Data Description Sheet

for

Labor Market Effects of Spatial Licensing Requirements: Evidence from CPA Mobility

Stefano Cascino, Ane Tamayo, and Felix Vetter

In this document, we provide additional information regarding the construction of our datasets and the sources of our data following the *Journal of Accounting Research* Data Policy. Please, find below a point-by-point response to the different questions asked.

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

Data collection and empirical analyses were carried out by Felix Vetter.

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.

We provide detailed information regarding data sources and access dates for each dataset used in our paper below. In addition, we provide detailed information regarding the collection of data for our control variables. This data description sheet complements variable definitions and sample construction information presented in the paper's Appendix and in Section 3 of the Online Appendix.

QCEW State-Level Dataset

This dataset is based on Bureau of Labor Statistics' Quarterly Census of Employment and Wages (available at: https://www.bls.gov/cew/datatoc.htm). To compute control variables, we obtained data on GDP from the Bureau of Economic Analysis (available at: https://www.bea.gov/national/index.htm#gdp) and state-level unemployment rates from the Bureau of Labor Statistics Local Area Unemployment Statistics program (available at: https://www.bls.gov/lau/#tables). To compute our migration controls, we obtained data from the American Community Survey (available at: https://usa.ipums.org/usa/). We obtained all these data in June 2019. We provide detailed sample construction information in Section 3.1 of the Online Appendix.

SUSB State-Level Dataset

This dataset is based on Census' Statistics of U.S. Businesses (available at: https://www.census.gov/programs-surveys/susb.html). To compute control variables, we obtained data on GDP from the Bureau of Economic Analysis (available at: https://www.bea.gov/national/index.htm#gdp) and state-level unemployment rates from the Bureau of Labor Statistics Local Area Unemployment Statistics program (available at: https://www.bls.gov/lau/#tables). To compute our migration controls, we obtained data

from the American Community Survey (available at: https://usa.ipums.org/usa/). We obtained all these data in June 2019. We provide detailed sample construction information in Section 3.2 of the Online Appendix.

QCEW Border-County Dataset

This dataset is based on Bureau of Labor Statistics' Quarterly Census of Employment and Wages (available at: https://www.bls.gov/cew/datatoc.htm). To identify individual border segments, we downloaded numerical border segment identifiers from Thomas Holmes' website (available at: http://users.econ.umn.edu/~holmes/data/BorderData.html). To identify neighboring counties, we accessed Census County Adjacency Files (available at https://www.census.gov/geo/reference/county-adjacency.html). We augmented these data with county-level unemployment rates provided by the Bureau of Labor Statistics Local Area Unemployment Statistics program (available at: https://www.bls.gov/lau/#tables). We obtained all these data in June 2019. We provide detailed sample construction information in Section 3.3 of the Online Appendix.

QCEW MSA-level Dataset

This dataset is based on Bureau of Labor Statistics' Quarterly Census of Employment and Wages (available at: https://www.bls.gov/cew/datatoc.htm). We obtained data on MSA-level GDP from the Bureau of Economic Analysis (available at: https://www.bea.gov/national/index.htm#gdp). We obtained these data in June 2019. We provide detailed sample construction information in Section 3.4 of the Online Appendix.

AICPA MAP Survey Dataset

This dataset is based on AICPA MAP Survey Sheets obtained from the AICPA. We provide information about our contacts at the AICPA to the editors as stipulated under point 3 of the *Journal of Accounting Research's* Data Policy. We obtained the raw AICPA MAP Survey data from the AICPA in May 2017. We provide detailed sample construction information in Section 3.5 of the Online Appendix.

Private Pension Plan Audit Dataset

This dataset is based on private pension plan data obtained from the Employee Benefits Security Administration of the Department of Labor (available at: https://www.dol.gov/agencies/ebsa/researchers/data/private-pension-plan-data). We obtained the raw data in April 2018. We provide detailed sample construction information in Section 3.6 of the Online Appendix.

AICPA Misconduct Dataset

Our AICPA Misconduct Dataset is based on data provided by Jack Armitage and Shane Moriarity. We obtained misconduct data from Jack Armitage and Shane Moriarity in June 2017. We provide contact information to the editors as stipulated under point 3 of the *Journal of Accounting Research's* Data Policy.We provide detailed sample construction information in Section 3.7 of the Online Appendix .

EBSA Deficient Filer Dataset

This dataset is based on EBSA Enforcement Data provided by the Department of Labor (available at: https://enforcedata.dol.gov/views/data_catalogs.php). We obtained these

data in April 2018. We provide detailed sample construction information in Section 3.8 of the Online Appendix.

