



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# » Measuring Monthly Net Incomes from linked Labour Force Survey Data with Administrative Records

**David Leite** | Sofia Rodrigues | Paula Cruz | Filipa Ribeiro | David Santos | António Portugal



WS on LFS Methodology, 25-26 may 2023



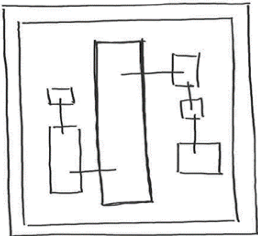
# Agenda

1. Introduction
2. Data
3. Methodology
4. Preliminary Results
5. Conclusions
6. Future Work

# 1. Introduction

- The aim of Statistics Portugal (SP) was to find out for how many of the individuals aged 15 and over in the 2019 and 2020 **Labour Force Survey (IE / LFS)** samples it would be possible to obtain, from administrative sources, namely from the **Monthly Income Declaration of the Tax Authority (DMR\_AT)**, information comparable with that collected on the net monthly income of employees.

## 2. Data – National Data Infrastructure

more intensive and integrated use of data	gain resources, space to intensify innovation	substantial extension of the domains covered
grasp the multidimensionality of data		application of new information extraction techniques
harness the entire production chain	stepping up the appropriation and use of administrative data and other sources	greater return to society

## 2. Data

- Labour Force Survey (IE) from SP
- Monthly Income Declaration of the Tax Authority (DMR\_AT) from Tax Authority
- Resident Population Database (BPR) administrative data, managed at SP, enriched with dwelling id, ALOJ\_COD from National Household Database (BNE), census and administrative data, managed at SP, based on a previous record linkage on address data.

### 3. Data - sources

## IE

ANO 2019

AREA\_AM

N\_ALOJ\_AM

N\_INDIVIDUO

SEX

AGE

CPT

SIT\_PROF

NAME\_IE

## BPR 2019

EDIF\_COD

ALoj\_COD

ENCRYPTED VAT NUMBER (NIF)

SEX

AGE

PLACE OF BIRTH

HOUSEHOLD ID CODE

NAME\_BPR

## DMR

YEAR

ENCRYPTED VAT NUMBER (NIF)

ECONOMIC VARIABLES

...

## 4. Data – structures and fields

1/2 VARIABLE ...	DESCRIPTION ...
ANO	Data reference year
AREA_AM	Sample area
N_ALOJ_AM	No. of accommodation in the sample area
N_INDIVIDUO	Individual number
EDIF_COD	Building code
ALOJ_COD	Dwelling code
NIF	Tax identification number encrypted
HID_IRS	Household identification code
NOME_BPR	Name at BPR
CPT_IE	IE occupation code
CPT_BPR	BPR occupation code
SIT_PROF_IE	IE profession status
SIT_PROF_BPR	BPR profession status

## 4. Data – structures and fields

2/2	...	...
	VARIABLE	DESCRIPTION
	RENDIMENTO_IE	IE Income
	SEXO_IE	IE Sex
	SEXO_BPR	BPR Sex
	DATA_NASCIMENTO_IE	Date of birth IE
	DATA_NASCIMENTO_AT	Date of birth BPR
	COMPARA_DTNASC	Compare date of birth (1: Equal; 2: Not equal)
	IDADE_IE	Age at IE
	IDADE_BPR	Age at BPR
	DIFF_IDADE	Absolute age difference between IE and BPR
	RANK	Ranking based on the variable DIFF_IDADE, by individual and accommodation.



## 4. Methodology – record linkage

- Of the 25,344 distinct **households** in the **SAMPLE\_IE\_2019**, it was immediately possible to verify that 20,919 households are **present in the 2019 BPR** linked by **dwelling code** (ALOJ\_COD).
- At the level of **individuals**, the following **linkage criteria** were considered:
  - (a) Restriction of the possible linkages (via ALOJ\_COD) to **individuals with the same SEX**;
  - b) Calculation of the **(absolute) difference between the ages** (reported to 2019) in each of the tables and creation of a ranking variable of increasing ordination.
  - c) Inclusion, in the possible links returned, of the **name of the individual as it appears in the BPR**, as well as the **date of birth and gender of the individual**, and the **age difference** between the table of the **IE sample** and the one in the **BPR**.
  - d) Allocation of the individuals to each dwelling, based on the **household ID**.

