MAJOR GUIDELINES TOWARDS THE DEVELOPMENT OF A PORTUGUESE SAM[(*)]

PRINCIPAIS ORIENTAÇÕES PARA O DESENVOLVIMENTO DE UMA MATRIZ DE CONTABILIDADE SOCIAL

Autor: António Machado Lopes
- Director-Adjunto do Gabinete de Estudos e Conjuntura - Instituto Nacional de Estatística

ABSTRACT:

- This article presents the general framework adopted to build a SAM for the Portuguese economy within the architecture of the national statistical system. The paper is divided in four sections. The first is an assessment of the availability of data to develop a SAM in the near future. The second and the third sections discuss the main breakdown, links and sources, as well as some shortcomings of the available data. The last section concludes with some examples of links between different accounts.

KEY-WORDS:

- *National Accounts; Social Accounts; Household Budget Survey; Household Consumption Patterns; LEG-SAM.*

RESUMO:

- Este artigo apresenta a estrutura geral de desenvolvimento de uma Matriz de Contabilidade Social para Portugal, no quadro do sistema estatístico nacional. O artigo está dividido em quatro secções. Na primeira são avaliadas as condições para a criação de uma SAM num futuro próximo. Na segunda e na terceira secções são discutidas as principais desagregações a considerar, as articulações entre as diferentes contas e as fontes disponíveis, bem como algumas das suas principais limitações. Na última secção são dados alguns exemplos de ligações entre diferentes contas

PALAVRAS-CHAVE:

- *Contas Nacionais; Contabilidade Social; Inquérito aos Orçamentos Familiares; Padrões de Consumo das Famílias; LEG-SAM.*

[(*)] Este texto resulta da participação do INE no LEG-SAM (Leader Group - Social Accounting Matrix), grupo de projecto liderado pelo CBS da Holanda (Central Bureau of Statistics) e apoiado pelo Eurostat. Neste grupo participam ainda os institutos de estatística do Reino Unido, Itália, Grécia, Bélgica, Finlândia, Noruega. Formalmente, o texto é principalmente uma "composição" de alguns materiais produzidos até ao momento, resultantes da colaboração interdepartamental entre o Gabinete de Estudos e Conjuntura (GEC) e o Departamento de Coordenação e Contas Nacionais (DCCN). O texto muito beneficiou das sugestões e comentários do Dr. Daniel Santos, bem como da reflexão feita pelo grupo interdepartamental, no qual cabe destacar a participação da Dra. Isabel Quintela do DCCN. Todas as imprecisões, erros e outros lapsos são, porém, da minha responsabilidade.

1. GENERAL REMARKS:

The Portuguese statistical system has not a set of reconciled data on labour supply and demand, expenditure and income of households, in the framework of a system of satellite accounts: Labour Accounts and Social-Economic Accounts. Only National Accounts (NA) are available, for the moment on ESA 79 basis. So, the statistical information system excludes interim integrated information systems, with the exception of NA. Had a global integrated system been available in a regular basis and a SAM would appear as the piece of information that would close the system. However, being the major objective to build a pilot SAM as soon as possible, this implies for the near future:

A) A **top down approach**, in which the results of the National Accounts (NA) are taken as constraints (in general, this implies to take the values and structures of the NA as a *"fait accompli"*, adjusting the estimates to the NA aggregates).

B) A **low level of assessment of the quality of the other sources**, namely those related to expenditure and income of households and supply of the different types of labour force. Note that this does not mean an uncritical view of existing data, but only the assumption of its use without any major changes.

C) A **pilot SAM for the year 1995** on ESA 79 basis.

D) A "**learning by doing**" process, in order to achieve the expertise in building a SAM on ESA 95 basis, as soon as data in this format is available from the national accounts information system.

In the medium term, one can imagine there will be some kind of network connecting data needs for compiling a SAM with the development of the Labour Accounts and the improvement of NA quality. In fact, this network should be one of the outputs of the pilot exercise. For instance, the needs for a SAM built in a regular basis could be reflected into Labour Accounts and National Accounts frameworks, adjusting the architecture of these accounts to the input needs of the SAM. The same applies to he questionnaires of Household Budget Survey (HBS) and Labour Force Survey (LFS) which could be improved too.

