Silvia Mendolia

Labour market, health and well-being: economic analysis using panel data

Plans for today

- What is Panel Data?
- Why is panel data relevant in labour/health economics?
- Cross section vs Panel data
- Important Longitudinal Household Surveys
- Empirical questions that can be addressed using panel data
- Balanced and Unbalanced Panel Data
- Using Stata to describe Panel Data
- Time to start organising your group projects

Useful references

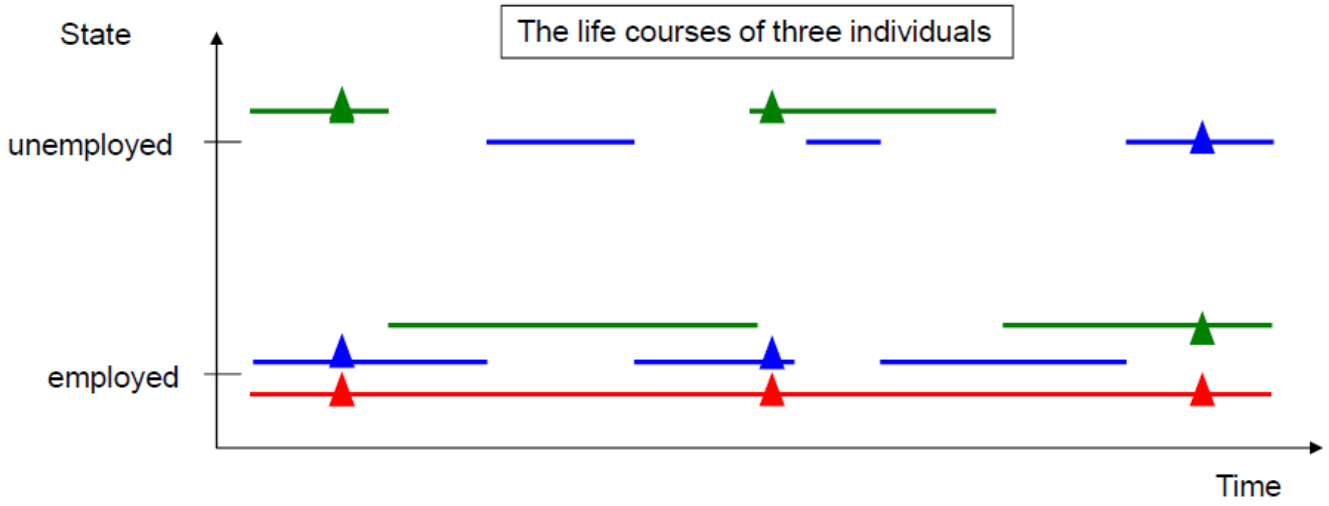
- Cameron, A. Colin and Pravin K. Trivedi (2010). Applied Microeconometrics Using Stata. Stata Press.

- Recommended introductory textbooks that provide an introduction to panel data analysis are:
  - Wooldridge, Jeffrey M. (2015). Introductory Econometrics. Cengage Learning Services
  - Kennedy, Peter (2008). A Guide to Econometrics. John Wiley & Sons, 6th edition.

What is Panel Data?

- A cross-section (of people, firms, countries, etc.) is observed over time.
- Panel data provides observations on the same units in several time periods (unlike independently pooled cross sections).
-
- Another word for this: Longitudinal data

Panel Data vs. other data



Source: Brüderl/Ludwig, Panel Analysis, April 2019

- Cross-sectional data
  - "Snapshot" at one time point
- Panel data
  - Repeated measurement
- Event history data
  - Information on the complete life course

Cross section and panel data

- **Cross-sectional research design**
  - For identifying the causal effect of a treatment, we compare the outcome of people in the treatment group with the outcome of different people in the control group
  - We call this design "**between estimation**"

- **Panel data research design**
  - For identifying the causal effect of a treatment, we investigate how the outcome changes, if the same people change from the control to the treatment condition over time.
  - We call this design "**within estimation**"

The origins of Panel Data

- **Princeton "radio project" (1937-1939)**
- Research question: Effect of radio ownership on political attitudes- Will the Americans become communist?
- Inference from cross-sectional (control group) or panel data?

- "Most of the control groups available for social research are 'self-selected'."

- "If we give radios to a number of farmers and then notice considerable differences without any great external changes occurring at the same time, it is safer to assume that these differences are caused by radio than it would be, if we were to compare radio owners with non-owners."

Lazarsfeld/Fiske (1938) The "panel" as a new tool for measuring opinion. Public Opinion Quarterly 2: 596-612.

