# **A Measure of Firm Complexity**

**Explanation of Variables and their Measurement**

We propose that firm complexity is best measured with accounting disclosures. Accounting is the “language of business”, and accounting disclosures of most business activities are mandated. Therefore, relying on accounting disclosures is the best approach for capturing a wide range of firm activities in a consistent manner for a large cross-section of firms. Accounting Reporting Complexity (ARC), proposed by Hoitash and Hoitash (2018) is one of several measures used to measure firm complexity.

The data provided on this website is sourced from annual and quarterly XBRL filings. We rely on data from the SEC’s website (available at <https://www.sec.gov/dera/data/financial-statement-and-notes-data-set.html>), to construct the variables.

The files contain several measures that are based on: [Hoitash, R., Hoitash, U., 2018. Measuring Accounting Reporting Complexity with XBRL. *The Accounting Review*, 2018, Vol. 93, No. 1, pp. 259-287.](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2433677)

[A detailed description of the data construction and SAS code can be found in: Hoitash, R., Hoitash, U., 2022. “A Measure of Firm Complexity- Data and Code”.](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4098929)

If you have questions, please contact Udi Hoitash ([u.hoitash@northeastern.edu](mailto:u.hoitash@northeastern.edu))

or Rani Hoitash ([rhoitash@bentely.edu](mailto:rhoitash@bentely.edu)). We are happy to help!

There are two sets of files:

1. Containing Firm Complexity, covering annual measure of firm complexity for the years 2011-2022 based on 10-K filings.

2. Containing various measures of Accounting Reporting Complexity (ARC), covering the years 2009-2022, including annual and quarterly measures of firm complexity and several permutations of ARC.

Both files share similar firm identifiers as described below

***ARC (Firm Complexity) included in both files***- is the count of unique XBRL tags per financial statement disclosure and is constructed as follows: The instance document contains all XBRL facts (a fact is an instance of a tag) included in the filings.

1. First, we remove all facts that are not monetary. These include facts that are textual, shares, percent, etc.
2. Second, we keep non-repeating facts within each disclosure (disclosure is a statement, a financial statement note, or a table in a financial statement note).
3. After removing non- monetary facts and facts that repeat within a disclosure we count the number of facts in each filing. The count of remaining facts is ARC.

Below we highlight a few ARC measures included only in the ARC files and present the complete list of variables and their definitions.

***ARC\_ALL\_NUMS***- There are several alternatives for constructing ARC. As mentioned above, ARC includes non-repeating tags within each disclosure. We also create measures based on ALL numerical tags (monetary and other numbers) reported in the filings. These include tags that repeat within a disclosure and tags that repeat because of comparable financial statement disclosure. This approach assumes that complexity is a combination of the number of tags used and their frequency of use.

***ARC\_EXTENSIONS* -** Each tag is either a standard taxonomy tag or a customized extended (company-specific) tag. Taxonomy tags are tags approved by FASB and appear on the XBRL taxonomy at the time of the filing. Extensions are tags created by companies when firms determine that the available taxonomy tags cannot accurately capture their company-specific economic transactions. *ARC* can be divided into two components, *ARC\_TAXONOMY* and *ARC\_EXTENSIONS*, together the two sum up to *ARC*. The dataset includes ARC and *ARC\_EXTENSIONS*. To construct *ARC\_TAXONOMY* you should subtract *ARC\_EXTENSIONS* from *ARC*.

***ARC\_PCT\_EXTENDED -*** is the percent of custom (extended) monetary tags in an SEC filing. It is calculated as the ratio of *ARC\_EXTENSIONS* to *ARC*.

**Important Considerations when using the data:** Users of the data should take note of the following. First, most smaller filers were not required to use XBRL to tag their financial statement notes until 2012, and hence, the use of data prior to 2012 should be limited to larger firms. The variable *ARC\_NOTES* (in the ARC files) can be used to aid researchers in cleaning the data. Second, while we provide ARC based on interim filings (e.g., 10-Q) in the ARC files, firms are not required to disclose all financial statement notes in these filings (see [Hoitash, Hoitash, Morris, and Yezegel 2021](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3786081)). Therefore, when aiming to capture annual ARC you should use ARC measures that are based on the 10-K (i.e, *Form* = '10-K') .

