Preparation of the earnings variable in the GGS data Wave 1

Joanne Sophie Muller[12]

Correspondence should be addressed to

Joanne Muller, MSc

Netherlands Interdisciplinary Demographic Institute

Lange Houtstraat 19

P.O. box 11650

2502 AR The Hague

The Netherlands

T: +31619181096

E: muller@nidi.nl

Affiliations

[1] Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands

[2] University of Groningen, The Netherlands

GGP data used: Wave 1, V.4.3

Document Version 2, April 10th, 2017
Document first published online October 28th, 2016

Related Stata datafile: EarningsImputation_Muller.dta

INTRODUCTION

This document describes how the variable of personal earnings in the Generations and Gender Surveys (GGS) data Wave 1 was created. This process involved harmonization across countries and imputation of missing values. I created three respondent's earnings variables: (1) earnings in local currencies, (2) earnings in PPP euro's, (3) imputed log earnings (in PPP euro's). And the same three for the partner's earnings.

Currently the earnings variables are available for 14 countries:

Belgium
Bulgaria
Czech Republic
Estonia*
France
Georgia
Germany
Lithuania
Netherlands
Norway
Poland
Romania
Russia
Sweden

*The Austrian, Italian and Estonian data do not contain the partner's earnings.

Hungary not included because only household income is available (i.e. no personal earnings)

CODEBOOK VARIABLES

Variable name	Description
acountry	Country id
id	Respondent ID variable. Created with stata code: gen double id=arid+(acountry/100) format id %24.2f
Earnings_local	Respondent's annual personal net earnings from a job or self-employment. Unit = local currency Contains missings.
Earnings_ppp	Respondent's annual personal net earnings from a job or self-employment. Unit = 2008 PPP euro's Contains missings.
_1_log_earnings_imputed to _5_log_earnings_imputed	Respondent's log annual personal net earnings from a job or self-employment. Unit = 2008 PPP euro's Missing values imputed using multiple imputation.
Earnings_imputed_flag	Respondent's earnings imputed (1) or not (0)
P_earnings_local	Partner's annual personal net earnings from a job or self-employment. Unit = local currency Contains missings.
P_earnings_ppp	Partner's annual personal net earnings from a job or self-employment.[1] Unit = 2008 PPP euro's Contains missings.
_1_p_log_earnings_imputed to _5_p_log_earnings_imputed	Partner's log annual personal net earnings from a job or self-employment. Unit = 2008 PPP euro's Missing values imputed using multiple imputation.
P_earnings_imputed_flag	Partner's earnings imputed (1) or not (0)

---

[1] For Norway, the partner's earnings are the GROSS amount. Therefore the partner's earnings are relatively higher than the respondent's (which are NET). The GGP data contain the respondent's GROSS earnings as well (variables a866_2_2000 with label "annual GROSS amount of payment of each type of income").

RECODING (Earnings_local)

In all countries earnings information were collected with survey questions, except in Sweden and Norway where this information was derived from register data. The GGP questionnaires asked about the respondent's average monthly net earnings from a job during the last 12 months. Respondents were also asked how often they received that amount during the last 12 months (in most cases this was 12 times) to adjust for seasonal or otherwise not-year round work.[2] By multiplying the monthly earnings the appropriate times, I estimated the annual net earnings. If respondents were not able or willing to indicate an exact amount, they were asked in which range their earnings lay. I replaced such indicated bands by the median of the earnings of respondents who did indicate an exact value within that band. I used the median rather than the mean, because earnings are not normally distributed within bands but positively skewed.

Four GGP countries (Poland, Estonia, Sweden and Norway) did not provide monthly earnings information, but the net total annual income received. This measure also includes other income sources than earnings from a job or self-employment, such as social benefits. In order to minimize measurement difference between datasets I made two adjustments. Using additional survey questions, I first checked whether respondents were employed. These questions regard current activity status (employed or not), number of working hours per week and income payment types received. If respondents were not employed, I assumed their earnings to be zero.