Panel data in Macro/Micro Economics

- Use in Macroeconomics:
  - Unit of analysis: countries
  - N small, T large
  - Cross-sectional time series

- Use in Microeconomics
  - Unit of analysis: persons
  - N large, T small
  - (Micro) panel data
- **This class is about Micro panel data**

How does Panel Data look?

Panel data (also known as longitudinal or cross sectional time-series data) is a dataset in which the behavior of entities are observed across time.

- It looks like this!

country	year	Y	X1	X2	X3
1	2000	6.0	7.8	5.8	1.3
1	2001	4.6	0.6	7.9	7.8
1	2002	9.4	2.1	5.4	1.1
2	2000	9.1	1.3	6.7	4.1
2	2001	8.3	0.9	6.6	5.0
2	2002	0.6	9.8	0.4	7.2
3	2000	9.1	0.2	2.6	6.4
3	2001	4.8	5.9	3.2	6.4
3	2002	9.1	5.2	6.9	2.1

Source: Reyna (2007) Panel Data Analysis Fixed and Random Effects using Stata

A panel of individuals

- Units are people
- Time is years
- A row is “person-year”

Notation: y_{it}

$i = 1, \ldots, N$: units

$t = 1, \ldots, T$: time

Example: panel data with $T = 2$

id	time	Y	X
1	1	y_{11}	x_{11}
1	2	y_{12}	x_{12}
2	1	y_{21}	x_{21}
2	2	y_{22}	x_{22}
⋮	⋮	⋮	⋮
N	1	y_{N1}	x_{N1}
N	2	y_{N2}	x_{N2}

Source: Brüderl/Ludwig, Panel Analysis, April 2019

Panel Data in Labour/Health Economics

- Panel data allows you to control for variables you cannot observe or measure like:
  - cultural factors, psychological traits, ability, attitudes, values
- These factors are often called "**unobserved heterogeneity**" and can cause bias in the estimation with cross-sectional data

Panel Data in Labour/Health Economics

- With panel data you can include variables at different levels of analysis (i.e. families, students, schools, districts, states) suitable for multilevel or hierarchical modeling.

- Panel data allows to analyse changes within a small unit such as a family, a class, a school

Panel Data in Labour/Health Economics

- Panel data allows to study individual trajectories:
  - in/out labour market; development of health issues; changes in structure of family

Panel Data in Labour/Health Economics

- Panel data allows to look at the effects of earlier life circumstances on later outcomes in a person's life (e.g. impact of parental job loss on later outcomes in adult life)

- Panel data usually has a large size and it allows to investigate differences between regional areas (eg determinants of COVID-19 vaccination take up in different areas of the UK)

Important Longitudinal Surveys

Household panels

- Panel Study of Income Dynamics (PSID) [since 1968]
  - The role model for all household panels
- German Socio-Economic Panel (SOEP) [since 1984]
- Understanding Society (UKHLS) [since 1991]
- Household Income and Labour Dynamics in Australia (HILDA) [since 2001]

Cohort panels

- British Cohort Studies: children born 1958, 1970, 2000
- National Longitudinal Survey of Youth (NLSY79): U.S. cohort born around 1960

Important Longitudinal Surveys

Panels on special populations:

- Survey of Health, Ageing and Retirement in Europe (SHARE)
- Millenium Cohort Study (MCS) [since 2000]
- Next Steps (LSYPE) [since 2004]
- English Longitudinal Study of Ageing (ELSA)[since 2002]
- Longitudinal Study of Australian Children (LSAC) [since 2003]
- Invalsi (Italian students in primary and secondary schools) [since 2009]

Online panel surveys

- – LISS panel: A Dutch online panel survey
- – German internet panel (GIP)
- – GESIS Panel

Important Longitudinal Surveys

- Understanding Society (UKHLS)
- https://www.understandingsociety.ac.uk/
- 40,000 households living in the UK
- 11 waves available (from 2009-2010)
- Can be linked to BHPS (data since 1991)
- Topics: family, health, labour market, social issues, values, and more!
- Special COVID-19 survey: monthly from April 2020

Important Longitudinal Surveys

- Next Steps
- https://cls.ucl.ac.uk/cls-studies/next-steps/
- Representative of English teenagers (born in 1989/1990)
- Year 9 in 2004 – 15,000 young people from more than 700 schools
- Parents and children are interviewed in the first 3 waves
- Wave 8: age 25 (early adulthood)

Important Longitudinal Surveys

- Invalsi
- https://invalsi-serviziostatistico.cineca.it/
- Data on Italian students in grade 2, 5, 8 and 10
- Recently organised as panel data
- Detailed information on students' achievements, socio-economic status, and some parents' characteristics
- Data on achievements and attitudes
- Teachers' questionnaires

How to get to know your data?