## 4. Methodology – record linkage

- e) From the total of 54,541 individuals in the **LFS sample**, a connection of **66.6%** was obtained with the **2019 BPR**.
- f) Using the new data made available from the **LFS sample**, that included the new variables **NAME and PLACE OF BIRTH**, several tests and different connections (Date of birth, Sex, Name, Parish) were made to try to connect and fill the largest number of records, the following table shows a summary of all connections.

## 4. Preliminary Results

- It was possible to connect 91% to BPR and there are 91% records with NIF.

	Total	NIF	No NIF	Link to BPR
Records Number	54,541	49,412	5,129	49,424
%		<b>91%</b>	<b>9%</b>	<b>91%</b>

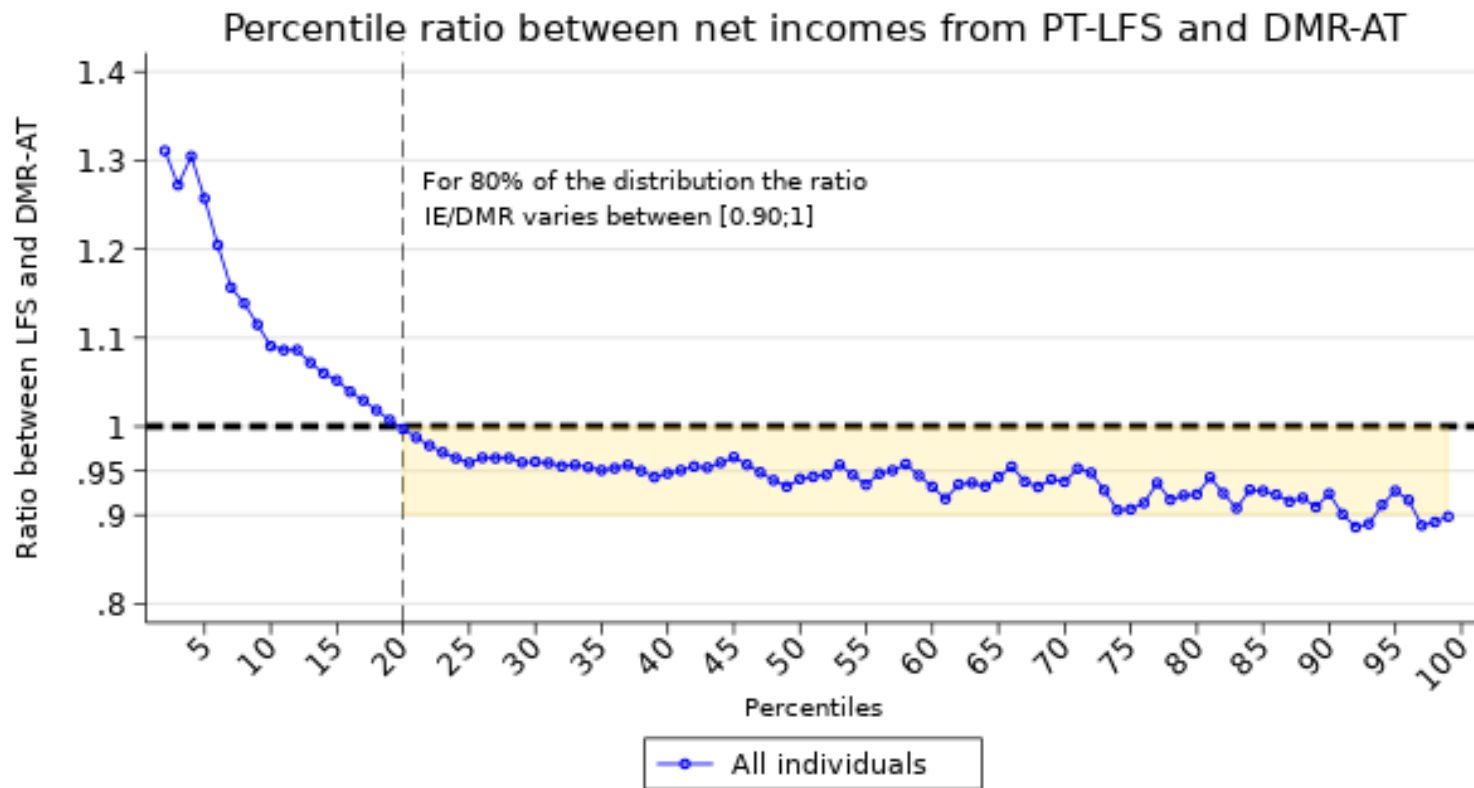
- These results point to the fact that the link between data from administrative sources, or registries of statistical units and data from statistical operations, in this case (LFS), have a very high degree of association.

