

Pseudo Panels Part II

Bert Van Landeghem, Department of Economics and InstEAD

April 23, 2014

Introduction

- ▶ As seen in previous part of presentation, pseudo panel data are obtained by aggregating data from repeated cross-sections in $x * y$ cells, where y is generally the calendar year and x another variable such as year-of-birth, state, country, city.
- ▶ One can first-difference data within year-of-birth, city, country
...

Why a Pseudo Panel Is not Necessarily Worse than Genuine Panel?

- ▶ Both suffer of course from nonresponse.
- ▶ Nonresponse might, both in genuine panel data and pseudo panel data, cause estimated means or estimated regression coefficients not to be representative for the whole population.
- ▶ In genuine panel data, we have the additional problem of nonrandom attrition during the course of the panel.
- ▶ One can of course work with a balanced panel, but again it is not necessarily representative for the population.

Why a Pseudo Panel Is not Necessarily Worse than Genuine Panel?

- ▶ Genuine panel data can suffer from panel conditioning or panel effects.
- ▶ Recently, people have been making progress in convincingly identifying such panel effects by using refreshment samples or by randomly surveying one part of their sample.

Why a Pseudo Panel Is not Necessarily Worse than Genuine Panel?

- ▶ Das et al. (2011) find that, by comparing refreshment samples with more experienced samples, there is a panel effect for knowledge questions but not for attitudinal questions.
- ▶ In three field studies, Zwane et al. (2011) randomly allocate part of their sample to a survey on health and/or household finances.
- ▶ They find that being surveyed changes people's behaviour, and that it can affect the mean of outcome variables as well as estimated parameters of regression equations.
- ▶ For example, being surveyed about health increases the use of water treatment products and take-up of medical insurance.
- ▶ Crossley et al. (201') do a similar exercise and find that surveying people about pension savings has a huge impact on their saving behaviour.

Case Study: Macroeconomics of Loss Aversion

- ▶ In our work-in-progress, it was very natural to choose for pseudo panel data.
- ▶ First, we are interested in relationships at the macrolevel (not individual level).
- ▶ Moreover, repeated cross-sections with the necessary information is available for almost every country.

Case Study: Macroeconomics of Loss Aversion

- ▶ Some might know the debate in the literature about economic growth and happiness.
- ▶ Some argue that the relationship is flat over time (cfr. Easterlin paradox).
- ▶ Others find a positive gradient.

Case Study: Macroeconomics of Loss Aversion

- ▶ In our case study, we focus however on asymmetries between economic upturns and downturns.
- ▶ Our global evidence so far suggests that a decrease in GDP lowers life satisfaction, and that an increase in GDP weakly increases life satisfaction.
- ▶ In absolute terms, a 1% decrease in GDP has an impact which is several times larger than a 1% increase.

Case Study: Macroeconomics of Loss Aversion

- ▶ We hence perform a country-level analysis (and a state-level analysis for the USA).
- ▶ The regressions include state-fixed effects (or country-fixed effects) since there are reasons to control for country/state specific unobservables which might influence the results.

Case Study: Macroeconomics of Loss Aversion

- ▶ We have used three large repeated cross-sectional datasets that contain life satisfaction measures.
- ▶ We matched those with administrative data on unemployment rates, GDP and inflation rates.
- ▶ Gallup World Poll provides us with data from 151 countries over 7 years.
- ▶ The Eurobarometer documents life satisfaction in 15 countries over four decades.
- ▶ The Behavioural Risk Factor Surveillance System (BRFSS) gives a detailed picture of the states of the USA from 2005 to 2010.

Behavioral Risk Factor Surveillance System

- ▶ Three nice datasets with their own strengths.
- ▶ Given the audience, I would like to focus a bit more on the Behavioral Risk Factor Surveillance System (BRFSS).
- ▶ The data are free to download (no paperwork required !) and it might contain useful information for the type of research carried out at our institutes.

Behavioral Risk Factor Surveillance System

- ▶ BRFSS was started in 1984 and conducted by telephone in 15 states.
- ▶ Its aim is to monitor risky behaviour and health issues.
- ▶ Monthly data collection, collected through telephone surveys.
- ▶ Monitored by the Centers for Disease Control and Prevention (CDC)
- ▶ Data allow state-level analysis, but some states implemented stratification that allows us to monitor certain regions.

Behavioral Risk Factor Surveillance System

- ▶ BRFSS became nation-wide in 1993.
- ▶ Exists of a core questionnaire (which is implemented by all states) and optional modules (among which states can choose).
- ▶ In 2008, a cell phone pilot was started to be able to reach a broader segment of the population.

Behavioral Risk Factor Surveillance System

- ▶ One can create state-level (pseudo) panel data.
- ▶ The time frequency can be one year, but also one quarter or one month.
- ▶ The fact that surveys are carried out on a monthly basis creates opportunities (e.g. helped to analyse the effects of Hurricane Katrina).
- ▶ Sample is huge: 500000 individuals in 2012.

Behavioral Risk Factor Surveillance System

- ▶ Data are not available in Stata format.
- ▶ But SAS Export files can be read into Stata.
- ▶ There is a PDF with the questionnaire for each year.
- ▶ However, you need to look at the codebook to see whether a question has been asked.
- ▶ Sometimes, you find a question in the questionnaire, the associated variable (with all missing values) in the datafile.
- ▶ If you go to the codebook for that year, you will see that the data is not available for that variable.

BRFSS: Overview of Questionnaire

- ▶ Health status (self-reported health etc.)
- ▶ Healthy days (mental, physical)
- ▶ Health care access
- ▶ Sleep (did you get enough rest)
- ▶ Exercise (do you do sports?)

BRFSS: Overview of Questionnaire

- ▶ Diabetes (ever told having diabetes?)
- ▶ Oral health (dentist visits, removed teeth etc.)
- ▶ Cardiovascular Disease Prevalence
- ▶ Asthma
- ▶ Disabilities (very general)

BRFSS: Overview of Questionnaire

- ▶ Tobacco use (current use, history, trying to quit)
- ▶ Demographics (race, marital status, income bracket)
- ▶ Alcohol use
- ▶ Immunization
- ▶ Falls (how often fell, did it cause injury)

BRFSS: Overview of Questionnaire

- ▶ Seatbelt use, drinking and driving
- ▶ Women's health (breast cancer screening and prevalence etc.)
- ▶ Prostate cancer screening/prevalence and other cancers
- ▶ HIV/AIDS (questions about various tests as well as risky behaviours)
- ▶ Emotional support and satisfaction with life.

Conclusion

- ▶ Pseudo panel data are constructed from repeated cross-sections and can be useful to control for time-invariant unobservables.
- ▶ Pseudo panel data are not necessarily a second best solution to genuine panel data.
- ▶ Sometimes it is very natural to use pseudo panels, and many repeated cross-sections are available.

References

- Crossley, T., J. de Bresser, L. Delaney and J. Winter (2014) Can Survey Participation Alter Household Saving Behaviour? *IFS Working Paper* No 14/06.
- Das, J., A. van Soest and V. Toepoel (2011) Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys. *Sociological Methods and Research* 40, 32-56.
- Zwane, A., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. giné, E. Duflo, F. Devoto, B. Crepon and A. Banerjee (2011) Being Surveyed Can Change Later Behavior and Related Parameter Estimates. *Proceedings of the National Academy of Sciences* 108, 1821-1826.