- Read, read, read!
- Read academic papers using the dataset you are interested in
- Read policy reports summarising the data
- Check the relevant websites for documentation and data dictionary
- Read the questionnaires and think of interesting variables

Useful resources on Understanding Society

- Publications catalogue:
  - https://www.understandingsociety.ac.uk/research
- Key variables:
  - https://www.understandingsociety.ac.uk/content/understanding-society-key-variables
- Questionnaires:
  - https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires

Empirical questions

- Can we think of examples of empirical questions in the areas of health and labour economics that can be addressed using longitudinal data?
- From the literature:
  - Impact of job changes on life satisfaction
  - Impact of commuting on health
  - Impact of childhood adverse conditions on economic outcomes
  - ….

Empirical questions

- Think of interesting variables in the areas of health and labour market
- Think of how different conditions/situations can be correlated
- Think of impact of specific events on individuals' lives trajectories
- Think of the policy implications of your research question

Empirical questions – Some examples

LABOUR MARKET

- Unexpected job loss
- Job insecurity
- Labour supply within the family
- Early/Late retirement
- Overtime work/ Hours of work
- Part time/Full time
- Job satisfaction

HEALTH

- Fertility choices
- Mental health
- Life satisfaction
- Physical health limitations
- Chronic conditions
- Family members' health
- Smoking
- Alcoholism

Why are these research questions important?

- All these questions have important policy implications

- The identification of life events that have a large and significant impact on health and, more generally, on individual and family well-being may be useful in the **elaboration of policies** that focus on the occurrence of such events

- **Specific care may be intensified if such events are observed**

Example: L market and health (1/2)

- What are the main transmission channels of the effects?

- There are various transmission channels underlying the relationship between health and labour market

  - **Income effect**: employment is a provider of income, so this analysis is strictly linked with the analysis of the relationship between health and income

Example: L market and health (2/2)

- Employment is a provider of **social relationship and identity in the society**
- Working conditions affect people other than the individual concerned
- The balance between **work and family life** is an important determinant of individual and family well-being

Balanced and Unbalanced panel

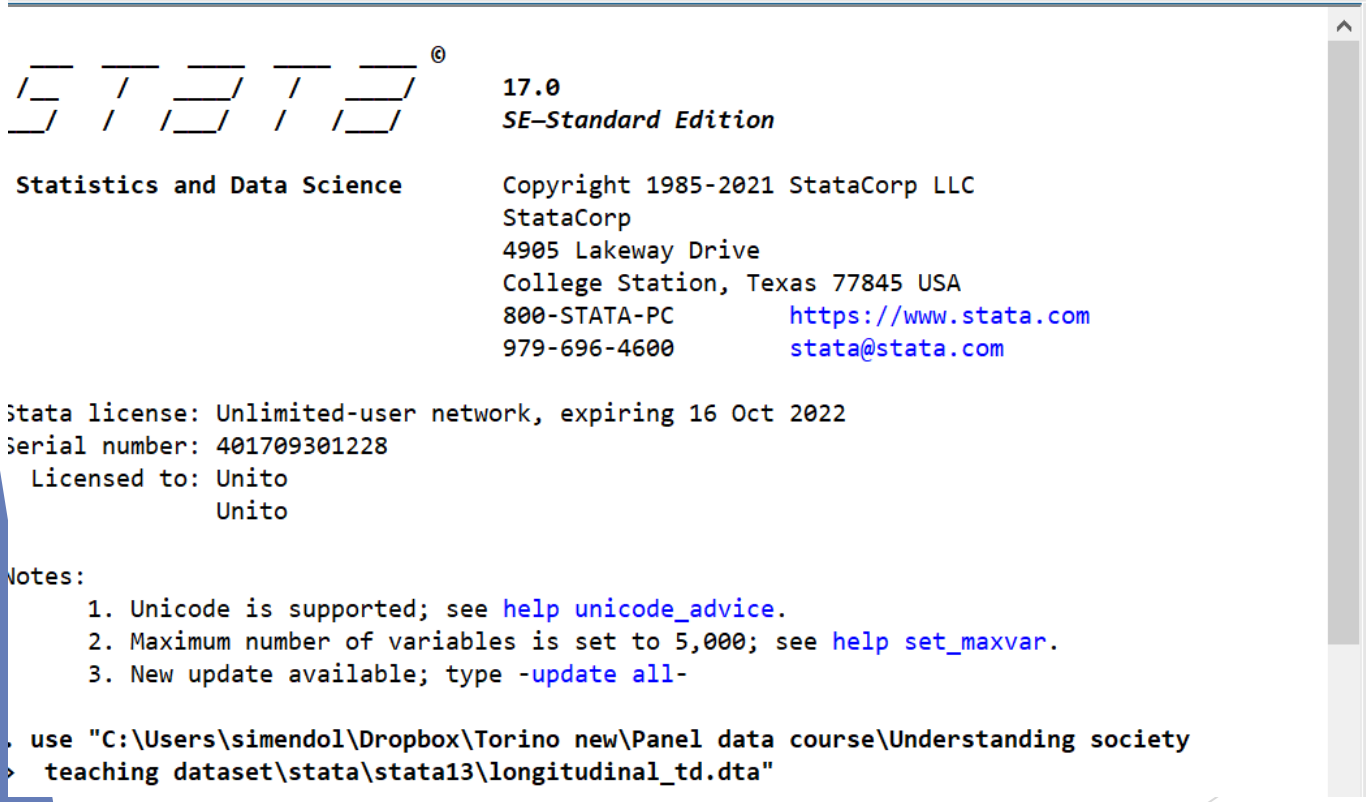
- Balanced panel: all individuals are in the dataset at all waves
- Unbalanced panel: individuals are in the dataset for a different number of waves
- A moderate level of attrition is normal, but high levels can be problematic