# 5. Preliminary Results: Harmonization of concepts

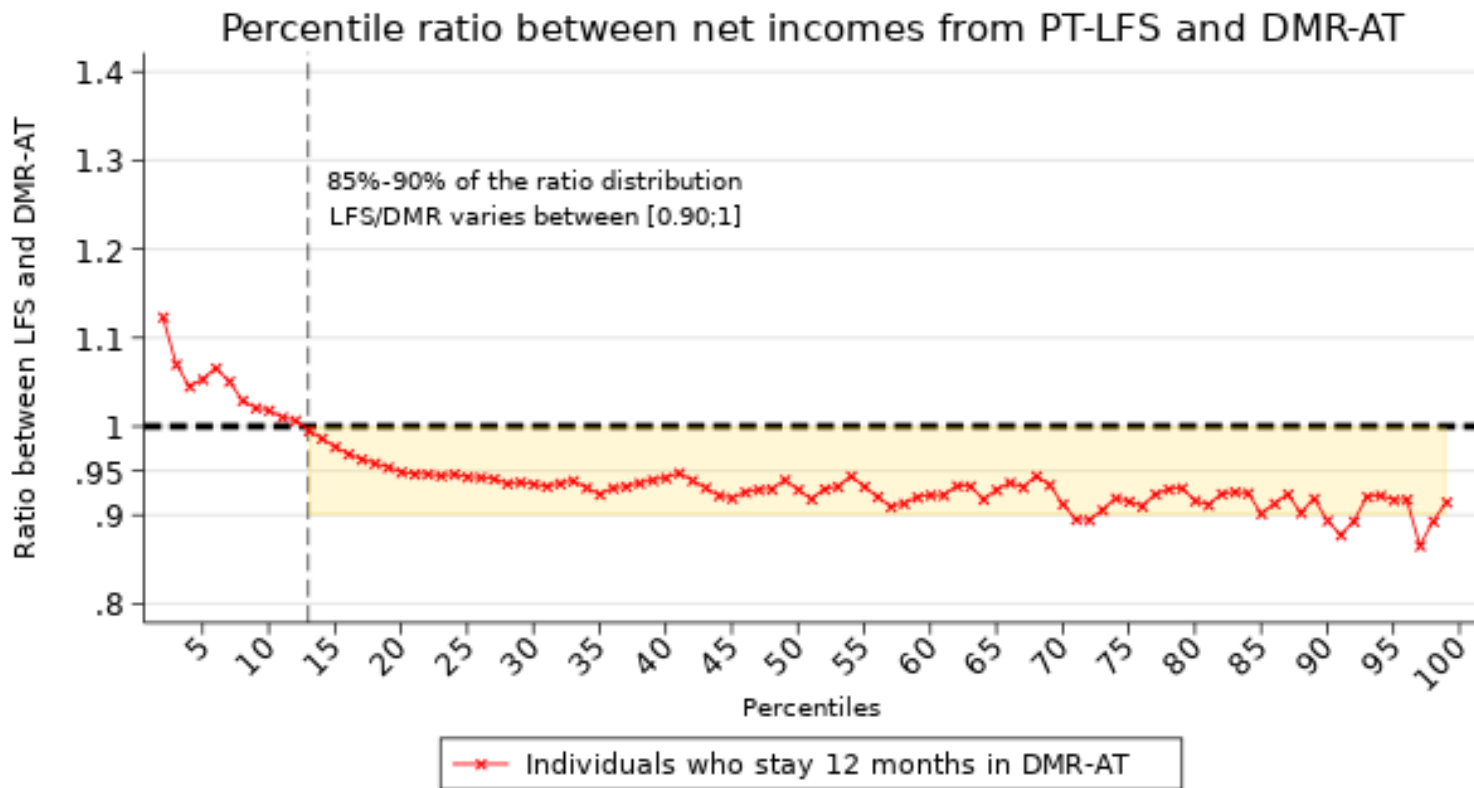
- After the integration of the survey data with administrative sources, we carefully harmonize the concept of “net income” of the LFS with the administrative data.