**Variable Definitions**

|  |  |
| --- | --- |
| **Variables About an EDGAR Submission** | |
| *ADSH* | Accession Number. The 20-character string formed from the 18-digit number assigned by the SEC to each EDGAR submission. |
| *CIK* | Central Index Key (CIK). Ten digit number assigned by the SEC to each registrant that submits filings. |
| *Fiscal\_Period\_End* | Fiscal period end (day, month, year) equivalent to the *DATADATE* in Compustat |
| *SIC* | Standard Industrial Classification (SIC). Four digit code assigned by the SEC as of the filing date, indicating the registrant's type of business. |
| *Filer\_Status* | The type of filer, accepting the following values: 1-LAF=Large Accelerated, 2-ACC=Accelerated, 3-SRA=Smaller Reporting Accelerated, 4-NON=NonAccelerated, 5-SML=Smaller, Reporting Filer, NULL=not assigned |
| *Form* | The type of form filed with the SEC (10-K, 10-Q, 20-F, 40-F and their respective amended filings) |
| *Date\_Filed* | The actual date of company filing |
| *Date\_Accepted* | The date the file accepted by the SEC (Filings accepted after 5:30pm EST are considered filed on the following business day.) |
| *NCIKS* | Number of registrants (i.e., business units) included in the consolidating entity's submitted filing. |
| **Measures of ARC** |  |
| *ARC* | The count of distinct (within each disclosure) monetary XBRL tags in an SEC filing |
| *ARC\_ALL\_NUMS* | The count of XBRL all tagged numbers in an SEC filing |
| *ARC\_ALL\_MONETARY* | The count of monetary XBRL tags in an SEC filing (allowing repeating tags) |
| *ARC\_UNIQUE\_MONETARY\_TAGS* | The count of unique monetary XBRL tags in an SEC filing |
| *ARC\_FACE* | The count of distinct (within each disclosure) monetary XBRL tags on the face (e.g., income statement, balance sheet) of the financial statements in an SEC filing |
| *ARC\_NOTES* | The count of distinct (within each disclosure) monetary XBRL tags in the financial statement notes in an SEC filing |
| *ARC\_IS* | The count of distinct monetary XBRL tags in the income statement in an SEC filing |
| *ARC\_BS* | The count of distinct monetary XBRL tags in the balance sheet in an SEC filing |
| *ARC\_CF* | The count of distinct monetary XBRL tags in the statement of cash flows in an SEC filing |
| *ARC\_EQ* | The count of distinct monetary XBRL tags in the statement of shareholder equity in an SEC filing |
| *ARC\_UN* | The count of distinct monetary XBRL tags in uncategorized financial statement in an SEC filing |
| *ARC\_ALL\_TEXTBLOCK\_TAGS* | The count of Textblock tags in an SEC filing |
| *ARC\_ALL\_FACTS* | The count of all facts (the sum of *ARC\_ALL\_NUMS* and *ALL\_TEXTBLOCK\_TAGS*) in an SEC filing |
| **Measures of ARC Extensions** |  |
| *ARC\_EXTENSIONS* | The count of distinct (within each disclosure) extended (custom) monetary XBRL tags in an SEC filing |
| *ARC\_ALL\_NUMS\_EXTENSIONS* | The count of all extended (custom) XBRL tagged numbers in an SEC filing |
| *ARC\_ALL\_MONETARY\_EXTENSIONS* | The count of extended (custom) monetary XBRL tags in an SEC filing (allowing repeating tags) |
| *ARC\_UNIQUE\_MONETARY\_EXTENSIONS* | The count of extended (custom) unique monetary XBRL tags in an SEC filing |
| *ARC\_FACE\_EXTENSIONS* | The count of extended (custom) distinct (within each disclosure) monetary XBRL tags on the face of the financial statements in an SEC filing |
| *ARC\_NOTES\_EXTENSIONS* | The count of extended (custom) distinct (within each disclosure) monetary XBRL tags in the financial statement notes in an SEC filing |
| **Measures of ARC Percent Extensions** |  |
| *ARC\_PCT\_EXTENDED* | The ratio of *ARC\_EXTENSIONS* to *ARC* |
| *ARC\_ALL\_NUMS\_PCT\_EXTENDED* | The ratio of *ARC\_ALL\_NUMS\_PCT\_EXTENDED* to *ARC\_ALL\_NUMS* |
| *ARC\_ALL\_MONETARY\_PCT\_EXTENDED* | The ratio of *ARC\_ALL\_MONETARY\_PCT\_EXTENDED* to *ARC\_ALL\_MONETARY* |
| *ARC\_UNIQUE\_MONETARY\_PCT\_EXTENDED* | The ratio of *ARC\_UNIQUE\_MONETARY\_PCT\_EXTENDED* to *ARC\_UNIQUE\_MONETARY\_TAGS* |