Data Description Sheet

Evidence on the Use and Efficacy of
Internal Whistleblowing Systems

By Stephen R. Stubben and Kyle T. Welch

March 2020

1. Description of which author(s) handled the data and conducted the analyses

Both Stephen Stubben and Kyle Welch signed non-disclosure agreements and agreed to follow a procedure for securely managing the internal whistleblowing data provided by NAVEX Global in December of 2017. NAVEX securely provided an update to the dataset (with an earlier start date, later end date, and additional variables) on July 31, 2018.

The Corporate Research Project of Good Jobs First provided data on corporate fines from its Violation Tracker database to Stephen Stubben on April 15, 2019. Kyle Welch accessed MSCI's KLD scores and shared them with Stephen Stubben on May 18, 2019. All other data was directly accessed through Wharton Research Data Services (WRDS) by Stephen Stubben. Stephen Stubben wrote and executed all the SAS and Stata code for the analyses presented in the paper.

2. Detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments).

Data	Source	Most Recent Download
Internal whistleblowing reports	NAVEX Global	July 31, 2018
Corporate fines	Violation Tracker	April 15, 2019
KLD scores	MSCI (WRDS)	May 18, 2019
Accounting data	Compustat (WRDS)	August 3, 2019
Stock market data	CRSP (WRDS)	August 3, 2019
Institutional ownership	Thomson Financial (WRDS)	August 3, 2019
Board of directors	BoardEx (WRDS)	August 3, 2019
Litigation	AuditAnalytics (WRDS)	August 3, 2019
Internal control weaknesses	AuditAnalytics (WRDS)	August 3, 2019

3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors.

Proprietary data on internal whistleblowing reports was provided by NAVEX Global, and proprietary data on corporate fines was provided by the Corporate Research Project of Good Jobs First. Both Stephen Stubben and Kyle Welch signed non-disclosure agreements with the two data providers. We have privately provided the editors with the contact information of company representatives who can vouch for the data sharing. All other data were obtained from publicly available sources listed above.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper.

Sections 3.1, 3.2, and 3.3 of the published paper provide a comprehensive overview of the data sources, variable measurement, and models estimated. Further detail on each step taken to collect, process, and use the data can be found in the SAS and Stata code that accompanies this data sheet.

5. The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program.

The SAS and Stata code used to process the data and produce the tables that appear in the published paper accompany this data sheet in a separate attachment titled "JAR Code.pdf". The attachment includes an overview that explains how the code was used.

In accordance with the non-disclosure agreement we signed with NAVEX Global, we are not able to provide a list of identifiers for the firms in our sample.

6. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

The authors agree to maintain the data and programs used in this paper for the six-year time period.