Inquérito ao Emprego		DMR - Modelo 22			
	Rendimentos Líquidos do Emprego Principal	Quadro	Campo	Código(s)	Observações
<b>Valores Incluídos</b>	Rendimentos do trabalho dependente	5	3	A	artº 2 nº 1 do CIRS
	Comissões	5	3	A	artº 2 nº 1 e 2 do CIRS
	Diuturnidades	5	3	A	artº 2 nº 1 e 2 do CIRS
	Prémios de antiguidade	5	3	A	artº 2 nº 1 e 2 do CIRS
	Prémios por assiduidade	5	3	A	artº 2 nº 1 e 2 do CIRS
	Produtividade	5	3	A	artº 2 nº 1 e 2 do CIRS
	Abonos para falhas	5	3	A	artº 2 nº 11 c) do CIRS
	Horas extraordinárias	5	3	A	artº 2 nº 1 e 2 do CIRS
	Subsídios de refeição, alojamento de transporte	5	3	A, A21, A22	Subsídio de refeição (parte sujeita e não sujeita) + Ajudas de custo
	Trabalho nocturno, aos domingos e feriados	5	3	A	artº 2 nº 1 e 2 do CIRS
<b>Valores a Deduzir</b>	Gratificações e similares decorrentes da condição do trabalhador	5	3	A, A2, A62	artº 2 nº 11 g) do CIRS ("em razão da prestação de trabalho" não atribuídas pela entidade patronal)
	Impostos	5	6	Retenção IRS	Retenções de IRS
	Contribuições para a Segurança Social	5	7	Contribuições Obrigatórias	Contribuições para SS, CGA e outros sistemas obrigatórios de segurança social + Contribuições para subsistemas de saúde (ex. ADSE)
	Quaisquer outros subsistemas de fins análogos			Quotizações Sindicais	Quotizações para Sindicatos
<b>Valores Excluídos</b>	Duodécimos dos subsídios de férias e de Natal	5	3	A3, A4, A61	Subsídios de férias e natal
	Pagamentos em géneros	5	3	A5, A63, A66, A24	Utilização da casa casa, vales educação, utilização de viatura
	Rendimentos de investimentos-ativos, poupanças ou ações	5	3	A5, A65, A67	Empréstimos sem juros, aquisição de viatura por preço inferior ao de mercado, ganhos sobre valores mobiliários em benefício dos trabalhadores
	Reembolsos feitos pelo empregador para despesas relacionadas com o trabalho	5	3		
	Pagamentos de indemnizações	5	3	A25, A30	Indemnizações deslocação, lesão corporal, etc.