2. CLASSIFICATIONS, LINKS AND SOURCES (EXPANSION TOWARDS AN ORIENTED EMPLOYMENT/HOUSEHOLD SAM)

2.1. PRODUCTION:

The **level A6** for the activity classification would be a good choice, due to the weakness of reconciled data on labour force/wage income on one hand, and data on income/expenditure of households on the other. Nevertheless, a more detailed breakdown, such as A17, should be used in the working process of the pilot SAM (and this applies to other kind of classifications). This will allow assessing the dimension of discrepancies between data sources, improving the quality of the pilot. Besides, it approximates this exercise to the real situation of producing a SAM according to users needs. The source is 1995 data of National Accounts, on ESA 79 basis.

2.2. GENERATION OF INCOME (1):

It is possible to consider **six types of labour for each productive branch**, defined by gender and level of education (low, medium and high levels). **The objective** is the

distribution of total wage income generated in each branch, which was evaluated in the NA, to the six types of labour.

The **main sources** are the Labour Force Survey and Quadros de Pessoal (these are administrative data; enterprises employing at least one paid employee have the legal obligation to send their list of personnel to the Ministry of Employment). The former source would be used to estimate the total annual worked hours for each type of labour (eventually, the Earnings Survey can also be considered). The latter would allow estimating the average hour compensation per type of labour.

2.3. GENERATION OF INCOME (2):

The following **classification of Households** into sub-sectors is possible, according to:

1. The main source of income (wage income, self-employment income, income property, pension income, other sources).

2. The dimension of "wage-households" (one member, two without child, two or more with one or more children).

3. At least on a preliminary basis, the "self employment income households" can be divided by number of employees (one, more than one). In addition, and on the same basis, both the "self employment income households" and the "income property households" can be classified using the same categories of dimension as those of "wage-households".

4. The "pension-households" related or not with age. At least on a preliminary basis, both can be classified by dimension, (one member, two or more members for the first kind of household; one member, two without child, two or more with one or more children, for the second one).

5. At least on a preliminary basis the "Other-households" can be classified by the same categories of dimension as those of "wage-households".

The **objective** is to establish the link between these groups of households and the value of each kind of income they receive. Note that the wage income should be distributed to all types of household, and the same applies to the other kinds of income.

The **main sources** are the Labour Force Survey and the Quadros de Pessoal for wage-income and the Household Budget Survey for the allocation of other income. Concerning wage-income, the Labour Force Survey and the Quadros de Pessoal provide estimates that should be constrained by the global results of wage income of NA (see 2.2.Generation of income (1)). The allocation of wage income should be done according to the structure of wage income by types of households, estimated from the results of the Household Budget Survey. The final result should be consistent with branch generation of wage income (NA), labour factor distribution of wage income (LFS, QP) and household sub-sector distribution of wage income (HBS).

In addition, the HBS will be the main source for the distribution of gross operating surplus to the households.

2.4 ALLOCATION AND SECONDARY DISTRIBUTION OF INCOME

The **first objective** is the breakdown of in/outflows of income property to/from each types of household, obtaining as a balance the total primary income. **Secondly**, it is necessary

to estimate the outflow of social contributions, direct taxes and other income transfers and inflow of social benefits and other income transfers, in order to obtain the disposable income of each type of household.

Once again, the results of the Household Budget Survey could be used as the **main source**, restricted to National Accounts global results. Other sources, such as data from tax and social security files can be used as well.

The principal drawback of this approach, based on the Household Budget Survey, is the strong probability that property income (level and weight) is underestimated, owing to the reluctance of households to answer these questions.

Another drawback is the "asymmetry" of the sub-matrices. In fact, is the breakdown by recipient households and the breakdown by payer households is possible, but the breakdown of correspondent sectors cannot be done so easily. The subdivision of these ones will be no more detailed then General Government/Other sectors.

2.5.   USE OF INCOME

The breakdown of final consumption by types of products and types of households (sub-sector) will be done taking the previous referred sources (NA and HBS). The consumption weight of each type of household in total consumption will be given by the HBS and should be taken as a constraint. The weights of each product on the total consumption, given by NA, should also be considered a constraint. The value of the total consumption (NA) will be distributed to the relevant matrix. An interactive process, based on an algorithm developed by Richard Stone will reconcile the level of each element of that matrix both with the figures of consumption of each product (NA) and the weight consumption of each type of household (HBS) on the total consumption (NA). See the example in 4.1.

