

Data Sources

Prof. Tzu-Ting Yang

楊子霆

Institute of Economics, Academia Sinica

中央研究院經濟研究所

March 5, 2023

Data Sources

- Survey data
- Administrative data
- Firm-level data
- Country-level data
- Government Open data
- Geographical data
- Journal data archives
- Other data source

Survey Data: Cross-sectional Data

Survey Data

- General information for Taiwan survey data
 - Survey Research Data Archive (Academia Sinica): <https://srda.sinica.edu.tw/>
- General information for US survey data
 - IPUMS (University of Minnesota): <https://www.ipums.org/>

Survey Data

- Taiwan (cross-sectional data)
 - Population and Housing Census: https://srda.sinica.edu.tw/datasearch_detail.php?id=205
 - Manpower Utilization Survey: https://srda.sinica.edu.tw/datasearch_detail.php?id=2826
 - Survey of Family Income and Expenditure: https://srda.sinica.edu.tw/datasearch_detail.php?id=2850
 - Taiwan Social Change Survey: https://srda.sinica.edu.tw/datasearch_detail.php?id=2822

Population and Housing Census

- **Website:**

https://srda.sinica.edu.tw/datasearch_detail.php?id=205

- **Observation unit:** individual

- **Frequency:** 1960-2010, every 10 year

- **Key variables:** age, occupation, living township, marital status, disability status, housing size

- It does not have English version.

- Only faculty member at Academia Sinica can download it. Let me know if you want to use it

Manpower Utilization Survey

- **Website:**

https://srda.sinica.edu.tw/datasearch_detail.php?id=2826

- **Observation unit:** individual (above age 15)

- **Frequency:** 1976-now, every year (May)

- **Key variables:** age, occupation, industry, monthly wage, working hour, part-time job, firm size, number of children by children's age, employment status, unemployment reason

- Labor force survey

- It has English version.

Survey of Family Income and Expenditure

- **Website:** https://srda.sinica.edu.tw/datasearch_detail.php?id=2850
- **Observation unit:** household
- **Frequency:** 1976-now, every year
- **Key variables:** household income, household income by family members, household expenditure
- It has English version.

Taiwan Social Change Survey

- **Website:** https://srda.sinica.edu.tw/datasearch_detail.php?id=2822
- **Observation unit:** individual
- **Frequency:** 1980-now, every year
- **Key variables:** marital status, occupation, industry, monthly wage, birth year, birth month, many questions about an respondent's opinion on politics, social issue, religion
- It has English version.

Survey Data

- United States (cross-sectional data)
 - US Census Data: <https://usa.ipums.org/usa>
 - Current Population Survey (CPS): <https://cps.ipums.org/cps>
 - American Time Use Survey (ATUS): <https://www.atusdata.org/atus>
<https://www.ahtusdata.org/ahtus>
 - National Health Interview Survey (NHIS): <https://ihis.ipums.org/ihis>
 - IPUMS Higher Education Survey: <https://highered.ipums.org/highered>

Current Population Survey (CPS)

- **Website:**
<https://cps.ipums.org/cps>
- **Observation unit:** individual and household
- **Frequency:** 1962-now, every year
- **Key variables:** age, occupation, marital status, migration, wage, disability, tax, health insurance, education

American Time Use Survey (ATUS)

- **Website:**
<https://www.atusdata.org/atus>
<https://www.ahtusdata.org/ahtus>
- **Observation unit:** activity, individual and household
- **Frequency:** 2003-now, every year
- **Key variables:** time use on each activity
- Select one family member in CPS household and ask them about their time use (24 hour) on specific date
- This survey can link with CPS to get more information

National Health Interview Survey (NHIS)

- **Website:**
<https://ihis.ipums.org/ihis>
- **Observation unit:** individual and household
- **Frequency:** 1960-now, every year
- **Key variables:** health condition, mortality, health behavior, healthcare use

IPUMS Higher Education Survey

- **Website:**
<https://ihis.ipums.org/ihis>
- **Observation unit:** individual and household
- **Frequency:** 1993-now, every year
- **Key variables:** education, employment, income, work history

