

25-26 May 2023, Lisbon

16th Workshop on Labour Force Survey Methodology

Italian strategy to obtain gross monthly pay for LFS employees

Index

- Mixed strategy
- Operative definition of gross wage from administrative data
- Record linkage and validation
- Statistical model
- Results
- Gross vs. Net income
- Conclusion

Mixed strategy

- The strategy to obtain gross monthly pay perceived by the employees has been designed after having conducted an **extensive experimental study**.
- In Italy, **directly asking interviewees for their gross salary is not feasible**, as they are not aware of the value (employees receive their net salary directly, after deducting taxes and contributions). The information has a high rate of non-responses and low reliability.
- On the other hand, **administrative information can be used** to this aim. In recent years Istat is implementing a harmonized set of statistical registers integrating administrative data. The Labour Register (LR) contains information on employment, including wages.
- The variable INCGROSS is obtained adopting a **mixed strategy based on administrative data and model prediction**:
 - i. LFS collected data are linked with the LR and, subject to a validation step, information on the wage is imputed from the LR;
 - ii. Model prediction is applied in cases where LR information is not validated (inconsistent work characteristics between LFS and LR, outliers) or not available (missing value in the LR, missing linkage).

Operative definition of gross wage from administrative data

- The LR covers all **regular paid jobs**, active in the national territory and in all sectors of economic activity and the statistical unit is job position.
- We use LR to obtain monthly gross pay for the main position of LFS employees using also **information related to the reference week**.
- LR is based on a **plurality of administrative sources** each with different coverage, variables, units, reference periods, metadata, timeliness etc. This implies that the LR is produced through a complex process of harmonization and integration of these data sources.
- The defining aspects of gross wage also raise relevant issues. The information have been partially treated to derive a **new statistical variable taking into account the requirements of regulations of LFS**:
 - additional monthly payments (e.g. yearly payments such as 13th month) have been hived off and added in proportion to the monthly wage received.
 - in public sector, unlike the private sector, maternity and parental leave are included in declared gross wages and cannot be separately and clearly identified as social benefits. To harmonize the estimate of gross wage between private and public sector, the allowances payable by social security institutions were added to the gross wage also in the private sector.

Linkage and validation with RL

This process consists of **two steps**:

1. Each **LFS employee is linked with one or more job positions** active during the reference week in LR (possible thanks a unique pseudo-anonymized person ID which accounts of privacy requirements);
2. The **validation of consistency between the LFS and LR work characteristics is made** using 3 variables: full time/part time, temporary/permanent, economic activity (NACE). The possible combinations resulting from the comparison of linked records between LFS and LR can be:
 - fully validated: all variables are consistent
 - partially validated no one: two variables are consistent
 - partially validated no two: only one variable is consistent
 - not validated: no variables are consistent between LFS and LR.

The **working time** is the variable considered more relevant because determines different pay levels.