3.   SOME SHORTCOMINGS OF AVAILABLE DATA

The relevant sources show some fragility but they still have a sufficient degree of quality to allow its use in the exercise of building a pilot SAM, at least with the 1995 as a reference year.

Labour Force Survey: Regarding the year of 1995, the one that is considered as reference for the pilot SAM, available data is of acceptable quality. However, for 1996 and 1997, the results of preliminary data exploration do not advise the use of the Labour Force Survey without any changes of its raw data. This is a severe constraint for the building of a SAM referring to 1996 or 1997.

Household Budget Survey: There are two main drawbacks related to the use of this source. The first is the reluctance of households to answer to questions concerning their level of income, which implies additional problems to evaluate the weights of each source of income. This can produce some distortions in the breakdown by type of household. The other drawback is that national accounts underused this source to estimate the levels of consumption of each product. Again, this can produce significant discrepancies between the structures of consumption given by National Accounts estimates and by the results of Household Budget Survey (see the correspondent example of consumption breakdown).

Quadros de Pessoal: This is an administrative which is not representative of household small businesses (the very small enterprises), although it will be used to estimate the wage rates. As a consequence, it is probable that wage rates, especially in some activities, will be overestimated, implying overestimates of wage-incomes. In addition, there are some branches that are not well covered, namely those where public services have major weights, such as education and wealth care. This will distort some estimates resulting from the combination of this data source and the LFS (see the correspondent example of breakdown of compensation by branch, gender and level of education, in 4.2.).

Income tax and Social Security files: The unit household in the fiscal sources is not compatible with the institutional unit implicit in the NA and in the HBS. The fiscal source would only enable the breakdown of the households into three groups of income earners: Compensations of employees, Income from individual employers and own-account workers, Income from retirement pensions. All the other categories of income (such as property income and transfers) are excluded. Besides, all households with a level of income below a limit in which there is an exemption in taxation are also excluded. The fiscal evasion is also a difficult problem to deal with. As a consequence income tax files can be used only on a complementary way. Concerning the social security files, there is no way to establish the link between the individual files and the type of household.

National Accounts: National Accounts will be the fundamental source, restraining strongly the building of the pilot SAM. On the other hand, 1995 data of National Accounts, although produced in a well-defined framework, with a standard methodology and sources, has a data basis comparatively out of date. As a result, some deviation from the true economic structures may exist, although it may not be easily noticed (since the "true economic structures" are not observable at least on a coherent framework).

4. SOME EXAMPLES OF EXPANSION TOWARDS THE SAM

4.1. CONSUMPTION BREAKDOWN

The following exercise combines two sources, the NA estimates of consumption by products, classified at the level A6 of ESA 95 (the figures in the last columns of the table) and the Household Budget Survey (HBS). Take the next table: according to the main source of income, there are five types of households and the level of consumption of each one is broken down by product (there are six products, defined according to classification A6 of ESA 95). Summing the elements of each row, the consumption by product is obtained (this is the first column labelled with "Total"). Summing the elements of this column, the HBS estimates of total consumption is also obtained. Each element of the last but two rows represents the level of consumption by type of household (the row labelled "total"), according to the HBS data. The next row shows the correspondent structure of consumption.

HOUSEHOLD CONSUMPTION PATTERNS AND TOTAL CONSUMPTION (NON-RECONCILED DATA)

Unit: 10^6 escudos

PRODUCTS	HBS-HOUSEHOLDS							Constraints	
	1	2	3	4	5	Total		NA total	
A	215081	64943	16994	95880	11902	**404800**	0.053	**624319**	0.064
B	1990397	609941	167101	659186	131297	**3557922**	0.467	**5655241**	0.580

