5. CONCLUSION

We examine the use of the SEC's FSDS to measure disaggregation − a type of disclosure quality. While prior studies have attempted to fill the void for a proper measure of disclosure quality, most have limitations. In a related study, CMS measure disaggregation using S&P's *Compustat* by counting the number of non-missing balance sheet and income statement variables. We complement their study by counting the number of line-items on the balance sheet and income statement from the FSDS (*ITEMS*). *ITEMS* is very intuitive because it captures the direct financial reporting practices of companies without being constrained by S&P and it does not require a data aggregator subscription.

We validate *ITEMS* by testing whether it is explained by firm fundamentals, and whether it explains disclosure consequences: analyst EPS forecast error and dispersion, bid-ask spread, and cost of equity capital. We find that *ITEMS* is positively associated with firm fundamentals in the

predicted direction: restructuring, M&A, special items, size, number of segments and age. We also find *ITEMS* is negatively associated with disclosure consequences: forecast error, forecast dispersion, bid-ask spread, and cost of equity capital, as predicted, providing evidence that *ITEMS* captures firm disclosure quality. Our results are robust to several tests, including CMS's *DQ* measure of disclosure quality and HH's *ARC* measure of accounting reporting complexity with *ITEMS* in the tests of disclosure consequences. We find that the effect of *ITEMS* on disclosure consequences is not subsumed by *DQ* or *ARC*. Our study contributes to the disclosure quality literature in its quest to develop a new and improved measure of disaggregation. Future studies can examine parenthetical items and footnote items in more depth.

There are three caveats to using *ITEMS*. First, it is not available before 2010 and may not be useful for longer-term studies at present, but it is available for most of the past decade and will be more useful in later years. Second, unlike CMS's measure, *ITEMS* does not have a scaler and might not be helpful to compare across industries. Instead, researchers may size-industry-adjust *ITEMS* or add industry and size controls in multivariate tests.[36] Third, prior studies (e.g., Debreceny et al. 2011) have alerted that early XBRL filings have high error rates. However, consistent with the learning curve theory, the error rate of XBRL filings is decreasing as companies complete more XBRL filings, suggesting improved data quality following the introduction of the XBRL filing requirement (Vasarhelyi et al. 2012; Du et al. 2013). Further, the recent move towards the inline XBRL reporting requirements could mitigate some of these issues.

[36] We recommend size-industry adjusting *ITEMS* because there is more variation of *ITEMS* between industries (σ=14.08) than within (σ=4.89), suggesting that industry practices effect disclosure similarity. To illustrate, recall that Ryder System, Inc. reports three revenue categories. Industry peers, such as Hertz Global Holdings, Inc. (cik 1364479) and Avis Budget Group, Inc. (cik 723612), for the same fiscal year (FY), report three and two categories, respectively (σ=0.58). However, comparing the mean number of revenue categories of the three vehicle rental companies (μ=2.67) to other industries such as a retailer (two categories - Walmart, Inc., cik 1041690, FY 01/31/2016), a technology company (one category - Microsoft Corp., cik 789019, FY 06/30/2015), and a petroleum company (three categories - Exxon Mobil Corp., cik 34088, FY 12/31/2015), there is more variation between the four industries (σ=0.88) than within the four vehicle rental companies.

APPENDIX 1– PANEL A

Extract from Form 10-K filings

Ryder System, Inc. for the fiscal year ended December 31, 2015 (filed February 12, 2016, cik 85961)

Consolidated Statements of Earnings

		12 Months Ended		
	Amounts in thousands	**Dec. 31, 2015**	**Dec. 31, 2014**	**Dec. 31, 2013**
1	Lease and rental revenues*	$ 3,121,553	$ 2,939,422	$ 2,770,026
2	Services revenue*	2,912,063	2,911,465	2,819,673
3	Fuel services revenue*	538,277	787,887	829,586
4	Total revenues	6,571,893	6,638,774	6,419,285
5	Cost of lease and rental*	2,153,450	2,036,881	1,925,546
6	Cost of services	2,413,156	2,447,867	2,359,880
7	Cost of fuel services*	519,843	768,292	814,058
8	Other operating expenses*	135,038	126,572	131,659
9	Selling, general and administrative expenses	844,497	816,975	790,681
10	Pension lump sum settlement expense*	0	97,231	0
11	Gains on vehicles sales, net*	(117,809)	(126,824)	(96,175)
12	Interest expense	150,434	144,739	140,463
13	Miscellaneous income, net	(10,156)	(13,613)	(15,372)
14	Restructuring and other charges (recoveries), net*	14,225	2,387	(470)
15	Total expenses*	6,102,678	6,300,507	6,050,270
16	Earnings from continuing operations before income taxes	469,215	338,267	369,015
17	Provision for income taxes	163,226	118,042	125,740
18	Earnings from continuing operations	305,989	220,225	243,275
19	Loss from discontinued operations, net of tax	(1,221)	(1,884)	(5,404)
20	Net earnings	$ 304,768	$ 218,341	$ 237,871
Abstract	Earnings (loss) per common share — Basic			
Per share	Continuing operations (in dollars per share)	$ 5.78	$ 4.18	$ 4.67
Per share	Discontinued operations (in dollars per share)	(0.02)	(0.04)	(0.10)
Per share	Net earnings (in dollars per share)	5.75	4.14	4.57
Abstract	Earnings (loss) per common share — Diluted			
Per share	Continuing operations (in dollars per share)	5.73	4.14	4.63
Per share	Discontinued operations (in dollars per share)	(0.02)	(0.03)	(0.10)
Per share	Net earnings (in dollars per share)	$ 5.71	$ 4.11	$ 4.53

* denotes an extended tag, all others, numbered 1 to 20, are standard tags. *ITEMS* is constructed from the count of extended and standard tags (excluding abstract and per-share items).

