MIMOSIS: a microsimulation model for social policy in Belgium.

Description of the model with units sampled on January 1st 2009, income 2008

6 October 2011

Technical note FPS Social Security

Version: MIMOSIS_Steekproef1januari2009_Eng.doc

Author of correspondence: guy.vancamp@minsoc.fed.be

Content

1 INTRODUCTION

When policy makers consider certain reforms in socio economic legislation, often the need arises to make an ex ante assessment of the possible cost and the impact on the income distribution (who gains, who looses) of these reforms. This text describes the microsimulation model MIMOSIS (MIcrosimulation MOdel for Belgian Social Insurance Systems)[1], that allows to make such impact estimates for reforms in a number of policy areas. 1st

The policy areas on which MIMOSIS focuses in the first place are: social security contributions, pensions, unemployment benefits, means tested benefits, sickness- and disability benefits and family allowances. For reforms in each of these domains, also the effect on personal income taxes is taken into account.

The MIMOSIS model uses a sample of coded personal social data, coming from the Datawarehouse Labour Market and Social Protection (DWH LM&SP) of the Crossroads Bank for Social Security (CBSS). It is because these microdata are used, that an assessment of the impact on both the budget and the income distribution is possible.

The original version of MIMOSIS was developed for a sample of administrative data of 2001 (sample of January 1st 2002).[2] This text describes the update of the model to 2008 data (sample of January 1st 2009).[3]

In section 2 of this note we describe the sample on which the model has been built. In this section we discuss, among other things, how weights have been constructed to extrapolate the sample to the level of the population. Section 3 gives more detail on how the model works. In this section it is explained how input data are manipulated to obtain the impact estimates. In section 4 we describe some output indicators, produced by the model.

Descriptions in this text are always based on the sample of all available observations. Apart from that the model can be used to process the data of one, fictitious individual or household. We do not deal with this possibility in this text.

2 DESCRIPTION OF THE MIMOSIS SAMPLE

The sampling frame[4] encompasses all individuals that, according to the National Register file of the CBSS, are alive on January 1st 2009 and who live in a Belgian community.[5] This group mainly contains people that are physically present in Belgium. Belgians that are abroad can

---

1 MIMOSIS is one of the many models that exist for Belgium. For an overview of other models see Decoster (2010) and Decoster et al (2008).

2 This version was developed as part of a series of AGORA projects, funded by Belgian Science Policy. For a description of the first version of the model see Decoster et al (2008).

3 This update was made possible thanks to support from the Belgian Science Policy and the Crossroads Bank for Social Security. The help of Chris Brijs, Hans Knapen, Anke Mutsaerts and Patrick Lusyne in the processing of the data is greatly appreciated. They are obviously not responsible for the choices made.

4 See Molenberghs en Gaens (1994), p. 11 for a description of the concept sampling frame.

5 See DOCDWH (2011), document "fiches_beschrijving_rr_ksz.doc" for more information on the content of the data in the National Register in DWH LM&SP.

be part of this group, but only if their residential address in Belgium is maintained during their leave. This mainly concerns people who only left the country on a temporary basis because they are on holiday or on a temporary mission abroad. Hence, Belgians who reside permanently abroad are not taken up in the sample frame.

Each individual in the sketched sampling frame can be identified on the basis of a unique identification number. From the set of all possible numbers, a sample is drawn. The sampling procedure can be divided in three main steps:

- Step 1: It was tested how many individuals that were randomly sampled for the earlier MIMOSIS-sample were still alive on January 1st 2009.[6] This concerned 90.641 individuals.
- Step 2: A complement of 9.359 individuals was sampled from the sketched sampling frame minus the 90.641 individuals from the earlier sample that are still alive. The sampling procedure, applied to sample these 9.359 extra individuals, is systematic sampling taking into account the possible non random attrition within different age classes.[7]
- Step 3: Of the total number of 100.000 individuals, obtained after step 2, the identification numbers of all other household members are selected if the individual that is originally sampled, is part of a private household.

Households can be defined here as all individuals who have the same residential address.[8] In addition, one can distinguish private and collective households.

A private household, as defined in the National Register, comes close to what a sociological household is. Sociological households are generally defined as all individuals who share a residence and jointly decide on the majority of their expenses.[9] For the remainder of this paper we therefore use the terms private and sociological household as synonyms.

Collective households encompass communities such as retirement homes, hospitals, prisons, monasteries,….[10]

When drawing the sample in step 1 and step 2 we do not take into account the distinction between private and collective households. When the identification numbers of household members are collected in step 3, we do.

Of all the individuals in the final sample, we then collect a number of variables from the DWH LM&SP. A complete list of the requested variables can be found in FOD SZ (2011a).[11]

---

6 For the previous MIMOSIS sample, 100,000 individuals were drawn from a similar sampling frame on January 1st 2002. See Vleminckx (2008) for a description of the content of this sample.

7 The applied principles of systematic sampling are explained in Vleminckx (2011), section 9. The entire mechanism is implemented by the program SampleMIMOSIS_2009.sas (see FOD Sociale Zekerheid (2011b)).

8 See Rijksregister (2011).

9 See Atkinson et al (1995), p. 16 for more information on possible definitions of the term household.

10 See APS (2011).

11 This not necessarily concerns variables that are directly present in the DWH LM&SP, but it concerns manipulated variables. The full package of programs that convert the data from the

2.1 CONSTRUCTION OF HOUSEHOLD NUMBER AND ESTIMATE OF HOUSEHOLD SIZE

The sample, constructed according to the principles described above, comprises 299,257 individuals. Of these individuals, we know whether they a) stay in a collective household and b) if not, who the reference person of the private household was on January 1st 2009 and January 1st 2002.[12]

In total, 1,172 individuals within the sample reside in a collective household. These individuals get a unique household number and will be considered as singles in what follows.

Of 136 individuals, who do not reside in a collective household, the variable reference person is not completed on January 1st 2009. If the variable reference person on January 1st 2002 was completed, this value was taken. Otherwise, the individual is regarded as a single. All individuals from private households, that have the same identifier for the reference person, will further be treated as members of the same household.

Verification of the data showed that two families in the sample consisted of respectively 27 and 28 individuals. Although this number of household members is not impossible in principle, the chances of errors in the household composition become larger as the number of members increases. Therefore these two households were removed from the sample that will be analyzed further.

After this operation, the sample still encompasses 299,202 individuals. 298,030 of these individuals live in 97,760 different private households.[13]

2.2 CONSTRUCTION OF A NUMBER OF CHARACTERISTICS BASED ON DATA FROM THE NATIONAL REGISTER

Within the set of variables that were retrieved for all individuals of the sample[14] the variables from the National Register file take a special place, because they allow to identify the household units and the relationship between the household members.

One of these variables is the variable NAREGNIS_RELATION that expresses the relationship of each household member with respect to the household head in a code. The possible values that the variable NAREGNIS_RELATION can take, are shown in Table 1.

DWH LM&SP into the variables included in the data demand, can be found in FOD Sociale Zekerheid (2011c).

[12] The reference person is, in principle, the person who defends the interests of the household or who provides the largest amount of household income. In practice, however, it may also be the person dealing with the administrative affairs of the household (see APS (2011)).

[13] These manipulations are performed with the program CreatePopCharAndWeights.sas (see FOD Sociale Zekerheid (2011c)).

[14] See Vleminckx (2011) for the full list of variables.

Table 1: Possible values of the variable NAREGNIS_RELATION

Possible values	Description
1	Head of the household
2	spouse
3	son, daughter
4	son in law, daughter in law
5	Grand son, grand daughter
6	father, mother
7	father in law, mother in law
8	grandfather, grandmother
9	brother, sister
10	brother in law, sister in law
11	relative
12	non relative
13	stepson, stepdaughter
14	great grandson, great granddaughter
15	uncle, aunt
16	cousin (relative to the third degree)
17	cousin (relative to the fourth degree)
20	communities, homes

In principle, the variable NAREGNIS_RELATION should always have a value for the reference person within a given household. In 136 cases the value for the variable NAREGNIS_RELATION was missing. In these cases the variable was manually set equal to 1.

In principle, there should also be at maximum one household head per household. However, there were nine households for which this variable had a value of 1, 2 times within the household. In seven of the nine cases, one of the two cases was already observed as household head in 2002. For these cases, the value of the variable NAREGNIS_RELATION was not adjusted. The other case was manually adjusted to a value of 12 for the variable NAREGNIS_RELATION, i.e. non relative.[15]

For the remaining two cases, the value 1 for NAREGNIS_RELATION was retained in the line for which the identification number of the individual was equal to that of the reference person. In the other case, the value was manually set equal to 12.

In principle, the value 2, i.e. partner, should occur only once within a given household as well. This was not always the case. In 7 households two partners were found and in 1 case there were 3 partners. These cases were manually corrected after inspection of the households. In 4 cases there were always two partners that came up with exactly the same data. Presumably, this is so because their lines have been selected twice with different

[15] The 7 observations for which no adjustment was executed, were the lines for which the individual identification number was equal to that of the reference person.

identification numbers in the selection of the sample. One of these two lines was therefore removed from the sample that will be analyzed further.

A variable that we observed directly in the previous MIMOSIS-sample but not in the current sample, is the variable NAREGNIS_CIVIL. We give the possible values of this variable in Table 2.

Table 2: Possible values of the variable NAREGNIS_CIVIL

Possible values	Description
1	unmarried
2	Married
3	widower, widow
4	divorced
5	divorced of table and bed

Based on other available data, we try to reconstruct the values of NAREGNIS_CIVIL_2009 on January 1st 2009. If two individuals occur within a household with a value of 1 and 2 respectively for the variable NAREGNIS_RELATION, we give them both the value 2 for the variable NAREGNIS_CIVIL_2009. If after this operation the variable NAREGNIS_CIVIL_2009 is not yet completed, we take the value of NAREGNIS_CIVIL_2002 on January 1st 2002 if this value is larger than 2. If after this second step, the value of NAREGNIS_CIVIL_2009 is not yet completed, we consider someone as a widow or widower if an individual is older than 18 and receives a survivor pension. A person is considered to be divorced when he was married according to the variable NAREGNIS_CIVIL_2002 on January 1st 2002, does not receive a survivor pension in 2008 and there is no identification of a possible partner on January 1st 2009. If two individuals were married on January 1st 2002 and if for none of these individuals a value was entered after the previous operations, the variable NAREGNIS_CIVIL_2009 was then set equal to 2. All other individuals, for which in previous operations no value has been entered for NAREGNIS_CIVIL_2009, are considered to be unmarried.[16]

Besides the unique identification number of the individual, the household to which an individual belongs, the relationship of an individual with the head of the household (in NAREGNIS_RELATION) and the marital status (in NAREGNIS_CIVIL) we observe for the sampled individuals also the month and year of birth, gender, a code of the municipality in which the individual resides, the LIPRO position of the individual, whether the individual is randomly selected or not in the first or second stage of the sampling process and a typology of the household to which each individual belongs.[17]

2.3 CONSTRUCTION AND EVALUATION OF SAMPLE WEIGHTS

In the sample we observe the region where the observed individuals live. In Table 3 we give the distribution of the unweighted results (i.e. the percentage share of individuals in the total

[16] All operations in this section were performed with the program CreatePopCharAndWeights.sas (see FOD Sociale Zekerheid (2011c)).

[17] More information on the content of these data can be found in DOCDWH (2011), see document "fiches_beschrijving_rr_ksz.doc".

sample) and compare these results with those based on population statistics on January 1st 2009.

Table 3: Non weighted percentage share of observations by region[18]

	Real share based on population statistics 2009	Unweighted share based on sample observations	Percentage difference between unweighted and real share
Region Brussels Capital	9.94	8.52	-14.26
Region Flanders	57.74	58.49	1.30
Region Wallonia	32.32	32.98	2.03

As shown in the column "Percentage difference between unweighted and real share" in Table 3, the percentage share without weights differs, especially in Brussels, considerably from the real share. Given the sampling features, this should not surprise. If we want to use the sample results to make statements for the whole population, it seems appropriate to construct weighting factors that correct for the non-random way of selecting the observations.

An abstract description of the procedure followed to construct these sample weights is included in Appendix 1 of this text. This procedure is applied both at the individual and household level to construct sample weights. In both cases, the procedure is applied separately for each region.[19]

In Table 4 we give the distribution of the weighted number of observations by gender and region. If we would add the lines for men and women, we see that the total numbers per region are rather well approximated. This should not hide the fact that for the Region Brussels Capital there is an underestimation of the number of men and an overestimation of the number of women. Both deviations more or less cancel each other. In Appendix 2 we further disentangle these groups into age classes of 5 years. This decomposition shows that the deviations, observed for the Region Brussels Capital, mainly come from deviations for men and women aged 20 to 40 years.[20]

[18] The data in the column real share are based on data from ADSEI, see ADSEI (2011a): Population by gender, age and age classes.

[19] This procedure is implemented with the program CreatePopCharAndWeights.sas (see FOD Sociale Zekerheid (2011c)).

[20] The reference point for determining the age in all applications for this note is January 1st 2009.

Table 4: Weighted number of observations on individual level by gender and region[21]

		Real number based on population statistics 2009	Weighted amount based on sample observations	Percentage difference between weighted and real amount
Men	Region Brussels Capital	516,250	503,135	-2.6
	Region Flanders	3,064,169	3,051,317	-0.4
	Region Wallonia	1,688,232	1,690,329	0.1
Women	Region Brussels Capital	552,282	565,460	2.3
	Region Flanders	3,144,708	3,157,629	0.4
	Region Wallonia	1,787,439	1,785,410	-0.1

In Table 5 we give the distribution of weighted observations by gender and marital status in Belgium. As this table shows, the number of married and divorced individuals is significantly underestimated. Presumably people are classified as unmarried because we lack information for the construction of the variable NAREGNIS_CIVIL. The number of widowers is significantly underestimated, but the group of widows, by far the largest subgroup among the widowers/widows, is well approximated. We consider the observed deviations for this variable more as an indicator of the limited quality of the reconstructed variable NAREGNIS_CIVIL than an indicator on the quality of the weights themselves.

Table 5: Weighted observations on individual level by gender and civil status for Belgium[22]

		Real number based on population statistics 2009	Weighted amount based on sample observations	Percentage difference between weighted and real amount
Men	Unmarried	2,495,157	2,698,598	7.5
	Married	2,239,410	2,112,464	-6.0
	Widower, widow	137,892	81,254	-69.7
	Divorced	396,175	352,368	-12.4
Women	Unmarried	2,195,730	2,417,431	9.2
	Married	2,238,798	2,112,415	-6.0
	Widower, widow	575,477	578,781	0.6
	Divorced	474,421	399,765	-18.7

It should also be emphasized that for further modelling in MIMOSIS, the distinction between the status unmarried and divorced is not important. On the contrary, the identification of a marital relationship is of importance for future applications.

In Appendix 3 we decompose the results, reported in Table 5 further in age classes of 5 years. This decomposition shows that the deviation between the sampling distribution and the actual distribution for widowers mainly occurs in the group of widowers of 70 years or older. To further explore the combination of survival pensions and allowed labour income, this civil status identifier should therefore be reliable.

[21] The data in the column real share are based on data from ADSEI, see ADSEI (2011a): Population by gender, age and age classes.

[22] The data in the column real share are based on data from ADSEI, see ADSEI (2011a): Population by gender, age and age classes.

Finally, we give the breakdown of household units, weighted by the reconstructed household weights, by region and the number of household members in Table 6. Both at the aggregate level of the region as at the Belgian level, the actual distribution is relatively good approximated, except for the small group of households with seven or more household members. This underestimation of the very large households occurs in all regions. It is also striking that, in line with the individually weighted results, the number of single men is underestimated and that of single women is overestimated for the Region Brussels Capital.

Table 6: Weighted observations for private households by region and number of household members[23]

		Single men	Single women	2 persons	3 persons	4 persons	5 persons	6 persons	7 persons	8 persons and more	Total private households
Weighted amount based on sample observations	Belgium	725,182	822,762	1,463,483	701,405	583,921	214,335	64,400	19,283	11,773	4,606,544
	Region Brussels Capital	121,054	128,787	120,790	59,909	45,394	21,032	8,855	3,398	2,239	511,458
	Region Flanders	360,678	417,671	890,689	406,879	354,812	121,376	33,421	9,595	6,145	2,601,266
	Region Wallonia	243,450	276,304	452,004	234,617	183,715	71,927	22,124	6,290	3,389	1,493,820
Real number based on population statistics 2009	Belgium	710,551	837,803	1,471,120	690,936	569,434	207,817	62,850	17,196	10,271	4,577,978
	Region Brussels Capital	108,782	135,663	123,981	59,618	44,059	22,481	9,334	3,140	2,049	509,107
	Region Flanders	356,344	429,001	894,278	401,032	342,465	116,693	31,281	8,652	4,970	2,584,716
	Region Wallonia	245,425	273,139	452,861	230,286	182,910	68,643	22,235	5,404	3,252	1,484,155
Percentage difference between weighted and real amount	Belgium	-2.0	1.8	0.5	-1.5	-2.5	-3.0	-2.4	-10.8	-12.8	-0.6
	Region Brussels Capital	-10.1	5.3	2.6	-0.5	-2.9	6.9	5.4	-7.6	-8.5	-0.5
	Region Flanders	-1.2	2.7	0.4	-1.4	-3.5	-3.9	-6.4	-9.8	-19.1	-0.6
	Region Wallonia	0.8	-1.1	0.2	-1.8	-0.4	-4.6	0.5	-14.1	-4.0	-0.6

[23] The data in the column real share are based on data from ADSEI. Unlike other population data, data on private households were not yet publically available for January 1st 2009 at the time of writing. They were supplied on request by ADSEI.

2.4 DESCRIPTION OF A NUMBER OF VARIABLES FORM OTHER SOURCES THAN THE NATIONAL REGISTER

Apart from the variables from the National Register, already cited above, also a number of variables from other Public Institutions of Social Security (IPSS) were collected from the DWH LM&SP. In Table 7 we give a list of the abbreviations used to describe the different sources consulted and a brief description of these sources. Unless otherwise stated, data from a file are from the observation year 2008.

Table 7: Abbreviations and description of the different data sources that have been consulted for the selected individuals[24]

Name file	Description
CIMIRe	A source with career data on individual level for employees. The file contains data on periods that the employees have been active on the labour market and periods on which they have not been active. The possible years of observation in the file range from 1954 to 2001.
NIC	A source with data on individual level on the period of sickness of those insured as employee or self employed.
PODMI	A source with data on individual level about the interventions of OCMW concerning social assistance.
PensCad	A source with data on individual level about paid first and second pillar pensions.
RKW_RG	A source with data on individual level that indicates whether the individual is a beneficiary child in the employee scheme, in the scheme of guaranteed child benefits and the scheme of civil servants.
RKW_RH	A source with data on individual level that indicates whether the individual grants the right on child benefits in the employee scheme, in the scheme of guaranteed child benefits and the scheme of civil servants.
RKW_BT	A source with data on individual level that indicates whether the individual receives the child benefits in the employee scheme, in the scheme of guaranteed child benefits and the scheme of civil servants.
RSVZ_KB_RG	A source with data on individual level that indicates whether the individual is a beneficiary child in the scheme of self employed.
RSVZ_KB_RH	A source with data on individual level that indicates whether the individual grants the right on child benefits in the scheme of self employed.
RSVZ_KB_BT	A source with data on individual level that indicates whether the individual receives the child benefits in the scheme of self employed.
RVA	A source with data on individual level on people in charge of the RVA. This group comprises among other things the full benefit unemployed but also people on career break or in early retirement.
FAO	A source with data on individual level monitored by the Fonds voor ArbeidsOngevallen.
FBZ	A source with data on individual level monitored by the het Fonds voor Beroepsziekten.
RIZIV	A source with data on individual level of individuals whose disability is recognised by the Geneeskundige Raad voor Invaliditeit (GRI).
FOD_SZ	A source with data on individual level of individuals whose disability is recognised to grant a means tested allowance for disabled people.
LABM97	A source with data on individual level about the labour market status of individuals during the period 1997 until 2008.
RSVZ	A source with data on individual level about people insured as self employed
RSZPPO_QUART	A source with labour market information of individuals monitored by the local and provincial authorities in the fourth quarter of 2008.
RSZPPO_20022008	A source with labour market information of individuals monitored by the local and provincial authorities during the period 2002 until 2008.
RSZ_QUART	A source with labour market information of individuals active as wage earner, except those monitored by the local and provincial authorities, in the fourth quarter of 2008.
RSZ_20022008	A source with labour market information of individuals active as wage earner, except those monitored by the local and provincial authorities, during the period 2002 until 2008.