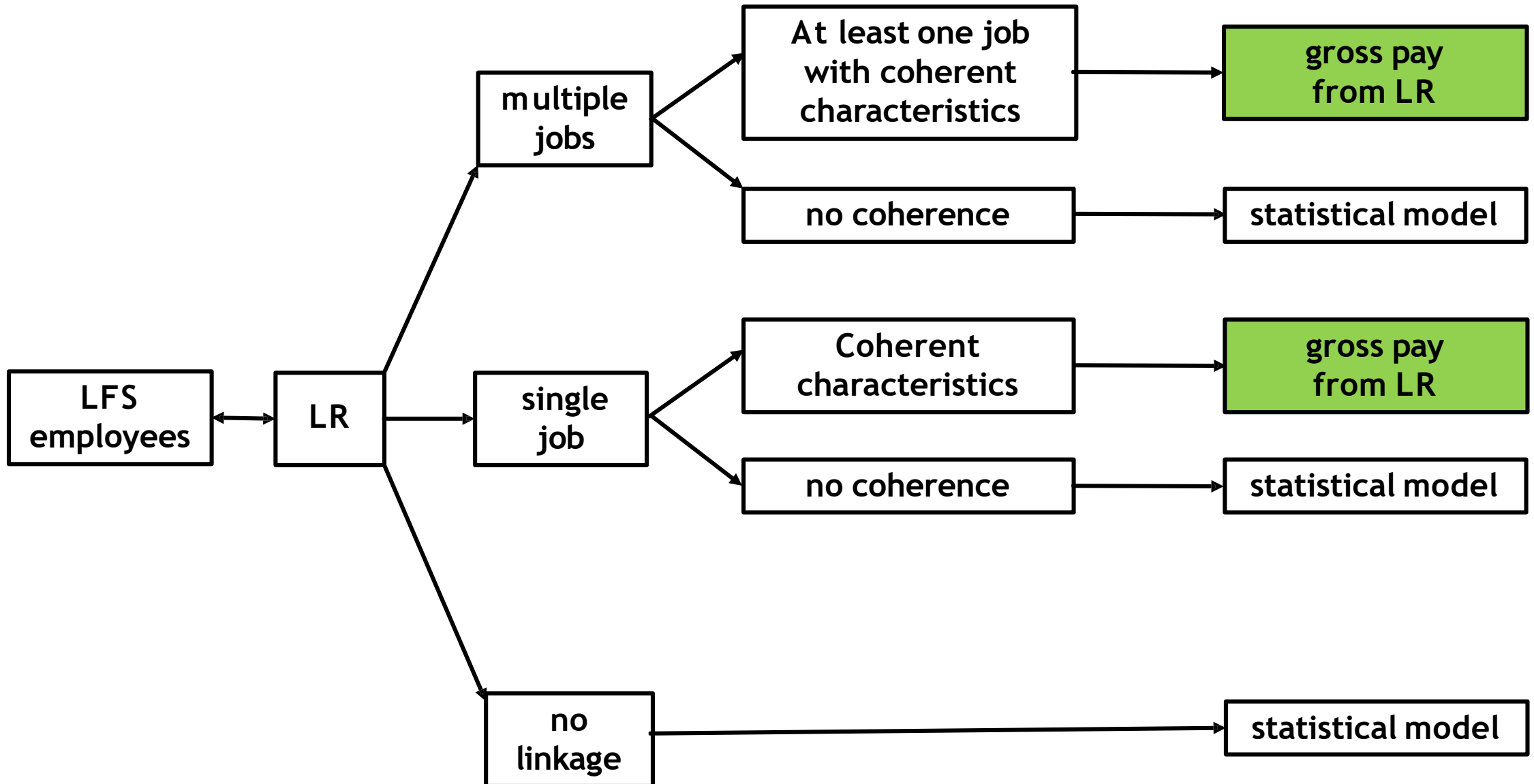
Linkage and validation with RL

The **order** of variables reliability:

- **full time/part time** is a reliable variable, and it is fundamental to guarantee consistency in terms of worked hours, and net or gross work income;
- **temporary/permanent**, it may happen that the respondent's perception differs from what is stated in the employment contract (for instance in sectors in which it is usual renewing several temporary contracts one after the other, or in sectors where the permanent contract does not guarantee job stability as in construction);
- **economic activity**, the respondent may have difficulty classifying the activity being performed according to the company's NACE classification

In **2019 the record linkage** between LFS (around 154 thousand employees) and LR is around 94% of the LFS employees, having at least one LR job position (81% have only one job position and the remaining 13% have two or more job positions). Of these, 89% are validated (fully or partially).

The methodological approach



Statistical model

To develop the **statistical models** to predict gross wage in cases in which it is not possible to impute administrative data from LR, it is preventively necessary identifying **four subsamples**:

- 1) LFS employees linked with LR and validated (consistent working time) → information on gross wage is available and usable.
- 2) LFS employees not linked with LR or linked but not validated; irregular job is excluded, i.e. public sector.
In this group also are included:
 - ✓ LFS employees linked with LR and validated; information on gross wage is missing or outlier
 - ✓ LFS part-time workers linked with LR but partially validated because full-time in LR; all activity sectors
 - ✓ LFS full-time workers linked with LR but partially validated because part-time in LR; irregular job is excluded, i.e. public sector
- 3) LFS full-time workers linked with LR but partially validated because part-time in LR; “grey” has to be verified (partially regular job, in which part of the wage is undeclared in administrative sources)
- 4) LFS employees not linked with LR or linked but not validated; the presence of irregular job is presumed, and a comparison has also been made with the official estimates of National Accounting

Statistical model

The adopted **strategy is different in each of the four groups:**

From Register (84%)

- 1) Administrative information on gross wage from LR is imputed to LFS employees (validated information).

From Model (16%)

- 2) Administrative information on gross wage from LR is unavailable or unusable; a statistical model (GROSS_MODEL) is estimated on the subsample in group 1 (linear regression, where $Y = \text{LR gross wage}$, $X = \text{a large set of LFS variables}$); this model is used to predict gross wage in group 2.
- 3) To estimate if there is part of the wage undeclared in administrative sources, the net wage from LFS referred to a LFS full-time job is compared with NET-GROSS function.
- 4) In this group irregular job is assumed (no social security contributions and no taxes are paid), so the amount of income assigned to these LFS employees reflects the net wage.