C	14353	2104	726	13879	106	**31168**	**0.004**	**12642**	**0.001**
D	934568	261229	60508	190085	58977	**1505367**	**0.198**	**1498510**	**0.154**
E	663253	195646	137295	191554	42300	**1230050**	**0.161**	**997579**	**0.102**
F	525911	122325	55318	149748	38592	**891893**	**0.117**	**956110**	**0.098**
Total	**4343564**	**1256187**	**437942**	**1300333**	**283174**	**7621200**	**1.000**	**9744401**	**1.000**
Constraints HBS	**0.570**	**0.165**	**0.057**	**0.171**	**0.037**	**1.000**			
Constraints HBS/AN	**5553643**	**1606150**	**559949**	**1662595**	**362064**	**9744401**			

It can be seen that the HBS underestimate the NA total consumption, although some components of the former are higher than the latter. In other words, the structures are significantly different. Nevertheless, it is possible to reconcile these two sets of data. The underlying hypotheses are, of course, that both the NA levels of consumption and the HBS structure are sufficiently good estimates of the true values.

Firstly, note that the last row results from the breakdown of NA total consumption by type of household, according to their weight in the total consumption. It means that, by construction, each element of this row vector reconciles the NA total consumption and the HBS structure of consumption by household. Next, using the RAS procedure, the matrix is filled with new estimates of the breakdown of consumption by products and type of households:

HOUSEHOLD CONSUMPTION PATTERNS AND TOTAL CONSUMPTION (RECONCILED DATA)

Unit: 10^6 escudos

PRODUCTS	HBS-HOUSEHOLDS							Constraints	
	1	2	3	4	5	Total		N Accounts	
A	333786	99530	28150	144420	18433	**624319**	**0.064**	**624319**	**0.064**
B	3177975	961741	284780	1021536	209209	**5655241**	**0.580**	**5655241**	**0.580**
C	5895	853	318	5532	44	**12642**	**0.001**	**12642**	**0.001**
D	933341	257638	64500	184251	58780	**1498510**	**0.154**	**1498510**	**0.154**
E	537425	156556	118743	150649	34206	**997579**	**0.102**	**997579**	**0.102**
F	565221	129832	63458	156207	41392	**956110**	**0.098**	**956110**	**0.098**

HBS/AN	4343564	1256187	437942	1300333	283174	9744401	1.000	9744401	1.000
HBS	0.570	0.165	0.057	0.171	0.037	1.000			

## 4.2 BREAKDOWN OF COMPENSATION BY BRANCH, GENDER AND EDUCATION LEVEL

The objective is to distribute the compensation by branch (roughly at level A17) by gender and level of education (four levels were considered: G1-level 1 of ISCED-97, G2-level 2 of ISCED-97, G3-level 3 of ISCED-97 and G4-level 5 of ISCED). The procedure has two stages: the aim of the first is to obtain preliminary estimates and the aim of the other is to reconcile them with the NA figures.

Stage one: preliminary estimates

Data was taken from Quadros de Pessoal, namely figures of earnings by gender, level of education and branch and organised in a matrix [C*] with a (8,14) dimension. This matrix was adjusted in order to incorporate the employer's actual and imputed social contributions. Since this data is available in the NA only for the total of each branch, it was assumed that the implicit tax of actual social contribution was the same in each branch. This means that the tax was considered the same, independently of the kind and amount of labour force used in the branch. The imputed social contribution was assumed to be proportional to the amount of wage and salaries and uniform in each branch. The first assumption is very close to reality. The second assumption must be seen as a simplification derived from the availability of data but with no important effects considering the value of the imputed contributions. Consequently, the corrected earnings were defined according to $W_{ij}^g = W_{ij}^l * (1 + t_{isc}) / (1 - t_{asc})$, where i refers to the type of labour and j refers to the branch. Each estimate was also adjusted to an annual scale, $W_{ij} = 48 * W_{ij}^g$.

The matrix of hours worked in each branch, by gender and education level, was taken from the same source (Quadros de Pessoal). This matrix, with the general element H_{ij}, was also scaled to an annual basis. Combining the two set of data, the average hourly earning, $w_{ij} = (W_{ij}/H_{ij})$, was estimated.