Some information on the size of the various files is included in Table 8.

The column Total in Table 8 gives the share of all different individuals in a given file as a percentage of the total population.[25] Among other things, one can infer from this column

[24] More information about most of these files can be found in DOCDWH (2011).

[25] The results are weighted by the individual weighting factors. All results that follow from now on, are weighted by these constructed weights.

that for more than half of the population we observe some data on the career info (see the line CIMIRe) or on the labour market history during the period 1997 to 2008 (see the line LABM97).

In the column Total Unique of Table 8 we give the share of the individuals that are only observed in the given file as a percentage of all individuals in the given file. From this column it follows that more than 70% of the children entitled to child benefits and known by the RKW (see RKW_RG line) and over 60% of the children entitled to child benefits and known by the RSVZ (see RSVZ_KB_RG line), only occur in the file with children that generate the right on child benefits.

In the subsequent columns of Table 8 we give the proportion of individuals who occur in the file mentioned in the line and column of the table, expressed as a percentage of the number of individuals in the file mentioned in the line. The data in Table 8 illustrate that many of the children that generate the right on child benefits are observed as well in the RSZ file (see, respectively, the line RKW_RG and column RSZ_20022008 and the line RSVZ_KB_RG and column RSZ_20022008).

From Table 8 it can also be inferred that most individuals in the sample have at least once been registered as employee in the period 1997 to 2008 (see column LABM97). Exceptions occur mainly for children that generate the right on child benefits (see lines RKW_RG and RSVZ_KB_RG and the column LABM97), pensioners (see the line PensCad and column LABM97) and persons with disabilities (see the line LABM97 and column FOD SZ).

It can also be remarked that 64% of the individuals listed in the RVA file are simultaneously observed in the file with quarterly data from the RSZ (see the line RVA and column RSZ_Quart in Table 8).

Table 8: Share of individuals in line- and column data sources as percentage of the number of individuals in the source mentioned in the line

Name file	Total	Total unique	CIMIRe	NIC	PODMI	PensCad	RKW_RG	RKW_RH	RKW_BT	RSVZ_KB_RG	RSVZ_KB_RH	RSVZ_KB_BT	RVA	FAO	FBZ	RIZIV	FOD_SZ	LABM97	RSVZ	RSZPPO_QUART	RSZPPO_20022008	RSZ_QUART	RSZ_20022008
CIMIRe	53.4	4.5	100.0	7.5	1.0	26.8	0.7	15.4	16.4	0.1	1.1	1.1	21.3	2.2	1.1	4.3	7.0	70.4	11.0	5.1	8.2	47.1	60.1
NIC	6.3	0.0	64.0	100.0	2.1	0.9	0.9	30.3	32.8	0.1	0.5	0.6	46.6	9.0	1.1	7.2	2.6	95.5	8.0	6.8	13.1	74.1	89.5
PODMI	1.2	7.7	46.3	11.0	100.0	3.7	16.1	14.0	28.2	0.8	0.3	1.1	26.9	1.8	0.0	1.8	5.2	67.6	3.0	7.1	19.1	20.9	56.1
PensCad	19.2	16.4	74.6	0.3	0.2	100.0	0.2	1.4	1.8	0.0	0.1	0.1	2.7	0.1	2.3	0.7	14.8	31.8	4.1	0.4	2.5	7.9	21.0
RKW_RG	22.8	73.5	1.7	0.3	0.8	0.2	100.0	0.1	1.2	0.7	0.0	0.0	1.2	0.4	0.0	0.1	3.1	22.8	0.3	0.3	3.3	9.1	22.0
RKW_RH	11.3	0.0	73.1	16.9	1.5	2.4	0.1	100.0	37.8	0.0	0.3	0.4	31.6	4.3	0.5	4.0	2.4	95.8	8.0	7.6	11.0	78.9	89.2
RKW_BT	12.6	3.3	69.5	16.3	2.7	2.7	2.1	33.7	100.0	0.0	0.2	0.5	30.2	2.2	0.5	3.6	2.5	86.4	7.6	8.6	13.5	64.0	77.4
RSVZ_KB_RG	2.0	62.2	2.2	0.2	0.5	0.0	8.2	0.1	0.1	100.0	0.4	0.9	1.0	0.3	0.0	0.1	2.3	28.6	1.0	0.2	2.6	10.8	28.0
RSVZ_KB_RH	1.0	0.0	60.7	3.1	0.3	1.8	0.1	3.6	3.2	0.9	100.0	77.3	1.6	0.0	0.0	1.8	0.8	36.3	97.6	0.5	1.6	4.8	22.7
RSVZ_KB_BT	1.0	1.4	61.4	3.9	1.3	2.9	0.1	4.6	6.0	1.8	77.1	100.0	3.2	0.2	0.1	1.9	1.2	39.5	87.7	0.6	2.2	7.9	26.2
RVA	15.7	0.4	72.2	18.6	2.1	3.3	1.7	22.6	24.2	0.1	0.1	0.2	100.0	4.5	0.8	1.3	2.3	92.2	3.9	4.6	10.8	64.2	84.3
FAO	1.8	0.0	65.2	30.9	1.2	1.2	5.2	26.7	15.4	0.3	0.0	0.1	38.5	100.0	0.6	1.8	1.5	99.8	4.3	1.4	6.4	95.3	99.7
FBZ	0.7	0.0	83.8	10.1	0.1	63.6	0.6	8.7	9.0	0.0	0.1	0.1	18.6	1.5	100.0	6.4	19.3	85.0	2.5	3.0	5.4	80.8	82.4
RIZIV	2.7	0.9	85.2	16.7	0.8	5.2	1.0	16.6	17.0	0.1	0.6	0.7	7.5	1.2	1.6	100.0	32.6	67.5	8.4	2.2	4.9	34.8	54.0
FOD_SZ	5.7	4.0	65.7	2.9	1.1	50.3	12.4	4.7	5.5	0.8	0.1	0.2	6.3	0.5	2.3	15.6	100.0	25.4	2.3	0.9	1.9	12.9	19.1
LABM97	55.8	0.1	67.3	10.8	1.5	10.9	9.3	19.3	19.5	1.0	0.6	0.7	26.0	3.3	1.0	3.3	2.6	100.0	9.0	6.6	12.3	65.3	88.0
RSVZ	8.9	9.8	65.7	5.6	0.4	8.7	0.8	10.1	10.8	0.2	10.6	9.6	6.9	0.9	0.2	2.5	1.5	56.2	100.0	2.4	4.7	24.5	45.2
RSZPPO_QUART	3.7	0.0	73.9	11.7	2.3	1.9	2.2	23.3	29.7	0.1	0.1	0.2	19.9	0.7	0.6	1.6	1.3	100.0	5.8	100.0	100.0	13.5	45.0
RSZPPO_20022008	6.8	0.0	64.1	12.0	3.4	7.1	11.0	18.1	24.9	0.8	0.2	0.3	24.8	1.7	0.5	1.9	1.6	100.0	6.1	53.5	100.0	31.4	59.7
RSZ_QUART	36.5	0.0	69.0	12.8	0.7	4.1	5.7	24.3	22.2	0.6	0.1	0.2	27.7	4.8	1.5	2.6	2.0	100.0	6.0	1.4	5.9	100.0	100.0
RSZ_20022008	49.1	0.0	65.3	11.5	1.4	8.2	10.2	20.4	19.9	1.2	0.4	0.5	27.0	3.7	1.1	3.0	2.2	100.0	8.2	3.4	8.3	74.2	100.0

In the remainder of this section we briefly sketch the contents of the available files and describe the manipulations carried out to rearrange these different files in such a way that they can be used as input for the simulation model.[26]

A general rule followed in the processing of the different files, is that per individual only one line was retained.[27] In some files it appeared that individuals were observed with multiple observation lines. This appeared among other things for individuals with multiple social security rights in a given domain during the year or for individuals with multiple jobs as employee or multiple statuses as self employed.

This line reduction was usually done by taking the characteristics of the line with the highest income amounts and aggregating the nominal variables, such as labour or replacement incomes, over all the observation lines of a given individual. Where it was desirable multiple rights were retained by converting observation lines into column variables. This was done for pensions to be able to distinguish retirement and survivors' pensions. Also for sickness and disability benefits the distinction was made between disability benefits and maternity benefits.

Other aspects on the file treatment are discussed file by file or by a group of files, in what follows.

2.4.1 SOURCES WITH DATA ON CHILD ALLOWANCES (RKW_RG, RKW_RH, RKW_BT, RSVZ_KB_RG, RSVZ_KB_RH, RSVZ_KB_BT)

These files contain essentially only variables with individual identification numbers. These identification numbers allow us to identify who is the child that generates the child benefit, who is the beneficiary of the child benefit right (i.e. the individual that opens the right on child benefits) and who is the recipient of the child benefits (i.e. the individual that receives the child benefits).

Since we work with a sample of data, it happens that either data on the beneficiary or the recipient of a child that is in the sample, are outside the sample. In these cases we observed in the identification variables of the beneficiary or the recipient the value "niet in steekproef".

Besides the identification numbers, the child benefit data files also include an identifier that indicates whether benefits were paid within the system of guaranteed benefits.

The files with information on the child benefits do not contain variables that identify the rank or statuses that determine either the right on a social supplement or single parent supplement. When computing the child allowances later on, these concepts will therefore have to be reconstructed.

---

[26] In this section we can obviously not treat the detail of all the data manipulations performed. For this we refer to the programs in Federal Social Security (2011b) and Federal Social Security (2011c).

[27] The documentation of observation units used in different base files can be found in DOCDWH (2011).

2.4.2 SOURCES WITH ALLOWANCES FOR SICKNESS, DISABILITY, INDUSTRIAL ACCIDENTS AND OCCUPATIONAL DISEASES (NIC, RIZIV, FAO EN FBZ)

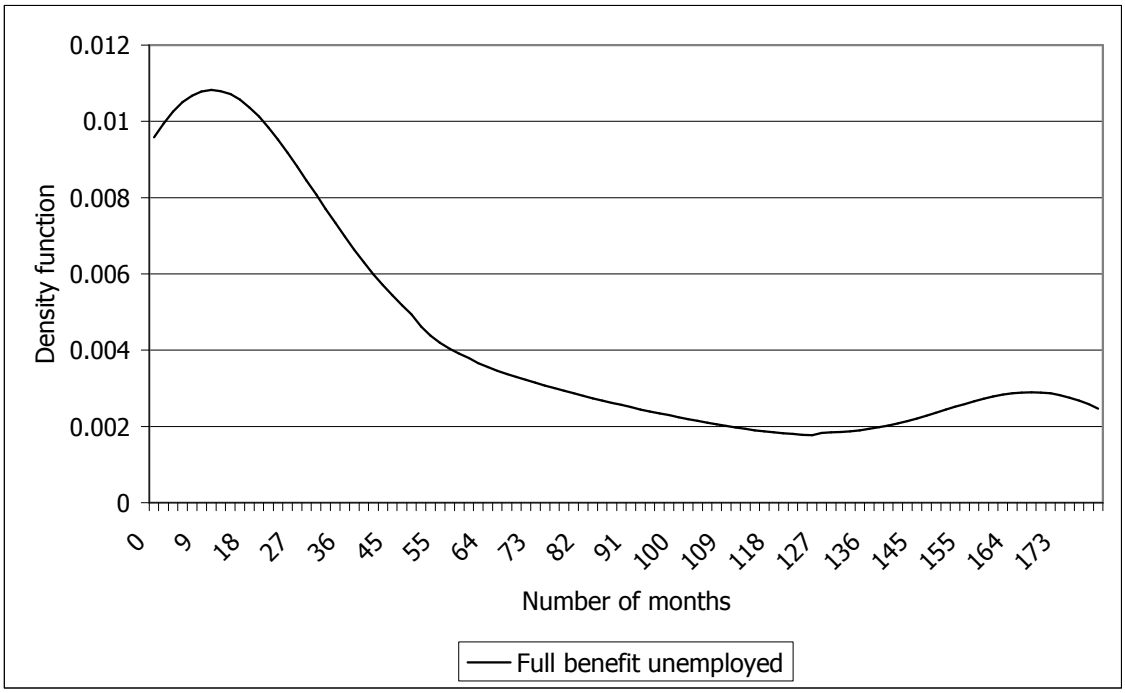
In these files, we observe for the year 2008 a) some non-monetary characteristics of the rights that were opened due to the cited diseases, b) an indication of the period in which the payment was done and c) an indication of the extent of the amount paid. All amounts are usually expressed in classes of 10 Euro. Time indications are, as for all other sources, available in the form of a month and year registration. Hence, the exact date when the right or the payment starts or stops, is never known.

2.4.3 SOURCES WITH ALLOWANCES FROM THE RVA (RVA)

The database of those in charge of the RVA in 2008 includes a variable that allows to characterize the person who receives the benefit and the number of months during which the person receives the benefit. The amount paid, divided into classes of 10 Euros, is also included in the file.

With these data it is possible to identify the full benefit unemployed (FBU). We draw, in Figure 1, the density function of the number of months during which members of this FBU group received a benefit.[28] This figure shows that between 12 and 136 months there is a steady outflow from the unemployment benefit system. After 136 months, there is an accumulation of very long-term unemployed.

Figure 1: Density function of the number of months during which a full benefit unemployed receives unemployment benefits



[28] All density functions in this publication are estimated using the SAS procedure, KDE, option Univar and the default method for estimating the bandwidth (i.e. the Sheather Jones Plug In method) (see SAS (2011)). All results are always weighted by the individual weighting factors.

2.4.4 SOURCE WITH PAID PENSIONS (PENSCAD)

The pension register file (Penscad) contains data on pensions paid in 2008. Besides the amounts themselves, again divided into classes of 10 Euro per year, the file also contains per right, the date, expressed in months and years, when payment of the right started. Based on these data it is possible to distinguish payments for retirement and survivors' pensions, respectively, in the systems of employees, self-employed and civil servants. It can also be identified, who during the year was entitled to an income guarantee for the elderly (IGO).

In Table 9 we present the weighted amount of individuals, differentiated according to retirement status, and compare these results with those from an external source.

It is striking that for all lines the size of the group of beneficiaries is approximated quite well, but that both the size of those receiving a retirement or a survival pension as employee are underestimated as compared to the external source. On the contrary, the number of beneficiaries with retirement or survival pensions as self employed or civil servant are slightly overestimated.

Table 9: Number of beneficiaries that receives a retirement- or survival pension in comparison with an external source[29]

	Number of beneficiaries on 31 December 2008 in external source	Weighted amount of beneficiaries based on sample observations	Percentage difference between weighted and real amount
Retirement pension employee	1,442,487	1,403,737	-2.7
Retirement pension self employed	440,907	449,792	2.0
Retirement pension civil servant	393,634	410,637	4.3
Survival pension employee	491,860	460,909	-6.3
Survival pension self employed	145,301	149,765	3.1
Survival pension civil servant	117,142	120,679	3.0

2.4.5 SOURCE WITH DATA ON DISABLED PEOPLE WITH MEANS TESTED BENEFITS (FOD SZ)

With the 2008 data, the individuals can be identified for whom their disability and/or entitlement to means tested benefits is recognised. Besides the actual amounts paid, the file also contains a number of features that allow to characterize the type of benefit.

In Table 10 we give the estimated number of beneficiaries with an Integration benefit and/or Income Replacement Allowance (IT/IVT) or the Allowance for assistance to the elderly (THB) The comparison of the weighted sample data with those of an external source, shows that we overestimate the size of the group of beneficiaries on an IT/IVT allowance by about 5% and that we underestimate the size of the group of people with THB by about 1%.

[29] For producing the data in this sample we use the variable soort_pensioen (see also the Appendices of Berghman et al (2007)). The external data used here are from the basistoepassingen of the DWH LM&SP (see KSZ (2011)).

Table 10: Number of beneficiaries of Integration benefit of Income Replacement Allowance or the Allowance for assistance to the elderly[30]

	Number of beneficiaries on 31 December 2008 in external source	Weighted amount of beneficiaries based on sample observations	Percentage difference between weighted and real amount
Entitled to Integration benefit of Income Replacement Allowance	143,037	149,734	4.7
Entitled to Allowance for assistance to the elderly	130,455	129,273	-0.9

2.4.6 SOURCE WITH MEANS TESTED SOCIAL BENEFITS (POD MI)

Apart from data on the leefloon and the equivalent leefloon, the 2008 file with the means tested social benefits, also contains a number of variables that contain information on a number of the other interventions made by the OCMW's (i.e. the local authorities that govern these social benefit interventions). For the further development of the module that simulates these social benefits, we only focus on the data that cover the leefloon and equivalent leefloon.[31] Next to a number of characteristics that allow to identify these rights, the file with 2008 data also contains the amounts paid, divided into classes of 10 Euros per year.

As shown in the line of Belgium of Table 11, the total number of cases with a leefloon or equivalent leefloon is approximated quite well with the weighted sample data. However, this aggregate figure hides some regional variation, as is also illustrated in Table 11. Especially the underestimation of the total number of cases in Brussels and the overestimation of the total number of cases in Henegouwen, is striking.

[30] The data from the external source come from FOD Sociale Zekerheid (2010b).
[31] These selection rules are specified more precisely in Knapen (2011).

Table 11: Weighted number of individuals with right on the Leefloon or equivalent Leefloon benefit[32]

Region	Province	Real number observed in external source	Weighted amount based on sample observations	Percentage difference between weighted and real amount
Brussels		29,393	24,080	-18.1
	Antwerpen	9,896	9,624	-2.7
	Limburg	2,483	2,945	18.6
	Oost-Vlaanderen	8,919	8,296	-7.0
	Vlaams-Brabant	4,262	3,674	-13.8
	West-Vlaanderen	4,483	6,082	35.7
Flanders		30,043	30,621	1.9
	Henegouwen	15,798	20,429	29.3
	Luik	18,346	18,769	2.3
	Luxemburg	1,918	2,467	28.6
	Namen	4,742	6,358	34.1
	Waals-Brabant	2,232	1,857	-16.8
Wallonia		43,035	49,881	15.9
Belgium		102,471	104,582	2.1

2.4.7 SOURCE WITH DATA ON SELF EMPLOYED (RSVZ)

For people, insured as self employed in 2008, we observe information on their contribution status and the annual income amount, divided into classes of 10 Euro, that serves as the basis for determining the contributions they have to pay as self-employed.

Usually we do not impute income amounts in the context of this exercise. However, for those insured as self employed we make an exception to this rule, since for certain self employed a registered income amount is lacking. We do not impute an income amount for all cases with a lacking observation. For spouses helping another self employed, for instance, it might be that we observe a code that indicates that the income of both spouses is captured by the registered income amount of only one of the two spouses. There are also codes that point to the fact that the self employed has put his or her activities on hold, but that he or she maintains the status of "insured as self employed" during this period.[33] For these groups no income is imputed. The cases for which we impute income, are self employed that start their activity. For these cases one normally does not observe income, the first 3 years of their acitivity, although they presumably have earned income during this period.

To impute missing income amounts, we first estimate the average of the observed incomes per group of people with the same contribution code and belonging to the same age class. If possible we impute this average for cases with a missing value but with a matching value for the contribution code and the age class. If no imputation according to this rule was possible, we try to impute the average of people with a matching contribution code, then with a matching age class value and finally the overall average is imputed. In Table 12 we give an

[32] The data from the external source come from POD Maatschappelijke Integratie (2011).
[33] The exact codes to identify these cases are listed in Knapen (2011).

overview of the number of persons for whom an income amount was imputed, according to the various rules.

Table 12: Rules used to impute income of self employed

	Weighted amount of self employed in the sample
No income observed and no income imputed	136,646
Income observed	606,753
Imputed income is the mean of the observed incomes of the individuals age and contribution group	154,065
Imputed income is the mean of the observed incomes of the individuals contribution group	17,185
Imputed income is the mean of the observed incomes of the individuals age group	6,853
Imputed income is the global mean	250

2.4.8 SOURCE WITH DATA ON WAGE AND LABOUR TIME DATA FOR THOSE WHO PAY SOCIAL SECURITY CONTRIBUTIONS TO THE RSZ IN THE 4TH QUARTER OF 2008 (RSZ_QUART)

The file RSZ_Quart contains for the 4th quarter of 2008, wages and labour time details of those who have to pay contributions to the RSZ (i.e. employees in the private sector and those in the public sector at the federal level). In this file individuals can appear with multiple employment lines (employment lines cover activities for the same employer with a similar status). These employment lines can be grouped in employee lines (these are all employment lines with the same contribution regime) and into jobs (these are all employment lines with the same value for the Paritair Comité).[34]

For later use in the model it is of particular importance to indentify the contribution regime and the reductions applied on these contributions. For further processing of the data, we only retain one contribution regime per individual. The reduction of multiple lines per individual is done here by retaining the characteristics of the most important job. Values of variables that can be aggregated over jobs, like income earned and time performed, are aggregated and linked to the characteristics of the most important job. As compared to the base file, this means a loss of information, but all in all this loss should be limited.