Statistical model

The statistical model implements a **linear regression**. The strength of this model derives from the very **large set of covariates**, highly correlated with the dependent variable (gross monthly pay);

✓ **Socio-demographic variables:**

Gender, age, age-squared, NUTS1, big municipality (more than 250,000 resident persons), degree of urbanization, citizenship, highest level of education, number of family members, household typology, family role (derived from the relationship with the reference person).

✓ **Job characteristics:**

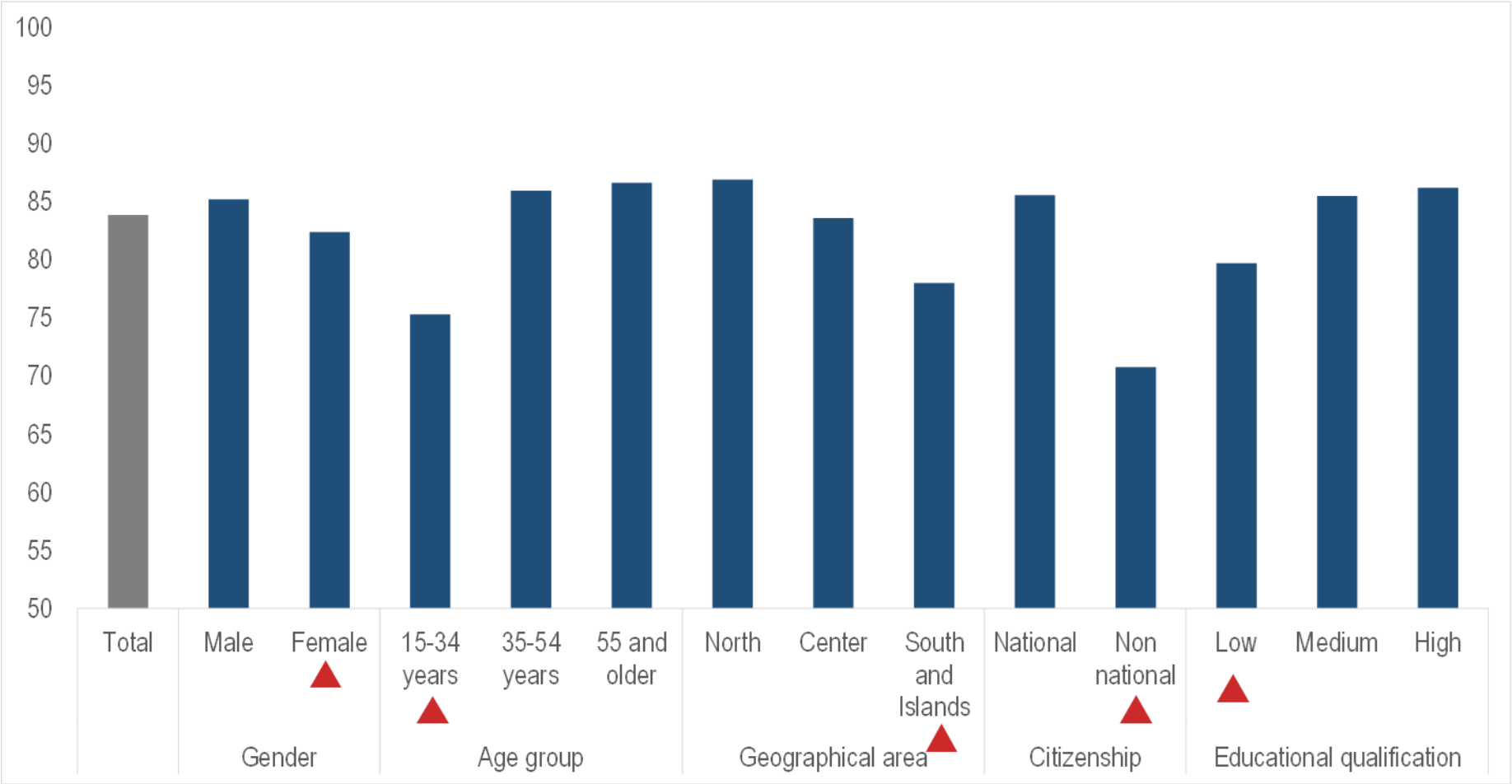
Professional role (Italian variable reflecting contractual status), supervisor responsibility, occupation (ISCO1 digit), temporary/permanent job, number of employees in the local unit, economic activity sector, full-time/part-time job, usual worked hours, actual worked hours, absence or reduced hours, overtime, years of work in this job, years of work in this job-squared, net monthly pay (only in GROSS_MODEL).