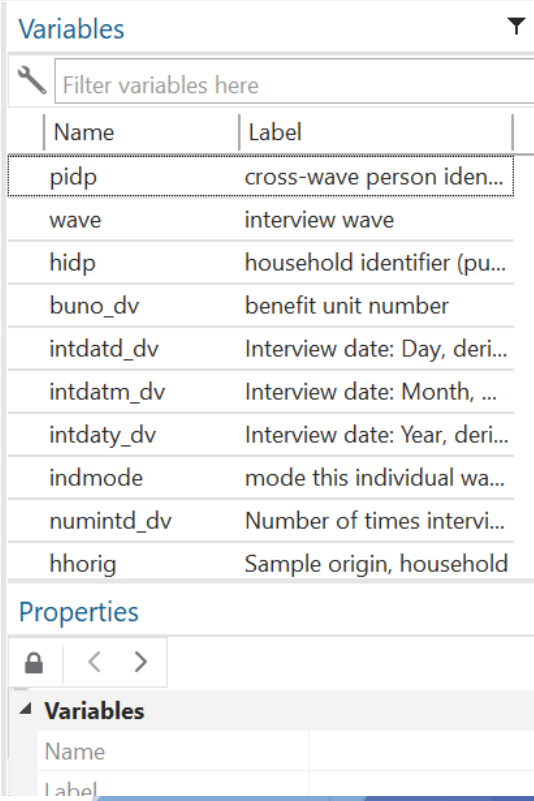
Complete and incomplete person-wave data

pid	wave	sex	age	mastat	jbstat	fihhmn
10019057	1	female	59	never ma	retired	780
10019057	2	female	60	never ma	retired	759.14
10019057	3	female	61	never ma	retired	923.5
10019057	4	female	62	never ma	retired	62.5
10019057	5	female	63	never ma	retired	663
10019057	6	female	64	never ma	retired	missing o
10019057	7	female	65	never ma	retired	1254.963
10019057	8	female	66	never ma	retired	1270.432
10019057	9	female	67	never ma	retired	1364.555
10019057	10	female	67	never ma	retired	1479.74
10019057	11	female	68	never ma	retired	1328.25
10019057	12	female	69	never ma	retired	1371.49
10019057	13	female	71	never ma	retired	missing o
10019057	14	female	71	never ma	retired	1372.333
10019057	15	female	73	never ma	retired	1475.812
10028005	1	male	30	never ma	employed	1501.155
10028005	2	male	31	never ma	employed	1636.259
10028005	3	male	32	never ma	employed	1943.283
10028005	6	male	35	never ma	employed	2001.54
10028005	7	male	36	never ma	employed	1634.33
10028005	9	male	38	never ma	employed	1587.945