Survey Data

Developed Countries

- Other countries (cross-sectional data)
 - Census data from other countries: <https://international.ipums.org/international/>
 - North Atlantic Population Project: <https://www.nappdata.org/napp/>
 - Census microdata from Canada, Denmark, Great Britain, Germany, Iceland, Norway, Sweden, and the United States from 1703 to 1911
 - Multinational Time Use Study: <https://www.mtusdata.org/mtus/>

Survey Data

Developed Countries

- Other countries (cross-sectional data)
 - Luxembourg Income Study Database (LIS)
<http://www.lisdatacenter.org/our-data/lis-database/>
 - The largest available income database of harmonised microdata collected from about 50 countries spanning five decades.
 - Codebook:
<http://www.lisdatacenter.org/wp-content/uploads/files/data-lis-variables.pdf>
 - Labour income, capital income, pensions, public social benefits (excl. pensions) and private transfers, as well as taxes and contributions, demography, employment, and expenditures

Survey Data

Developed Countries

- Other countries (cross-sectional data)
 - Luxembourg Wealth Study Database (LWS)
<http://www.lisdatacenter.org/our-data/lws-database/>
 - The first cross-national wealth database in existence. Harmonized into a common framework
 - Codebook:
<http://www.lisdatacenter.org/wp-content/uploads/files/data-lws-variables.pdf>
 - Assets and debt, market and government income, household characteristics, labour market outcomes and, in some datasets, expenditures and behavioural indicators

Survey Data

Developing Countries

- Other countries (cross-sectional data)
 - World Bank Survey database:
<http://microdata.worldbank.org/index.php/home>
 - Demographic and Health Survey (DHS):
<https://dhsprogram.com/data/>
 - DHS collect and disseminate accurate, nationally representative data on health and population in developing countries.

Survey Data

Standardized Test Data

- Other countries (cross-sectional data)
 - PISA test database:
<http://www.oecd.org/pisa/data/>
 - TIMSS test database:
<https://timssandpirls.bc.edu/>
 - PISA and TIMSS are databases for international comparative assessments of student achievement

Survey Data: Panel Data

Survey Data

- Taiwan (panel data)
 - Panel Survey of Family Dynamics (PSFD): http://psfd.sinica.edu.tw/web/data_en.htm
 - Taiwan Education Panel Study (TEPS): https://srda.sinica.edu.tw/datasearch_detail.php?id=2755

Panel Survey of Family Dynamics (PSFD)

- **Website:**
http://psfd.sinica.edu.tw/web/data_en.htm
- **Observation unit:** follow the same households over time and follow their offspring's households over time
- **Frequency:** 1999-now, every year
- **Key variables:** household income, education, working experience, marriage, family value and attitudes, household expenditure, child care
- It has English version.

Taiwan Education Panel Study (TEPS)

- **Website:** https://srda.sinica.edu.tw/datasearch_detail.php?id=2755
- **Observation unit:** follow the same students from their junior/senior high school and also have questions for their parents, teachers, and schools
- **Frequency:** 2001-2007 and 2010-2014
- **Key variables:** test score, friends, part-time work, learning attitude, extracurricular activities, wage
- It has English version.

Survey Data

- United States (panel data)

- Panel of Income Dynamics (PSID): <http://psidonline.isr.umich.edu>
- Health and Retirement Study (HRS): <https://hrs.isr.umich.edu/data-products>
- National Longitudinal Surveys (NLS): <https://www.nlsinfo.org/content/getting-started/accessing-data>
- Survey of Income and Program Participation (SIPP): <http://www.nber.org/data/survey-of-income-and-program-participation.html>
- Medical Expenditure Panel Survey (MEPS): <https://www.meps.ahrq.gov/mepsweb/>

Survey Data

- Other countries (panel data)
 - Cross-National Equivalent File (CNEF): <https://www.cnefdata.org/about/about-cnef>
 - The Cross-National Equivalent File is a user package of equivalently defined variables for 9 general population household panel surveys.
 - Codebook:
<https://www.cnefdata.org/documentation/codebooks>