APPENDIX 1- PANEL B

Extract of PRE file for Ryder System, Inc. 2015 Q4

Income Statement and Comprehensive Income Statement

adsh	report	line	Tag	version	plabel
0000085961-16-000079	2	1	LeaseAndRentalRevenue	0000085961-16-000079	Lease and rental revenues
0000085961-16-000079	2	2	ServiceRevenue	0000085961-16-000079	Services revenue
0000085961-16-000079	2	3	FuelServiceRevenue	0000085961-16-000079	Fuel services revenue
0000085961-16-000079	2	4	Revenues	us-gaap/2014	Total revenues
0000085961-16-000079	2	5	CostOfLeaseAndRental	0000085961-16-000079	Cost of lease and rental
0000085961-16-000079	2	6	CostOfServices	us-gaap/2014	Cost of services
0000085961-16-000079	2	7	CostOfFuelServices	0000085961-16-000079	Cost of fuel services
0000085961-16-000079	2	8	OtherOperatingExpenses	0000085961-16-000079	Other operating expenses
0000085961-16-000079	2	9	SellingGeneralAndAdministrativeExpense	us-gaap/2014	Selling, general and administrative expenses
0000085961-16-000079	2	10	Definedbenefitplanlumpsumsettlementchargepretax	0000085961-16-000079	Pension lump sum settlement expense
0000085961-16-000079	2	11	GainsOnVehicleSalesNet	0000085961-16-000079	Gains on vehicles sales, net
0000085961-16-000079	2	12	InterestExpense	us-gaap/2014	Interest expense
0000085961-16-000079	2	13	OtherNonoperatingIncomeExpense	us-gaap/2014	Miscellaneous income, net
0000085961-16-000079	2	14	RestructuringAndOtherChargesRecoveriesNet	0000085961-16-000079	Restructuring and other charges (recoveries), net
0000085961-16-000079	2	15	Expenses	0000085961-16-000079	Total expenses
0000085961-16-000079	2	16	IncomeLossFromContinuingOperationsBeforeIncomeTaxesMinorityInterestAndIncomeLossFromEquityMethodInvestments	us-gaap/2014	Earnings from continuing operations before income taxes
0000085961-16-000079	2	17	IncomeTaxExpenseBenefit	us-gaap/2014	Provision for income taxes
0000085961-16-000079	2	18	IncomeLossFromContinuingOperations	us-gaap/2014	Earnings from continuing operations
0000085961-16-000079	2	19	IncomeLossFromDiscontinuedOperationsNetOfTaxAttributableToReportingEntity	us-gaap/2014	Loss from discontinued operations, net of tax
0000085961-16-000079	2	20	NetIncomeLoss	us-gaap/2014	Net earnings
0000085961-16-000079	2	21	EarningsPerShareBasicAbstract†	us-gaap/2014	Earnings (loss) per common share \x14 Basic
0000085961-16-000079	2	22	IncomeLossFromContinuingOperationsPerBasicShare◊	us-gaap/2014	Continuing operations (in dollars per share)
0000085961-16-000079	2	23	IncomeLossFromDiscontinuedOperationsNetOfTaxPerBasicShare◊	us-gaap/2014	Discontinued operations (in dollars per share)
0000085961-16-000079	2	24	EarningsPerShareBasic◊	us-gaap/2014	Net earnings (in dollars per share)
0000085961-16-000079	2	25	EarningsPerShareDilutedAbstract†	us-gaap/2014	Earnings (loss) per common share \x14 Diluted

adsh	report	line	Tag	version	plabel
0000085961-16-000079	2	26	IncomeLossFromContinuingOperationsPerDilutedShare◊	us-gaap/2014	Continuing operations (in dollars per share)
0000085961-16-000079	2	27	IncomeLossFromDiscontinuedOperationsNetOfTaxPerDilutedShare◊	us-gaap/2014	Discontinued operations (in dollars per share)
0000085961-16-000079	2	28	EarningsPerShareDiluted◊	us-gaap/2014	Net earnings (in dollars per share)
0000085961-16-000079	3	1	NetIncomeLoss	us-gaap/2014	Net earnings
0000085961-16-000079	3	2	OtherComprehensiveIncomeLossBeforeTaxPortionAttributableToParentAbstract†	us-gaap/2014	Other comprehensive (loss) income:
0000085961-16-000079	3	3	OtherComprehensiveIncomeForeignCurrencyTransactionAndTranslationAdjustmentBeforeTaxPortionAttributableToParent	us-gaap/2014	Changes in cumulative translation adjustment and other
0000085961-16-000079	3	4	OtherComprehensiveIncomeLossActuarialAndPriorServiceCostsNetPeriodicBenefitCostBeforeTax	0000085961-16-000079	Amortization of pension and postretirement items
0000085961-16-000079	3	5	OtherComprehensiveIncomeLossReclassificationnetgainlosspriorservicecredit	0000085961-16-000079	Income tax expense related to amortization of pension and postretirement items
0000085961-16-000079	3	6	Othercomprehensiveincomelosspensionreclassificationnetoftax	0000085961-16-000079	Amortization of pension and postretirement items, net of tax
0000085961-16-000079	3	7	OtherComprehensiveIncomeLossReclassificationAdjustmentFromAOCIPensionAndOtherPostretirementBenefitPlansBeforeTax	us-gaap/2014	Reclassification of net actuarial loss from pension settlement
0000085961-16-000079	3	8	OtherComprehensiveIncomeLossPensionAndOtherPostretirementBenefitPlansAdjustmentBeforeReclassificationAdjustmentsAndTax	us-gaap/2014	Change in net actuarial loss and prior service credit
0000085961-16-000079	3	9	OtherComprehensiveIncomeLossPensionAndOtherPostretirementBenefitPlansTax	us-gaap/2014	Income tax benefit (expense) related to change in net actuarial loss and prior service credit
0000085961-16-000079	3	10	OtherComprehensiveIncomeLossPensionAndOtherPostretirementBenefitPlansAdjustmentNetOfTax	us-gaap/2014	Change in net actuarial loss and prior service credit, net of taxes
0000085961-16-000079	3	11	OtherComprehensiveIncomeLossNetOfTaxPortionAttributableToParent	us-gaap/2014	Other comprehensive (loss) income, net of taxes
0000085961-16-000079	3	12	ComprehensiveIncomeNetOfTax	us-gaap/2014	Comprehensive income

Column Names: adsh – the accession number of the filer, **report –** numeric statement type (2=income statement, 3=Statement of Comprehensive Income, 4=Balance Sheet), **line** – The line number of the report, **tag –**the tag name, **version –** standard tag (us-gaap/2014), or an extended tag (accession number for extended tags), and **plabel –** the line-item text presented on the financial statement. For brevity, we do not show columns **stmt (**alphanumeric statement type (IS = Income Statement, CI = Comprehensive Income Statement, BS = Balance Sheet)), **inpth (**parenthetical disclosure indicator (1 = yes, 0 = no)), and **rfile –** file type on the EDGAR website (H = .htm file, X = .xml file)). † indicates an abstract tag (row heading). ◊ Indicates a per-share item.