CPA Firm Disciplinary Action Dataset

This dataset is based on CPA firm license data collected from the Colorado State Board of Accountancy (available at: https://www.colorado.gov/pacific/dora/Accountancy) following Vetter (2020). We obtained these data in June 2019. We provide detailed sample construction information in Section 3.9 of the Online Appendix.

Additional Data Sources

In addition to the above-mentioned data sources, we collected the following data for our CPA Mobility adoption prediction model (Section 2.2, Table 1, Panel B). We list data source and access date for each of the predictor variables in our prediction model below.

Variable	*Source and Access Date*
$CPABoardMembers_s$	This variable is based on survey data obtained from Gary Colbert and Dennis Murray. We obtained these survey data in May 2018.
$LocalCPABoardMembers_s$	This variable is based on survey data obtained from Gary Colbert and Dennis Murray. We obtained these survey data in May 2018.
$MobilityTaskForce_s$	This variable is based on hand-collected data from State Board of Accountancy's' websites and statutes. We obtained these data in May 2018.
$FundingAutonomy_s$	This variable is based on survey data obtained from Gary Colbert and Dennis Murray. We obtain these survey data in May 2018.
$FirmBirth_{s,t-1}$	This variable is based on Census Business Dynamics Statistics data (available at: https://www.census.gov/ces/ dataproducts/bds/data_estab2015.html), which was accessed in May 2018.
$JobBirth_{s,t-1}$	This variable is based on Census Business Dynamics Statistics data (available at: https://www.census.gov/ces/ dataproducts/bds/data_estab2015.html), which was accessed in May 2018.
$SenateDemocrats_{s,t-1}$	This variable is based on hand-collected data from the Book of States Archive at the Council of State Governments (available at: http://knowledgecenter.csg.org/kc/category /content-type/bos-archive), which was accessed in May 2018.
$HouseDemocrats_{s,t-1}$	This variable is based on hand-collected data from the Book of States Archive at the Council of State Governments (available at: http://knowledgecenter.csg.org/kc/category /content-type/bos-archive), which was accessed in May 2018.

(continued)

(continued)

Variable	*Source and Access Date*
$HouseDemocrats_{s,t-1}$	This variable is based on hand-collected data from the Book of States Archive at the Council of State Governments (available at: http://knowledgecenter.csg.org/kc/category /content-type/bos-archive), which was accessed in May 2018.
$BillsIntroduced_{s,t-1}$	This variable is based on hand-collected data from the Book of States Archive at the Council of State Governments (available at: http://knowledgecenter.csg.org/kc/category /content-type/bos-archive), which was accessed in May 2018.
$BillsEnacted_{s,t-1}$	This variable is based on hand-collected data from the Book of States Archive at the Council of State Governments (available at: http://knowledgecenter.csg.org/kc/category /content-type/bos-archive), which was accessed in May 2018.

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 obtained proprietary data from the AICPA on May 8, 2017. We received these data from Shelly Guzetta (email: Shelly.Guzzetta@aicpa-cima.com). We have signed a data sharing agreement with the AICPA that prevents us from publicly sharing these data. In addition, we agreed to report aggregate figures only.

We obtained AICPA misconduct case data from Jack Armitage and Shane Moriarity on June 30, 2017. We received these data from Jack Armitage (email: jarmitage@unomaha.edu). We agreed not to share these data publicly.

We obtained survey data on State Board of Accountancy member characteristics from Gary Colbert and Dennis Murray on May 14, 2018. We received these data from Gary Colbert (email: Gary.Colbert@ucdenver.edu). We agreed not to share these data publicly.

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 provide detailed sample construction procedures in the paper's Appendix and in Section 3 of the Online Appendix.

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 our raw data into the datasets described under point 2 of this data description sheet and to perform our empirical analyses. The Stata program "00_CPAMob_Master.do" calls individual Stata programs that, for each set of empirical tests, separately prepare and combine the raw data to form estimation samples and produce the results of our empirical analysis. We provide the mapping between individual Stata programs, datasets, and empirical results presented in the paper in "00_CPAMob_Master.do". The identifiers for our QCEW and SUBS datasets are presented in the "QCEW_indentifiers_counties.txt", "QCEW_indentifiers_msa.txt", "QCEW_indentifiers_state.txt", and "SUSB_indentifiers_state.txt" 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.**

   We will maintain all data and programs for at least six years.