The contribution regime is mainly determined by the employee and employer class. About 95% of all individuals appear in the base file of the 4th quarter of 2008 with only one employee and one employer code.

For what the reduction codes is concerned, we do observe for all individuals one code at maximum that captures the reduction of the personal contributions. About 99.9% of all individuals have at maximum 3 different codes that capture the reductions of the contributions of the employer. Therefore we make three variables that capture the possible employer reductions codes per individual.

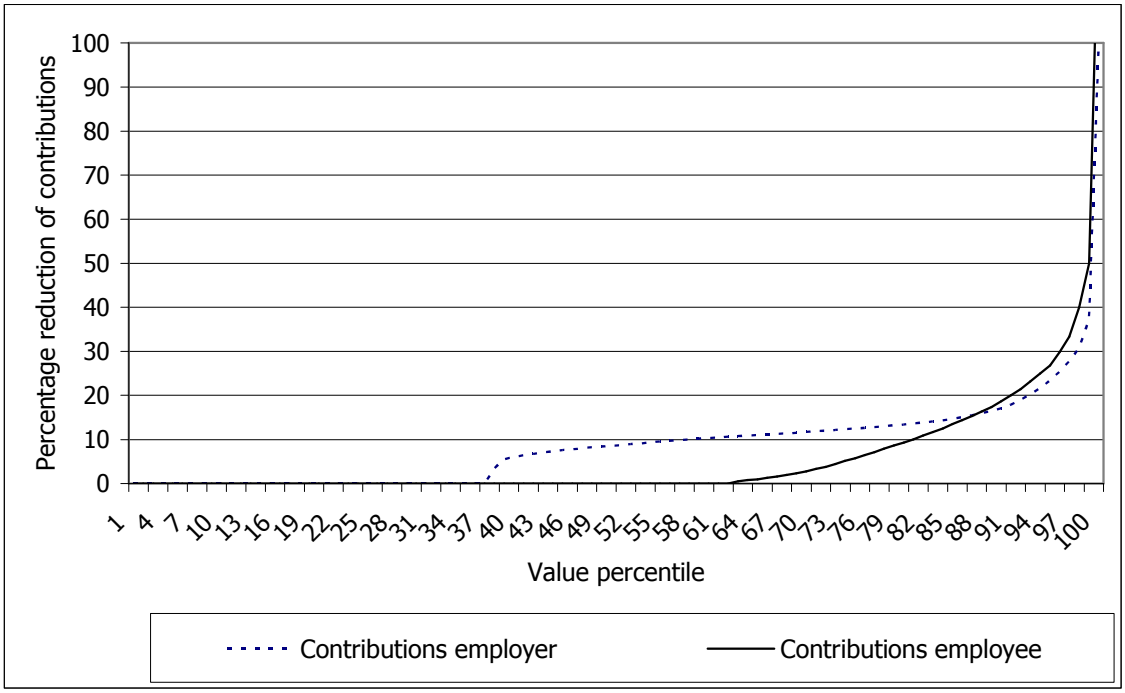
[34] More information about these concepts can be found in DOCDWH (2011), documents "fiches, beschrijving RSZ, Nederlands, versie 3-1-2011.doc" and "fiches, beschrijving RSZPPO, Nederlands, versie 20-5-2011.doc".

Apart from these reduction codes, we also observe that amounts of contributions due before reductions and the reductions that can be applied on it.[35]

In Figure 2 we give additional information on the distribution of these amounts. First, we expressed the reductions as a percentage of the contributions due before reduction. Then the distribution of the constructed percentages was ordered from low to high and divided into percentiles. In Figure 2 we print the average value of the percentage reduction of employer and employee contributions, per percentile.

One can deduct from Figure 2 that about 2/3 (start at 38th percentile) of the observed cases, enjoy a reduction in employers' contributions. For the vast majority of these cases, i.e. the 38th to 97th percentile, this reduction is below 20% of the contribution amount due before reduction. Approximately one third of the observed cases are entitled to a reduction on the personal contributions. Here again the bulk of the reduction percentages is below 20%.[36]

Figure 2: Percentiles of percentage reduction of employers and employees contributions in the RSZ data base



2.4.9 SOURCE WITH DATA ON WAGE AND LABOUR TIME DATA FOR THOSE WHO PAY SOCIAL SECURITY CONTRIBUTIONS TO THE RSZPPO IN THE 4TH QUARTER OF 2008 (RSZPPO_QUART)

The file RSZPPO_Quart contains data that are similar to those in the file RSZ_Quart file, but for those active in the local public sector (i.e. not at the federal level). Again, multiple jobs

---

As is the case for the majority of the nominal amounts available in the sample, these data are classified in income classes of 10 Euros. In this case the boundary is applied on a quarterly basis.

[36] The reduction rates expressed here are based on the proportion of all reductions and all contributions in the given quarter. It is therefore possible that this distribution differs from the one at the level of the jobs.

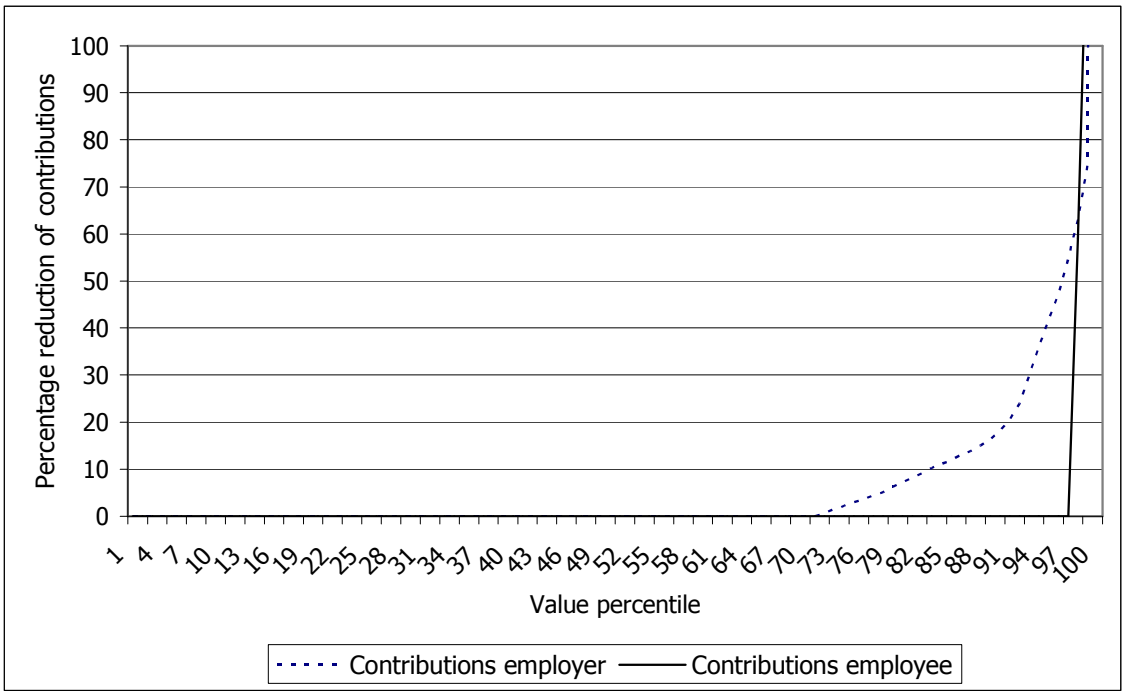
per person can occur. A job is basically defined in the same way as within the RSZ-file, but to reduce the information we were forced to work in a different way as for the RSZ-data since for the RSZPPO-data we did not dispose of the same identifiers as for the RSZ. In case of the RSZPPO, we selected the most important job as the job that still existed on the last day of the quarter and generated the largest income amount.[37]

Again, the loss of information because of reducing the observation lines, should be limited. Over 95% of the individuals appear with only one employer and employee code in the file of the 4th quarter of 2008. Hence, it seems defensible to assume that individuals in a given quarter are generally in a single contribution regime.

For the contribution reductions, we are obliged to follow another approach as for the RSZ-data as well. Codes that capture these reductions, are available in the RSZPPO file, but were not included in the data demand, and therefore not available for further analysis.

To model these reductions, we could rely on the amounts for those who pay their contributions to the RSZPPO. In Figure 3 we plot, as for the RSZ data, by percentile the average percentage reduction of employer and employee contributions. The resulting distribution is in this case very different from the one based on the RSZ data. Of the observed cases in the file, about 30% of those paying contributions to the RSZPPO are entitled to a reduction of the employers' contributions and only 1% to a reduction in the employee contributions.

Figure 3: Percentiles of the percentage reduction of employers and employee contributions in the RSZPPO data base



[37] As compared to what we did with the RSZ data, we follow a different rule here, because the RSZPPO complement of the variable we used at RSZ level, was not included in the data demand. When we apply both criteria for the RSZ data, we find that in 9.5% of the cases the two criteria did not result in a selection of the same line.

2.4.10 SOURCES WITH DATA ON LABOUR MARKET HISTORY OF WAGE EARNERS (LABM97, RSZPPO_20022008, RSZ_20022008 EN CIMIRE)

For all individuals in the sample, which have been active as wage earner in the period 1997 to 2008, we can identify whether they were mainly active in the private or in the public sector. For individuals that were mainly active on the private labour market, we can discern whether they mainly had a worker or an employee status. For those active in the public sector we can distinguish civil servants from those in contractual employment.

The files RSZ_20022008 and RSZPPO_20022008 contain much less variables than the files RSZ_Quart and RSZPPO_Quart, but still allow to quantify some aspects of the labour market activity of wage and salary earners for the years 2002 to 2008. The variables that were identified for each of the years 2002 to 2008, are listed in Table 13.

Table 13: Abbreviations and descriptions of variables that quantify the labour market activity of wage earners for the years 2002 until 2008[38]

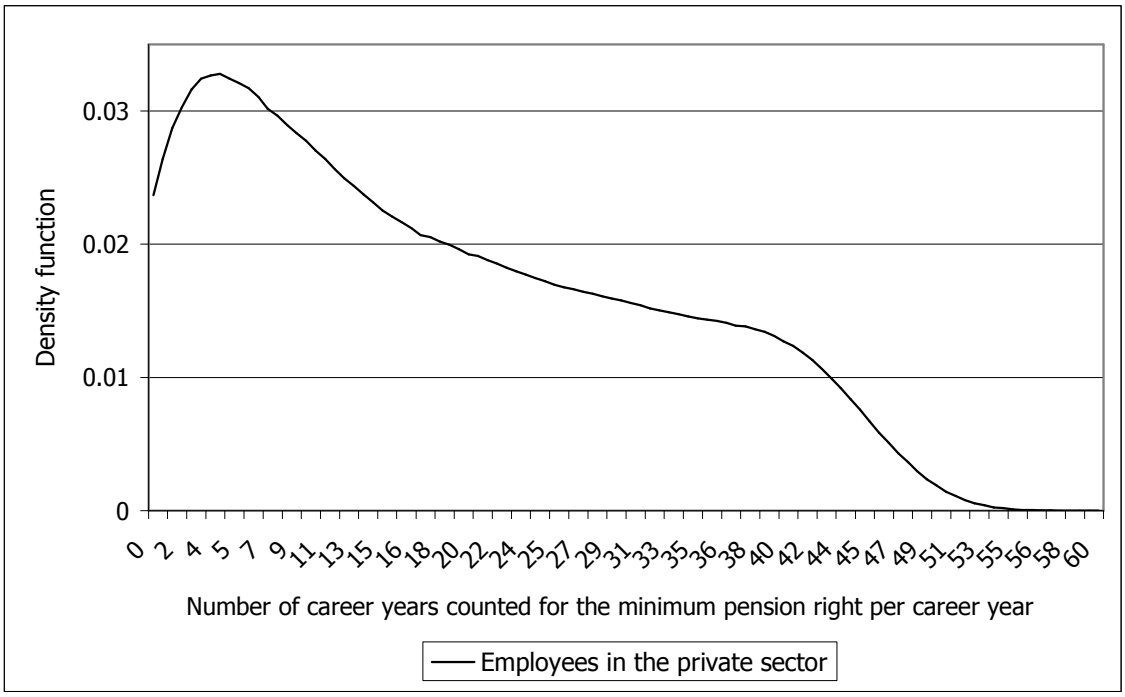
Name variable	Description
ASSIMDAYS	Number of assimilated days during the year
REMDAYS	Number of remunerated days during the year
REMHOURS	Number of remunerated hours during the year
REMUN	Total amount of labour income on annual basis (in classes of 10 Euro)
DAYS_FULL	Number of days full time work during the year
DAYS_PART	Number of days part time work during the year
LABMSTAT_PRIV	Most important status on the private labour market during the year (worker, employee)
LABMSTAT_PUB	Most important status on the public labour market during the year (contractual, civil servant)
SALJRS	Daily wage

The first 4 variables in Table 13 could also be completed for the year that individuals were active on the private labour market in the period 1954 to 2001.

With these historic data, we can among other things sketch the number of career years that result in a minimum pension right per career year. Figure 4 gives the density function of the number of these years per individual on the basis of the CIMIRe data.

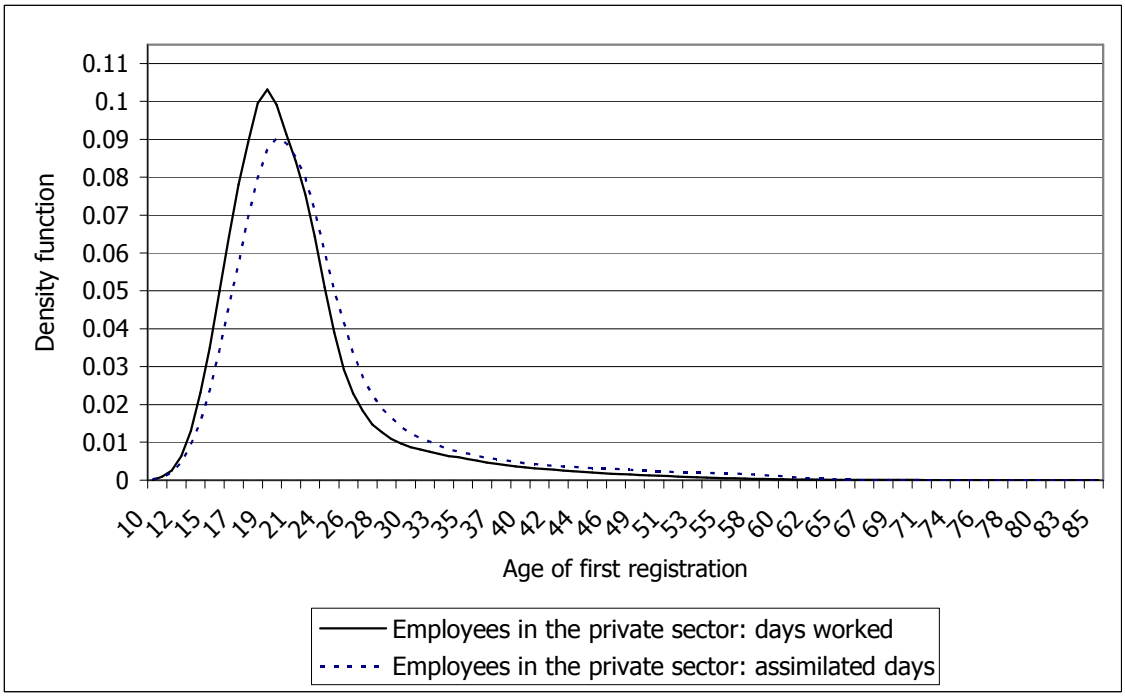
[38] For the further elaboration of these concepts, we rely on the work described in Mutsaerts (2011).

Figure 4: Density function of the number of career years that are counted for the minimum pension right per career year for employees in the private sector.



With these CIMIRe data we can also examine what the age of first labour market contact was. In Figure 5 we plot the density function of the ages on which observed individuals first entered the labour market and on which they had their first registration of assimilated days.

Figure 5: Density function of age of the first labour market experience and first assimilation of days for employees in the private sector



As shown in Figure 5, most individuals perform their first labour market activity between the age of 15 and 28, with a peak around 20-21 years. Ages on which the first assimilated days are registered, are distributed in a similar way as the ages of the first labour market activity, but the peak of the assimilated day's age is less high and the first assimilation occurs in general at a slightly higher age than the first performance.

2.4.11 CONSTRUCTION OF THE LAST EARNED DAILY WAGE

A variable on which we have direct observations for some individuals and for which we reconstruct a value for some other individuals, is the last earned daily wage. These values are stored under the variable name SALJRS_LAST_CORR. This variables is important for analyses later on, since it will serve as input for the recalculation of certain benefits and because it is an identifier of the earning power of individuals.

If either the RSZ or RSZPPO provides an observed value of the daily wage, this value is generally produced by dividing the remunerations paid by the labour time performed during the same period.[39] In what follows we call this variable daily wage concept 1.[40]

With information of the latest year for which wages and working time data are available, we can also calculate own daily wage concepts. This was done in two ways:

- The observed remunerations were divided by the number of days for which remunerations have been paid (daily wage concept 2),
- The observed remunerations multiplied with 7.6 divided by the number of hours worked (daily wage concept 3).

The latter two concepts are calculated only under certain conditions. An individual must have been active at least 24 days per year or 182 hours (i.e. 4 weeks in a 6 days week or the hour's equivalent with days of 7.6 hours per day). The observed performance time should not exceed 52*6 days or 52*6*7.6 hours. The wage and labour time information of the lines that do not comply with these rules, were not used and the constructed daily wage concepts were set equal to 0.

For about 71% (7,635,230 units at a total of 10,753,080) of the weighted number of units in the sample, we observe wages and working performance data in at least one year ranging from 2008 to 1954.

For approximately 60% (4,646,508 out of 7,635,230 units) of this group we observe daily wages that are directly delivered by the RSZ or RSZPPO in 2008. For 9% of all cases there is no wage for 2008 but still an observed daily wage concept that is directly delivered by the RSZ or RSZPPO for a year preceding 2008.

For all cases for which the daily wage concept was directly observed (i.e. daily wage concept 1), SALJRS_LAST_CORR was completed on the basis of this concept. For other

---

[39] More information about these concepts can be found in DOCDWH (2011), documents "fiches, beschrijving RSZ, Nederlands, versie 3-1-2011.doc" and "fiches, beschrijving RSZPPO, Nederlands, versie 20-5-2011.doc".

[40] Unlike the other nominal amounts within our sample, this daily wage concept is not divided into classes.

cases, first daily wage concept 2 was selected and if this concept was not available daily wage concept 3 was selected, if available. If the SALJRS_LAST_CORR variable was completed on the basis of data preceding 2008, the observed value was revalued on the basis of revaluation coefficients included in Appendix 4 of this document.

After these interventions we check whether the value of SALJRS_LAST_CORR has no outliers compared to the observed values of 2008. We consider a value as an outlier if the reconstructed value is greater than the mean plus two times the standard deviation of the observations of 2008. Concrete, the upper limit used is equal to 309.93 Euros.[41] If the daily wage was identified to be an outlier according to concept 1 but not according to concept 2 or 3, the value of SALJRS_LAST_CORR was replaced by one of these latter values.

The daily wage concept, that we obtained after the previous steps, turned out to be equal to 0 in about 3% of the cases where labour market information is observed. In some cases we also must go back into time quite far to construct a value for SALJRS_LAST_CORR. Whether the value is missing or possibly outdated is of particular importance for those who are still potentially active on the labour market since for labour market simulations of this group the reconstructed daily wage will be brought into action. Therefore we report in Table 14 data on the reconstructed daily wage for the group of individuals who are 18 years or older and younger than 65.

Table 14 provides more information on the source (i.e. the source that was used to recover the observation), the performance year (i.e. the year for which the values that were used, have been registered) and the height of the reconstructed daily wage.

For the source we distinguish five possible values:

- Either the observation is no outlier and based on data from the RSZ, RSZPPO or CIMIRe. The value for the source is, in these cases, the abbreviation of the file.
- Either it was not possible to construct a daily wage. The value for the source in this case is then "Unknown".
- Either the observed value is greater than 309.93 Euros. The value for the source in this case is then "Aftop".

When analyzing the amount of the estimated daily wage, we focus on four areas:

- Either the estimated daily wage is equal to 0.
- Either the estimated daily wage is less than the minimum daily wage. The limit value we use is 57.97 Euro.[42]

[41] The average of the daily wage that we observe for 2008 is 122.05 Euro and the standard deviation is 93.94 Euro. This implies that we classify outliers from 309.93 Euro on.