APPENDIX 1- PANEL C
Extract of PRE file for Ryder System, Inc. 2015 Q4 – Balance Sheet

adsh	report	line	Tag	version	plabel
0000085961-16-000079	4	1	AssetsAbstract†	us-gaap/2014	Assets:
0000085961-16-000079	4	2	AssetsCurrentAbstract†	us-gaap/2014	Current assets:
0000085961-16-000079	4	3	CashAndCashEquivalentsAtCarryingValue	us-gaap/2014	Cash and cash equivalents
0000085961-16-000079	4	4	ReceivablesNetCurrent	us-gaap/2014	Receivables, net
0000085961-16-000079	4	5	InventoryNet	us-gaap/2014	Inventories
0000085961-16-000079	4	6	PrepaidExpenseAndOtherAssetsCurrent	us-gaap/2014	Prepaid expenses and other current assets
0000085961-16-000079	4	7	AssetsCurrent	us-gaap/2014	Total current assets
0000085961-16-000079	4	8	RevenueEarningEquipmentNetOfAccumulatedDepreciation	0000085961-16-000079	Revenue earning equipment, net
0000085961-16-000079	4	9	PropertyPlantAndEquipmentNet	us-gaap/2014	Operating property and equipment, net
0000085961-16-000079	4	10	Goodwill	us-gaap/2014	Goodwill
0000085961-16-000079	4	11	IntangibleAssetsNetExcludingGoodwill	us-gaap/2014	Intangible assets
0000085961-16-000079	4	12	DirectFinancingLeasesAndOtherAssets	0000085961-16-000079	Direct financing leases and other assets
0000085961-16-000079	4	13	Assets	us-gaap/2014	Total assets
0000085961-16-000079	4	14	LiabilitiesAndStockholdersEquityAbstract†	us-gaap/2014	Liabilities and shareholders equity:
0000085961-16-000079	4	15	LiabilitiesCurrentAbstract†	us-gaap/2014	Current liabilities:
0000085961-16-000079	4	16	DebtCurrent	us-gaap/2014	Short-term debt and current portion of long-term debt
0000085961-16-000079	4	17	AccountsPayableCurrent	us-gaap/2014	Accounts payable
0000085961-16-000079	4	18	AccountsPayableAndOtherAccruedLiabilitiesCurrent	us-gaap/2014	Accrued expenses and other current liabilities
0000085961-16-000079	4	19	LiabilitiesCurrent	us-gaap/2014	Total current liabilities
0000085961-16-000079	4	20	LongTermDebtAndCapitalLeaseObligations	us-gaap/2014	Long-term debt
0000085961-16-000079	4	21	OtherLiabilitiesNoncurrent	us-gaap/2014	Other non-current liabilities
0000085961-16-000079	4	22	DeferredTaxLiabilitiesNoncurrent	us-gaap/2014	Deferred income taxes
0000085961-16-000079	4	23	Liabilities	us-gaap/2014	Total liabilities
0000085961-16-000079	4	24	StockholdersEquityAbstract †	us-gaap/2014	Shareholders equity:

adsh	report	line	Tag	version	plabel
0000085961-16-000079	4	25	PreferredStockValue	us-gaap/2014	Preferred stock, no par value per share authorized, 3,800,917; none outstanding, December 31, 2015 or 2014
0000085961-16-000079	4	26	CommonStockValue	us-gaap/2014	Common stock, $0.50 par value per share authorized, 400,000,000; outstanding, December 31, 2015 53,490,603; December 31, 2014 53,039,688
0000085961-16-000079	4	27	AdditionalPaidInCapitalCommonStock	us-gaap/2014	Additional paid-in capital
0000085961-16-000079	4	28	RetainedEarningsAccumulatedDeficit	us-gaap/2014	Retained earnings
0000085961-16-000079	4	29	AccumulatedOtherComprehensiveIncomeLossNetOfTax	us-gaap/2014	Accumulated other comprehensive loss
0000085961-16-000079	4	30	StockholdersEquity	us-gaap/2014	Total shareholders equity
0000085961-16-000079	4	31	LiabilitiesAndStockholdersEquity	us-gaap/2014	Total liabilities and shareholders equity

Column Names: adsh – the accession number of the filer, **report –** numeric statement type (2=income statement, 3=Statement of Comprehensive Income, 4=Balance Sheet), **line** – The line number of the report, **tag –**the tag name, **version –** standard tag (us-gaap/2014), or an extended tag (accession number for extended tags), and **plabel –** the line-item text presented on the financial statement. For brevity, we do not show columns **stmt (**alphanumeric statement type (IS = Income Statement, CI = Comprehensive Income Statement, BS = Balance Sheet)), **inpth (**parenthetical disclosure indicator (1 = yes, 0 = no)), and **rfile –** file type on the EDGAR website (H = .htm file, X = .xml file)). † indicates an abstract tag (row heading).

APPENDIX 2
Variable Definitions

$ITEMS_{i,t}$ = The number of items reported by firm *i* on the income statement and the balance sheet for year *t*. For brevity, in all variable definitions, year *t* is fiscal year *t*.

$DQ_{i,t}$ = The measure of disclosure quality from CMS that includes the number of line-items reported by firm *i* from the financial statements and footnotes using *Compustat's* balancing model in year *t* scaled by the number of possible reporting items, following CMS.

$ARC_{i,t}$ = An HH based measure of accounting reporting complexity that includes the number of items reported by firm *i* from the income statement, balance sheet and footnotes in year *t*.

Firm Fundamentals

$Restructure_{i,t}$ = An indicator variable equal to one if firm *i* had non-zero restructuring charges (*Compustat* mnemonic *RCP*) in year *t*, and zero otherwise.

$M\&A_{i,t}$ = An indicator variable equal to one if firm *i*'s sales reflected mergers and acquisitions activity in year *t* as indicated by the footnote on *SALE* within *Compustat*, and zero otherwise.

$SPI_{i,t}$ = The absolute value of special items, scaled by total assets (*AT*) for firm *i* in year *t*.

$Std(RET_{i,t})$ = The standard deviation of daily returns from the CRSP's daily stock file for firm *i* in year *t*.

$Log(AT_{i,t})$ = The log of total assets (*AT*) for firm *i* in year *t*.

$Log(NSEG_{i,t})$ = The log of the number of business segments firm *i* reports in year *t*.

$Log(AGE_{i,t})$ = The log of the number of years firm *i* has had accounting data available in *Compustat*.

BIG4 = An indicator variable equal to one when the firm's auditor is a big four auditor, and zero otherwise.

Tier2 = An indicator variable equal to one when the firm's auditor is Grant Thornton or BDO, and zero otherwise.

Analyst Information Model

$FE_{i,t}$ = The average absolute analyst earnings per share forecast error for year *t+1* that were issued or reviewed in the 90 days after the filing of firm *i*'s Form 10-K for year *t*, scaled by price per share at the end of year *t*.

$DISP_{i,t}$ = The standard deviation of analyst earnings per share forecasts for year *t+1* that were issued or reviewed in the 90 days after the filing of firm *i*'s Form 10-K for year *t*, scaled by price per share at the end of year *t*.

