

# Technical Notes:

## Understanding the Industry Financial Ratios

---

### Description

The Industry Financial Ratios (IFR) that are included with the Industry Reports provide financial statement and ratio benchmarks within the industry context. These IFR values are designed to support industry research efforts that require an analysis of how a specific business's financial position compares with its peers. The NAICS industry-level context of the IBISWorld reports allows for granular comparison that can enhance analysis beyond data that may be available on public companies or with internal data. IFR data is produced using the IRS Statistics of Income as a base reference, then industry-level values are estimated using IBISWorld's database of industry information. The resulting set provides a time series of 72 values (statement items and ratios) that are segmented by size. This series supports analysis of individual years, changes, trends, and averages, which can support cross-sectional comparisons as well as efforts to normalize, evaluate and forecast business financials.

IFR data is produced using an algorithmic modeling process that produces estimates of industry-level data and forecasts data up to the latest year. Input data from the IRS provides the model with higher-level NAICS hierarchy values that act as the model foundations. This data includes corporate income tax returns sampled from the universe of total returns to the IRS, which provides a materially large sample set to underpin the model. Data is taken as is and converted to common-size terms for modeling purposes. IBISWorld's high-dimensional industry data is reduced to various key metrics that are related to the task of estimating financials. This data is part of IBISWorld's holistic approach to industry analysis, so information from various sources and the broad industry context influences this data. The resulting estimates are stated in common size terms and ratios are calculated from these estimates. The modeling process includes dynamic outlier handling as well as post-processing of the data. This model instance represents Industry Financial Ratios: Model Version 1.1.0. The model release uses three-part semantic versioning. The first place represents the data set and model algorithm. The second place represents the secondary parameters of the model, which include outlier thresholds and post-processing. The third place represents and patches for data mapping, missing values, front-end presentation fixes. Model 1.1.0 has relatively lenient outlier criteria in the modeling process with hard limits during the post-processing. To support a full data set where data is present in other cases, rolling averages have been used in post-processing to gap-fill missing data or data restricted during outlier handling.

## About the Data

The model input data can be segmented into two categories: Baseline Inputs and Industry Inputs.

### Baseline Inputs:

- IRS Statistics of Income
- US Census Quarterly Financial Report

### Industry Inputs:

- IBISWorld Industry Database
  - Economic Census
  - Annual Survey of Manufacturers
  - Annual Retail Trade Survey
  - Annual Wholesale Trade Survey
  - Annual Business Survey
  - Statistics of US Businesses
  - County Business Patterns
  - BLS Quarterly Census of Employment & Wages

The April 2021 Model Version 1.1.0 release includes data for the period 2006-2020 across 72 unique items and 11 asset size classes. This data is provided at the IBISWorld Industry detail level except for NN codes and the “492” sub-sector. The final dataset includes 9,725,475 year-industry-asset-item values. Then three averages are added to each IFR report value: a three-, five- and ten-year historical average. IFR values are presented as a final chapter in the industry report that serves as an appendix for financial value analysis. The data in this section is updated independently from the general industry report. While the data in the report is used at the time of estimation, the data publishing date will not always align. The IFR data is scheduled for annual updates in the Fall of each calendar year. Each update will include the latest data available from each source. Additionally, the modeling procedure will be tested and updated as appropriate. So, historical and latest year data are subject to change as new data comes in, a new target year is added, and the modeling process is updated. The size of the database makes it more likely than not that output errors will be present. These can be the result of inconsistencies in the inputs; cases of outlier handling at the margin, producing strange ratio outputs; and cases where the modeling algorithm is lacking for a particular tuple in the dataset. As a result of model estimates, no representation is made to the accuracy of the data set. No research has been performed to assess the accuracy of the data external to the training, validation,

and test procedures in the model development. The true population of which this model is providing estimates is unknown.

## Limitations

### *“All models are wrong; some are useful.” - George E.P. Box*

There are known limitations of the data that should inform how they are interpreted. The data is based on aggregated financials and each value in the set is estimated independently. So, each value should be interpreted as an expected value for that industry-year-asset size-financial item, rather than a representation of a financial statement for an average industry company. This process does not produce ranges or errors for each estimate, so each value must be taken as is. Rather than an upper and lower value, varied averages are provided to support the construction of bounded values. Cases where the initial IRS data is insufficient to produce a base value are removed. Then outliers are restricted from the data set during the estimation process. This creates cases where values will be missing in the final data set. This is done to remove unreliable values. However, there are cases where the estimation process produces values that exist within a theoretically allowable range, but may not be effective representations for benchmarking or other efforts.

Discretion should be taken when using the IFR as direct inputs into a model. In addition to scrutiny of the values, attention should be paid to the industry context and size. The dataset “Ratio Reference Count.csv” is available upon request and provides the base count of returns in the reference IRS data. Industries constructed from a smaller sample may be less representative of the true underlying population. In the case of industries defined by a small population, aggregated data is likely a poor source of comparison. Direct comparison with competitors is likely a sounder approach in these cases. Industry context and the context of the comparison target should also be considered. Industries with a few large key operators and a very large set of small operators may be misrepresented. Comparison targets are likely to have idiosyncratic factors that impact their results and may have operations that differ from an average industry operator. So, the specifics of the business must be considered in the context of benchmarking. In cases where IFR data differ significantly from expectations or displays an extreme value or change from year to year, the magnitude and direction may provide information value where the specific value may be lacking. An explosive value that approaches an outlier criterion, but is not restricted, may represent a flaw in the model. However, the magnitude and direction provide information about the underlying industry conditions and how they are expected to influence financial results. In addition to analysis of the magnitude and direction, the averages provided in the IFR reports should provide a baseline comparison to ground any other value.