The independent variables selection has been conducted through a **stepwise procedure**, most of the variables included in the model are significant. The fitness of the model is good: **$R^2 > 70-80\%$** in the model excluding monthly net income, it becomes even higher including monthly net income.

Statistical model

MODEL	LOG(GROSS_INCOME)=BETA*X									
R ²	0,79									
Variable: X	beta	Pr > t		Variable: X	beta	Pr > t		Variable: X	beta	Pr > t
Intercept	5,5823	<.0001		xpospro1	0,0387	0,0002	CONTRACT_LEV=1	xc185	0,0601	<.0001
xrip1	0,0701	<.0001	NUTS1=1	xpospro2	0,0894	<.0001	CONTRACT_LEV=2	xc121	0,0530	<.0001
xrip2	0,0772	<.0001	NUTS1=2	xpospro3	0,0766	<.0001	CONTRACT_LEV=3	xc122	0,3057	<.0001
xrip3	0,0393	<.0001	NUTS1=3	xc10	0,0093	0,0002	SUPVISOR=1	xc123	0,2986	<.0001
xdegurba1	0,0091	<.0001	DEGURBA=1	xprof11	-0,0984	<.0001	ISCO=1	xc124	0,3192	<.0001
xcesso	0,0421	<.0001	SEX=M	xprof12	-0,0217	0,0223	ISCO=2	xc125	0,2305	<.0001
xeta	0,0074	<.0001	AGE	xprof13	-0,0925	<.0001	ISCO=3	xc126	0,2832	<.0001
xeta2	-0,0001	<.0001	AGE^2	xprof14	-0,1152	<.0001	ISCO=4	xc127	0,2512	<.0001
xcittad	0,1057	<.0001	CITIZENSHIP=IT	xprof15	-0,1343	<.0001	ISCO=5	xc128	0,3664	<.0001
xsg4	-0,0071	<.0001	HOUSEHOLD_MEMBERS	xprof16	-0,1599	<.0001	ISCO=6	xc129	0,2376	<.0001
xtn21	0,0078	0,0217	COUPLE_WITH_CHILDREN	xprof17	-0,1406	<.0001	ISCO=7	xc1210	0,2977	<.0001
xtn22	0,0093	0,0020	COUPLE_WITHOUT_CHILDREN	xprof18	-0,2217	<.0001	ISCO=8	xc1211	0,2572	<.0001
xtitstu1	0,0601	<.0001	HATLEVEL=5,6,7,8	xdetind	0,0670	<.0001	TEMP=1	xpiepar	0,2309	<.0001
xtitstu2	0,0373	<.0001	HATLEVEL=3,4	xc181	0,1281	<.0001	SIZEFIRM>=250	xc31	0,0089	<.0001
xtitstu3	0,0269	<.0001	HATLEVEL=2	xc182	0,1226	<.0001	50<=SIZEFIRM<=249	xc37	0,0016	<.0001
xduratt	0,0008	<.0001	MONTHS_STARTWK	xc183	0,0999	<.0001	20<=SIZEFIRM<=49	xorariopiu	0,0309	<.0001
xduratt2	0,0000	<.0001	MONTHS_STARTWK^2	xc184	0,0812	<.0001	16<=SIZEFIRM<=19	xfol_netta	0,0005	<.0001
										11<=SIZEFIRM<=15
										NACE_SECTOR=1
										NACE_SECTOR=2
										NACE_SECTOR=3
										NACE_SECTOR=4
										NACE_SECTOR=5
										NACE_SECTOR=6
										NACE_SECTOR=7
										NACE_SECTOR=8
										NACE_SECTOR=9
										NACE_SECTOR=10
										NACE_SECTOR=11
										FTPT=1
										HWUSUAL
										HWACTUAL
										EXTRAHRS
										NET_INCOME