$STD(E)_{i,t}$ = The decile rank of the standard deviation of income before extraordinary items (*IB*) divided by adjusted shares outstanding (*CSHO*AJEX*) scaled by price per share at the end of year *t*.

$Growth_{i,t}$ = The average percentage growth in sales (*SALE*) taken over year *t*-4 to year *t*.

$ROA_{i,t}$ = Income before extraordinary items (*IB*) divided by total assets (*AT*).

$Log(COV_{i,t})$ = The log of the number of brokerage firms issuing or reviewing a forecast of year *t*'s earning within year *t*.

$Log(PRICE_{i,t})$ = The log of the average closing price over the 90 days after the filing of firm *i*'s Form 10-K for year *t*.

Bid-Ask Spread Model

$QBAS_{i,t}$ = The average daily quoted bid-ask spread over the 90 days after the filing of firm *i*'s Form 10-K for year *t*. Daily quoted bid-ask spread is (*ASK*-*BID*)/MIDPOINT, where *ASK* and *BID* are the closing ask and bid from the CRSP daily stock file and *MIDQUOTE* is (*BID*+*ASK*)/2.

$EBAS_{i,t}$ = The average daily effective bid-ask spread taken over the 90 days after the filing of firm *i*'s Form 10-K for year *t*. The daily effective bid-ask spread is the absolute value of 2*(*PRICE* - *MIDQUOTE*) divided by *MIDQUOTE*, where *MIDQUOTE* is (*BID*+*ASK*)/2, and *PRICE*, *ASK*, and *BID* is the closing price, ask, and bid from the CRSP daily stock file, respectively.

$Log(PRICE_{i,t})$ = The log of the average closing price over the 90 days after the filing of firm *i*'s Form 10-K for year *t*.

$Log(VOL_{i,t})$ = The log of the average trading volume over the 90 days after the filing of firm *i*'s Form 10-K for year *t*.

$BTM_{i,t}$ = The book value of equity (*CEQ*) divided by the market value of equity (*PRCC_F*CSHO)* at the end of year *t*.

Cost of Equity Capital Model

$COEC_{i,t}$ = The average cost of capital (*COEC*) for three months following the Form 10-K filing date. *COEC* is the average of *r_MPAG*, *r_GLS*, *r_CT* using the most recent estimated earnings, and prices at the end of the month, and multiplied by 100. We use the formula used by Li and Mohanram (2014) and the method used by Hou et al. (2012) to estimate expected future earnings.

$Log(MV_{i,t})$ = The log of market value (*PRCC_F*CSHO)* of firm *i* at the end of year *t*.

$BTM_{i,t}$ The book value of equity (*CEQ*) divided by the market value of equity (*PRCC_F*CSHO)* at the end of year *t*.

$Beta_{i,t}$ = The Scholes-Williams beta of firm *i* for year *t*. This method adjusts for non-synchronous trading, and uses daily returns and market returns from CRSP daily stock and index files.

Additional Item Variables

$ITEMS_{BS,i,t}$ = The number of items reported by firm *i* on the balance sheet in year *t*.

$ITEMS_{IS,i,t}$ = The number of items reported by firm *i* on the income statement in year *t*.

$10\text{-}K\ File\ Size_{i,t}$	=	The gross Form 10-K filing size for firm *i* in year *t* as provided by Loughran and McDonald available here: https://sraf.nd.edu/data/.
$Log(NSEG_{i,t})$	=	The natural log of the number of geographic segments reported by firm *i* in year *t*.
$ITEMS(FS)_{i,t}$	=	The number of items reported on the financial statements for firm *i* in year *t*.
$ITEMS(NOTES)_{i,t}$	=	The number of items reported in the footnotes to the financial statements for firm *i* in year *t*.

REFERENCES

Alles, M., and G. L. Gray. 2012. A relative cost framework of demand for external assurance of XBRL filings. *Journal of Information Systems* 26 (1): 103-126.

American Institute of Certified Public Accountants (AICPA). 2013. *Performing Agreed-Upon Procedures Engagements that Address the Completeness, Mapping, Consistency, or Structure of XBRL-Formatted Information.* Statement of Position 13-2. New York, NY: AICPA. Available from http://www.aicpa.org/interestareas/frc/accountingfinancialreporting/xbrl/pages/sop13-2performingagreed-uponprocedures.aspx.

AICPA. 2017. *SEC Rules for Reporting Financial Statements in XBRL Format.* New York, NY: AICPA. Available from: https://us.aicpa.org/interestareas/frc/accountingfinancialreporting/xbrl.

Barth, M. E., and K. Schipper. 2008. Financial reporting transparency. *Journal of Accounting, Auditing & Finance* 23 (2): 173-190.

Basoglu, K. A., and C. E. White, Jr. 2015. Inline XBRL versus XBRL for SEC reporting. *Journal of Emerging Technologies in Accounting* 12: 189-199.

Beyer, A., D. A. Cohen, T. Z. Lys and B. R. Walther. 2010. The Financial Reporting Environment: Review of the Recent Literature. *Journal of Accounting & Economics* 50 (2/3): 296-343.

Blackwell, D. 1951. *Comparisons of Experiments*. Edited by J. Neyman. Berkeley and Los Angeles: University of California Press.

Blankespoor, E. 2019. The impact of information processing costs on firm disclosure choice: Evidence from the XBRL mandate. *Journal of Accounting Research* 57 (4): 919-967.

Blankespoor, E., E. deHaan, and I. Marinovic. 2020. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics* 70 (2-3): 1-46. https://doi.org/10.1016/j.jacceco.2020.101344.

Bonner, S. E., S. M. Clor-Proell, and L. Koonce. 2014. Mental accounting and disaggregation based on the sign and relative magnitude of income statement items. *The Accounting Review* 89 (6): 2087-2114.

Boritz, J. E., and W. G. No. 2016. Computer-Assisted Functions for Auditing XBRL-Related Documents. *Journal of Emerging Technologies in Accounting* 13 (1): 53-83.

Boritz, J. E., and W. G. No. 2020. How significant are the differences in financial data provided by key data sources? A comparison of XBRL, Compustat, Yahoo! Finance, and Google Finance. *Journal of Information Systems* 34 (3): 47-75.

Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72 (3): 323-349.

Chen, S., B. Miao, and T. Shevlin. 2015. A new measure of disclosure quality: The level of disaggregation of accounting data in annual reports. *Journal of Accounting Research* 53 (5): 1017-1054.

Cheong, F. S., and J. Thomas. 2011. Why do EPS forecast error and dispersion not vary with scale? Implications for analyst and managerial behavior. *Journal of Accounting Research* 49 (2): 359-401.