[42] To estimate the minimum daily wage we take the guaranteed average minimum monthly salary (GGMMI) of a 22 year old with 1 year seniority. In 2008, the GGMMI was 1360.71 Euros (see Put (2008), p. 1043). This amount is multiplied by 12/13 assuming that this amount contains the monthly equivalent of a end of the year bonus. Subsequently, this amount is divided by (52 * 5) / 12 or 21.66667, the assumed number of working days per month in a full-time regime.

- Either the estimated daily wage is greater than or equal to the minimum daily wage but less than the limit value for outliers (i.e. 309.93 Euro).
- Either the estimated daily wage is greater than or equal to the limit value for outliers.

For later analyses it is particularly important to examine to what extent values below the minimum wage or above the maximum occur. If this happens either these values will be increased later to the minimum daily wage or capped at the upper ceiling. To determine whether these interventions differ among the various subpopulations with a replacement income or a social assistance benefit, we analyse the reconstructed daily wage for the groups of beneficiaries known in the various files with data on the subpopulations.

Table 14: Origin, performance year and height of the estimated daily wage (EDW) for individuals known in source with replacement incomes or social benefits and with age 18 or older but younger than 65

	POD_MI		NIC		PensCad		RVA		FAO		FBZ		RIZIV		FOD SZ	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Source EDW																
Aftop	31	0.0	1,699	0.3	4,225	1.3	4,711	0.3	431	0.2	107	0.4	526	0.2	319	0.1
CIMIRe	20,184	20.7	21,127	3.4	110,944	33.3	204,207	12.7	143	0.1	7,512	26.3	123,241	47.2	132,488	55.8
Unknown	16,115	16.6	6,063	1.0	19,963	6.0	28,166	1.8	74	0.0	142	0.5	11,183	4.3	26,590	11.2
RSZPPO	8,325	8.6	43,657	7.0	28,554	8.6	84,436	5.3	515	0.3	2,191	7.7	9,157	3.5	7,190	3.0
RSZ	52,737	54.2	548,587	88.3	169,403	50.9	1,285,532	80.0	184,034	99.4	18,579	65.1	117,097	44.8	70,661	29.8
Total	97,392	100.0	621,133	100.0	333,089	100.0	1,607,053	100.0	185,197	100.0	28,531	100.0	261,203	100.0	237,248	100.0
Performance year EDW																
Value is 0	16,147	16.6	7,762	1.3	24,188	7.3	32,877	2.1	505	0.3	249	0.9	11,708	4.5	26,909	11.3
Before 1960	0	0.0	0	0.0	233	0.1	0	0.0	0	0.0	0	0.0	0	0.0	126	0.1
Before 1960 until 1969	1,311	1.4	0	0.0	13,452	4.0	339	0.0	0	0.0	53	0.2	517	0.2	3,379	1.4
Before 1970 until 1979	3,643	3.7	552	0.1	27,658	8.3	6,531	0.4	53	0.0	249	0.9	12,174	4.7	18,059	7.6
Before 1980 until 1989	5,115	5.3	3,570	0.6	24,704	7.4	34,287	2.1	0	0.0	1,647	5.8	27,610	10.6	26,273	11.1
Before 1990 until 1999	6,878	7.1	9,818	1.6	35,697	10.7	108,747	6.8	89	0.1	4,347	15.2	64,283	24.6	52,840	22.3
Before 2000 until 2007	25,686	26.4	69,352	11.2	136,016	40.8	354,901	22.1	1,610	0.9	7,142	25.0	107,177	41.0	67,741	28.6
2008	38,613	39.7	530,078	85.3	71,141	21.4	1,069,371	66.5	182,939	98.8	14,842	52.0	37,735	14.5	41,922	17.7
Total	97,392	100.0	621,133	100.0	333,089	100.0	1,607,053	100.0	185,197	100.0	28,531	100.0	261,203	100.0	237,248	100.0
Height EDW																
EDW is 0	20,317	20.9	10,206	1.6	26,699	8.0	48,597	3.0	74	0.0	794	2.8	21,591	8.3	68,404	28.8
EDW < minimum	18,195	18.7	12,977	2.1	55,011	16.5	68,022	4.2	1,621	0.9	1,123	3.9	38,772	14.8	49,280	20.8
minimum < EDW < allowed maximum	58,849	60.4	596,252	96.0	247,154	74.2	1,485,723	92.5	183,070	98.9	26,506	92.9	200,315	76.7	119,246	50.3
EDW >= allowed maximum	31	0.0	1,699	0.3	4,225	1.3	4,711	0.3	431	0.2	107	0.4	526	0.2	319	0.1
Total	97,392	100.0	621,133	100.0	333,089	100.0	1,607,053	100.0	185,197	100.0	28,531	100.0	261,204	100.0	237,248	100.0

For what the source is concerned, we can conclude from Table 14 that capping of values above the upper limit happens all in all not very often. For all files, except for individuals known in the RIZIV or FOD SZ files, more than 60% of the estimated daily wages was recovered from files of the RSZ or RSZPPO. For individuals known by the RIZIV or FOD SZ, the daily wage is, to a relatively large extent, constructed on the basis of CIMIRe data.

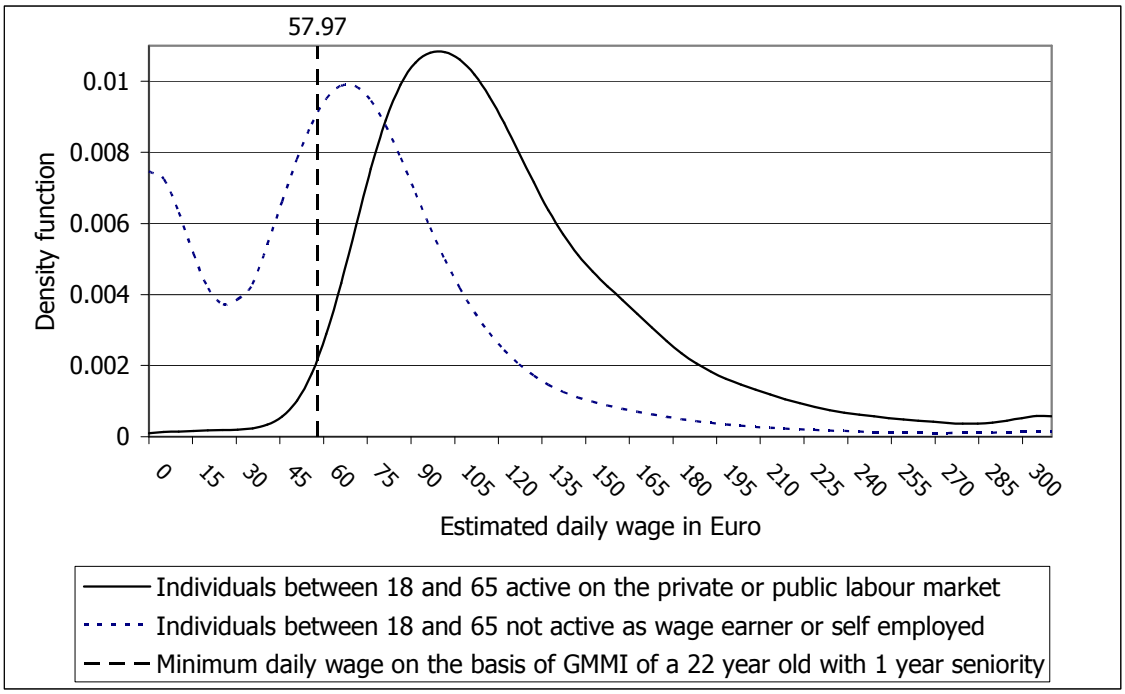
The results in Table 14 also show that the number of cases for whom the daily wage is reconstructed with records of performance years before 1990, is relatively limited. Exceptions are again the files of individuals known by the RIZIV and FOD SZ and the file with pensions paid. In these cases, more than 15% of the observations turn out to have observation years before 1990. For individuals in other files, even those known by the RIZIV and in the Pension Register, the share of the last 9 performance years, i.e. 2000 to 2008, was consistently higher than 50%. Especially for the population known by the FOD SZ, the share of recent employment contacts is limited as compared to other populations.

For the population known by the POD MI and the FOD SZ, the proportion of estimated daily wages below the minimum threshold is relatively high. This is both a consequence of a relatively high proportion of zeros (i.e. not reconstructable values) as of a significant proportion of estimated values below the limit.

Based on the source, the performance year and the amount of the estimated daily wage one can, not surprisingly, conclude that the files of the NIC, the RVA, the FAO and FBZ contain subpopulations that are close to the labour market as compared to the populations of the POD MI and FOD SZ.

Finally, we draw in Figure 6 the density function of the estimated daily wage for the group of individuals 18 to 65 years old, that respectively, a) performed paid employment in the last quarter of 2008 and b) were not active in paid employment in the last quarter of 2008 or registered as self-employed in 2008.

Figure 6: Density function of the estimated daily wage for wage earners active on the labour market and non active individuals



From Figure 6 one can conclude that within the group of people that are not active on the labour market, those with low earning power are much more present than within the group of active people. This figure also suggests that if the non active group should be reactivated, the vast majority of this group will enter at a wage that is somewhere between the minimum wage and twice the minimum wage (i.e. 115 Euros).

3 How MIMOSIS works

The input file of the model can be seen as a matrix with 299,198 lines and many columns. The columns contain variables that are completed with information from the various sources, discussed in the previous section. The data from this input file are the exogenous variables of the model. These exogenous variables together with a set of parameters are read in by the model. The parameters are elements from the different legislative areas covered by the model and constructed in such a way that they can be adapted in a flexible way. For the unemployment benefits, for example, these parameters are the percentages and the upper and lower ceilings applied on that last earned income, differentiated by family type. The combination of a number of exogenous factors, such as a typology of the unemployed and his last earned wage with the parameters allows us to model the endogenous variable "unemployment benefits". Because the parameters can be changed easily, various scenarios can be simulated relatively easily.

The source code that processes the exogenous data and the parameters, is written in FORTRAN. The conversion of data from the DWH LM&SP into an input file for the model, is done with SAS.

To keep an overview over the various parts of the model, the source code is spread out over different modules. In this section we can not go into the detail of each module.[43] We try to give a snapshot of what each module does and the output generated by each module. Table 15 gives us a list of the 9 module names that make up the MIMOSIS model, and a brief description of what these modules do.

Table 15: MIMOSIS-module names and short description

Module name	Brief description
FAMREL	Marking off of families (i.e. smaller units within the sociological household) and determination of relations between household members
PENSWELF	Welfare adaptations of pensions but no recomputation of pensions according to the pension rules
FAMAL	Computation of family allowances
CONTRIB	Computation of contributions on income from labour and withholdings on replacement income
PIT	Personal income taxes paid by fiscal unit
SICK	Computation of sickness and disability benefits, allowances for industrial accidents and occupational diseases
UNEM	Computation of unemployment benefits, career break benefits and early retirement pensions
SOCBEN	Computation of means tested benefits (subsistence minima, IGO and means tested allowances for disabled people)
EVAL	Evaluation of the simulated output (computation of budgetary aggregates and distributional indicators)

Within a complete run of MIMOSIS, most modules run twice. Once in a pre-run and once in a normal run. The pre run generally creates files that serve as an input for the module run in the normal run. The module CONTRIB (contributions and certain withholdings), for example, needs certain disability benefits and pensions as input, to compute withholdings on it. Therefore the modules SICK (including sickness and disability benefits) and PENSWELF (pensions) generate certain output variables in the pre run which then subsequently are used as input in the normal run by the CONTRIB module.

In total, we execute 17 operations sequentially with the 9 modules listed. The sequence in which the MIMOSIS modules run in a full run, are presented in Table 16.

[43] For more details see FOD Sociale Zekerheid (2011d).

Table 16: Running sequence of different MIMOSIS-modules

	Pre run		Normal run
1	FAMREL		
2	PENSWELF	9	PENSWELF
3	FAMAL	10	FAMAL
4	CONTRIB	11	SICK
5	PIT	12	UNEM
6	SICK	13	CONTRIB
7	UNEM	14	PIT
8	SOCBEN	16	SOBEN
		17	EVAL

The principle to determine the sequence of the different modules is that the output of previous runs serves as input for modules that follow. If possible the input for the normal run is generated within the normal run, but where not possible this input is generated in the pre run. The input needed for the pre run, is read in from external files. In principle, there is no reason to limit a run of the whole model to a normal and a pre run. One could launch for the pre run a pre pre run as well. The choice to limit a complete run to two sub-runs, is mainly inspired by the need to limit the run time of a complete run.[44]

In the remainder of this section, we discuss for the different modules, except for the EVAL module, how they work and what kind of output they generate.

3.1 FAMREL

The module FAMREL mainly processes data from the National Register in order to determine the relationships between household members.[45] Unlike the other modules of the model, this module does not process rules that are included in the socio-economic regulation. In fact, the output of this module mainly serves as input for the other modules.

This module determines, among other things, the affinity to the third degree between household members. For the head of the household, this is more or less obvious to do, since the variable NAREGNIS_RELATION serves as direct input for this job. For other members of the household the determination of these affiliations is based on a conversion of the relationship with the head of the household. This necessarily requires some assumptions.

Furthermore, we also define families in the module FAMREL. Families are smaller units than the sociological household. These smaller family structures are a more relevant starting point than the broader concept of "sociological household" for certain applications, such as the construction of tax units. The number of families that is determined on the basis of this module is, in comparison with the number of sociological households, quite limited. In total

[44] A full run of the model on a PC with the following features: 2.40 GHz CPU and 2.0 GB of RAM takes about 20 minutes.

[45] A description of this module is given in Decoster et al (2007b). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

we observe, a weighted amount of 4,577,973 sociological households in the sample and 4,606,377 families.

We have no external sources to compare the output of the module. To give an idea of the output produced by the FAMREL module, we present in Table 17 some of the results generated by this module. The results reported in this table cover a) a typology of the family to which an individual belongs, b) the position of the family member within the family and a characterization of the married and de facto couples.

Table 17: Position of individuals within the family (output of the FAMREL module)

Description	Number	%
Individual belongs to		
Family of the household head	10,676,659	99.3
Family of children of the household head	62,306	0.6
Family of parents of the household head or partner of the household head	10,888	0.1
Family of grand parents of the household head or partner of the household head	31	0.0
Family of brother or sister of the head or partner of the household head	1,378	0.0
Family of grandchildren of the head or partner of the household head	1,752	0.0
Family of uncle or aunt of the head or partner of the household head	66	0.0
Total individuals	10,753,080	100.0
Position within the family		
Head of the household	4,804,976	44.7
Partner of the head of the household	2,584,767	24.0
Child of the head of the household head or the partner of the household head	3,206,207	29.8
Ascendant in the first degree of the household head or the partner of the household head	41,432	0.4
Ascendant in the second degree of the household head or the partner of the household head	865	0.0
Relatives to the third degree of the household head or the partner of the household head	49,371	0.5
Other household members	65,462	0.6
Value unknown	0	0.0
Total individuals	10,753,080	100.0
Type of relation of married or de facto couples		
Individual belongs to		
Married couples: partners with opposite gender	4,211,386	81.5
Married couples: partners are both males	4,513	0.1
Married couples: partners are both females	4,823	0.1
De facto couples: partners with opposite gender	897,299	17.4
De facto couples: partners are both males	28,729	0.6
De facto couples: partners are both females	22,785	0.4
Total individuals	5,169,534	100.0

3.2 UNEM

In the UNEM module, the unemployment benefits and some other allowances, paid by the RVA, are modelled.[46] This is done by combining a number of status variables from the RVA and the necessary elements from the legislation.

If the income position of a household member would change, this could influence the family charge position of the unemployed and as a consequence the benefit he receives. Therefore, the family charge position of the unemployed is defined in a flexible way within this module.

In some exceptional cases, such as for a number of activation measures, the rules are not implemented in this module but the actual amounts paid, are used.

We can compare the simulation results with data from external sources, but because the individual amounts paid are also in the base file we can also compare with these amounts. We do this in Table 18 for a number of different aggregates.

Table 18: Expenses unemployment, early retirement and career break benefits in 2008 in 1,000 Euro based on external source, simulated with UNEM and registered values in sample[47]

	External source	Simulated amounts	Observed amounts in the sample
Unemployed			
Full benefit unemployed	5,425,924	4,805,130	5,022,082
Temporary unemployed	430,638	460,744	473,886
Activation of unemployed	310,153	578,766	635,327
Early retirement			
Conventional early retirement	1,435,299	1,536,843	1,515,158
Career break	703,674	708,396	726,475
Total	8,305,688	8,089,880	8,372,928

The results in Table 18 learn that there are significant discrepancies between the data from the external source and the simulated amounts. Presumably this is more a result of classification problems (i.e. the assignment of the simulated amounts to the appropriate comparison group) rather than a result of the way the benefits are modelled. In general, the simulated amounts coincide better with the actual amounts observed than with the data from the external source.

Nonetheless, we observe a total difference of about 300 million Euros between the simulated and the real amount. This is mainly due to an underestimation of the sum of the simulated benefits for full benefit unemployed. In this context it should be noted that simulations here, as well as for the other modules, are carried out with parameters that capture the state of affairs on January 1st of the year 2008. It is possible that a number of the registered

[46] The basic principles of this module are explained in Rombaut et al (2007a). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

[47] The data from the external source come from FOD Sociale Zekerheid (2011a), p 188 (column 2008).

amounts have been paid according to amounts that have been indexed or adapted outside the index mechanism, as compared to January 1st 2008.

In general, we consider the affiliation between the simulated amounts on the one hand and the amounts recorded on the other hand, to be sufficiently reliable to continue to use the module for simulation exercises. The amounts reported in Table 18 can of course still mask individual deviations. Therefore we draw in Figure 7 the density function of the simulated and registered annual benefits for full benefit unemployed.

Figure 7: Density function of simulated and observed unemployment benefits on an annual basis for full benefit unemployed

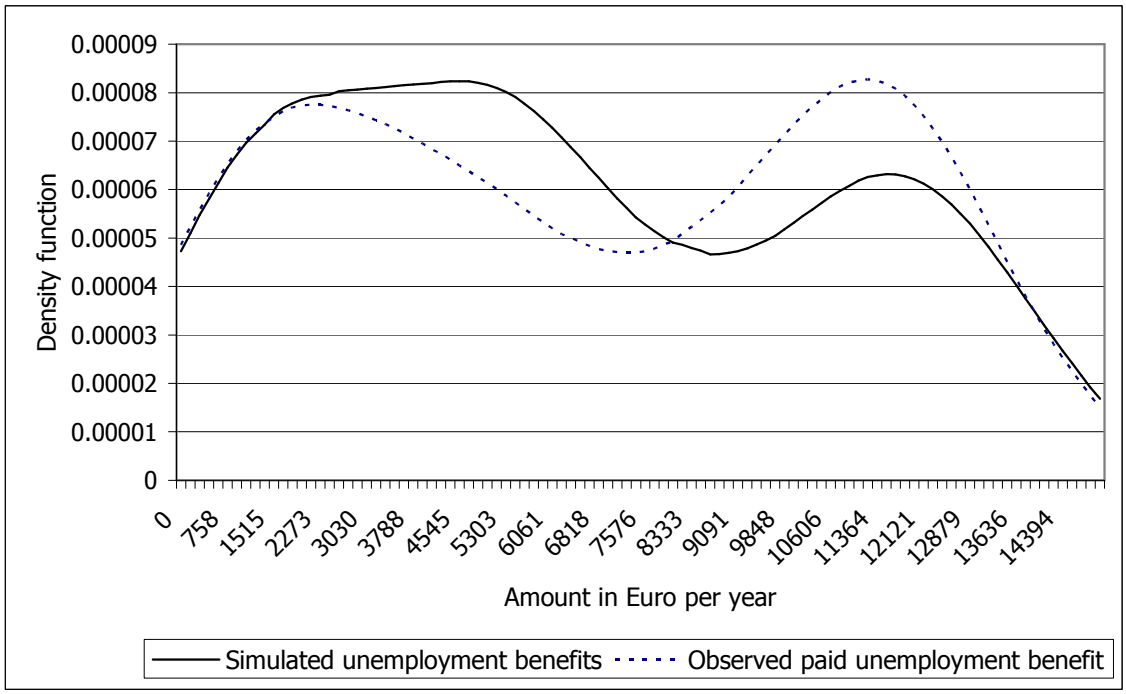


Figure 7 illustrates that within the UNEM module we observe a larger number of cases with low unemployment benefits and fewer cases with high benefits, as compared to the actual amounts recorded.