It was admitted that data of labour Force Survey should be used to build the relevant matrix of hours worked. This matrix was constructed on a four-term average of actual hours by the same breakdown, scaled to an annual basis. Actually three matrices were calculated, one for wage-employment, another for total employment and the last one for the rest of (Other) employment, each one with the general element H_{ij}^w, H_{ij}^t and H_{ij}^o, respectively. Preliminary estimates were calculated, by multiplying each w_{ij} by the correspondent H_{ij}^w. The matrix of compensation [C], with the general term C_{ij}, was generated, as a result. Knowing that each row refers to a kind of labour, and each column to a branch, summing the elements of a given column of the matrix [C] the compensation paid by the branch is obtained. Summing the elements of a given row the compensations of a given kind of labour is obtained. See chart 1 to evaluate the "quality" of the first estimates.
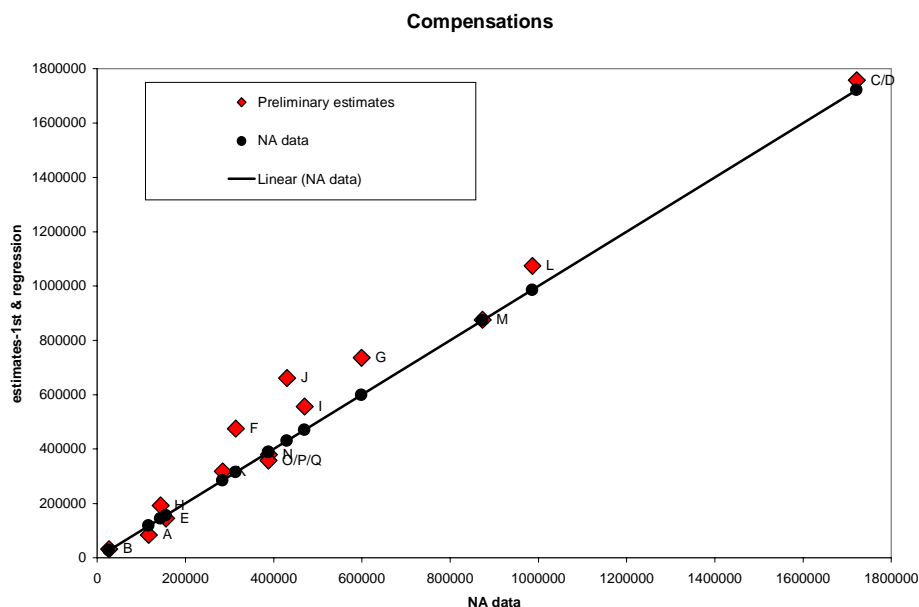
Compensations

Stage two: Reconciliation with NA data

The next step has the objective of reconciling the first estimates with data from National Accounts. An interactive process is developed, taking as constraint the NA figures of compensations by branch. Another (weak) constraint is given by the weights of each type of labour in the total compensation, obtained from the matrix [C*] of the first estimates. See the next two table, the first showing the final estimates of compensation breakdown, the second presenting the correspondent estimates of monthly compensation per worker. See also chart 2, for a preliminary view of the distribution of monthly compensation per worker.

COMPENSATIONS

UNIT: 10 ^9 escudos

		A	B	C/D	E	F	G	H	I	J	K	L	M	N	O/P/Q	Total
M	G1	77.8	21.6	680.8	52.2	239.3	203.8	52.2	182.1	29.4	32.8	157.0	20.6	25.5	65.2	1840.5
	G2	0.7	1.9	202.0	32.4	26.2	95.2	16.2	73.5	82.0	28.0	77.1	10.2	10.9	24.4	680.8
	G3	1.8	0.6	136.5	23.3	18.9	65.0	6.6	43.4	101.8	36.6	64.6	17.6	8.5	24.5	549.7
	G4	5.5	0.7	162.7	23.1	17.4	40.9	1.7	70.0	108.9	82.8	406.1	249.1	105.2	43.4	1317.5
F	G1	27.2	1.4	354.9	3.7	1.4	72.9	43.8	12.9	11.1	18.5	45.2	51.9	65.5	157.3	867.6
	G2	1.1	0.3	76.4	5.9	3.9	53.0	12.0	32.5	22.2	50.5	29.7	34.3	26.7	379.4
	G3	2.4	0.2	60.8	5.8	3.7	39.6	6.4	26.8	37.4	29.7	47.3	28.4	21.2	23.7	333.4
	G4	0.3	0.0	47.4	9.4	3.7	28.7	4.7	30.4	27.3	33.4	138.2	466.2	117.7	23.0	930.6
	M	85.9	24.9	1182.0	131.0	301.8	404.8	76.7	368.9	322.2	180.3	704.8	297.5	150.1	157.6	4388.5
	F	31.0	2.0	539.4	24.7	12.7	194.2	67.0	101.1	108.4	103.8	281.2	576.1	238.7	230.7	2511.0
	TOTAL	116.9	26.8	1721.4	155.8	314.5	599.1	143.7	469.9	430.6	284.1	986.1	873.6	388.8	388.3	6899.5