Christie, A. A. 1982. The stochastic behavior of common stock variances: Value, leverage, and interest rate effects. *Journal of Financial Economics* 10 (4): 407-432.

Chychyla, R., and A. Kogan. 2015. Using XBRL to conduct a large-scale study of discrepancies between the accounting numbers in Compustat and SEC 10-K filings. *Journal of Information Systems* 29 (1): 37-72.

D'Souza, J. M., K. Ramesh, and M. Shen. 2010. The interdependence between institutional ownership and information dissemination by data aggregators. *The Accounting Review* 85 (1): 159-193.

De Franco, G., H. Fogel-Yaari, and H. Li. 2020. MD&A textual similarity and auditors. *Auditing: A Journal of Practice and Theory* 39 (3): 105-131. https://doi.org/10.2308/ajpt-18-016.

Debreceny, R. S., S. M. Farewell, M. Piechocki, C. Felden, A. Gräning, and A. d'Eri. 2011. Flex or break? Extensions in XBRL disclosures to the SEC. *Accounting Horizons* 25 (4): 631-657.

Dichev, I. D., and V. W. Tang. 2009. Earnings volatility and earnings predictability. *Journal of Accounting & Economics* 47 (1/2): 160-181.

Du, H., M. A. Vasarhelyi, and X. Zheng. 2013. XBRL mandate: Thousands of filing errors and so what? *Journal of Information Systems* 27 (1): 61-78.

Elliott, W. B., J. L. Hobson, and K. E. Jackson. 2011. Disaggregating management forecasts to reduce investors' susceptibility to earnings fixation. *The Accounting Review* 86 (1): 185-208.

Farewell, S., and R. Pinsker. 2015. Does assurance on XBRL-derived financial statements impact the decisions of nonprofessional investors? *Management Accounting Quarterly* 16 (3): 13-20.

Feng, M., C. Li, S. E. McVay, and H. Skaife. 2015. Does ineffective internal control over Financial reporting affect a firm's operations? Evidence from firms' inventory management. *The Accounting Review* 90 (2): 529-557.

Forker, J. J. 1992. Corporate governance and disclosure quality. *Accounting & Business Research* 22 (86): 111-124.

Francis, J., D. Nanda, and P. Olsson. 2008. Voluntary disclosure, earnings quality, and cost of Capital. *Journal of Accounting Research* 46 (1): 53-99.

Francis, J.R., M. Pinnuck, and O. Watanabe. 2014. Auditor style and financial statement comparability. *The Accounting Review* 89 (2): 605–633.

Ghai, P., and A. Rapp. 2016. Value of XBRL for financial analysis. *Accounting Today.* Available from https://www.accountingtoday.com/news/value-of-xbrl-for-financial-analysis.

Hayes, R. M., and R. Lundholm. 1996. Segment reporting to the capital market in the presence of a competitor. *Journal of Accounting Research* 34 (2): 261-279.

Healy, P. M., and K. G. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31 (1-3): 405-440.

Hirshleifer, D., and S. H. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36 (1-3): 337-386.

Hirst, D. E., L. Koonce, and S. Venkataraman. 2007. How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research* 45 (4): 811-837.

———. 2008. Management earnings forecasts: A review and framework. *Accounting Horizons* 22 (3): 315-338.

Hodge, F. D., J. J. Kennedy, and L. A. Maines. 2004. Does search-facilitating technology improve the transparency of financial reporting? *The Accounting Review* 79 (3): 687-703.

Hoitash, R., and U. Hoitash. 2018. Measuring accounting reporting complexity with XBRL. *The Accounting Review* 93 (1): 259-287.

Hoitash, R., and U. Hoitash, and L. Morris. 2020. eXtensible Business Reporting Language: A Review and Directions for Future Research. Available at *SSRN*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3595401.

Hou, K., M. A. van Dijk, and Y. Zhang. 2012. The implied cost of capital: A new approach. *Journal of Accounting & Economics* 53 (3): 504-526.

Howard, J. K., and J. Zhou. 2021. The timeliness of XBRL filings: An empirical examination. *Journal of Information Systems* 35 (1): 65-77.

Johnston, J. A., and J. H. Zhang. 2021. Auditor style and financial reporting similarity. *Journal of Information Systems.* 35(1): 79-99.

Kinney, M. R., and E. P. Swanson. 1993. The accuracy and adequacy of tax data in Compustat. *Journal of the American Taxation Association* 15 (1): 121-135.

Kothari, S.P. 2019. *Policy Challenges and Research Opportunities in the Era of Big Data.* U.S. Securities and Exchange Commission (July 13). Available at https://www.sec.gov/news/speech/policy-challenges-research-opportunities-era-big-data.

Kothari, S.P., K. Ramanna, and D. Skinner. 2010. Implications for GAAP from an analysis of positive research in accounting. *Journal of Accounting and Economics* 50 (2-3): 246–286.

Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research* 31 (2): 246-271.

Lang, M. H., and R. J. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4): 467-492.

Lansford, B., B. Lev, and J. W. Tucker. 2013. Causes and consequences of disaggregating earnings guidance. *Journal of Business Finance & Accounting* 40 (1/2): 26-54.

Leone, A. J., M. Minutti-Meza, and C. E. Wasley. 2019. Influential observations and inference in accounting research. *The Accounting Review* 94 (6): 337–364.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2-3): 221-247.

Li, K., and P. Mohanram. 2014. Evaluating cross-sectional forecasting models for implied cost of capital. *Review of Accounting Studies* 19 (3): 1152-1185.

Li, E. X., K. Ramesh, and M. Shen. 2011. The role of newswires in screening and disseminating value-relevant information in periodic SEC reports. *The Accounting Review* 86 (2): 669-701.

Lim, S. C. 2014. The information content of disaggregated accounting profitability: Operating activities versus financing activities. *Review of Quantitative Finance & Accounting* 43: 75-96.

Loughran, T. I. M., and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69 (4): 1643-1671.

Moffitt, K. C., V. J. Richardson, N. M. Snow, M. M. Weisner, and D. A. Wood. 2016. Perspectives on past and future AIS research as the Journal of Information Systems turns thirty. *Journal of Information Systems* 30 (3): 157-171.

Myers, S. C. 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5 (2): 147-175.

Myers, S. C. 1984. The capital structure puzzle. *Journal of Finance* 39 (3): 575-592.

Myers, S. C., and N. S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13 (2): 187-221.

Richardson, V. J. 2000. Information asymmetry and earnings management: Some evidence. *Review of Quantitative Finance and Accounting* 15: 325-347.

Sanmiguel, J. G. 1977. Reliability of R&D data in Compustat and 10-K reports. *The Accounting Review* 52 (3): 638-641.