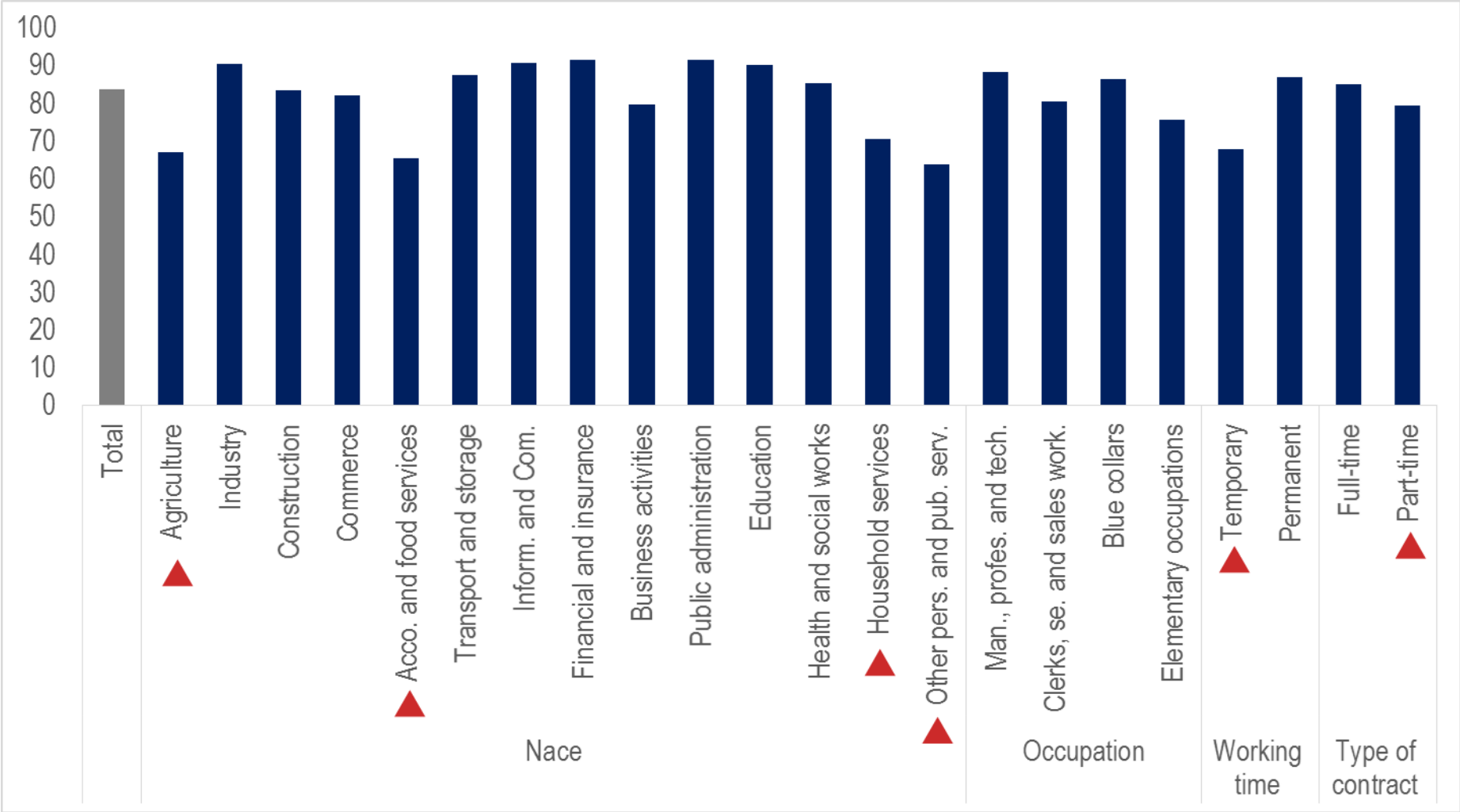
Percentage of validated (gross wage from LR)



The **percentage with gross wages imputed from LR is lower** for aged 15-34 years, women, residents in the southern regions of Italy, foreigners, and individuals with a low level of education.

In general, these groups are more likely to experience situations of **work vulnerability**

