

# **Economic Growth and Financial Statement Verification**

**Petro Lisowsky**

University of Illinois at Urbana-Champaign and  
Norwegian Center for Taxation  
lisowsky@illinois.edu

**Michael Minnis**

University of Chicago Booth School of Business  
michael.minnis@chicagobooth.edu

**Andrew Sutherland**

Massachusetts Institute of Technology Sloan School of Management  
ags1@mit.edu

## **Compliance with Data Policy for the Journal of Accounting Research**

January 2017

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

All authors were involved in the collection and analysis of the data. Data handling and analyses were performed as described in more detail below (see #2 in particular).

2. *A detailed description of how the raw data were obtained or generated, including data sources, the 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.*
  - Our main analyses are based on Risk Management Association's Annual Statement Studies (RMA). RMA collects financial statements from member banks each year and aggregates the data from these financial statements into its Annual Statements Studies. See the paper (Appendix A, specifically) for additional details. The aggregated data is available publicly from the RMA. Michael Minnis received approval from the RMA to confidentially receive the Annual Statement Studies data at the bank level. Minnis received the calendar year 2002 to 2010 data from the RMA in November 2013 and the 2011 data in December 2013. Andrew Sutherland performed all data analyses using this data.
  - We obtained bank-level information from the Federal Reserve Bank of Chicago dataset of Reports of Condition and Income ("Call Reports"). Sutherland downloaded the data in May 2014 via WRDS. Sutherland and an administrative assistant hand-matched the RMA bank names to the call report data to merge these two datasets.
  - Minnis privately received firm-level data from Sageworks in October 2008. Minnis handled and performed the related analyses with this data. See Minnis (2011—reference in the paper) for a complete description of the dataset.

- Petro Lisowsky obtained firm-level tax return data between February and June, 2013, from the Large Business & International Division of the IRS. The data use parent-level consolidated Forms 1120 (C corporations), 1120S (S corporations), and 1065 (partnerships and LLCs). Data from the main portion of the forms were combined with each firm's respective consolidated Schedules M-3 by an IRS data analyst for Lisowsky. The Schedules M-3 are only available for firms with assets of \$10 million or more that e-file their tax returns; the relevant GAAP and audit data are located on the Schedules M-3. Lisowsky performed all tax return data analyses.

All authors vouch for the stated sources of 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 agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).*

We have provided contact information to the editors for the RMA, Sageworks, and IRS datasets.

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 papers, we require information about subject eligibility and/or selection, as well as any exclusion criteria.*

We describe our data collection processes in various places throughout the paper for the various datasets used. Also, see #2 above and our Stata do-files.

5. *Prior to final acceptance of the paper, the computer program used to convert the raw data into the dataset used in the analysis plus a brief description that enables other researchers to use this program. Instead of the program, researchers can provide a detailed step-by-step description that enables other researchers to arrive at the same dataset used in the analysis. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the sample was formed, including the treatment of outliers, Winsorization, truncation, etc. This programming is in most circumstances not proprietary. However, we recognize that some parts of the data generation process may indeed be proprietary or otherwise cannot be made publicly available. In such cases, the authors should inform the editors upon submission, so that the editors can consider an exemption from this requirement.*

We use Stata to convert the raw data and perform all analyses (see, *LMS\_code\_RMA 2016-01-19.txt*, *LMS\_code\_Sageworks 2016-01-19.txt*, and *LMS\_code\_IRS 2016-01-19.txt* for do-files which processes the RMA data, Sageworks data, and tax data, respectively, and generates the respective tables). Some pre-processing of data (e.g., converting raw data to Stata datasets) occurred prior to using Stata, but the authors did all data manipulation as

recorded in the Stata do-files. The only exceptions are steps which are redacted to protect the confidentiality of the underlying data. These steps are minor, but are discussed in the do-files.

6. *Data and programs should be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.*

The authors will maintain the data for at least six years. The only exception is the IRS data. The contract between the IRS and Lisowsky stipulates the following clause: “The contractor certifies that the IRS data processed during the performance of this contract shall be completely purged from all data storage components of its computer facility and no output will be retained by the contractor at the time the IRS work is completed. If immediate purging of all storage components is not possible, the contractor certifies that any IRS data remaining on any storage component will be safeguarded to prevent unauthorized disclosures.” Lisowsky is continuing to work with the data set on one other project, so data access continues. However, once the other project is completed, the data must be purged. Lisowsky anticipates that access to this data set will be terminated by the end of 2017.