Taub, S. 2012. Best chance of fixing disclosure overload is here. *Compliance Week* 9 (103): 1.

Tysiac, K. 2018. FASB takes big steps on disclosure effectiveness. *Journal of Accountancy*. Available at: https://www.journalofaccountancy.com/news/2018/aug/fasb-disclosure-effectiveness-201819617.html.

U.S. Securities and Exchange Commission (SEC). 2016a. Final rule: Interactive data to improve financial reporting (Release Nos. 33-9002; 34-59324; 39-2461; IC-28609; File No. S7-11-08; January 30). Available from https://www.sec.gov/rules/final/2009/33-9002.pdf.

———. 2016b. Financial statement data sets. *Office of Structured Disclosure.* Available from https://www.sec.gov/dera/data/financial-statement-data-sets.html.

———. 2020. Inline XBRL. April 9. Available from: https://www.sec.gov/structureddata/osd-inline-xbrl.html.

Vasarhelyi, M. A., D. Y. Chan, and J. P. Krahel. 2012. Consequences of XBRL standardization on financial statement data. *Journal of Information Systems* 26 (1): 155-167.

Williamson, O. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

Willis, M. 2019. *Insights from 'as Reported' disclosures*. 2019 AAA Intensive Data & Analytics II Summer Workshop, at Orlando, Florida.

Wooldridge, J. M. 2016. *Introductory Econometrics: A Modern Approach*. 6th ed. Mason, OH: Thomson/South-Western.

Wu, Y. A., and M. Wilson. 2016. Audit quality and analyst forecast accuracy: The impact of forecast horizon and other modeling choices. *Auditing: A Journal of Practice & Theory* 35 (42): 167-185.

XBRL International. 2016. *An Introduction to XBRL*. Available from: https://www.xbrl.org/the-standard/what/an-introduction-to-xbrl/.

XBRL U.S. 2016. *2016 US GAAP Financial Reporting Taxonomy*. Available from https://xbrl.us/xbrl-taxonomy/2016-us-gaap/.

Table 1
Sample Selection

	Sample Size
SEC FSDS Form 10-K filings in fiscal years 2009 to 2017	47,509
ITEMS not calculable	-512
Not matched to Compustat	-9,787
Not matched to CRSP	-7,376
ITEMS more than 100 or less than three items on the income statement and the balance sheet	-282
Missing data for firm fundamentals	-1,874
Base sample	27,678
Studentized residual magnitude greater than five	-47
Firm fundamentals sample	27,631
Base sample	27,678
Missing data for analysts' model	-13,286
Studentized residual magnitude greater than five	-382
Analyst sample	14,010
Base sample	27,678
Missing data for the bid-ask spread model	-150
Studentized residual magnitude greater than five	-133
Bid-ask spread sample	27,395
Base sample	27,678
Missing data for the COEC model	-6,600
Studentized residual magnitude greater than five	-296
Cost of capital sample	20,782

Table 2
Summary Statistics

Panel A: Firm Fundamentals Sample

	Mean	Std	Q1	Median	Q3
ITEMS	57.509	14.351	48.000	56.000	66.000
Restructure	0.283	0.450	0.000	0.000	1.000
M&A	0.170	0.376	0.000	0.000	0.000
SPI	0.018	0.077	0.000	0.002	0.012
Std(*RET*)	0.109	0.080	0.061	0.089	0.133
Log(*AT*)	6.943	2.155	5.490	7.028	8.402
Log(*NSEG*)	0.278	0.449	0.000	0.000	0.693
Log(*AGE*)	2.884	0.757	2.398	2.944	3.401
Big4	0.703	0.457	0.000	1.000	1.000
Tier2	0.097	0.296	0.000	0.000	0.000

Panel A reports summary statistics for the firm fundamentals sample (n=27,631) covering 2009 through 2017. Appendix 2 defines all variables.

Panel B: Analyst Sample

	Mean	Std	Q1	Median	Q3
ITEMS	59.326	14.023	50.000	57.000	67.000
FE	0.034	0.030	0.020	0.028	0.038
DISP	0.003	0.006	0.000	0.105	0.374
STD(*E*)	5.624	2.733	3.000	6.000	8.000
Growth	0.191	0.575	0.020	0.078	0.175
ROA	0.020	0.131	0.007	0.034	0.072
Log(*COV*)	2.478	0.644	1.946	2.485	2.996
Log(*PRICE*)	3.387	0.956	2.819	3.469	4.022
BIG4	0.856	0.351	1.000	1.000	1.000
Tier2	0.069	0.253	0.000	0.000	0.000
Log(*AT*)	7.856	1.797	6.644	7.830	8.987

Panel B reports summary statistics for the analyst sample (n=14,010) covering 2009 through 2017. Appendix 2 defines all variables.

Panel C: Bid-Ask Spread Sample

	Mean	Std	Q1	Median	Q3
ITEMS	57.476	14.418	48.000	56.000	66.000
QBAS	54.021	100.862	4.015	9.944	47.274
EBAS	48.412	78.402	6.119	12.182	47.766
Log(*PRICE*)	2.887	1.243	2.160	3.059	3.788
Log(*VOL*)	12.545	1.924	11.362	12.732	13.895
BTM	0.625	0.540	0.270	0.494	0.810
Log(*AT*)	6.953	2.115	5.509	7.041	8.409
BIG4	0.707	0.455	0.000	1.000	1.000
Tier2	0.096	0.294	0.000	0.000	0.000

Panel C reports summary statistics for the bid-ask spread sample (n=27,395) covering 2009 through 2017. Appendix 2 defines all variables.

Table 2 (cont.)

Panel D: Cost of Capital Sample

	Mean	Std	Q1	Median	Q3
ITEMS	58.005	14.498	48.000	56.000	67.000
COEC	11.695	13.417	2.900	6.850	15.130
Log(*MV*)	6.766	2.061	5.312	6.831	8.174
BTM	0.665	0.570	0.293	0.530	0.856
Beta	1.020	0.641	0.601	0.961	1.374
BIG4	0.707	0.455	0.000	1.000	1.000
Tier2	0.092	0.289	0.000	0.000	0.000
Restructure	0.303	0.459	0.000	0.000	1.000
M&A	0.176	0.381	0.000	0.000	0.000
SPI	0.017	0.073	0.000	0.002	0.012
STD(*Ret*)	0.108	0.078	0.062	0.089	0.130
Log(NSEG)	0.300	0.461	0.000	0.000	0.693
Log(AGE)	2.972	0.725	2.565	2.996	3.466

Panel D reports summary statistics for the cost of capital sample (n=20,782) covering 2009 through 2017. Appendix 2 defines all variables.

