Business and Finance Data Quality Problems:
What Economic Researchers Need to Know

Grace Liu
Assistant Professor, Business Librarian
West Chester University, FHG Library

November 2022
Beyond the Numbers
Data quality problems troubling business and financial researchers: A literature review and synthetic analysis

Grace Liu

Pages 315-371 | Published online: 19 Nov 2020

Abstract

The data quality of commercial business and financial databases greatly affects research quality and reliability. The presence of data quality problems can not only distort research results, destroy a research effort but also seriously damage management decisions based upon such research. Although library literature rarely discusses data quality problems, business literature reports a wide range of data quality issues, many of which have been systematically tested with statistical methods. This article reviews a collection of the business literature that provides a critical analysis on the data quality of the most frequently used business and finance databases including the Center for Research in Security Prices (CRSP), Compustat, S&P Capital IQ, I/B/E/S, Datastream, Worldscope, Securities Data Company (SDC) Platinum, and Bureau van Dijk (BvD) Orbis and identifies 11 categories of common data quality problems, including missing values, data errors, discrepancies, biases, inconsistencies, static header data, standardization, changes in historic data, lack of transparency, reporting time issues and misuse of data.
Common Data Quality Problems: 1. Missing Values

- It is the most prevalent data quality problem and is found in almost all examined databases.
- This happens more commonly to data reflecting complex accounting concepts, complex transactions, or data in financial statement footnotes.
- Excluding missing values from the sample may create omission or selection biases and misleading results.
Common Data Quality Problems: 2. Data Errors

- It can happen in many different ways, from typos and arithmetic errors to improperly including or excluding certain accounting items in the computation.
- Errors are more likely to happen to complex financial concepts, complex transactions, and to data in footnotes.
- It can happen due to misunderstanding or improper translation or interpretation.
- It can distort the research results.
Common Data Quality Problems: 3. Discrepancies

- It may result from differences in coverage, definitions, coding policies, classifications, etc. among databases.
- It can happen due to standardization or when databases choose different units, scales, or formats to represent the same firm’s data.
- It can lead to the “database effect.”
- It would be risky to solely rely on a single database for research.
4. Biases

- The concerns about biases in databases are widely discussed.
- Examples include selection bias, survivorship bias, incubation bias, etc.
- Biases are likely to occur to new firms, small firms, foreign firms, delisting firms, and firms with poor performance.
- Biases can distort empirical research results and produce unreliable research conclusions.
Common Data Quality Problems: 5. Inconsistencies

• Describe a situation where the data is not treated consistently with the same rule, particularly in the same database.

• Examples include the same accounting items defined differently in different financial statements or in different time periods.

• Databases from the same vendor may have different reporting policies; Information from different sections of a database may be inconsistent as well.

• Inconsistent data can create comparability problems.
Common Data Quality Problems: 6. Static Header Data

- Happens when the database only provides the latest available information and lacks time-series records.
- Happens mostly in the data field including “company name,” “ticker,” “address,” headquarter,” “ownership,” “stock exchange,” and “industry code.”
- It restricts researchers from including these data variables in time-series analysis.
- It may induce incorrect samples, omissions, or selection biases.
Common Data Quality Problems:
7. Standardization

- Make it possible to compare companies from different countries.
- It can understate or overstate its original financial outcome and affect the accuracy of certain prediction models.
- Extra efforts are needed to understand the standardization process and changes made to the original datasets.
Common Data Quality Problems: 8. Change in Historic Data

- Historical data can be revised due to the correction of errors or updates from restatement.
- Systematic changes in historic data can be a problem and it may raise concerns over data integrity.
Common Data Quality Problems: 9. Lack of Transparency

- Is a fundamental issue for many other data problems.
- Database vendors are often not transparent about their data collection and management practices and are reluctant to warn researchers about potential data problems and biases.
Common Data Quality Problems: 10. Reporting Time Issues

- It can be a problem when database providers are not transparent about their data reporting time and data update schedule.
- It can affect research when the dates that the data is available to the public are different from the date that a study assumes it to be.
- Improperly recorded reporting time can induce look-ahead bias or selection bias.
Common Data Quality Problems:  
11. Misuse of Data

- Researchers may sometimes improperly use the data as proxies or measurements.
- A database only covering public firms in an industry, is a poor proxy for calculating actual industry concentration and can lead to incorrect conclusions.
Practical Advice for Researchers

- Researchers should always evaluate data quality and check its accuracy and completeness.
- Researchers should not solely rely on a single data source and should err on the side of caution for the “database effect.”
- Researchers should understand how the data is defined or calculated, especially for standardized or adjusted data.
- Researchers should be cautious of using databases as a screening tool to identify data samples.
Practical Advice for Researchers

• Researchers need to seek proper procedures to mitigate the biases and clearly explain the impact of potential biases and the limitations of their research.

• Researchers should not treat data acquisition as a one-time transaction.

• Researchers need to be cognizant of the reporting time, reporting lag, update schedule, or embargo period of the data sources.

• Researchers should keep open communications with vendors on data problem.
What Should We Do?
Business and Finance Data Quality Problems: What Economic Researchers Need to Know

Grace Liu
Assistant Professor, Business Librarian
West Chester University, FHG Library
yliu@wcupa.edu

November 2022
Beyond the Numbers