

Imputation of income in household questionnaire LISS panel

date 01-09-2008
author(s) Klaas de Vos

version 4
classification

© CentERdata, Tilburg, 2008

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher.



1 Introduction

This note describes the procedure used to impute missing income information in the household questionnaire of the LISS panel. So far, for the calculation of household income missing income data were recoded to zero. This could lead to implausible results. In particular, in many cases gross household income was below net household income. To prevent this from happening, we want to add an imputed value for gross income if net income is available and gross income is not, and vice versa.

2 Gross-net regressions

In the (original) household questionnaire dataset (June 2008) gross and net income information is available as follows (table 1).

Table 1. Gross and net income in household questionnaire (avars)

		gross			total
		missing	0	>0	
net	missing	610	10	43	663
	0	542	3241	13	3796
	bracket	1017	8	678	1703
	> 0	2428	0	4084	6512
total		4597	3259	4818	12674

Proposal: positive values and zeroes are accepted. When net = zero and gross = missing gross income is set to zero and vice versa. For the 43 persons with missing net income and positive gross income net income is imputed. For the $1017+2428 = 3445$ persons with missing gross income and positive net income gross income is imputed. The imputations are based on the 4084 observations where gross and net income are positive and filled out exactly (i.e. the 678 observations where only bracketed information on net income is available are not used in the imputation regression).

For the 610 persons where both gross and net income are missing we still have a problem. In the households in question household income will have to be registered as missing. In principle income information from the core module on income could be used but this is not feasible as a quick solution.

For the persons for whom only bracketed information on net income (up to 500, 501-1000, 1001-1500 etc) is available, we use the midpoint of the bracket. This is plausible because the average income of persons for whom net income is known exactly is close to the middle in each bracket (see table 2). An exception is formed by the incomes above 7500 Euro per month. In this category 10000 Euro was used so far, but the average is more than 100000 Euro. It is likely that some of the respondents have reported amounts in Eurocents or yearly instead of monthly amounts. It seems rather risky to attribute a net income of 100000 Euro to all respondents with a net income in the bracket above 7500 Euro (and the exact amount unknown) on the basis of these observations. If gross income is available (4 out of 9 observations) we could use the imputation equation to be estimated for persons with missing net income. For the time being we will stick to 10000 Euro for all 9 observations.



Tabel 2. Average personal net monthly income, by bracket

Personal net monthly income in brackets	Mean	N	Std. Deviation
1 EUR 500 or less	252.93	628	146.688
2 EUR 501 - EUR 1000	782.77	1260	142.939
3 EUR 1001 - EUR 1500	1293.54	1560	141.507
4 EUR 1501 - EUR 2000	1765.51	1579	148.450
5 EUR 2001 - EUR 2500	2260.02	751	152.290
6 EUR 2501 - EUR 3000	2793.11	353	158.778
7 EUR 3001 - EUR 3500	3291.89	161	153.166
8 EUR 3501 - EUR 4000	3809.74	99	164.347
9 EUR 4001 - EUR 4500	4289.77	30	151.682
10 EUR 4501 - EUR 5000	4852.15	26	180.152
11 EUR 5001 - EUR 7500	6095.54	37	644.978
12 More than EUR 7500	102263.39	28	70319.936

For the imputation of net income on the basis of gross income the following equation can be used ($r^2 = 0.972$):

$$\text{nettoink}_f = \exp(-0.26356 + \text{lbruto} * 1.12750 + \text{lbruto}^2 * -0.01850 + \text{oud} * 0.12893 + \text{pos1} * 0.03209 + \text{belbezig1} * 0.46602 + \text{bel1lbruto} * -0.06101) \quad (1)$$

with $\text{lbruto} = \ln(\text{brutoink})$, where brutoink stands for gross monthly income
 $\text{lbruto}^2 = \text{lbruto}^2$,
 $\text{oud} = (\text{leeftijd} > 64)$: age > 64
 $\text{pos1} = (\text{positie} = 1)$: head of household
 $\text{belbezig1} = (\text{belbezig} = 1)$: in paid employment
 $\text{bel1lbruto} = \text{belbezig1} * \text{lbruto}$

For the estimation of this equation we only use observations with net income known exactly (not in brackets). We have dropped outliers with net incomes above 15000 Euro or gross income above 30000 Euro.

For the imputation of gross income on the basis of net income we have estimated the following equation ($r^2 = 0.970$)

$$\text{brutoink}_f = \exp(0.47784 + \text{lnetto} * 0.82146 + \text{lnetto}^2 * 0.02167 + \text{olnetto} * -0.02004 + \text{pos1} * 0.01782 + \text{belbezig1} * -0.14874 + \text{bel1lnetto} * 0.02372) \quad (2)$$

with $\text{lnetto} = \ln(\text{netinc})$, where netinc stands for net monthly income
 $\text{lnetto}^2 = \text{lnetto}^2$,
 $\text{olnetto} = \text{oud} * \text{lnetto}$,
 $\text{bel1lnetto} = \text{belbezig1} * \text{lnetto}$.



For persons who have reported an exact value of gross income but net income in brackets, the predicted value of net income based on equation (1) could be used to check the plausibility of the bracketed income information.