Using Stata to describe Panel data

- Let's open the Understanding Society teaching dataset on Moodle





Using Stata to describe Panel data

- Begin by browsing your data

	pidp	wave	hidp	buno_dv	intdatd_dv	intdatm_dv	intdaty_dv	indmode	numintd_dv	hhorig	psu	strata	sampst
1	68001367	1	68001363	1	19	january	2009	f2f	1	ukhls g…	2012	2006	OSM
2	68004087	1	68004083	1	8	january	2009	f2f	9	ukhls g…	2012	2006	OSM
3	68004087	2	68013602	1	16	february	2010	f2f	9	ukhls g…	2012	2006	OSM
4	68004087	3	68013604	1	9	february	2011	f2f	9	ukhls g…	2012	2006	OSM
5	68004087	4	68013606	1	24	january	2012	f2f	9	ukhls g…	2012	2006	OSM
6	68004087	5	68013608	1	18	january	2013	f2f	9	ukhls g…	2012	2006	OSM
7	68004087	6	68013610	1	21	march	2014	f2f	9	ukhls g…	2012	2006	OSM
8	68004087	7	68013612	1	20	april	2015	f2f	9	ukhls g…	2012	2006	OSM
9	68004087	8	68006814	1	24	may	2016	f2f	9	ukhls g…	2012	2006	OSM
10	68004087	9	68006816	1	22	march	2017	web	9	ukhls g…	2012	2006	OSM
11	68006127	1	68006123	1	9	january	2009	f2f	9	ukhls g…	2012	2006	OSM
12	68006127	2	68020402	1	21	january	2010	f2f	9	ukhls g…	2012	2006	OSM
13	68006127	3	68020404	1	24	january	2011	f2f	9	ukhls g…	2012	2006	OSM
14	68006127	4	68020406	1	5	september	2012	f2f	9	ukhls g…	2012	2006	OSM
15	68006127	5	68020408	1	19	march	2013	f2f	9	ukhls g…	2012	2006	OSM
16	68006127	6	68020410	1	2	june	2014	f2f	9	ukhls g…	2012	2006	OSM
17	68006127	7	68020412	1	18	march	2015	f2f	9	ukhls g…	2012	2006	OSM
18	68006127	8	68013614	1	30	march	2016	f2f	9	ukhls g…	2012	2006	OSM
19	68006127	9	68013616	1	12	march	2017	web	9	ukhls g…	2012	2006	OSM

Using Stata to describe Panel data

- Panel data is usually organised in a long format (individual-time pair)
- In our case, the cross-section and time variables are pidp and wave. Respectively
- Inform Stata of the structure of the data and the individual and time id using xtset

- xtset pidp wave

Using Stata to describe Panel data

Cases not observed for every time period

```
. xtset pid wave

       panel variable:  pid (unbalanced)
        time variable:  wave, 1 to 15, but with gaps
                delta:  1 unit
```

Period between observations in units of the time variable

Using Stata to describe Panel data

- Use all xt… commands to analyse panel data
- xtdescribe

```
. xtdescribe

    pidp:  68001367, 68004087, ..., 1.635e+09               n =      36619
    wave:  1, 2, ..., 9                                     T =          9
           Delta(wave) = 1 unit
           Span(wave)  = 9 periods
           (pidp*wave uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       2         5         9       9       9

     Freq.  Percent    Cum.  |  Pattern
-------------------------------+---------
     13694    37.40   37.40  |  111111111
      8539    23.32   60.71  |  1........
      4508    12.31   73.02  |  11.......
      2411     6.58   79.61  |  111......
      1956     5.34   84.95  |  11111....
      1588     4.34   89.29  |  11111111.
      1498     4.09   93.38  |  1111.....
      1332     3.64   97.02  |  111111...
      1093     2.98  100.00  |  1111111..
-------------------------------+---------
     36619   100.00          |  XXXXXXXXX
```

Using Stata to describe Panel data

- Use xttab and tab (and see the differences)

```
. ta sex

     Sex,
 derived |      Freq.     Percent        Cum.
---------+-----------------------------------
    Male |     81,167       42.24       42.24
  Female |    110,979       57.76      100.00
---------+-----------------------------------
   Total |    192,146      100.00

. xttab sex

                    Overall                 Between                 Within
  sex_dv |     Freq.   Percent        Freq.   Percent                Percent
---------+-----------------------------------------------------------------
    Male |      81167     42.24        15816     43.19                 100.00
  Female |     110979     57.76        20802     56.81                 100.00
---------+-----------------------------------------------------------------
   Total |     192146    100.00        36618    100.00                 100.00
                                  (n = 36618)
```