Since the distributions, drawn in Figure 7, are payments on an annual basis, the observed deviations may be due to a different number of days during which a daily allowance is paid as well as a difference in the daily benefit itself. Discrepancies on the level of the daily benefit can be due to differences in the lost wage used or the family status of the unemployed when simulating as compared to the real computation. The exact cause of these differences is not investigated further in this text.

3.3 SICK

In the module SICK, benefits for sickness and disability, industrial accidents and occupational diseases, are simulated.[48] The calculation for a number of benefits were already

[48] The basic principles of this module are explained in Rombaut et al (2007b). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

programmed in this module but the identification of the different number of days during which the benefits are paid, is insufficiently developed to estimate the annual distributions of these benefits reliably. Therefore, we focus temporarily on the reconstruction of a number of statuses and we process the associated observed paid amounts consequently.

In Table 19 we report the total number of beneficiaries on a number of benefits and, if available, the data we have for a similar group found in external sources.

Table 19: Beneficiaries of sickness and disability benefits, industrial accidents and occupational diseases on the basis of an external source and SICK module[49]

	External source	Simulated amount	Percentage difference between simulated and real amount
General scheme: temporary incapacity to work, workers	NA	291,748	
General scheme: temporary incapacity to work, employees	NA	113,138	
General scheme: disability, workers	172,565	181,701	5.3
General scheme: disability, employees	59,588	60,158	1.0
General scheme: maternity leave, first period	NA	164,493	
General scheme: maternity leave, disability	NA	1,266	
Self employed scheme: temporary incapacity to work	NA	16,489	
Self employed scheme: disability	18,552	20,304	9.4
Self employed scheme: maternity leave, first period	NA	5,279	
Self employed scheme: maternity leave, disability	NA	0	
Allowances temporary disablement in case of industrial accidents	NA	98,439	
Allowances permanent disablement in case of industrial accidents	NA	8,278	
Allowances temporary disablement in case of occupational diseases	NA	57	
Allowances permanent disablement in case of occupational diseases	NA	55,909	

Table 19 shows that the number of times for which we have comparable data on the total number of beneficiaries, is very limited. Only for persons in disability, we observe data for which the definition of the status should coincide with the one used in the external source.

The size of the subgroups of individuals with disability benefits is, in the SICK module, compared to the external source, always overestimated. One explanation could be that the module SICK covers a total number of cases cumulated over a year, while the external source only covers the cases, at one point in time (i.e. December 31st 2008).

In Table 20 we report the corresponding reconstructed annual expenditures for the various items.

[49] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 139 and 250 (column 2008).

Table 20: Expenses on sickness and disability benefits, industrial accidents and occupational diseases on the basis of an external source and SICK module in 1,000 Euro[50]

	External source	Output SICK	Percentage difference between output SICK and real amount
General scheme: temporary incapacity to work, workers	761,431	801,474	5.3
General scheme: temporary incapacity to work, employees	346,219	310,970	-10.2
General scheme: disability, workers	1,914,635	1,990,606	4.0
General scheme: disability, employees	707,159	649,563	-8.1
General scheme: maternity leave, first period	492,409	476,740	-3.2
General scheme: maternity leave, disability	3,760	4,265	13.4
Self employed scheme: temporary incapacity to work	56,716	54,435	-4.0
Self employed scheme: disability	206,508	207,093	0.3
Self employed scheme: maternity leave, first period	14,979	15,838	5.7
Self employed scheme: maternity leave, disability	36	0	100.0
Allowances temporary disablement in case of industrial accidents	NA	137,713	
Allowances permanent disablement in case of industrial accidents	NA	78,991	
Allowances temporary disablement in case of occupational diseases	NA	1,284	
Allowances permanent disablement in case of occupational diseases	NA	157,550	

The comparison of the simulated annual expenses should suffer less from the comparability problem as the statuses, because both in the external source and in the SICK module, aggregates are produced on an annual basis.

Table 20 illustrates that there is no uniform pattern of over- or underestimation of the partial budgets. It is however striking that when classifying the individuals in groups of a) in temporary incapacity to work or b) disabled, there is probably a problem in dividing these groups in workers and employees. It the amounts of workers and employees are added per regime, the total annual budget is fairly well approximated.

The other reconstructed partial budgets, for which we have comparable data, seem to approximate the external source quite well, except for the maternity benefits during a period of disability when self-employed. There are no such cases in our sample that meet this criterion. Given the size of the actual budget, this should not really surprise.

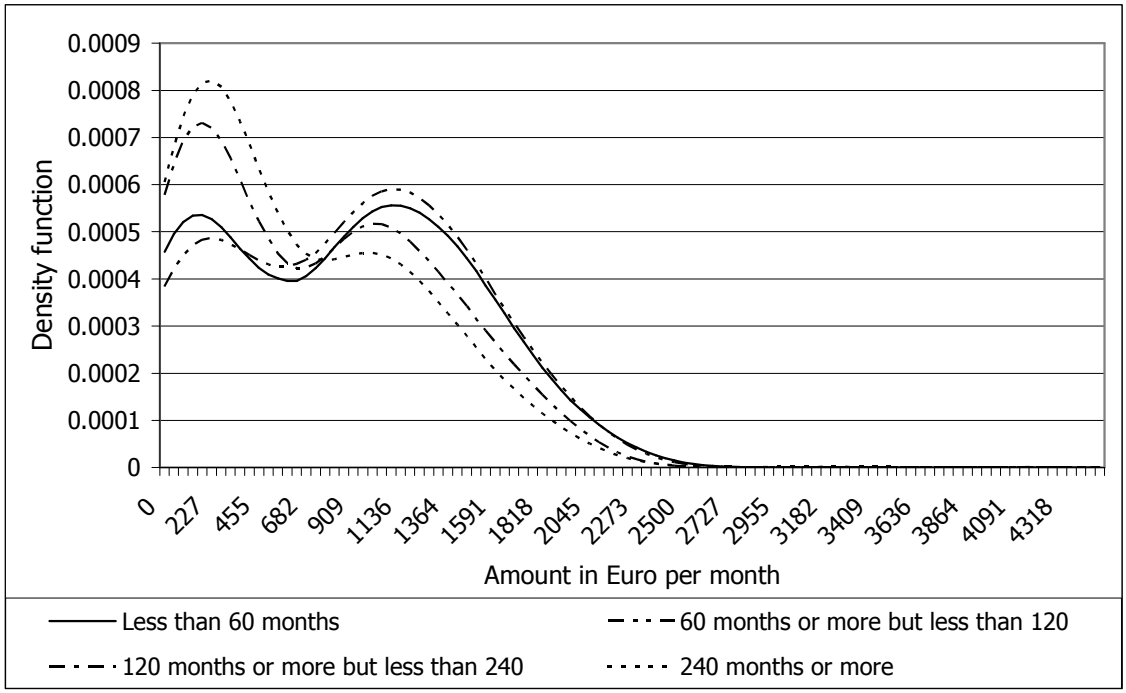
[50] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 138, 140 and 252 (column 2008).

3.4 PENSWELF

With the PENSWELF module one can adapt the registered pension benefits.[51] This can be done by adding or subtracting absolute amounts from the observed amounts or by adjusting this base amount in percentage terms. The adjustments can also be differentiated depending on the type of pension and in function of the "age" of retirement. Compared to the other modules that cover legislative areas, the PENSWELF module is an atypical one by the fact that the observed basic amounts are taken as given, while for other areas the basic amounts are recalculated by applying the relevant calculation rules in combination with a number of exogenous variables. For this module, it therefore makes little sense to compare the baseline data in the module with these in external sources.

But we can give an idea of the basic distributions on which the absolute or percentage changes could be applied. We do this in Figure 8, Figure 9 and Figure 10 by drawing the density functions of the retirement pensions of employees, self-employed and civil servants, differentiated by the number of months the amounts are already paid.

Figure 8: Density function of the monthly retirement pension paid to employees in function of the number of months since the first benefit was paid



[51] The basic principles of this module are explained in Maréchal et al (2007). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

Figure 9: Density function of the monthly retirement pension paid to self employed in function of the number of months since the first benefit was paid

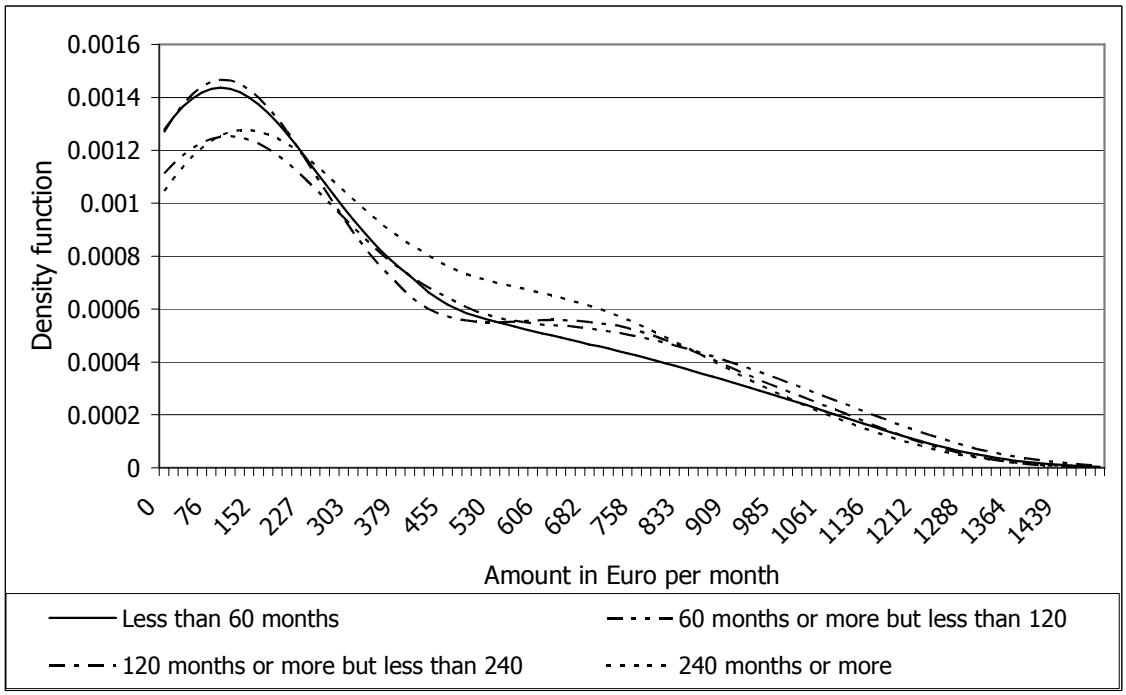
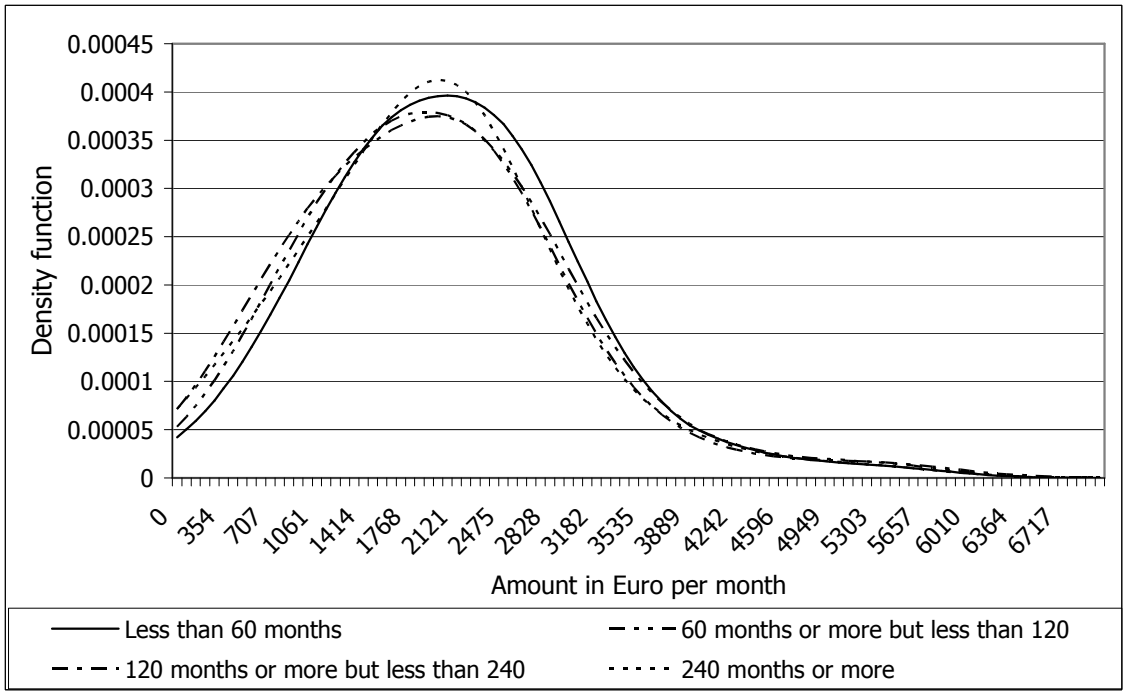


Figure 10: Density function of the monthly retirement pension paid to civil servants in function of the number of months since the first benefit was paid



The results in Figure 8, Figure 9 and Figure 10 suggest that only the pensions of employees vary considerably depending on the number of months elapsed since the first payment of the pension. In the case of employees the number of low pension amounts, increases as the first imbursement is further back in time.

3.5 FAMAL

The FAMAL module determines which children are eligible for child benefits, whom is the beneficiary and whom is the recipient of these benefits, as well as the rank of each child and the possible statuses that open the right to a social or single parent supplement.[52] For the calculation of the benefits the FAMAL module distinguishes four systems: the systems of employees, the system of self-employed, that of civil servants and the system of guaranteed benefits.

For children in the child benefit system of employees and that of self employed, we have data on the distribution of the children by scale and age in external sources. We give these external data and those produced using the FAMAL module in Table 21 and Table 22.

Table 21: Number of beneficiary children by scale and age class in the employee system on the basis an external source and the FAMAL module[53]

	0-6 year	6-12 year	12-18 year	18-25 year	+ 25 year	Total
External source						
Normal scale	478,228	442,491	435,407	234,807	16,636	1,607,569
Scale disabled	13,215	19,193	26,204	13,876	0	72,488
Orphan scale	1,550	5,719	13,011	11,730	0	32,010
Scale pensioners and unemployed	57,072	58,938	59,545	29,954	0	205,509
Total	550,065	526,341	534,167	290,367	16,636	1,917,576
Simulation results						
Normal scale	448,806	461,547	450,699	238,255	13,734	1,613,042
Scale disabled	14,752	22,220	29,716	15,367	752	82,807
Orphan scale	1,171	3,962	8,667	7,601	168	21,569
Scale pensioners and unemployed	56,906	64,901	61,781	33,448	312	217,348
Total	521,636	552,630	550,863	294,671	14,966	1,934,765

[52] The basic principles of this module are explained in Bay et al (2007b). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

[53] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 159 (tabel 30 juni 2008).

Table 22: Number of beneficiary children by scale and age class in the system of self employed on the basis an external source and the FAMAL module[54]

	0-6 year	6-12 year	12-18 year	18-25 year	+ 25 year	Total
External source						
Normal scale	37,446	50,678	60,864	43,096	578	192,662
Scale disabled	182	359	675	640	66	1,922
Orphan scale	213	895	2,222	2,675	3,435	9,440
Scale pensioners and unemployed	0	0	0	0	0	0
Total	37,841	51,932	63,761	46,411	4,079	204,024
Simulation results						
Normal scale	32,433	52,651	66,982	45,760	1,798	199,623
Scale disabled	308	522	1,202	939	0	2,971
Orphan scale	131	332	899	1,465	36	2,862
Scale pensioners and unemployed	474	579	944	776	9	2,783
Total	33,347	54,083	70,027	48,939	1,842	208,239

Both in the system of employees and that of self employed, the number of children in the FAMAL module are slightly overestimated (see the column and the line of Total in Table 21 and Table 22). This overestimation of the number of cases does not systematically come back across the different age groups or scales.

The number of children in the normal scale is overestimated by FAMAL in both cases, but here the differences between the module results and the external source seem all in all quite modest (see the lines normal scale in Table 21 and Table 22).

The number of cases in the disability scale and the scale of pensioners and the unemployed in the system of employees is significantly overestimated by FAMAL, which obviously has consequences for the subsequent calculation of the social supplements.

Finally, it is striking that the number of children being entitled to the orphan scale in both systems is significantly underestimated. This underestimation systematically returns in all age groups (see the lines orphan scale in Table 21 and Table 22). This may be a direct result of problems in the construction of the variable NAREGNIS_CIVIL (see Table 5) since the orphan status is derived from the widow/widower status of the surviving parent.

For civil servants, there are no external data by scale and age. In Table 23 we report only the breakdown by scale for these cases.

[54] The data from the external source are from FOD Sociale Zekerheid (2011a) p. 265 (tabel 30 juni 2008). Although the RSVZ does not report data on the pension scale, these data are still separately distinguished because this scale is still mentioned in the legislation (see Put (2008) p. 798).

Table 23: Number of beneficiary children by scale in the civil servant system on the basis an external source and the FAMAL module [55]

	External source	Simulation results
Normal scale	427,124	414,379
Scale disabled	7,870	18,533
Orphan scale	7,418	6,654
Scale pensioners and unemployed	28,923	13,933
Total	471,335	453,500

The total number of beneficiary children in the civil servant scheme is underestimated by the FAMAL module. This seems to be a direct consequence of a shortage of observations in the normal scale, because the sum of the individuals of all other scales in FAMAL comes close to the sum of the individuals observed in the external source for these scales (take the sum of all lines, other than the normal scale for both columns in Table 23). However, on the level of the individual scales, there is a substantial overestimation of the number of cases in the disability scale and an underestimation of the beneficiaries in the scale of pensioners and unemployed.

The total number of children that were entitled to a benefit in the system of guaranteed child benefits is 14,675 units. With the FAMAL module we identify 18,802 units in this system for 2008.

Presumably we will overestimate the total number of cases with a right on a single parent supplement. Based on the FAMAL module this right is assigned to 134,311 children. We only have external information on the number of beneficiary children in the system of employees. In 2008, 94,908 children were entitled to a single parent supplement.[56] Probably the number of children that opens this right in the system of self-employed, is less than 40,000.

Finally, we compare in Table 24 the total expenditures simulated with FAMAL for each system, with data from an external source. Contrary to the common practice in the rest of this text, we report in the column external source data from the year 2007 because most parameters in the family allowance system remained unchanged through 2007 while in 2008 a number of adjustments to the amounts were carried out. Since we simulate parameters of January 1st, 2008, we assume that the simulated costs best accord with spendings of 2007.

[55] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 382 (tabel 31 december 2009).

[56] See RKW (2011), p. 26.

Table 24: Child benefit expenses on the basis of an external source and the FAMAL module in 1,000 Euro[57]

	External source	Simulated amount	Percentage difference between simulated and real amount
Employees	3,693,798	3,674,163	-0.5
Self employed	355,031	342,131	-3.6
Civil servants	875,000	832,284	-4.9
Guaranteed child allowances	32,842	47,565	44.8

Despite the limited overestimation of the number of children in the system of employees and self employed and an overestimation of the number of beneficiaries on a social supplement in the employee system, the total budget is slightly underestimated in both cases. The possible causes for these findings need further investigation.

The underestimation of the budget paid to civil servants and the overestimation of the budget in the guaranteed child benefit system, is in line with expectations given the respective under and over estimation of the total number of children entitled to benefits in these systems.

3.6 CONTRIB

In the module CONTRIB, contributions on earned income and withholdings on certain replacement incomes, are modelled.[58]

For labour income, the contribution regime is closely linked to the labour market status. In Table 25 we give the number of individuals by activity status on the labour market, computed with CONTRIB and observed in an external source.

Table 25: Number of employees by status on the basis of an external source and the CONTRIB module[59]

	External source		Data CONTRIB	
	Private	Public	Private	Public
Workers	NA	NA	1,462,803	154,747
Employees	NA	NA	1,936,307	381,638
Civil servants	NA	NA		651,831
Self employed	923,946		921,752	

As shown in Table 25, we have very few external sources to compare the CONTRIB results with. The RSZ for example, publishes the statuses at the end of each quarter but not the number of individuals accumulated over a year by status.

Comparing the first three lines with status information with that at the end of a quarter, already shows that the aggregates of the CONTRIB module are significantly higher than the

[57] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 161 (colmun 2007), p. 267 (colmun 2007) and p. 381 (colmun 2006).