---

[2] The German dataset does not contain information about the frequency of payment (i.e. the number of times the amount received in the last year). Since the German questionnaire asked about monthly earnings, I assumed that the appropriate frequency is 12 times.

I assume that the lion's share of the total income of people with earnings from a job or self-employment depends on their earnings. Therefore, I expect that the ranking of the two income definitions is roughly the same. However, I do realize that especially at the bottom of the income distribution, the ordering between the net earnings and total net income could be different due to social benefits. Last, I 'undid' the GGS (original data) conversion using the following factors:

Country ID	Country	exratew1
11	Bulgaria	0.50982
12	Russia	0.02717
13	Georgia	0.43875
16	Hungary	0.00371
19	Romania	0.23833
20	Norway	0.1263
22	Estonia	0.06387
23	Belgium[3]	0.12561
24	Australia[4]	0.57651
25	Lithuania[5]	0.29040
26	Poland	0.25191
28	Czech Republic	0.0334
29	Sweden[6]	0.10826

---

[3] After robustness checks (e.g. comparison to the Worldbank GNI per capita in PPP international dollars values), I decided not to use the conversion factor for Belgium, Lithuania and Sweden.
[4] Not harmonized yet
[5] See note 3.
[6] See note 3.

4

CONVERSION

Earnings_ppp and p_earnings_ppp

The earnings variable is transformed from local currencies to 2008 euros using Purchasing Power Parity (PPP) in three steps. First, I convert to local currencies in the year 2008 (i.e. inflation correction) by using the consumer price index (The World Bank, 2016a). Next, these local 2008 currencies are converted to international comparable dollars using the PPP conversion factor (The World Bank, 2016b). Last, I convert to PPP euros using the annual average exchange rate in 2008 between dollars and euros (i.e. 1 euro = 1.4708 US dollar, Eurostat, 2015). Using PPPs allows us to make a more meaningful comparison between countries, because it 'controls for' the cost of living (ICP, 2011). I choose the year 2008 because most data were collected in 2008.

MULTIPLE IMPUTATION

Log_earnings_imputed and log_p_earnings_imputed

To retain as much information as possible, cases with missing earnings values were imputed using multiple imputation. In the multiple imputation model a large selection of control variables was used covering personal information (gender, age), educational level, family characteristics (e.g. young children), house characteristics (e.g. house size), health information, possessions, job characteristics (e.g. work hours per week) and financial indicators (e.g. 'difficulty to meet ends').

Control variables used for multiple imputation

- weekly workhours = number of hours respondent works per week (a835)
- living floor size (a120)
- educational level = educational level respondent (isled)
- mother's educational level (isled)
- father's educational level (isled)
- ends meet = difficulty to meet ends, from "with great difficulty" to "very easily" (a1002)
- depression = rowmean(a721_a a721_b a721_c a721_d a721_e a721_f a721_g)
- general health = general health status, 1 very good to 5 very bad (a701)
- number of rooms (a119)
- age = respondent's age in months
- age squared
- gender (dummy: male / female)
- partner = respondent has co-resident partner (1) or not (0) (ahg3*)
- children = number of children in hh younger than 18 (ahg5_`x' < 18)
- supervise (a840)
- disabled = respondent is disabled (ahg9_1)
- native = born in country of interview (1) or not (0) (a105)
- homeowner (1) or not (0) (a122)
- colortv
- dvdplayer
- microwave
- telephone
- car
- second car
- second home
- savings
- dishwasher
- holiday
- furniture
- clothes
- meat
- rent (a1004_1)
- mortgage (a1004_2)
- chronic ill (a702a)
- incpaidisei_mis
- area problems = rownonmiss(a144_1 a144_2 a144_3 a144_4 a144_5 a144_6)
- friends
- warm
- piped water (1) or not (0) (a138)
- bath or shower (1) or not (0) (a140)
- flush toilet (1) or not (0) (a142)
- refrigerator
- bathroom