MONTHLY COMPENSATIONS PER EMPLOYEE

UNIT: 10 ^3 escudos

		A	B	C/D	E	F	G	H	I	J	K	L	M	N	O/P/Q	Total
M	G1	88.9	100.3	123.0	320.8	100.5	117.9	110.2	198.9	353.6	168.1	116.3	154.2	189.9	114.3	124.7
	G2	109.3	120.9	181.8	324.3	121.1	148.2	135.3	265.9	397.3	175.6	112.3	177.3	241.4	161.1	179.4
	G3	145.9	104.4	201.5	352.7	141.4	173.7	124.4	267.8	362.8	183.5	132.4	184.9	257.6	201.1	203.4
	G4	189.5	102.7	411.5	528.7	283.7	343.5	355.0	545.6	554.4	398.7	889.8	414.2	449.1	377.9	506.8
F	G1	61.5	41.2	78.7	249.1	75.9	82.7	76.6	157.6	216.0	86.1	91.5	105.3	144.9	61.7	80.3
	G2	80.0	85.7	113.9	293.7	94.1	100.3	89.0	224.3	328.6	128.5	84.6	122.3	168.7	124.0	123.2
	G3	93.5	52.3	129.3	269.3	98.5	106.4	103.1	197.3	281.7	131.0	90.2	141.9	180.8	128.0	132.5
	G4	135.6	0.0	234.3	433.6	137.9	199.1	161.5	307.3	348.9	210.0	324.9	224.6	249.7	230.1	242.6
	M	92.9	101.8	153.1	351.8	108.1	141.3	117.8	248.9	420.2	236.6	236.4	335.1	336.0	164.3	184.0
	F	64.1	45.8	92.2	317.6	102.3	100.8	83.9	222.3	299.7	134.2	137.7	191.3	191.9	75.6	124.1
	TOTAL	83.0	93.4	126.9	345.9	107.8	125.0	99.1	242.7	381.6	185.0	196.3	224.1	230.0	96.8	156.5

These are just some illustrative estimates, with no other objective than to show the procedures that can be developed to build a SAM for Portugal. Much work has to be done to gather, reconcile data and test the results. One of the major issues is the consistence of estimates based on different sets of data. For instance, even if the compensation per worker estimates has some expectable characteristics, such as a positive correlation between the compensation per worker and the level of education, as well as relative higher values for male, the gap between male and female compensation should be checked with other sources.

Monthly Compensation per Worker

To briefly conclude, the building of a SAM is obviously facilitated when NA is taken as a constraint. The task would be more difficult, although more interesting, if the objective was the building of a system of integrated information, linking from the bottom data on labour, expenditure and output in a meaningful breakdown. In this framework much of intermediate reconciliation would have to be done and probably there would have to be also some changes in the procedures of raw data collection.

ANNEX: Branch/products classification and level of education

A: Agriculture, hunting, forestry

B: Fishing

C/D: Industry

E: Electricity, Gas and Water

F: Construction

G: Retail and Repair

H: Hotels and restaurants

I: Transport, Storage and Communication

J: Financial Activities

K: Real Estate, Renting and Business activities

L: Public Administration and Defence, Compulsory Social Security

M: Education

N: Health, Social Work

O/P/Q: Other Community, Social and Personal Services, Private Households With Employed Persons, International Organisms and Other Institutions

G 1: Level 1 of ISCED-97

G2: Level 2 of ISCED-97

G3: Level 3 of ISCED-97

G4: Level 5 of ISCED-97