[58] The basic principles of this module are explained in Bay et al (2007a). The latest version of the source code is part of the FOD Sociale Zekerheid (2011d).

[59] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 242 (tabel 31 december 2008).

end of the quarter amounts. The opposite exercise, i.e. the production of aggregates based on the jobs known at the end of the quarter, could also be done with the available sample data, but this has not been carried out so far.

For self employed we observe data in an external source that can be compared with the CONTRIB output. The total size of this group is quite good approximated.

The income masses that are subjected to contributions are reported in Table 26 for the various activity groups.

Table 26: Income mass subject to contributions on the basis of an external source and the CONTRIB module in 1,000 Euro[60]

	External source	Output CONTRIB	Percentage difference between CONTRIB and external source
Wage earners private sector	76,771,400	77,301,197	0.7
Wage earners public sector	33,256,900	34,014,882	2.3
Self employed	14,947,831	14,661,613	-1.9

With the available data we seem to approximate the income masses that are subjected to contributions, quite well.

For future use, we primarily aim at gross income masses minus the personal contributions paid on it, so that we consequently can apply the rules to determine the personal income taxes. Since for self employed the income mass for the contributions and personal income taxes is the same[61], we focus here on the estimation of personal contributions, and the possible reductions of this amount, levied on wage earner income.

In Table 27 we report the estimated contributions and the reductions based on the CONTRIB module and an external source.

Table 27: Personal contributions and reductions of contributions on the basis of external source and CONTRIB module in 1,000 Euro[62]

	External source	Output CONTRIB	Percentage difference between CONTRIB and external source
Personal contributions before deductions	12,586,858	12,080,776	-4.0
Amount Workbonus	698,863	754,754	8.0
Number of full time equivalents Workbonus	1,047,113	1,062,580	1.5

Contributions before reductions are underestimated by the CONTRIB module.[63] Spending on the Workbonus, the possible reduction of the personal contributions, is overestimated.

[60] The data from the external source are from FOD Sociale Zekerheid (2011a) p. 125 and p. 244 (column 2008).

[61] See Knapen (2011) for this point.

[62] The data from the external source come from different sources. For the line "Personal contributions before deductions" see FOD Sociale Zekerheid (2011a) p. 128 and p. 130. For lines "Amount Workbonus" and "Number of full time equivalents workbonus" see RSZ (2011), tabel 7a en 7b, Vermindering van de werknemersbijdragen met versterking van de koopkracht.

Presumably, this overestimation of the Workbonus is not a consequence of a misidentification of the individuals that are entitled to it. The reconstructed and the actual number of full-time equivalent jobs with a Workbonus, are close together.

Reporting on the aggregates in Table 27, says nothing about the distribution of the contributions. To see if the real and the simulated distribution of personal contributions after reduction, are different, we draw in Figure 11 the density functions of both concepts for employees that pay contributions to the RSZ and for whom we observe a reduction in the contributions.

Figure 11: Density function of the simulated and observed personal contributions after reductions (subgroup of employees of the RSZ with contribution reduction)

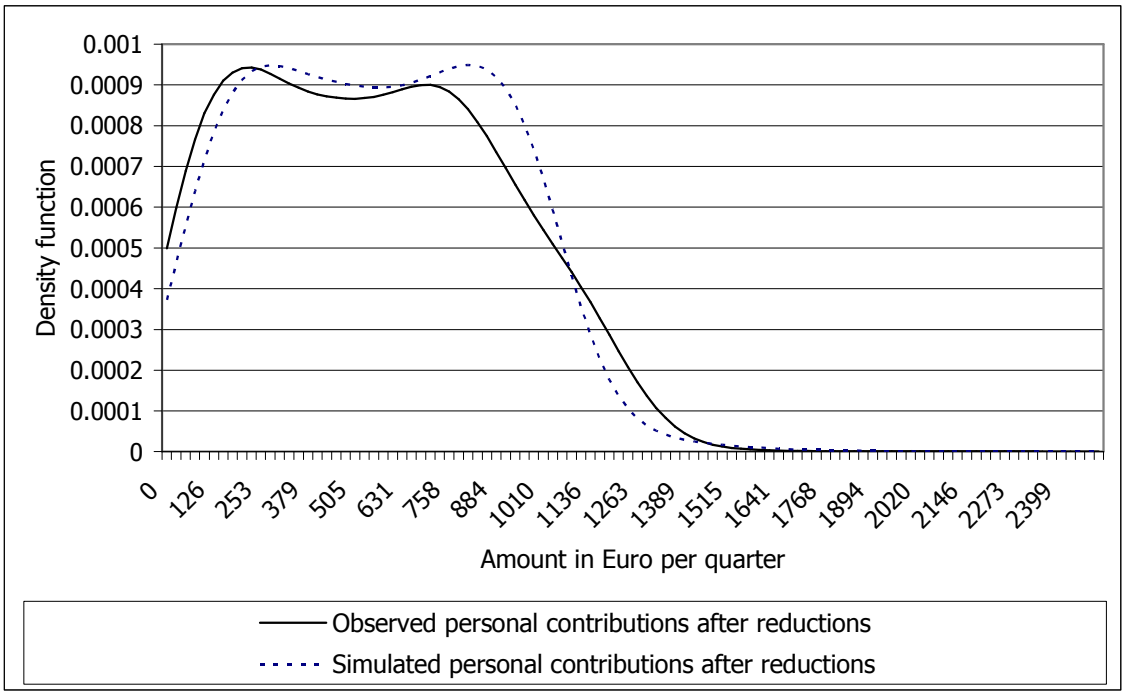


Figure 11 illustrates that the distribution of actual and simulated contributions after reductions, are very similar. As expected from the investigation of the aggregates, there are more simulated lower amounts as compared to the actual amounts.

3.7 PIT

With the PIT module we first reconstruct the tax units, i.e. the units that are supposed to submit a tax form.[64] In a next step we attribute the observed income amounts, on which personal income amounts have to be paid (i.e. the gross amounts minus the personal contributions), to these tax units. For each tax unit then the personal income taxes are computed, taking into account reductions for household composition, compensations for persons with disabilities and reductions for replacement income.

---

[63] For this aggregate from the external source it must be noted that further investigation is required to see whether this the most appropriate figure to compare with.

[64] The basic principles of this module are explained in Decoster et al (2007a). The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

It goes without saying that we, based on the available data, can not reconstruct the entire tax process. We miss a number of income components that are either subjected to the progressive rates or taxed separately. Components we are missing are those on movable and real estate income and "other" income sources. We also miss all expenditures which may be deducted from taxable income and spendings that may provide a tax credit. We therefore rely on a fixed scheme, to take into account the deductions that are allowed to compensate the costs made to earn professional income.

The weighted number of tax units that we obtain after applying these rules, is 5,973,954 units. This is an underestimation of the actual number of 6,143,173 units reported by ADSEI for the tax year 2009 (income 2008).[65] This underestimation could be a consequence of the fact that some individuals are not first classified as a separate tax unit because on their income, in the end, no taxes are due. In particular children who earn a small amount of income do not pay personal income taxes on this amount if the amount is limited. They might enter a tax form for this amount but have not been separated from the parents tax form in this exercise.

The results we obtain with the PIT module can not be assessed on an individual level because our sample does not contain personal income taxes. In Table 28 we therefore give a number of aggregates, reported by ADSEI for the tax year 2009 (income 2008) and the results for similar concepts based on the PIT module.

[65] See ADSEI (2011b).

Table 28: Total taxable net income and taxes of the Tax year 2009 – Belgium on the basis of an external source and PIT module (Decile distribution on the basis of total taxable net income)[66]

Decile	Percentile	Total amount of taxable net income	In % of total amount	Total amount of taxes	In % of total amount	Average tax rate (in %)
			External source			
1		881,090,524	0.5	13,646	0.0	0.0
2		5,185,634,199	3.2	72,588,052	0.2	1.4
3		7,491,804,651	4.7	161,670,093	0.4	2.2
4		9,190,184,310	5.6	708,429,848	1.8	7.7
5		11,257,736,991	6.9	1,459,065,907	3.7	13.0
6		13,558,257,618	8.3	2,535,803,998	6.5	18.7
7		16,276,426,069	9.9	3,520,213,331	9.0	21.6
8		20,367,817,827	12.4	5,018,103,088	12.8	24.6
9		27,493,259,624	16.8	7,718,778,101	19.7	28.1
	91	3,361,125,300	2.1	1,020,380,638	2.6	30.4
	92	3,520,440,662	2.1	1,090,161,732	2.8	31.0
	93	3,700,288,961	2.3	1,171,860,432	3.0	31.7
	94	3,904,114,254	2.4	1,265,583,744	3.2	32.4
	95	4,145,462,385	2.5	1,372,671,598	3.5	33.1
	96	4,447,023,528	2.7	1,510,664,604	3.9	34.0
	97	4,842,265,753	3.0	1,691,625,974	4.3	34.9
	98	5,417,005,578	3.3	1,951,061,372	5.0	36.0
	99	6,468,450,597	3.9	2,418,940,860	6.2	37.4
	100	12,650,775,385	7.7	4,423,280,218	11.3	35.0
10		52,456,952,401	32.0	17,916,231,173	45.8	34.2
Total		164,159,164,214	100.0	39,110,897,237	100.0	23.8
			Simulation results			
1		1,481,157,445	0.9	-78,381,271	-0.2	-5.3
2		6,015,172,194	3.7	56,799,800	0.1	0.9
3		7,710,496,535	4.7	203,224,593	0.5	2.6
4		9,393,800,038	5.7	818,103,518	2.0	8.7
5		11,728,615,659	7.2	1,864,432,608	4.5	15.9
6		14,028,393,487	8.6	2,916,179,877	7.0	20.8
7		16,683,328,420	10.2	4,118,625,210	9.9	24.7
8		20,525,635,276	12.6	5,592,131,103	13.5	27.2
9		27,445,514,717	16.8	8,155,868,551	19.6	29.7
	91	3,316,767,679	2.0	1,053,229,535	2.5	31.8
	92	3,453,291,998	2.1	1,115,829,454	2.7	32.3
	93	3,607,851,764	2.2	1,180,053,906	2.8	32.7
	94	3,789,911,459	2.3	1,261,876,021	3.0	33.3
	95	4,006,238,724	2.5	1,356,699,851	3.3	33.9
	96	4,287,523,526	2.6	1,493,248,292	3.6	34.8
	97	4,625,160,343	2.8	1,660,009,522	4.0	35.9
	98	5,139,204,547	3.1	1,908,638,226	4.6	37.1
	99	5,964,490,806	3.6	2,335,323,147	5.6	39.2
	100	10,294,527,980	6.3	4,505,272,672	10.9	43.8
10		48,484,968,826	29.7	17,870,180,626	43.0	36.9
Total		163,497,082,597	100.0	41,517,164,617	100.0	25.4

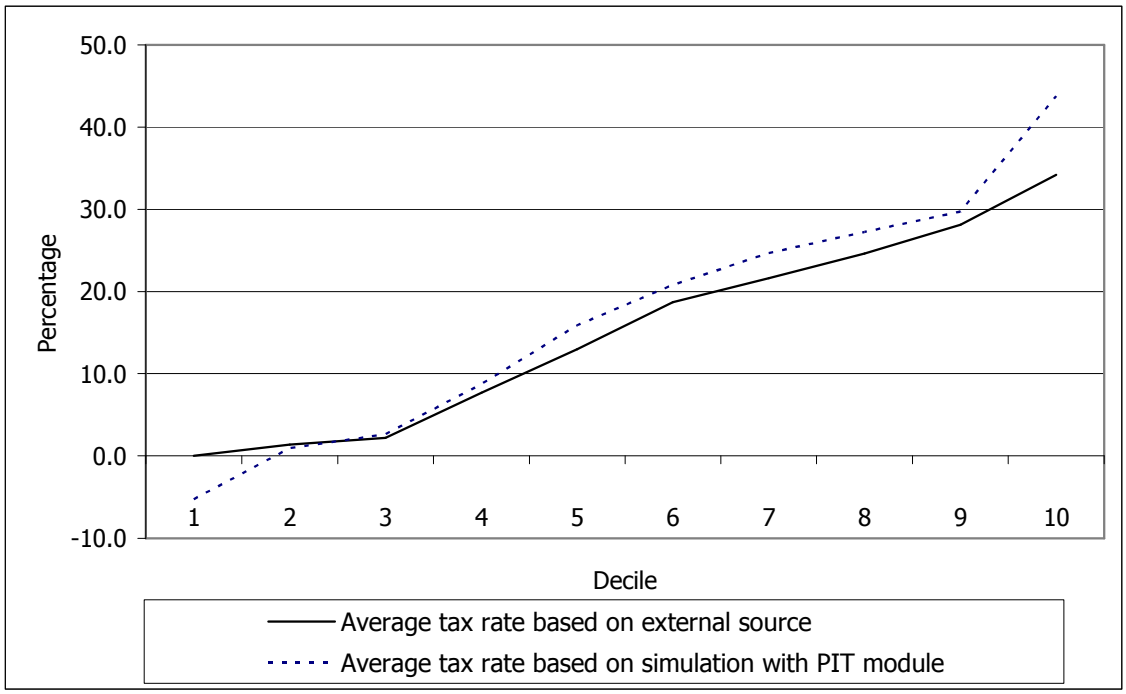
[66] The data from the external source come from ADSEI (2011b), Tabel B.1: Decielenverdeling van het totaal netto belastbaar inkomen, de totale belasting en de gemiddelde aanslagvoet.

Table 28 illustrates that we underestimate the total amount of taxable income and overestimate the income taxes paid, with the PIT module (compare the lines Total of the sub tables External source and Simulation Results).

The underestimation of the total amount of taxable net income does not automatically return in all deciles. For the first eight deciles, we observe an overestimation while for the 9th and the 10th decile we observe an underestimation (see column total amount of taxable net income in Table 28). This suggests that especially for the highest deciles the lacking of some taxable income components is more important than the missing deductible expenditures. The overestimation of the personal income taxes systematically comes back for all deciles except for the first two and last two deciles (see column Total amount of taxes in Table 28).

The influence of these deviations on the average tax rate is a priori unclear. Figure 12 shows that the average tax rate, except for the first two deciles, is systematically overestimated. From the fourth decile on, the deviation is between 1.0 and 3.1 percentage points. For the highest percentiles, the deviation between the two concepts is smaller that 3.1 percentage points except for the top percentile (compare for this the percentile lines of the column average tax rate of the sub tables External source and Simulation Results in Table 28).

Figure 12: Average tax rate in percent per decile on the basis of the External source and the PIT module[67]



The overestimation of the average tax rate, illustrated in Figure 12, is probably caused by the lacking of a number of deductible expenditures. These expenditures are usually unevenly distributed in function of income[68] and probably the share of it grows when income rises.

[67] The data from the external source come from ADSEI (2011b), Tabel B.1: Decielenverdeling van het totaal netto belastbaar inkomen, de totale belasting en de gemiddelde aanslagvoet.

[68] See Figure 37, p. 108 in Decoster e.a. (2011).

3.8 SOCBEN

With the SOCBEN module we first identify those entitled to a number of social welfare benefits.[69] In this identification process, a distinction is made between a) the right to the leefloon, b) to an income guarantee for the elderly (IGO) or guaranteed income for elderly (GIB) and c) the means tested benefits for people with disabilities. The identification of these rights is mainly based on status variables provided by the institutions that pay these respective rights.

Then the means test, that has to be satisfied to obtain the given benefit, is replicated by using as input the various income components delivered by the other modules. However, we do not dispose of all income components that should be taken into account. We do not observe the income on movable property and the real estate income. Income components on which we know nothing are put equal to zero in this application.

In Table 29 we report some of the aggregates, produced by the SOCBEN module and those observed for comparable statuses, in external sources.

Table 29: Beneficiaries by type of social benefit on the basis of external source and SOCBEN module[70]

	External source	Simulated amount	Percentage difference between simulated amount and external source
Leefloon			
Cohabitating with one or more persons	23,137	17,592	-24.0
Single	37,606	37,584	-0.1
Cohabitating with family charge	22,159	22,435	1.2
Total	82,902	77,611	-6.4
IGO/GIB			
IGO allowance and single	54,894	51,713	-5.8
IGO allowance and cohabitating	24,255	17,922	-26.1
GIB	14,468	15,729	8.7
Total	93,617	85,364	-8.8
Means tested allowances of disabled people			
Entitled to Integration benefit of Income Replacement Allowance (IT/IVT)	139,233	111,277	-20.1
Entitled to Allowance for assistance to the elderly (THB)	126,816	124,080	-2.2
Total	266,049	235,357	-11.5

From Table 29 it can be inferred that the number of beneficiaries of the various welfare benefits is always underestimated by a value ranging between 6% and 11.5% of the total number of beneficiaries.

[69] For this module no note with the basic principles was written. The latest version of the source code is part of FOD Sociale Zekerheid (2011d).

[70] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 319 (Leefloon) and p. 326 (IT/IVT en THB) and FOD Sociale Zekerheid (2010) p.357 (IGO/GIB).

For both the leefloon and the income guarantee for the elderly this is mainly due to an underestimation of the number of those who cohabit with other people but do not have family charge. For the means tested benefits for disabled people, this is mainly due to an underestimation of those entitled to an Integration benefit of Income Replacement Allowance (IT/IVT).

These underestimations are most likely due to the fact that we do not replicate the means test perfectly, since the observed data allow to identify the beneficiaries quite accurately. If we compare the group sizes of those entitled to a leefloon or equivalent leefloon and those entitled to means tested benefits for disabled people with figures from external sources, we can conclude that at least the size of these groups is replicated quite well (see Table 10 and Table 11).

One explanation for the "failure" of the means test could be that the means test that is executed in practice, is difficult to replicate with our data. To replicate these means tests we use, for example, the annual incomes. This implies that we also take into account all income amounts earned after the beneficiary has lost his means tested benefit. An individual who during the first two months of the year, received a leefloon, but then lost this right because he started to work will, based on our data and the way we handle them, be identified as a potential beneficiary, but will not receive the benefit based on the executed means-test.

This argument should apply mainly for those for which the income varies throughout the year, primarily due to changes in labour market activities. This would thus provide an explanation for the underestimation of the leefloon benefits and to a lesser extent for individuals that receive the IT/IVT. For the welfare benefits of the elderly this argument seems less plausible. It should therefore be investigated further which arguments could explain the recurrent underestimation of the number of beneficiaries of these allowances.

In Table 30 we give the total expenditures for each type of these welfare benefits, obtained with the SOCBEN module and the amounts reported in external sources.

Table 30: Expenses on social benefits by scheme on the basis of external source and SOCBEN module in 1,000 Euro[71]

	External source	Simulated amount	Percentage difference between simulated amount and external source
Leefloon	369,948	372,704	0.7
IGO/GIB	392,278	426,335	8.7
IT/IVT	1,013,536	802,635	-20.8
THB	416,412	347,092	-16.6

Despite the underestimation of the number of leefloon beneficiaries of and beneficiaries on an IGO/GIB and the absence of some income components in the application of the means test, the amount of expenditures on these items is overestimated. There is no immediate explanation for this observation. For the leefloon it might be remarked that we should

[71] The data from the external source come from FOD Sociale Zekerheid (2011a) p. 319 (Leefloon), p. 324 (IGO/GIB) and p. 327 (IT/IVT en THB).

further investigate whether or not the external source provides a good reference point. The reported external figure might also contain components paid by the POD Maatschappelijke Integratie to the OCMW's but not simulated here. It is currently unclear what the extent is of these non simulated components.

The underestimation of the IT/IVT budget is, based on the estimated number of beneficiaries, in line with expectations. Underestimation of the THB budget is again unexpected, given the limited deviation between the reconstructed and externally reported number of beneficiaries. The search for explanations for these different findings should therefore be subject of further research.

4 DESCRIPTION OF THE MODULE EVAL

The evaluation module (EVAL) produces a number of measures that summarize the situation before and after the reform. Two major blocks of information, produced by the EVAL module, can be distinguished.

First of all, a number of aggregates are reported to summarize the budgetary impact of the simulated reforms. These budgetary measures are direct output from the various modules. The performance of the model on this point has been discussed for the various modules, in the previous section.

Secondly, the model also produces a number of distributional measures. Among other things, the income inequality is summarized by an income inequality measure like the Gini coefficient before and after the reform. Furthermore, poverty rates, and also tables with winners and losers, divided into a number of different dimensions, are calculated.

In this distributional analysis the concept of the equivalent disposable income, takes an important place.[72] The At Risk Of Poverty rate that is most commonly used to compare the poverty risk in EU member states, for example, uses a poverty line that is set equal to 60% of the median of the national distribution of equivalised disposable income.