6

Country specific information

Respondent earnings

Country ID	Country	Number of variables (a864*)	# different income sources (a864*)	# band codes (a867*)	Euro	# Earnings sources
11	Bulgaria	13	13	9	No	Source1 + source2
12	Russia	6	14	12	No	Source1 + source2
13	Georgia	3	10	9	No	Source1 + source2
14	Germany	4	7	8	Yes	Source1
15	France	4	9	8	Yes	Source1 + source2
18	Netherlands	2	38	17	Yes	Source1
19	Romania	4	11	10	No	Source1 + source2
20	Norway	5	8	n.a.	No	Source1
22	Estonia	4	8	8	No	Source1 + source2
23	Belgium	4	13	8	Yes	Source1 + source2
25	Lithuania	4	11	19	No	Source1 + source2
26	Poland	10	11	13	No	Source1
28	Czech Republic	6	9	10	No	Source1 + source2
29	Sweden	4	7	4	No	Source1

Norway labels (individual, annual info)

a866_*_2000 (1, 3, 4, 5) label: "Net amount of payment type income * "

a866_2_2000 (2) label: "annual GROSS amount of payment of each type of income"

a866_2000 label: "annual NET amount of SUM of payment of ALL income types"

Estonia labels (annual info)

a866_2201 label: "sum of net amounts of payment from all mentioned income sources"

a867_2201 label: "range of sum of net amounts of payment from all mentioned income sources"

Poland labels

a866_2600 label "Net amount received from all income sources summed up (last 12 months)"

a867_2600 label: "Range amount received from all income sources summed up (last 12 months)"

Sweden labels

a867_2900 label: "Range amount received from all income sources"

7

Partner's earnings

Country ID	Country	Number of variables (a936*)	# different income sources (a936*)	# band codes (a939*)	Euro	# Earnings sources
11	Bulgaria	13	13	9	No	Source1 + source2
12	Russia	5	14	12	No	Source1 + source2
13	Georgia	3	10	9	No	Source1 + source2
14	Germany	3	7	8	Yes	Source1
15	France	4	9	8	Yes	Source1 + source2
18	Netherlands	2	8	17	Yes	Source1
19	Romania	3	11	10	No	Source1 + source2
20	Norway	4	8		No	Source1
23	Belgium	4	13		Yes	Source1 + source2
25	Lithuania	3	11	19	No	Source1 + source2
26	Poland[7]	-	-	-	No	-
28	Czech Republic	5	9	10	No	Source1 + source2
29	Sweden	4	6		No	Source1

Norway labels (individual, annual info)

a938_*_2000 (1, 3, 2, 4) label: "annual GROSS amount of payment of each type of income"[8]

Poland labels

a938_2600 label: "Net amount received from all income sources summed up (last 12 months)"

a939_2600 label: "Range amount received from all income sources summed up (last 12 months)"

Sweden labels

---

[7] Note that for Poland, the 'income source information' is missing. This implies that I was not able to correct for 'income from other sources' (i.e. change the partner's earnings to zero if the partner has no income from a job or self-employment). Therefore partner's without income from a job or self-employment might have a positive earnings value (i.e. it appears that they have income from a job).

[8] Note that for respondent's earnings in Norway both the NET and GROSS amount is available. For partner's in Norway, only the GROSS amount is available. Therefore the partner earnings in Norway are relatively high (i.e. gross amount rather than net amount)

a939_2900 label: "Range amount received from all income sources"

REFERENCES

Eurostat. (2015). ECU/EUR exchange rates versus national currencies. from http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tec00033&plugin=1

ICP. (2011). ICP Handbook. from http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/ICPEXT/0,,contentMDK:22407349~pagePK:60002244~piPK:62002388~theSitePK:270065,00.html

The World Bank. (2016a). Consumer price index (2010 = 100). from http://data.worldbank.org/indicator/FP.CPI.TOTL

The World Bank. (2016b). PPP conversion factor, private consumption (LCU per international $) from http://data.worldbank.org/indicator/PA.NUS.PRVT.PP?page=1