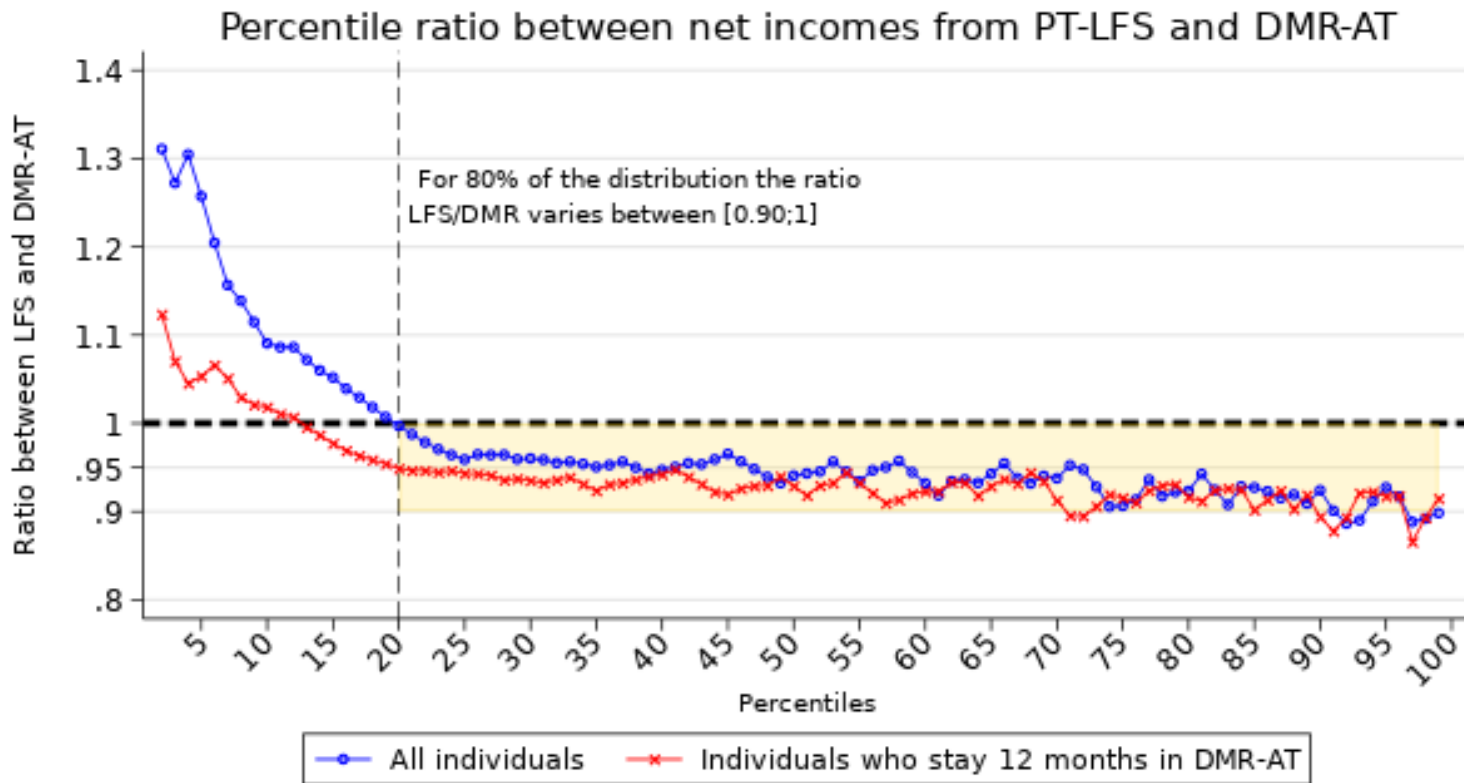
## 6. Preliminary Results: Percentile Ratios, LFS vs. DMR-AT



