Analysis and Graphical Representation of Health and Labour Force Participation among the Elderly in Europe

Humboldt-Universität zu Berlin
School of Business and Economics
Ladislaus von Bortkiewicz Chair of Statistics

Statistical Programming Languages
Winter 2017/18

Seminar Paper by

Claudia Günther, Phi Nguyen, Julian Winkel
576419, 526624, 562959

Berlin, 2018-03-15

List of Tables

List of Figures

Abbreviations

xx xxxxxxxxxx

Contents

1 Introduction

- relevance of exploring relationship between health and labour force participation due to changing demographic in Europe
- cite relevant study about ageing
- few sentences about relevant papers exploring relationship -> use paper from DIW
- very short literature overview on relationship between health and labour force participation
- share data set as rich data set for this purpose: 2 sentences about it
- introduction of journal article
- our approach: replicate results an enrich analysis
- especially: introduce graphical visualization tools for descriptive statistics -> ease interpretation of variables
- our aim: write code in a way that allows the user to work with easySHARE data set, even when working on different question

2 Panel data cleaning and subsetting

PHI

2.1 Theory and Design

2.2 Implementation

2.3 Empirical Results

The easySHARE dataset released in spring 2017 is a panel dataset of 108 variables of more than 100.000 individuals covering data from six survey waves carried out between 2004 and 2005. As we are only concerned with a small subset of observations, an important task was to define appropriate functions for subsetting. In oder to make our subsetting process understandable to readers, we decided on using pipe-operator. This allows us to apply the filter and select option in a clever way, where we can select different criteria at once.

-> code sniplet here

In this example, we first filter for participation in wave 1 and the age group between 50 and 64 and the select the desired variables as described in Kalwij and Vermeulen (2005).

Although the overall response rate in the SHARE are comparably high, the data set still has numerous missing values. The reason for this is due to the fact that the study was carried out on a crossnational scale, with some national survey institutions deciding not to participate in all survey modules. This means that the majority of missing values are to be found within observations that have missing values for entire survey modules or waves. The reason for the missing values are documented well in the "Guide to easySHARE release 6.0.0" and specifically coded. For example, the numbers -13 and -14 refer to "not asked in this wave" and "not asked in this country". Since this coding scheme is not useful for the purpose of our analysis, we decided on recoding all of the missing values as "NA". To this end, we defined a function based on the missing codes provided by SHARE that finds the NAs in the data and declares them as such.

-> code snipet here

Since the study carried out by Kalwij and Vermeulen (2005) is based on the use of mostly binary data, we needed to construct numerous dummies based on the orginal data.

-> code snipet here

The resulting dataframe containes XXX observations of YYY variables.

- coded countries in more readable manner
- use packe ddd -> match offical ISO code with country name
- create country list with all countries in study (not Israel)
- defined dummies

3 Multidisciplinary and crossnational summary statistics

JULIAN

3.1 Theory and Design

3.2 Implementation

3.3 Empirical Results

4 Crosssectional probit regression

JULIAN

4.1 Theory and Design

4.2 Implementation

4.3 Empirical Results

5 Wald Test

5.1 Theory and Design

In the context of probit regression, the Wald test can be used to test both single and multiple hypothesis on the model parameters. For example, it can be used to test whether the fit of the model is improved if a subset of regression coefficients are all set equal to zero. In the contect of probit regression, the Wald statistic for testing nonlinear and linear hypothesis of our $k \times 1$ parameters β is given by

$$W = c(\hat{\beta})'[\nabla_\beta c(\hat{\beta})\hat{V}\nabla_\beta c(\hat{\beta})]^{-1}c(\hat{\beta})$$

where $c(\hat{\beta})$ is a $m \times 1$ vector of linear or nonlinear restrictions, $\nabla_\beta c(\hat{\beta})$ is the $m \times k$ Jacobian of $c(\hat{\beta})$ evaluated at $\hat{\beta}$ and $\hat{V}$ is the estimated covariance matrix (Wooldridge 2010, 463). Under H0, the Wald test statistic is approximately $\chi 2_m$ distributed, with m being the number of specified restrictions. We can formulate our hypothesis in accordance with structure of $R\hat{\beta} = r$ to facilitate the derivation of our test statistics where XXX. If the necesaary conditions are satisfied, the Wald statistic can be rewritten as

$$W = (R\hat{\beta} - r)'[R\hat{V}R']^{-1}(R\hat{\beta} - r)$$

where R is the Jacobian matrix of size $m \times k$, r is restriction function size $m \times 1$. For the case where we are testing the single hypothesis of the joint significance of a subset of coefficients, the test statistic can be even more simplified:

$$W = \hat{\beta}_s'[\hat{V}]^{-1}\hat{\beta}_s$$

The nullhypothesis in this case is that all parameters in the vector β_s are jointly equal to zero, where the test statistic is approximately $\chi 2_1$.

5.2 Implementation

5.3 Empirical Results

6 Counterfactual exercise

CLAUDIA

6.1 Theory and Design

6.2 Implementation

6.3 Empirical Results

7 Graphical representation

PHI

7.1 Theory and Design

7.2 Implementation

7.3 Empirical Results

8 Conclusion

9 References

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* MIT press.

Declaration of Authorship

We hereby confirm that we have authored this Seminar paper independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

Berlin, 2018-03-15

Claudia Günther, Phi Nguyen, Julian Winkel