

**Does Social Capital Matter in Corporate Decisions?
Evidence from Corporate Tax Avoidance**

Iftekhar Hasan

Gabelli School of Business, Fordham University & Bank of Finland

Chun-Keung (Stan) Hoi

Saunders College of Business, Rochester Institute of Technology

Qiang Wu

Lally School of Management, Rensselaer Polytechnic Institute

Hao Zhang

Saunders College of Business, Rochester Institute of Technology

--Data Description Sheet--

Voluntary Compliance with Journal of Accounting Research Data Policy

July 10, 2017

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

The primary data collection and data analysis responsibilities were shared by Qiang Wu and Hao Zhang.

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.

The data sources and the years in which data were downloaded or obtained are as follows:

1. Social capital data were obtained from the Northeast Regional Center for Rural Development (NRCRD) at the Pennsylvania State University in 2013
2. Firm-level variables including tax avoidance and firm controls were obtained from Standard & Poor's Compustat database in 2014
3. County and state country controls data were obtained from Bureau of Economic Analysis and the US Census Bureau in 2014
4. Organ donation data were obtained from Organ Procurement and Transplantation Network in 2014
5. Religion data were obtained from The Association of Religion Data Archives in 2014
6. Corporate social responsibility data were obtained from Kinder, Lydenberg, Domini Research & Analytics (KLD) database in 2014
7. Corporate relocation information were collected from Securities and Exchange Commission (SEC) EDGAR 10-K filing in 2014
8. Corruption data were obtained from United States Department of Justice in 2015
9. Tax enforcement data (IRS auditing probability) were obtained from Syracuse university TRAC IRS database in 2015
10. Corporate state statutory tax rate were obtained from Council of State Governments Knowledge Center in 2015
11. Domestic state-level subsidies were obtained from Scott Dyreng's Exhibit 21 Data in 2015
12. International subsidies were obtained from Scott Dyreng's Exhibit 21 Data in 2013

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).

Not applicable

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.

A separate document titled, Detail Step by Step Description, provides detailed information and relevant codes for data processing and corresponding analyses performed. Below we provide a brief description of the steps.

1. We obtained company financial information for all publicly traded firms with headquarters located in US counties between 1990 and 2012 from Standard & Poor's Compustat database.
2. The Compustat source reports the latest location for which the firm's headquarter is located, creating a potential matching problem for firms that relocated their headquarters to other counties during the sample period. We obtained firms' historical headquarter addresses using electronic 10-K filings from the SEC Edgar database.
3. NRCRD provides data for four different years in 1990, 1997, 2005, and 2009 for US counties. We modified NRCRD data when it is necessary to do so and we estimated our own social capital indexes using factor analysis in each year in which NRCRD data are available. Section 3.2. (Social capital measure) and Appendix B (Constructing social capital measures) provide more detailed information on these procedures.
4. We estimated social capital indexes in 1990, 1997, 2005, and 2009 using NRCRD data. We filled in social capital indexes for the missing years using the estimated social capital index in the preceding year in which data are available; for example, we filled in missing values from 1991 to 1996 using the social capital index in 1990.
5. The dataset was created in two steps. First, we used county name of each firm's headquarter location to match county-year level data including social capital and county demographics with firm-level variables including tax avoidance and firm attributes. Second, we used state name and year to match the resulting data with state-level variable such as state statutory tax rates.
6. All continuous variables were winsorized, at the 1% and the 99% levels.
7. The final sample contains 63,807 firm-year observations for which data are available for all independent variables in the baseline regressions and at least one tax avoidance variable. It includes 8,702 unique firms with headquarters located in 823 unique US counties in the period 1990-2012.

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 used SAS to compile the raw data into datasets that are suitable for statistical analyses using Stata. We used Stata to perform the statistical analyses. Part 4 provides brief descriptions of these procedures. A separate document titled, Detail Step by Step Description, provides detailed information and relevant codes for data processing and corresponding analyses performed. Refer to the Excel file titled, GVKEY, for the sample of firms analyzed in this study.

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.

Qiang Wu and Hao Zhang will maintain the data and programs for at least six years.