The median of the distribution of equivalised disposable incomes, that we obtain on the basis of MIMOSIS 2009 (income 2008) is 1,486 Euro per month. With the 60% rule, this implies an estimated value of the poverty line of 892 Euros. The "official" standard for Belgium is 966 Euros (based on the EU_SILC 2009 (income 2008)).[73]

That we underestimate the median of the equivalent disposable income distribution, compared with the EU_SILC, should not surprise. Some components such as income from movable property, and transfers between households, which are used to calculate the EU_SILC concept, are not included in the MIMOSIS sample. Furthermore, we lack some data to replicate the personal income process, and therefore overestimate the personal income taxes that are due. The importance of these different components, in explaining the observed difference between both concepts, requires additional research that will not be carried out here.

---

[72] The notion of disposable income is explained in more detail in European Commission (2003).
[73] See Eurostat (2011).

To illustrate that the equivalised income concepts in both sources do differ systematically, we plot the average equivalent disposable incomes, by percentile, observed in the EU_SILC version 2009 (income 2008) and for the MIMOSIS baseline 2008 in Figure 13.[74]

Figure 13: Equivalised disposable income in Euro per month (average value per percentile)[75]

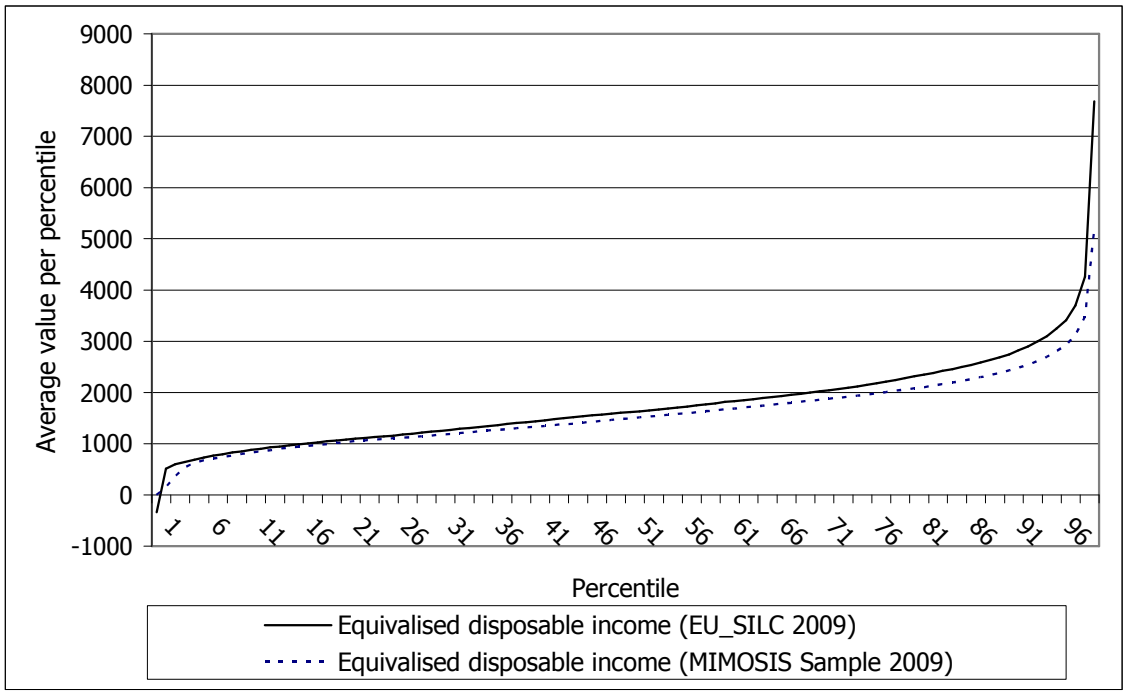


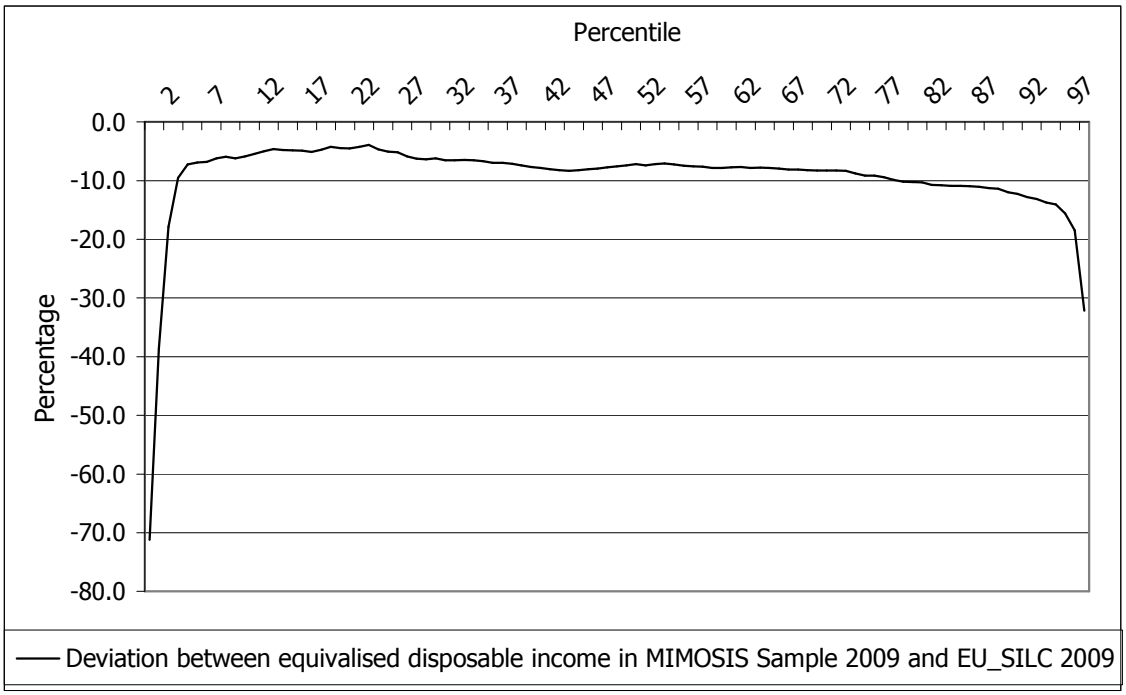
Figure 13 illustrates that with the MIMOSIS baseline, the average EU_SILC value of equivalised disposable income, is systematically underestimated at all percentiles, except the first percentile. For all but the first percentile, the averages are always positive within both sources. Within the EU_SILC, the value of the first percentile is negative, while it is slightly positive on the basis of the MIMOSIS baseline data.

Apparently, the distributions of the two concepts are close together. But, on the basis of Figure 13 it is hard to say to what extent the averages per percentile differ. That is why we plot in Figure 14 the percentage deviation between the two concepts for each percentile, except for the first percentile.

[74] The percentile in this and subsequent figures, is always created from the distribution of data in the source itself. The data source is always a completely independent production.

[75] The percentile values, based on the EU_SILC, are not publicly available. They were delivered by ADSEI on demand.

Figure 14: Percentage deviation between equivalised disposable income within MIMOSIS sample 2009 and EU_SILC 2009



As Figure 14 illustrates, the equivalent disposable income concept within MIMOSIS underestimates the EU_SILC variant, by 10% or less in most cases. Only for the first two and the last two percentiles, the deviation exceeds 20%. The concepts within both sources are closest to each other between the 10th and 20th percentile. From the 20th percentile, the percentage difference, more or less steadily increases over the different percentiles.

Because of the unequal deviation at percentile level and because there are a number of explanatory factors that could explain the underestimation, we correct the equivalised disposable incomes in MIMOSIS at percentile level, to obtain a concept that is comparable with the EU_SILC estimate. The correction, applied at percentile level in MIMOSIS, is equal to the percentage deviation between the EU_SILC value, based on the EU_SILC 2009 data, and the value of MIMOSIS that is obtained using the sample 2009 (income 2008) for the tax and contribution year 2008.

The median of the simulated equivalent income distribution that we obtain with MIMOSIS, after this correction procedure, is 1,609 Euro per month. This is also the amount obtained with the EU_SILC 2009 (income 2008) data.[76]

In Table 31 and Table 32 we report the poverty rates before and after the correction procedure of the equivalent disposable income distribution, described above, respectively by age class and region.

[76] See ADSEI (2011).

Table 31: At Risk of Poverty rate by age class before and after correcting equivalised disposable income

	Share of group in population	At risk of poverty rate before correcting the equivalised disposable income	At risk of poverty rate after correcting the equivalised disposable income
0 - 4 year	5.3	16.2	17.5
5 - 9 year	5.4	14.5	15.9
10 -14 year	5.7	14.6	16.1
15 - 19 year	6.1	15.9	17.5
20 - 24 year	5.9	12.5	13.6
25 - 29 year	6.3	10.5	11.2
30 - 34 year	6.1	10.4	11.4
35 - 39 year	6.8	10.8	11.7
40 - 44 year	7.4	11.0	12.1
45 - 49 year	7.6	10.2	11.2
50 - 55 year	7.0	10.5	11.3
56 - 59 year	6.5	12.1	12.8
60 - 64 year	6.0	14.2	15.3
65 - 69 year	4.4	15.7	17.5
70 - 74 year	4.5	18.1	20.6
75 - 79 year	4.0	18.3	20.9
80 - 84 year	2.9	17.6	20.0
85 - 89 year	1.6	14.4	17.0
90 - 94 year	0.4	11.5	14.0
>= 95 year	0.1	12.5	14.5
Belgium	100.0	13.2	14.5

Table 32: At Risk of Poverty rate by region before and after correcting equivalised disposable income

	Share of group in population	At risk of poverty rate before correcting the equivalised disposable income	At risk of poverty rate after correcting the equivalised disposable income
Region Brussels Capital	9.9	24.6	26.6
Region Flanders	57.7	9.7	10.8
Region Wallonia	32.3	16.0	17.3
Belgium	100.0	13.2	14.5

The overall AROP, computed as the percentage of the population with an equivalised disposable income of less than 60% of this median, is estimated to be 14.5% on the basis of the MIMOSIS data. On the basis of the EU_SILC data, this percentage is estimated to be 14.6%.[77]

A comparison of the subgroup results with those of the EU_SILC would give more insight into the extent to which the EU_SILC results can be replicated with these simulated data.

The age distribution, presented in Table 31, is not directly available for EU_SILC data. A different age classification, based on the EU_SILC, is available but has not been computed so far with the MIMOSIS-baseline.

[77] See Eurostat (2011).

For the breakdown by regions, only directly comparable data are available for Flanders and Wallonia.[78] Based on the MIMOSIS sample, the poverty rate for Flanders is slightly overestimated compared to EU_SILC (i.e. 10.8% based on MIMOSIS and 10.1% based the EU_SILC). The poverty rate in Wallonia is underestimated (17.3% based on MIMOSIS and 18.4% based on the EU_SILC). For the 26.6%, which is the estimated rate for Brussels, we do not have a reference point.

5 CONCLUSION

Throughout the text we presented a simulation model built on a sample of administrative data. By building a model on a sample of such administrative data, the model is situated between at least two alternative options, including a model that primarily uses survey data as input on the one hand and working with the entire administrative population on the other hand.

Compared to working with the entire population, as could be done by exploiting the complete Datawarehouse Labour Market and Social Protection, the approach presented here has the disadvantage that:

- One has to work with fewer observations than there are actually available,
- Not all available variables can be used because they could not all be included in a data demand,
- Most socio-economic conditions are measured only at a certain point in time and not in the more continuous way that could be followed with the full Datawarehouse.

Compared with surveys, the approach presented here has the disadvantage that not all variables, included in surveys, and relevant for the assessment of the socio-economic position of individuals or households, are available in administrative records.

Compared to surveys, the administrative data approach, has the following advantages:

- the sample size can be made much larger than the one of surveys and
- the concepts used are much more in line with the administrative situation that generated the reconstructed rights.

Compared to working with the entire administrative population, the approach also has the advantage that the computing time is reduced significantly. When working with a sample that is representative for the modelled phenomenon, the reduction in size should not come at the cost of losing insight in the modelled phenomenon.

Throughout the text we therefore focused on tests that could learn us to what extent the underlying sample is representative for the modelled sub-domains covered by the model, i.e. a) unemployment benefits, b) benefits on sickness and disability, industrial accidents and occupational diseases, c) pensions, d) family allowances, e) social assistance benefits, f) contributions and withholdings and g) personal income taxes. By disentangling the legislation

[78] See ADSEI (2011).

for each of these sub domains into a number of parameters, exogenous and endogenous variables, it is possible to execute already a number of simulations for the various domains.

However, as highlighted during the discussion as well, an additional effort is necessary to better exploit the simulation possibilities of the available data. This is especially true for the modules that treat the sickness benefits and contributions and withholdings.

6 REFERENCES[79]

[1.] ADSEI (2011a), Loop van de bevolking, *webpublicatie*, Brussel: Federale OverheidsDienst Economie, Algemene Directie Statistiek en Informatie. (http://statbel.fgov.be/nl/statistieken/cijfers/bevolking/loop/)

[2.] ADSEI (2011b), Fiscale Inkomens, *webpublicatie*, Brussel: Federale OverheidsDienst Economie, Algemene Directie Statistiek en Informatie. (http://statbel.fgov.be/nl/statistieken/cijfers/arbeid_leven/fisc/)

[3.] ADSEI (2011c), Inkomen en levensomstandigheden, *webpublicatie*, Brussel: Federale OverheidsDienst Economie, Algemene Directie Statistiek en Informatie. (http://statbel.fgov.be/nl/statistieken/cijfers/arbeid_leven/inkomens)

[4.] APS (2011), Huishoudens en familiekernen (ID: 1329), *webpublicatie*, Brussel: Studiedienst van de Vlaamse Regering. (http://aps.vlaanderen.be/sgml/langereeksen/1329.htm)

[5.] Atkinson, A., Rainwater, L. and Smeeding, T. (1995), *Income distribution in OECD Countries: Evidence from the Luxembourg Income Study*, Parijs: OECD.

[6.] Bay, C., Maréchal, C., Perelman, S. and Van Camp, G. (2007a), MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for the module on social security contributions, *webpublicatie*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming. *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wpcontrib.pdf*)*

[7.] Bay, C., Perelman, S. and Van Camp, G. (2007b), MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for the module on family allowances, *webpublicatie*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming. *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wpfamal.pdf*)*

[8.] Berghman, J., Curvers G., Palmans S. en Peeters H. (2007), De Belgische gepensioneerden in kaart gebracht, Deel 1: De wettelijke pensioenbescherming, *Working paper Sociale Zekerheid Nr. 6*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming. (http://www.socialsecurity.fgov.be/docs/nl/publicaties/workingpapers/WP_2007-6.pdf)

[9.] Cowell, F. and Jenkins, S. (2000), Estimating Welfare Indices: Household Weights and Sample Design, *Discussion Paper No. DARP 48*, London: London School of Economics. (http://eprints.lse.ac.uk/2160/Estimating_Welfare_Indices_Household_Weights_and_Sample_Design.pdf)

[10.] Decoster, A. en Valenduc, C. (eds.) (2011), *Belastingen en fiscaal beleid in België*, Leuven: Acco, pp. 291.

[11.] Decoster, A. (2010), Berekend Vlaanderen: workshop FLEMOSI – SBO-project 90044, *webpublicatie*, Flemosi-consortium. (http://www.flemosi.be/uploads/100/FLEMOSI%20WORKSHOP%202011-01-21%20André.pdf)

[12.] Decoster, A, De Swerdt, K, Orsini, K, and Van Camp, G. (2007a), MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for the module on personal income taxes, *webpublicatie*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming. *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wppit.pdf*)*

[13.] Decoster, A, De Swerdt, K, Orsini, K, and Van Camp, G. (2007b), MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for household composition and family relationships, *webpublicatie*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming. *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wpfamrel.pdf*)*

[79] Not all sources that are cited are currently publicly available. They will be made available on a reworked version of the website "socialsecurity.fgov.be".

[14.] Decoster, A, De Swerdt, K, Orsini, K, Lefèbvre, M., Maréchal, C., Paszukiewicz, A., Perelman, S., Rombaut, K., Verbist, G. and Van Camp, G. (2008), Valorisation of the microsimulation model for social security MIMOSIS, *Final Report Project AG/01/116*, Brussel: Federaal Wetenschapsbeleid.
*(*http://www.belspo.be/belspo/home/pub/pub_ostc/agora/raghh116_en.pdf)

[15.] DOCDWH (2011), Documentatie met betrekking tot het datawarehouse, *webpublicatie*, Brussel: Kruispuntbank van de Sociale Zekerheid
(http://www.ksz.fgov.be/nl/bcss/page/content/websites/belgium/statistics/statistics_01/statistics_01_05.html)

[16.] European Commission (2003), Commission regulation (EC) No 1980/2003 of 21 October 2003 implementing Regulation (EC) No 1177/2003 of the European Parliament and of the Council concerning Community statistics on income and living conditions (EU_SILC) as regards definitions and updated definitions, Official Journal of the European Union. (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:298:0001:0022:EN:PDF)

[17.] Eurostat (2011), Income, Social Inclusion and Living Conditions, *webpublicatie*, Luxemburg: Eurostat.
(http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/database)

[18.] FOD Sociale Zekerheid (2011a), *Vademecum van de financiële en statistische gegevens over de sociale bescherming in België: periode 2005-2011*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Sociaal Beleid, Domein Kerngegevens.
*(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/vademecum/2011/vademecum_2011_nl.pdf*)*

[19.] FOD Sociale Zekerheid (2011b), Programma's voor omzetting gegevens Datawarehouse Arbeidsmarkt en Sociale Bescherming naar bestanden aangevraagd in data-aanvraag*, mimeo*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming.

[20.] FOD Sociale Zekerheid (2011c), Programma's voor omzetting gegevens uit data-aanvraag naar input file MIMOSIS, *mimeo*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming.

[21.] FOD Sociale Zekerheid (2011d), Broncode MIMOSIS: Steekproef januari 2009, gegevens 2008, *mimeo*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming.

[22.] FOD Sociale Zekerheid (2010a), *Vademecum van de financiële en statistische gegevens over de sociale bescherming in België: periode 2004-2010*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Sociaal Beleid, Domein Kerngegevens.
*(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/vademecum/2010/vademecum_2010_nl.pdf*)*

[23.] FOD Sociale Zekerheid (2010b), *Jaaroverzicht 2009 in cijfers*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Personen met een Handicap.
*(*http://handicap.fgov.be/docs/vouwfolder_2009_overzicht_cijfers_nl.pdf*)*

[24.] Knapen, H. (2011), Constructie inkomensnotie Datawarehouse AM&SB, *tussentijds verslag project DOCDWH*, Leuven: K.U. Leuven, Centrum voor Sociologisch Onderzoek.

[25.] KSZ (2011), *"Overzicht belangrijkste cijfers basistoepassingen 2008 – verbeterde versie.xls: populatie pensioentrekkenden"*, bijlage bij mail met subject Beheersgroep datawarehouse arbeidsmarkt en sociale bescherming – nieuw overzicht van de belangrijkste cijfers een aangepast verslag vergadering en time Vrijdag 26 augustus 2011.

[26.] Maréchal, C., Paszukiewicz, A., Perelman, S. and Van Camp, G. (2007), MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for module on welfare adaptations of pensions,
*(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wppenswelf.pdf*)*

[27.] Ministerie van Financiën (2011), *Fiscaal Memento 2011*, Brussel: Ministerie van Financiën, Studie en Documentatiedienst.
(http://docfin.fgov.be/intersalgnl/thema/publicaties/memento/pdf/FM2011_V01_volledig.pdf)

[28.] Molenberghs, G. en Gaens, E. (1994), Gezondheidsenquête door middel van interview: Statistische beschouwing, *Werkdocument Limburgs Universitair Centrum*, Diepenbeek: Universitaire Campus.

[29.] Mutsaerts, A. (2011), Constructie Wstatus en Reconstructie CIMIRe op basis van het DWH AM&SB, *tussentijds verslag project SIPEBE*, Leuven: K.U. Leuven, Centrum voor Sociologisch Onderzoek.

[30.] POD Maatschappelijke Integratie (2011), Studies, publicaties en statistieken: statistieken, webpublicatie, Brussel: Programmatorische OverheidsDienst Maatschappelijke Integratie. (http://www.mi-is.be/be-nl/studies-publicaties-en-statistieken/statistieken)

[31.] Put, J. (2008), *Praktijkboek sociale zekerheid voor de onderneming en de sociale adviseur 2008*, Mechelen: Wolters Kluwer.