TABLE 3
Predictors of Disaggregation

Independent Variable	Predicted	Dependent Variable: *ITEMS*	Dependent Variable: *Log*(*ITEMS*)
Restructure	+	1.450 (4.22)***	0.030 (5.17)***
M&A	+	1.006 (2.53)**	0.024 (2.91)***
SPI	+	5.342 (2.87)***	0.105 (2.60)***
Std(*RET*)	+	3.365 (1.05)	0.036 (0.50)
Log(*AT*)	?	2.437 (14.33)***	0.044 (12.27)***
Log(*NSEG*)	?	2.405 (5.08)***	0.045 (4.86)***
Log(*AGE*)	+	1.917 (5.87)***	0.034 (5.10)***
BIG4	?	-1.482 (-2.29)**	-0.026 (-1.77)*
Tier2	?	0.459 (0.92)	0.014 (1.21)
Industry FE		Included	Included
Year FE		Included	Included
Observations		27,631	27,631
Adjusted R^2		0.539	0.495

This table reports the OLS results of regressing disaggregation (*ITEMS*) and *Log*(*ITEMS*)) on firm fundamentals using a sample of 27,631 firm-year observations (2009-2017). The dependent variable in the first column is the sum of the number of line-items presented on the balance sheet and the income statement (*ITEMS*), and in the second column is the log of *ITEMS*. Appendix 2 defines all other variables. Two-digit SIC industry and year fixed-effects are included but not reported for simplicity. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 4
Analyst Information and Line-items Reported on the Income Statement and Balance Sheet

Independent Variable	Predicted	Dependent Variable: *FE*	Dependent Variable: *DISP*
ITEMS	**-**	**-0.009** **(-3.61)*****	**-0.001** **(-2.61)*****
STD(*E*)	+	0.218 (15.48)***	0.038 (13.48)***
Growth	+	0.440 (5.96)***	0.069 (4.17)***
ROA	-	-4.570 (-8.13)***	-0.615 (-5.45)***
Log(*COV*)	-	-0.387 (-5.93)***	-0.022 (-1.59)
Log(*PRICE*)	-	-1.165 (-15.60)***	-0.239 (-10.92)***
Firm Fundamentals		Included	Included
Industry FE		Included	Included
Year FE		Included	Included
Observations		14,010	14,010
Adjusted R^2		0.324	0.309

This table reports the OLS results of regressing analyst forecast error and dispersion on the number of the line-items reported on the balance sheet and income statement (*ITEMS*) and other variables. The sample has 14,010 firm-year observations (2009-2017). The dependent variable in the first column is the analyst earnings per share forecast error (*FE*), and in the second column is the analyst earnings per share forecast dispersion (*DISP*). The variable of interest is *ITEMS*. Appendix 2 defines all other variables. Two-digit SIC industry and year fixed-effects are included but not reported for simplicity. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 5
Bid-Ask Spread and Line-items Reported on the Income Statement and Balance Sheet

Independent Variable	Predicted	Dependent Variable	
		QBAS	***EBAS***
ITEMS	**-**	**-0.254**	**-0.1457**
		(-3.31)***	**(-2.90)*****
Log(*PRICE*)	-	-31.918	-32.450
		(-14.18)***	(-18.92)***
Log(*VOL*)	-	-27.520	-17.346
		(-15.46)***	(-16.04)***
BTM	+	15.579	10.147
		(3.73)***	(3.60)***
Firm Fundamentals		Included	Included
Industry FE		Included	Included
Year FE		Included	Included
Observations		27,395	27,395
Adjusted R^2		0.593	0.645

This table reports the OLS results of regressing bid-ask spread measures on the number of line-items reported on the balance sheet and income statement (*ITEMS*) and other variables. The sample has 27,395 firm-year observations (2009-2017). The dependent variable in the first column is the quoted bid-ask spread (*QBAS*), and in the second column is the effective bid-ask spread (*EBAS*). The variable of interest is *ITEMS*. Appendix 2 defines all other variables. Two-digit SIC industry and year fixed-effects are included but not reported for simplicity. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 6
Cost of Equity Capital and Line-items Reported on the Income Statement and Balance Sheet

Independent Variable	Predicted	Dependent Variable *COEC*
ITEMS	**-**	**-0.036** **(-3.75)*****
Log(*MV*)	-	-2.625 (-17.31)***
BTM	+	5.905 (12.95)***
Beta	+	0.570 (1.64)
Firm Fundamentals		Included
Industry FE		Included
Year FE		Included
Observations		20,782
Adjusted R^2		0.396

This table reports the OLS results of regressing the cost of equity capital (*COEC*) on the number of line-items reported on the balance sheet and income statement (*ITEMS*) and other variables. The sample has 20,782 firm-year observations (2009-2017). The variable of interest is *ITEMS*. Appendix 2 defines all other variables. Two-digit SIC industry and year fixed-effects are included but not reported for simplicity. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 7
Decomposing Line-items by Statement

Independent Variable	Analyst Information		Bid-Ask Spread		Cost of Capital
	FE	***DISP***	***QBAS***	***EBAS***	***COEC***
$ITEMS_{BS}$	-0.001	-0.001	-0.521	-0.322	-0.098
	(-0.12)	(-1.35)	(-4.18)***	(-3.70)***	(-6.29)***
$ITEMS_{IS}$	-0.017	-0.001	-0.017	0.010	0.017
	(-3.23)***	(-0.94)	(-0.15)	(0.16)	(1.14)
Controls	Included	Included	Included	Included	Included
Firm Fundamentals	Included	Included	Included	Included	Included
Industry FE	Included	Included	Included	Included	Included
Year FE	Included	Included	Included	Included	Included
Observations	14,010	14,010	27,395	27,395	20,782
Adjusted R^2	0.325	0.309	0.593	0.646	0.397

This table reports the OLS results of regressing analyst information, the bid-ask spread, and the cost of equity capital on $ITEMS_{BS}$, $ITEMS_{IS}$, firm fundamentals and other variables using the respective samples for the dependent variables (2009-2017). $ITEMS_{BS}$ is the number of line-items on the balance sheet, and $ITEMS_{IS}$ is the number of line-items on the income statement. The first two columns report the analyst information model estimate (equation 2), where the dependent variable in the first and second columns are the analyst earnings per share forecast error (*FE*) and the forecast dispersion (*DISP*), respectively. The next two columns report the bid-ask spread model estimate (equation 3), where the dependent variable in columns four and five are the quoted bid-ask spread (*QBAS*) and the effective bid-ask spread (*EBAS*), respectively. The last column reports the results of the cost of capital model (equation 4), where the dependent variable is the cost of equity capital (*COEC*). Appendix 2 defines all other variables. Coefficients for the bid-ask spread models (columns three and four) are multiplied by 100 for ease of exposition. Two-digit SIC industry and year fixed-effects are included but not reported for simplicity. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 8: PANEL A
ITEMS, DQ, and ARC