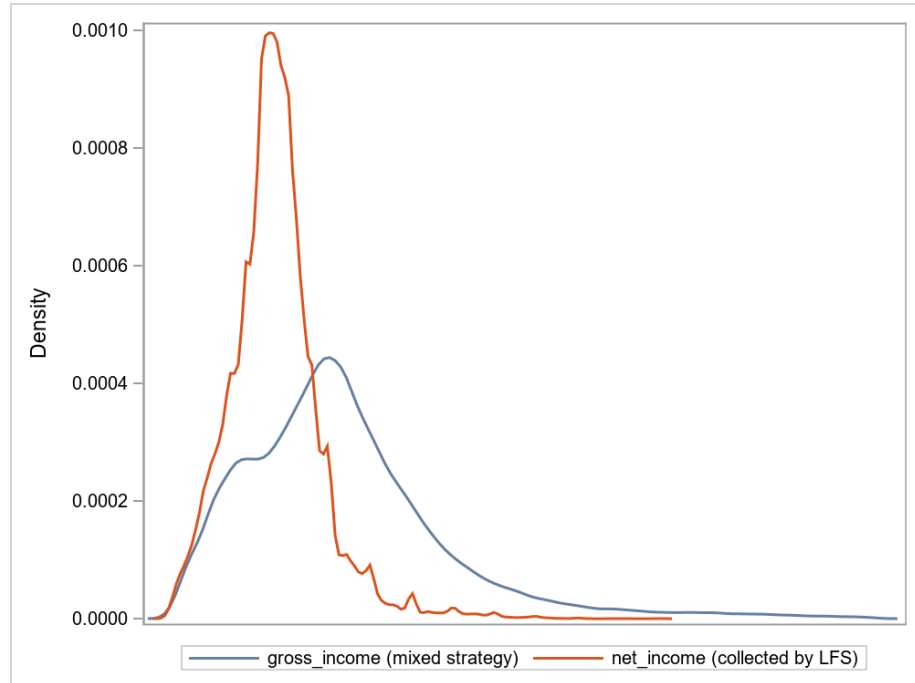
Percentage of validated (gross wage from LR)



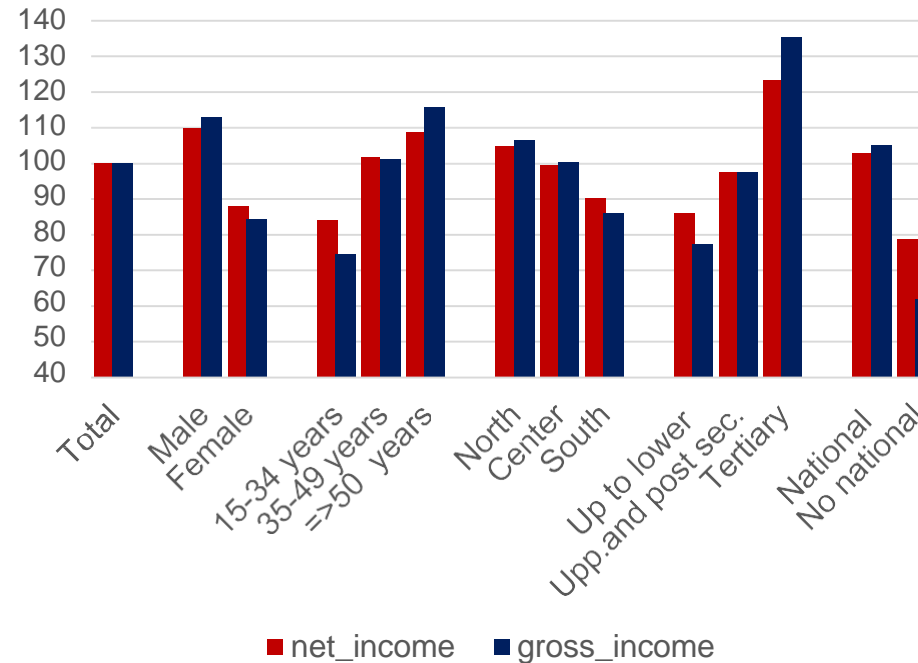
Considering **job characteristics**, in fact, the percentage with wages imputed from the registry is lower for temporary employees, part-time workers, in the sectors of Agriculture, Accommodation and food services, Household services and Other personal and public services, as well as among elementary occupations.

IT-LFS comparison of net and gross wage

Gross and net remuneration density distribution for employees. Year 2019



IT-LFS comparison of net and gross wage (index numbers). Year 2019



When analyzing **disparities among socio-demographic characteristics**, the most substantial variations between gross and net wages are observed in terms of citizenship and educational qualifications

As expected, **gross wage has a higher level and greater variability** compared to net wage due to the redistributive impact of the tax system. Additionally, the greater precision in estimating gross wages, unaffected by individuals rounding their incomes, also contributes to the observed differences.

The **coefficient of variation** increases from 15 for net wage to 27 for gross wage, and the ratio of the 90th to the 10th percentile increases from 2.9 to 4.6.

Thank you

Federica Pintaldi | pintaldi@istat.it