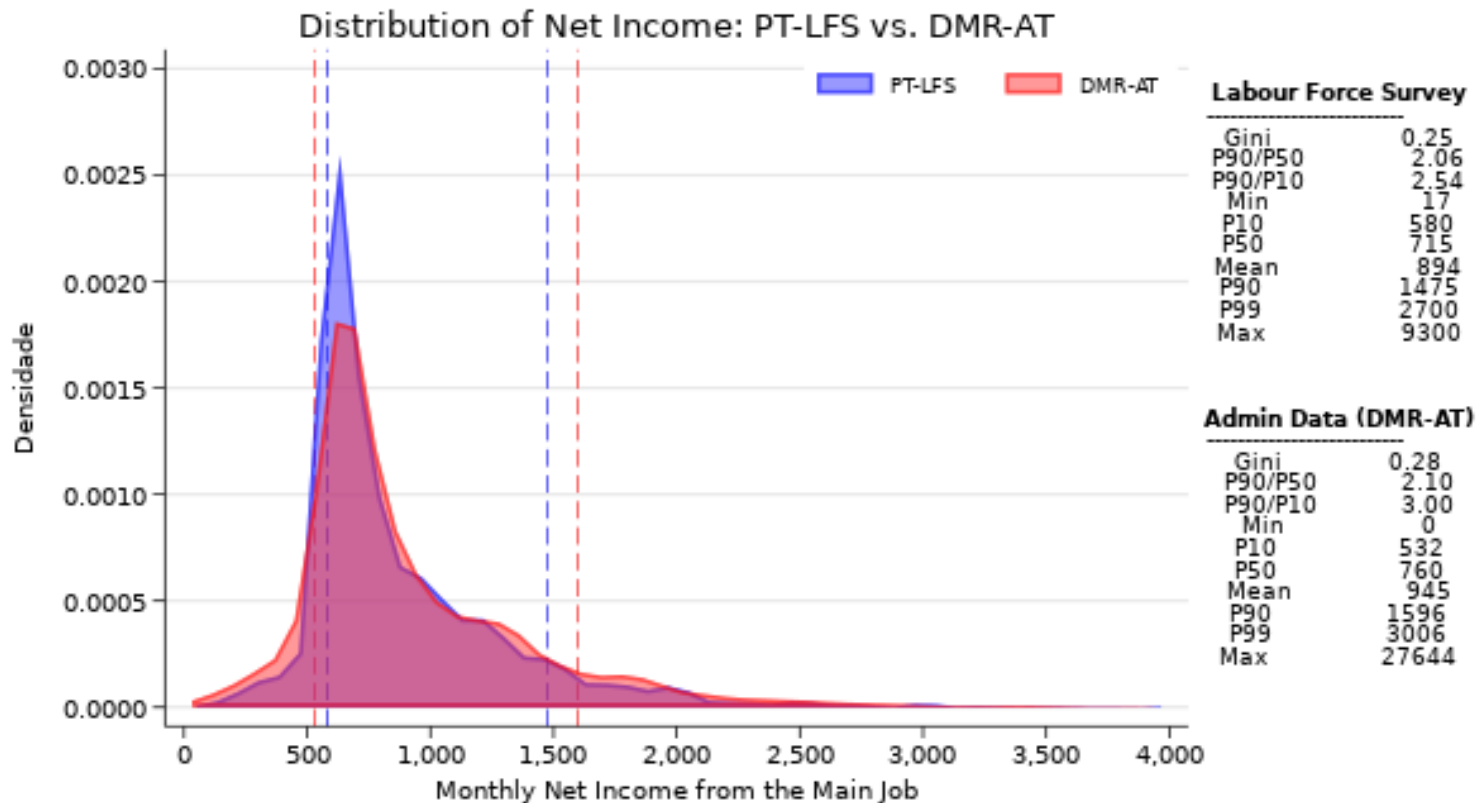
## 7. Preliminary Results: Percentile Ratios, LFS vs. DMR\_AT



## 8. Preliminary Results: Percentile Ratios, LFS vs. DMR\_AT



## 9. Preliminary Results: Income distribution, LFS vs. DMR\_AT





## 10. Conclusions (learnings)

- It was possible to link 91% of the employees in LFS to the admin data of the monthly income declaration (DMR\_AT).
- From DMR\_AT it was possible to attribute monthly net incomes to 1,590 individuals who did not report their monthly net incomes in the survey.
- Administrative sources seem to **underestimate monthly net incomes in the first quintile** (informality?).
- The percentile ratio between LFS and DMR\_AT is by about 94% for the top 4 quintiles of the distribution (highlights the impacts on income distribution measurement when using administrative sources).

# 11. Future Work

- Improvement the comparability of the dates of reference of the labour market status between the two data sources.
- Study the case of the individuals with more than one job who appear in the administrative data as employees, but declare that their main activity is other than employee.
- Extend this work to employees whose employer is an individual rather than an entity (e.g., housekeeping services)
- Extend this work to monthly earnings of self-employed people (e.g., platform workers) using electronic invoices from *e-Fatura*.



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# » Measuring Monthly Net Incomes from linked Labour Force Survey Data with Administrative Records

Obrigado | Thank You «

WS on LFS Methodology, 25-26 may 2023