Administrative Data

Administrative Data

- House transaction data (Taiwan): <http://lvr.land.moi.gov.tw/homePage.action>
 - I have bought this dataset. Let me know if you want to use this data for your term paper
- Data on mortgages (US): <https://www.fhfa.gov/DataTools/Downloads/Pages/Public-Use-Databases.aspx>

Administrative Data

- National Health Insurance Research Database (Taiwan): <http://nhird.nhri.org.tw/en/index.html>
 - It is a confidential dataset. Let me know if you want to use this data for your term paper
- Tax Return Data (Taiwan): <http://www.econ.sinica.edu.tw/FIA-Data>
 - It is a confidential dataset. I do not recommend you to use this data for your term paper because you need to spend a lot of time applying for it.

Administrative Data

- Sports Reference (US): <https://www.sports-reference.com/>
 - This website contains many statistics about sports

Administrative Data

- Yelp Open Dataset (US): <https://www.yelp.com/dataset>
 - Yelp is a company, which publishes crowd-sourced reviews about businesses
 - A subset of data on Yelp's businesses, reviews, and users

Administrative Data

- Google Mobility Data: <https://www.google.com/covid19/mobility/>
 - The data shows how visits to places, such as grocery stores and parks, are changing in each geographic region during COVID-19 pandemic

Administrative Data

- Apple Mobility Data: <https://covid19.apple.com/mobility>
 - Similar to Google Mobility Data

Firm-Level Data

Firm-Level Data

- Taiwan Economic Journal (TEJ): <http://www.finasia.biz/ensite/>
 - TEJ is a database of financial and market information on **listed company at stock exchange market** in Taiwan, Hong Kong or China
 - Business school should have bought this database
- Compustat (US): <https://marketintelligence.spglobal.com/client-solutions/>
 - Compustat is a database of financial and market information on active and inactive global companies throughout the world
 - Business school should have bought this database

Country-level data

Country-level data

- Penn World Table version 9.0:
<https://www.rug.nl/ggdc/productivity/pwt/>
- World Wealth & Income Database
<http://wid.world/>
- Work Bank Databank:
<https://databank.worldbank.org/data/home.aspx>
- OECD data
<https://data.oecd.org/>
- eurostat data
<https://ec.europa.eu/eurostat/data/database>
- FRED Economic data
<https://fred.stlouisfed.org/tags/series>

Government open data

Government open data

- US open data:
<https://www.data.gov/local>
- Taiwan open data:
<https://data.gov.tw/>
- UK open data:
<http://www.data.gov.uk>
- Japan open data:
<https://www.data.go.jp/>

Geographical data

Geographical data

- GADM

<https://gadm.org/index.html>

- It provides map (shapefile) for administrative areas of all countries, at all levels of sub-division

- Geographically based Economic data

<https://gecon.yale.edu/g-econ-project-yale-university->

- It develops a geophysically scaled economic data set

Geographical data

- Google Earth Engine

<https://earthengine.google.com/>

- It collects settlight image data from the following data sources

- Landsat

<http://landsat.usgs.gov/>

- Settligh image for land use, land cover, and land condition since 1972

- MODIS

<https://modis.gsfc.nasa.gov/>

- Settligh image for atmosphere or ocean since 1999

Geographical data

- DMSP-OLS

- <https://eogdata.mines.edu/products/dmsp/>

- Settlight image for nightlight from 1992-2013 (annual data)

- VIIRS

- <https://eogdata.mines.edu/products/vnl/>

- Settlight image for nightlight since 2012 (monthly data)

Journal Data Archives

Journal Data Archives

- American Economic Review:
<https://www.aeaweb.org/journals/aer/issues>
- AEJ Economic Policy:
<https://www.aeaweb.org/journals/pol/issues>
- AEJ Applied Economics:
<https://www.aeaweb.org/journals/app/issues>

Other data source

- Goodle Trend:

<https://trends.google.com/trends/>

- Kaggel:

<https://www.kaggle.com/datasets>

- Jerry Smith dataset collection:

<https://datascientistinsights.com/2013/02/02/data-monet>

- Our World in Data:

<https://ourworldindata.org/>