Independent Variable	Analysts' Information		Bid-Ask Spread		Cost of Capital
	FE	***DISP***	***QBAS***	***EBAS***	***COEC***
ITEMS	-0.010 (-3.47)***	-0.001 (-1.80)*	-0.305 (-2.45)**	-0.198 (-2.47)**	-0.063 (-5.50)***
ARC	0.000 (1.73)*	0.000 (0.35)	0.013 (0.59)	0.016 (1.03)	0.005 (3.95)***
DQ	-0.658 (-2.52)**	-0.121 (-2.66)***	23.965 (1.49)	8.353 (1.51)	0.576 (0.32)
10-K File Size	-0.000 (-1.71)*	-0.000 (-1.06)	-0.000 (-0.06)	-0.000 (-0.42)	0.000 (1.65)*
Log(*NSEG*)	-0.038 (-0.66)	-0.010 (-0.96)	-0.260 (-0.25)	-0.858 (-1.05)	-0.681 (-2.66)***
Controls	Included	Included	Included	Included	Included
Firm Fundamentals	Included	Included	Included	Included	Included
Industry FE	Included	Included	Included	Included	Included
Year FE	Included	Included	Included	Included	Included
Observations	14,006	14,006	27,387	27,387	20,780
Adjusted R^2	0.326	0.312	0.593	0.646	0.398

This table reports the OLS results of regressing analyst information, the bid-ask spread, and the cost of equity capital on the number of line-items reported on the balance sheet and income statement (*ITEMS*), HH's accounting reporting complexity measure (*ARC*), and CMS's disclosure quality measure (*DQ*). We use the respective samples for the dependent variables (2009-2017). The dependent variables in the first through the fifth columns are *FE*, *DISP*, *QBAS*, *EBAS*, and *COEC*, respectively. The regression estimates control for *10-K File Size*, and the log of business segments (*Log*(*NSEG*)). Appendix 2 defines all variables. For simplicity, we do not report the results of the control variables (shown in tables 4, 5, and 6), the firm fundamentals, and the two-digit SIC industry and year fixed-effects. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 8: PANEL B
ITEMS and Footnote Tags and Information Environment

Independent Variable	Analysts' Information		Bid-Ask Spread		Cost of Capital
	FE	***DISP***	***QBAS***	***EBAS***	***COEC***
ITEMS (*FS*)	-0.006	-0.001	-0.241	-0.156	-0.067
	(-1.65)*	(-1.63)	(-3.83)***	(-3.36)***	(-6.57)***
ITEMS (*NOTES*)	-0.000	-0.000	0.001	0.003	0.008
	(-1.19)	(-1.62)	(0.34)	(0.95)	(5.41)***
Controls	Included	Included	Included	Included	Included
Firm Fundamentals	Included	Included	Included	Included	Included
Industry FE	Included	Included	Included	Included	Included
Year FE	Included	Included	Included	Included	Included
F-test of FS = NOTES	5.02**	0.83	12.20***	11.02***	43.10***
Observations	11,295	11,295	23,539	23,539	17,991
Adjusted R^2	0.318	0.303	0.592	0.643	0.403

This table reports the OLS results of regressing analyst information, bid-ask spread, and cost of equity capital on the number of line-items reported on the balance sheet and income statement (*ITEMS* (*FS*)) and the number of items reported in the footnotes (*ITEMS* (*NOTES*)). We use the respective samples for the dependent variables (2009-2017) except for observations where we could not clearly identify the footnote sections. The dependent variables in the first through the fifth columns are *FE*, *DISP*, *QBAS*, *EBAS*, and *COEC*, respectively. The regression estimates control for the same variables as the main regressions reported in Tables 4, 5, and 6. Appendix 2 defines all variables. For simplicity, we do not report the results of the control variables (shown in tables 4, 5, and 6), the firm fundamentals, and the two-digit SIC industry and year fixed-effects. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.

TABLE 8: PANEL C
ITEMS and Footnote Tags and the Consequences of Reporting Complexity

Independent Variable	Dependent Variables				
	RESTATE	***MW***	***AQ***	***Audit-Delay***	***Audit-Fees***
ITEMS (*FS*)	0.179	0.566	0.007	6.032	-0.089
	(0.53)	(1.29)	(1.90)*	(4.86)***	(-1.38)
ITEMS (*NOTES*)	1.033	1.362	0.002	2.194	0.395
	(5.75)***	(5.90)***	(1.34)	(2.71)**	(8.79)***
Controls	Included	Included	Included	Included	Included
Industry FE	Included	Included	Included	Included	Included
Year FE	Included	Included	Included	Included	Included
Observations	7,054	6,914	4,153	7,077	7,077
Adjusted R^2	0.049	0.065	0.215	0.071	0.843

This table reports the OLS results of regressing the consequences of reporting complexity on the number of line-items reported on the balance sheet and income statement (*ITEMS* (*FS*)) and the number of items reported in the footnotes (*ITEMS* (*NOTE*S)). The dependent variables in the first through the fifth columns are misstatements (*RESTATE*), material weaknesses (*MAT_WEAK*), accrual quality (*AQ*), audit delay, and audit fees, respectively (2009-2017). The regression estimates control for *BUSINESS_SEG*, *GEO_SEG*, *FOREIGN*, *_10_K_LENGTH*, *SIZE*, *LOSS*, *GC*, *DISTRESS*, *EXT_GROWTH*, *STD_CFO*, *STD_SALES*, *LEVERAGE*, *INV_REC*, *RESTRUCTURE*, *ACQUISITION*, *SPECIAL_ITEMS*, *FIRM_AGE*, *LIT_IND*, *BIG4*, *MAT_WEAK*, and *RESTATE*. For the regression results in the first column, *RESTATE* is the dependent variable and *MAT_WEAK* is a control variable. Similarly, for the regression results in the second column, *MAT_WEAK* is the dependent variable and *RESTATE* is a control variable. In all other regression results, both *MAT_WEAK* and *RESTATE* are control variables. For the regression related to *AQ,* three additional control variables (*LOSS_PROP*, *STD_ROA*, and *MEAN_CYCLE*) are added, following HH. For simplicity, we do not report the results of the control variables and the two-digit SIC industry and year fixed-effects. We closely follow HH in constructing these variables. Please refer to HH's Appendix A for the variable definitions. We compute t-statistics (in parentheses) from standard errors clustered by firm and year. *, **, *** indicates significance at the 10%, 5%, and 1% levels, respectively.