[32.] Rijksregister (2011), Voostelling van het Rijksregister, webpublicatie, Brussel: Federale OverheidsDienst Binnenlandse Zaken. (http://www.ibz.rrn.fgov.be/index.php?id=2461&L=1)

[33.] RKW (2011), *Activiteitenverslag 2009*, Brussel: Rijksdienst voor Kinderbijslag voor Werknemers. (http://www.rkw.be/Nl/Documentation/Publication/AnnualReports/JV2009.pdf)

[34.] Rombaut, K., Van Camp, G. and Verbist, G. (2007a), *MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for the Unemployment Benefits Module*, *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wpunem.pdf*)*

[35.] Rombaut, K., Van Camp, G. and Verbist, G. (2007b), *MIMOSIS: MIcrosimulation MOdel for Belgian Social Insurance Systems, Modelling rules for the Sickness and Disability benefits Module*, *(*http://socialsecurity.fgov.be/NL/nieuws_publicaties/publicaties/mimosis/wpsick.pdf*)*

[36.] RSZ (2011), Verminderingen van de bijdragen voor sociale zekerheid: aantal betrokken werknemers en budgettaire impact, *webpublicatie*, Brussel: Rijksdienst voor Sociale Zekerheid. (http://www.rsz.fgov.be/nl/content/statistics/publications/reductions.html)

[37.] RVP (2011), Herwaardering van het loon voor de pensioenberekening, *webpublicatie*, Brussel: Rijksdienst voor Pensioenen. (http://www.onprvp.fgov.be/NL/profes/calculation/career/wages/revaluation/Pages/default.aspx#b)

[38.] SAS (2011), SAS/STAT 9.2 User's Guide, *webpublicatie*. (http://support.sas.com/documentation/cdl/en/statugkde/61796/PDF/default/statugkde.pdf)

[39.] Vleminckx, K. (2011), Aanvraag van gecodeerde sociale gegevens van persoonlijke aard voor het updaten van het microsimulatiemodel MIMOSIS en het bestuderen van de mogelijkheid tot armoedemeting met administratieve gegevens uit het DWH AM&SB, *mimeo*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming.

[40.] Vleminckx, K. (2008), Aanvraag van gecodeerde individuele sociale zekerheidsgegevens voor wetenschappelijk onderzoek in de projecten "Microsimulatiemodel Sociale Zekerheid" (fase II – uitbouw, validering en exploitatie), *mimeo*, Brussel: Federale Overheidsdienst Sociale Zekerheid, Directie-generaal Beleidsondersteuning, Domein Evolutie van de Sociale Bescherming.

APPENDIX 1: DESCRIPTION OF THE DETERMINATION OF THE SAMPLE WEIGHTS

In this section we describe how sample weights can be produced, given the sampling method. Two major steps can be distinguished in the sampling method. First, the creation of a list of 100,000 randomly selected individuals and then the supplementation of this list with all household members who belonged to the private household of the randomly drawn individual.

For the exposition, we introduce the following symbols:

NI : The number of individuals in the population (i.e. the size of the sampling frame)

ni : The number of individuals sampled

g_j : The number of household members in the household to which individual j belongs

P_j : The probability that individual j appears in the population

p_j : The probability that individual j is selected for the sample

In the case of simple random sampling, the probability that an individual is present in the sampling frame is equal to the probability that the individual is selected for the sample. The weighting factor w_j for each individual j in the sample, can in this case be determined as:

$$w_j = \frac{NI}{ni} \qquad (1).$$

Because in the second step all household members are included in the sample, the probability that an individual from the population ends up in the sample, is no longer equal for all individuals. The weighting factor w_j in equation (1) should be adjusted in this case by the ratio $\frac{P_j}{p_j}$, i.e. the probability that individual j is in the population divided by the probability that individual j is in the sample.[80] The adjusted weight w_j is in general terms given in equation (2):

$$w_j = \frac{NI}{ni} \cdot \frac{P_j}{p_j} \qquad (2).$$

The probability P_j, that an individual j is observed in the population is equal for all people, i.e. $\frac{1}{NI}$. The probability p_j that an individual j is observed in the sample is uneven and depends on size of the household to which j belongs. Each individual can be selected itself in the first step or one of his household members can be selected in the second step. The

[80] See Cowell, F. and Jenkins, S. (2000), p. 11.

selection probability of an individual j can therefore be written as follows $\frac{g_j}{\sum_{i=1}^{NI} g_i}$. The weight w_j in this case is defined as:

$$w_j = \frac{\sum_{i=1}^{NI} g_i}{ni \cdot g_j} \quad (3).$$

APPENDIX 2: WEIGHTED RESULTS BY REGION, AGE AND GENDER

Table 33: Weighted number of observations on individual level by region and age (men)[81]

		0-4 YEAR	5-9 YEAR	10-14 YEAR	15-19 YEAR	20-24 YEAR	25-29 YEAR	30-34 YEAR	35-39 YEAR	40-44 YEAR	45-49 YEAR
Weighted amount based on sample observations	Belgium	316,089	301,683	309,986	333,564	326,344	346,023	346,948	382,905	406,619	408,775
	Region Brussels Capital	40,224	32,539	29,578	29,669	33,947	44,661	47,037	45,015	40,495	35,102
	Region Flanders	172,438	164,267	174,577	188,554	182,132	196,081	192,433	215,687	240,219	245,510
	Region Wallonia	103,427	104,877	105,831	115,341	110,265	105,281	107,478	122,203	125,905	128,163
Real number based on population statistics 2009	Belgium	288,971	293,651	311,591	330,478	320,003	344,636	328,241	363,808	401,351	409,016
	Region Brussels Capital	38,530	32,626	29,884	29,167	32,488	38,906	39,027	38,833	38,034	35,564
	Region Flanders	153,874	158,023	173,986	189,444	175,099	198,596	186,108	206,833	238,479	245,965
	Region Wallonia	96,567	103,002	107,721	111,867	112,416	107,134	103,106	118,142	124,838	127,487
Percentage difference between weighted and real amount	Belgium	-9.4	-2.7	0.5	-0.9	-2.0	-0.4	-5.7	-5.2	-1.3	0.1
	Region Brussels Capital	-4.4	0.3	1.0	-1.7	-4.5	-14.8	-20.5	-15.9	-6.5	1.3
	Region Flanders	-12.1	-4.0	-0.3	0.5	-4.0	1.3	-3.4	-4.3	-0.7	0.2
	Region Wallonia	-7.1	-1.8	1.8	-3.1	1.9	1.7	-4.2	-3.4	-0.9	-0.5

[81] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, region and age classes.

Table 33 (continued): Weighted number of observations on individual level by region and age (men)[82]

		50-54 YEAR	55-59 YEAR	60-64 YEAR	65-69 YEAR	70-74 YEAR	75-79 YEAR	80-84 YEAR	85-89 YEAR	90-94 YEAR	95 YEAR AND MORE
Weighted amount based on sample observations	Belgium	377,220	339,739	301,676	219,183	206,333	171,250	109,706	52,235	9,952	2,421
	Region Brussels Capital	30,277	26,502	22,450	16,530	14,577	12,357	8,910	4,994	1,072	314
	Region Flanders	226,024	200,878	181,908	138,067	131,955	107,309	66,975	31,606	6,067	1,482
	Region Wallonia	120,919	112,359	97,318	64,586	59,801	51,584	33,821	15,635	2,813	625
Real number based on population statistics 2009	Belgium	377,759	346,092	318,169	227,248	219,340	174,977	118,318	57,826	10,691	2,516
	Region Brussels Capital	29,960	25,434	27,840	18,416	15,527	15,314	10,207	5,771	1,343	231
	Region Flanders	229,505	204,173	190,240	139,337	139,872	106,914	70,823	35,733	6,529	1,751
	Region Wallonia	118,294	116,485	100,089	69,495	63,941	52,749	37,288	16,322	2,819	534
Percentage difference between weighted and real amount	Belgium	0.1	1.8	5.2	3.5	5.9	2.1	7.3	9.7	6.9	3.8
	Region Brussels Capital	-1.1	-4.2	19.4	10.2	6.1	19.3	12.7	13.5	20.2	-35.9
	Region Flanders	1.5	1.6	4.4	0.9	5.7	-0.4	5.4	11.5	7.1	15.4
	Region Wallonia	-2.2	3.5	2.8	7.1	6.5	2.2	9.3	4.2	0.2	-17.0

[82] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, region and age classes.

Table 34: Weighted number of observations on individual level by region and age (women)[83]

		0-4 YEAR	5-9 YEAR	10-14 YEAR	15-19 YEAR	20-24 YEAR	25-29 YEAR	30-34 YEAR	35-39 YEAR	40-44 YEAR	45-49 YEAR
Weighted amount based on sample observations	Belgium	301,448	289,318	296,352	320,482	324,396	344,963	342,202	373,829	394,900	402,081
	Region Brussels Capital	38,498	31,473	28,084	29,146	37,916	49,069	46,334	41,571	36,836	34,406
	Region Flanders	164,444	157,490	167,109	180,576	178,634	192,199	189,040	211,129	232,243	238,548
	Region Wallonia	98,506	100,355	101,159	110,760	107,846	103,695	106,828	121,129	125,821	129,127
Real number based on population statistics 2009	Belgium	270,697	281,696	299,980	322,151	313,158	333,551	326,266	370,588	393,257	409,547
	Region Brussels Capital	36,383	30,992	29,447	30,205	33,125	41,043	40,253	38,853	36,714	35,410
	Region Flanders	146,578	150,920	169,917	180,361	174,380	189,936	182,113	211,521	235,028	243,728
	Region Wallonia	87,736	99,784	100,616	111,585	105,653	102,572	103,900	120,214	121,515	130,409
Percentage difference between weighted and real amount	Belgium	-11.4	-2.7	1.2	0.5	-3.6	-3.4	-4.9	-0.9	-0.4	1.8
	Region Brussels Capital	-5.8	-1.6	4.6	3.5	-14.5	-19.6	-15.1	-7.0	-0.3	2.8
	Region Flanders	-12.2	-4.4	1.7	-0.1	-2.4	-1.2	-3.8	0.2	1.2	2.1
	Region Wallonia	-12.3	-0.6	-0.5	0.7	-2.1	-1.1	-2.8	-0.8	-3.5	1.0

[83] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, region and age classes.

Table 34 (continued): Weighted number of observations on individual level by region and age (women)[84]

		50-54 YEAR	55-59 YEAR	60-64 YEAR	65-69 YEAR	70-74 YEAR	75-79 YEAR	80-84 YEAR	85-89 YEAR	90-94 YEAR	95 YEAR AND MORE
Weighted amount based on sample observations	Belgium	375,940	340,669	312,151	241,082	246,275	235,946	184,717	113,968	31,675	12,035
	Region Brussels Capital	31,661	28,836	25,597	19,948	19,341	19,277	16,891	12,028	3,829	1,541
	Region Flanders	220,609	196,478	183,263	147,278	151,193	140,758	105,786	63,391	17,699	6,841
	Region Wallonia	123,670	115,355	103,291	73,856	75,741	75,911	62,040	38,549	10,147	3,653
Real number based on population statistics 2009	Belgium	381,372	349,580	324,956	252,175	260,767	250,312	199,774	121,011	34,319	13,243
	Region Brussels Capital	36,134	33,529	27,783	25,531	23,795	24,078	21,750	13,460	5,137	1,805
	Region Flanders	217,944	198,685	190,744	150,802	159,747	148,347	113,128	67,109	19,204	7,404
	Region Wallonia	127,294	117,366	106,429	75,842	77,225	77,887	64,896	40,442	9,978	4,034
Percentage difference between weighted and real amount	Belgium	1.4	2.5	3.9	4.4	5.6	5.7	7.5	5.8	7.7	9.1
	Region Brussels Capital	12.4	14.0	7.9	21.9	18.7	19.9	22.3	10.6	25.5	14.6
	Region Flanders	-1.2	1.1	3.9	2.3	5.4	5.1	6.5	5.5	7.8	7.6
	Region Wallonia	2.8	1.7	2.9	2.6	1.9	2.5	4.4	4.7	-1.7	9.4

[84] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, region and age classes.

APPENDIX 3: WEIGHTED RESULTS BY CIVIL STATE, AGE AND GENDER

Table 35: Weighted number of observations on individual level by civil state and age (men)[85]

		0-4 YEAR	5-9 YEAR	10-14 YEAR	15-19 YEAR	20-24 YEAR	25-29 YEAR	30-34 YEAR	35-39 YEAR	40-44 YEAR	45-49 YEAR
Weighted amount based on sample observations	Unmarried	316,089	301,683	309,986	333,358	314,373	271,116	189,461	142,398	101,565	72,178
	Married	0	0	0	202	11,553	70,454	141,801	203,327	246,029	264,903
	Widower, widow	0	0	0	0	0	55	215	615	1,382	2,641
	Divorced	0	0	0	0	413	4,398	15,471	36,565	57,643	69,053
Real number based on population statistics 2009	Unmarried	288,972	293,650	311,590	330,443	314,492	292,118	204,782	166,226	139,034	110,538
	Married	0	0	0	36	5,511	51,180	114,932	175,996	225,849	249,720
	Widower, widow	0	0	0	0	0	0	0	196	995	1,368
	Divorced	0	0	0	0	0	1,337	8,528	21,391	35,473	47,390
Percentage difference between weighted and real amount	Unmarried	-9.4	-2.7	0.5	-0.9	0.0	7.2	7.5	14.3	26.9	34.7
	Married	0.0	0.0	0.0	-461.1	-109.6	-37.7	-23.4	-15.5	-8.9	-6.1
	Widower, widow	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-213.8	-38.9	-93.1
	Divorced	0.0	0.0	0.0	0.0	0.0	-228.9	-81.4	-70.9	-62.5	-45.7

[85] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, civil status and age classes.

Table 35 (continued): Weighted number of observations on individual level by civil state and age (men)[86]

		50-54 YEAR	55-59 YEAR	60-64 YEAR	65-69 YEAR	70-74 YEAR	75-79 YEAR	80-84 YEAR	85-89 YEAR	90-94 YEAR	95 YEAR AND MORE
Weighted amount based on sample observations	Unmarried	46,775	30,323	21,163	13,545	12,069	9,658	6,087	2,647	540	143
	Married	258,882	246,403	229,777	170,870	160,535	127,706	73,655	28,891	3,893	529
	Widower, widow	4,665	7,618	11,122	12,636	18,794	25,317	26,362	19,412	5,333	1,725
	Divorced	66,898	55,395	39,614	22,132	14,935	8,569	3,602	1,285	186	16
Real number based on population statistics 2009	Unmarried	80,695	56,226	39,324	22,751	21,326	12,856	7,934	4,571	936	134
	Married	246,285	235,794	231,964	167,832	165,803	127,459	79,453	29,816	4,275	559
	Widower, widow	2,087	3,765	5,086	7,955	11,640	16,480	14,855	11,681	3,751	1,395
	Divorced	48,693	50,307	41,795	28,710	20,572	18,181	16,077	11,757	1,729	428
Percentage difference between weighted and real amount	Unmarried	42.0	46.1	46.2	40.5	43.4	24.9	23.3	42.1	42.3	-6.7
	Married	-5.1	-4.5	0.9	-1.8	3.2	-0.2	7.3	3.1	8.9	5.4
	Widower, widow	-123.5	-102.3	-118.7	-58.8	-61.5	-53.6	-77.5	-66.2	-42.2	-23.7
	Divorced	-37.4	-10.1	5.2	22.9	27.4	52.9	77.6	89.1	89.2	96.3

[86] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, civil status and age classes.

Table 36: Weighted number of observations on individual level by civil state and age (women)[87]

		0-4 YEAR	5-9 YEAR	10-14 YEAR	15-19 YEAR	20-24 YEAR	25-29 YEAR	30-34 YEAR	35-39 YEAR	40-44 YEAR	45-49 YEAR
Weighted amount based on sample observations	Unmarried	301,448	289,318	296,352	318,311	288,328	221,915	143,612	99,930	66,904	47,020
	Married	0	0	0	2,160	34,622	113,411	172,652	221,749	253,876	267,154
	Widower, widow	0	0	0	0	48	336	1,037	2,680	5,420	10,376
	Divorced	0	0	0	9	1,397	9,301	24,901	49,470	68,700	77,531
Real number based on population statistics 2009	Unmarried	270,697	281,696	299,980	321,429	293,379	242,146	166,809	132,749	104,562	87,519
	Married	0	0	0	721	19,636	88,001	143,627	202,843	235,881	252,330
	Widower, widow	0	0	0	0	0	143	730	1,472	4,150	7,849
	Divorced	0	0	0	0	144	3,260	15,102	33,523	48,663	61,849
Percentage difference between weighted and real amount	Unmarried	-11.4	-2.7	1.2	1.0	1.7	8.4	13.9	24.7	36.0	46.3
	Married	0.0	0.0	0.0	-199.6	-76.3	-28.9	-20.2	-9.3	-7.6	-5.9
	Widower, widow	0.0	0.0	0.0	0.0	0.0	-135.0	-42.1	-82.1	-30.6	-32.2
	Divorced	0.0	0.0	0.0	0.0	-870.1	-185.3	-64.9	-47.6	-41.2	-25.4

[87] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, civil status and age classes.

Table 36 (continued): Weighted number of observations on individual level by civil state and age (women)[88]

		50-54 YEAR	55-59 YEAR	60-64 YEAR	65-69 YEAR	70-74 YEAR	75-79 YEAR	80-84 YEAR	85-89 YEAR	90-94 YEAR	95 YEAR AND MORE
Weighted amount based on sample observations	Unmarried	30,887	20,375	14,454	10,470	11,351	12,158	11,405	7,716	2,618	1,158
	Married	255,569	235,962	214,712	155,734	140,636	102,484	50,327	16,052	1,513	185
	Widower, widow	17,427	26,214	39,365	49,198	75,521	108,442	115,704	86,610	26,696	10,403
	Divorced	72,057	58,118	43,620	25,680	18,767	12,862	7,281	3,590	848	289
Real number based on population statistics 2009	Unmarried	62,880	44,315	32,482	19,744	16,013	15,667	11,325	9,592	3,428	1,019
	Married	246,336	229,120	213,258	159,308	143,776	107,894	52,292	15,517	1,786	89
	Widower, widow	17,177	21,726	36,698	48,259	77,912	108,957	123,969	90,163	27,763	11,813
	Divorced	54,977	54,419	42,516	24,864	23,065	17,794	12,188	5,739	1,341	321
Percentage difference between weighted and real amount	Unmarried	50.9	54.0	55.5	47.0	29.1	22.4	-0.7	19.6	23.6	-13.6
	Married	-3.7	-3.0	-0.7	2.2	2.2	5.0	3.8	-3.4	15.3	-107.9
	Widower, widow	-1.5	-20.7	-7.3	-1.9	3.1	0.5	6.7	3.9	3.8	11.9
	Divorced	-31.1	-6.8	-2.6	-3.3	18.6	27.7	40.3	37.4	36.8	10.0

[88] The data in column actual distribution are based on data from ADSEI, see ADSEI (2011a): Population by gender, civil status and age classes.

APPENDIX 4: REVALUATION FACTORS ESTIMATED DAILY WAGES

Table 37: Revaluation factors for estimated daily wages by year of observation[89]

Performance year	Revaluation factor	Performance year	Revaluation factor
1940	6.405881	1975	2.861783
1941	6.405881	1976	2.621545
1942	6.405881	1977	2.447630
1943	6.405881	1978	2.342965
1944	6.405881	1979	2.242618
1945	6.405881	1980	2.102809
1946	6.405881	1981	1.953805
1947	6.405881	1982	1.796984
1948	6.405881	1983	1.669099
1949	6.405881	1984	1.600875
1950	6.405881	1985	1.557084
1951	6.405881	1986	1.537166
1952	6.405881	1987	1.543957
1953	6.405881	1988	1.526233
1954	6.405881	1989	1.480237
1955	6.255990	1990	1.430892
1956	6.111607	1991	1.386377
1957	5.923573	1992	1.353496
1958	5.848569	1993	1.317196
1959	5.777393	1994	1.306018
1960	5.760077	1995	1.285983
1961	5.703567	1996	1.265062
1962	5.62469	1997	1.248598
1963	5.506432	1998	1.232904
1964	5.286357	1999	1.221376
1965	5.079292	2000	1.198833
1966	4.876030	2001	1.166827
1967	4.741017	2002	1.152043
1968	4.612922	2003	1.132669
1969	4.446323	2004	1.111100
1970	4.279467	2005	1.081046
1971	4.101257	2006	1.059012
1972	3.889245	2007	1.027111
1973	3.636347	2008	1.000000
1974	3.227094		

[89] The data in this table are based on data published by RVP (2011).