

**JAR Data Policy Disclosure Datasheet for**  
**“The Information Externality of Public Firms’ Financial Information**  
**in the State-Bond Secondary Market”**

**by Stephanie F. Cheng**

*1. A description of which author(s) handled the data and conducted the analyses.*

Cheng aggregated all data and conducted the analyses.

*2. A 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). We recommend that more than one author is able to vouch for the stated source of the raw data.*

I obtained data necessary for my analyses from publicly available sources and subscribed data sets.

On Sep 9, 2018, I obtained public firms’ financial information (e.g., earnings and announcement dates) from Compustat to construct state-level aggregate earnings signals, public firms’ monthly stock prices from CRSP as a control variable, and treasury indexes from CRSP to calculate treasury-adjusted bond returns..

On Nov 29, 2018, I obtained firms’ historical headquartered states from the WRDS SEC analytics Suite.

On Dec 4, 2018, I downloaded the secondary market trading data for municipal bonds from MSRB, including trading dates, transaction prices, volumes, and types.

On Dec 14, 2018, I accessed Thomson Reuters SDC Platinum to identify state governments that issued municipal bonds between 2005 to 2017. I downloaded the 6-digit CUSIP numbers associated with these bond issuers. I used these CUSIP numbers to merge with MSRB secondary market data.

On Dec 18, 2018, I collected bond characteristics as well as underwriters and legal advisors’ locations from Bloomberg.

On Jan 5, 2019, I collected state governments’ financial statements information from Bloomberg.

On Dec 2, 2018, I collected states’ Standard and Poor’s (S&P’s) credit ratings from the Pew Charitable Trusts. I supplemented missing observations with hand-collected data from various sources such as state governments’ websites.

On Dec 4, 2018, I obtained statewide aggregate tax-return data (first available in 2007) from Powerlytics to construct measures of public firms’ presence.

On Nov 20, 2018, I collected firms' press releases data from RavenPack Analytics.

On Sep 10, 2018, I collected states' quarterly gross domestic product (GSP) and unemployment rate (UR) data from the U.S. Bureau of Economic Analysis (BEA).

On Nov 28, 2018, I collected state coincident index data from the Federal Reserve Bank of St. Louis ([fred.stlouisfed.org](http://fred.stlouisfed.org)).

On Oct 10, 2020, I collected analyst forecasts from the IBES to construct the earnings surprise measure.

On Oct 10, 2020, I collected statistics on the investor base of U.S. municipal bond securities from the Securities Industry and Financial Markets Association (SIFMA). Data were originally provided to the SIFMA by the Federal Reserve System, Financial Accounts of the United States.

Cheng vouches for the stated source of the raw data.

*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. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).*

All of the source data for this paper are either publicly available or available with subscriptions to Thomson Reuters SDC Platinum, WRDS (Compustat, CRSP, MSRB), Bloomberg, RavenPack Analytics, and Powerlytics.

*4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.*

I describe my data in Section 5 of the paper and provide further details under point (5) below.

*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 purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such*

*cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.*

I prepare my variables using SAS, and I run regression analysis using Stata.

In the SAS codes, I first aggregate firm-level earnings at the state-month level, by firms' historical headquartered states and reporting months. The earnings aggregation follows the approach employed by Kothari et al. [2006].

Second, I separately prepare control variables, including gross state product growth, unemployment rate growth, coincident index growth, credit rating, rating changes, and states' general and government fund data.

Third, I calculate state and territory general obligation bonds' buy-and-hold treasury-adjusted monthly returns. I obtain trading data from MSRB. I download the issuers' CUSIPs from Thomson Reuters SDC Platinum, and obtain bond-level characteristics from Bloomberg. The returns calculation follows the approach used by Easton, Monahan, and Vasvari [2009].

Finally, I merge all datasets and export the final dataset as a stata file. The "cusip.dta" provides the identifiers of the study.

In the Stata codes, I winsorize continuous variables at the 1st and the 99th percentiles. I conduct regression analysis, using time and cusip fixed effects.

Please refer to the codes for more details.

The Sas file "sas code.sas" provides the code used to prepare variables and merge datasets into the final dataset.

The Do file "stata code.do" provides the code used for analysis.

*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.*

Cheng will maintain all the data and program used in this study until at least 2027.

## REFERENCES

- EASTON, P.; S. MONAHAN; AND F. VASVARI. "Initial Evidence on the Role of Accounting Earnings in the Bond Market." *Journal of Accounting Research* 47 (2009): 721-766.
- KOTHARI, S.; J. LEWELLEN; AND J. WARNER. "Stock Returns, Aggregate Earnings Surprises, and Behavioral Finance." *Journal of Financial Economics* 79 (2006): 537-568.