Part 3

Integrated Longitudinal Employer-Employee Data: a New Informative System on Employment in Italy

Caterina Viviano (ISTAT)

Workshop on Developing Entrepreneurship Statistics by Gender in the Republic of Moldova

Chisinau, Republic of Moldova, 30 June-1 July 2015
1. The new Informative structure on Employment – the Linked Employer-Employees Database (LEED)
2. The self-employment definition
3. The LEED data structure
4. The Administrative Data Sources
5. The Self-employment component: identification and measures
6. LEED: Contents and measures
The new Italian Business Register to measure employment (ASIA-Employment) has been realized in occasion of the first Economic “virtual” Census (CIS2011)

**Purpose**

reproducing census data entirely from administrative sources and existing statistical business register (data on the structure of enterprises, their local units and related employment)

The informative system on employment is realized according to a LEED structure.
Linked employer-employee data (LEED) is a database that assembles information from all workers and firms. LEED bring together information from both sides of the labor market: enterprise (or workplace) and individual (workers).

Feasible because:
- the existence of a consolidated BR (ASIA)
- degree of enhancement done in the statistical use of administrative sources and, on the other hand, on the quality improvements of each administrative body
- new set of available data for the system of business: Information available not only at firm level but also at individual one in that each person involved in the business production process is identified
The structure of longitudinal LEED contains observations about individuals and their employers linked by means of a work history.

The individuals (and their characteristics) are linked with firms by means of job position i.e. for each employer who employed individual i during period t

Those links are assured by the existence of unique identifiers - fiscal code- both for the business and individual side

The data are longitudinal because longitudinal administrative work history records exist for each individual and because longitudinal statistical data (SBR) exist for the employer.
Linked Employer-Employees Data

LEED has been realized with the purpose to revise, produce and enlarge the informative statistical database on enterprises and local units produced and disseminated every year from the BR system - (ASIA)

Focus: employment data – definition and measure

A) Revision of employment definition and classification
   - to guarantee better coherence for the whole national statistical system and with the international standard
   - to adopt a common conceptual framework

B) A revision of methodologies and operational rules for:
   o the identification and measure of the self-employment component
   o a more accurate measure of employees
   o identifying more typologies of workers used in the firm like outworkers, temporary workers;
   o producing detailed information on worker demographic characteristics and job variables (type of contract, job qualification)
conceptual reference frameworks:
- the “System of National Account – SNA 2008”
- the “International Classification of Status in Employment – ISCE 93”
- the “European System of Accounts - forthcoming ESA2010

Different **typologies** of workers employed by firms:
- a. Employees
- b. Self-employed/Contributing family workers
- c. Outworkers - external component
- d. Temporary workers - external component
**Self-employment**: Person who carries out its work inside an enterprise, without any formal contracts (not in the payroll), whose remuneration is included in mixed income (capital/labor). Self-employed persons include the following categories:

- Sole proprietor, free lance and own account workers;
- Contributing family workers working in the enterprise even if not paid, with paid contributions for social security or insurance against injuries at work;
- Members of partnerships, corporations and quasi-corporation under the condition of an actual working activity carried out in the same enterprise of which they are shareholders.
In the adopted Italian interpretation of the self-employed definition, the main difference with the standards lies in the inclusion of shareholders of a corporation as possible self-employed in addition to the sole owner and to owners of the unincorporated enterprises.

**Old approach** is based on the concept of “the self-employed exists in the business because there should always be a function of coordination and management.”

**New approach:**

a) definition is linked to the typology of remuneration for the type of process (mixed income)

b) identification of the self-employed person is different according to the legal form of enterprise:
   - sole proprietorship – the owner
   - partnership - similar in every way to a sole proprietorship (joint owners)
   - corporations: different, legal entity, individual is not personally responsible

In the adopted Italian interpretation of the self-employed definition, the main difference with the standards lies in the inclusion of shareholders of a corporation as possible self-employed in addition to the sole owner and to owners of the unincorporated enterprises.
The problem of identification of self-employed persons (SE) is a relevant issue. (OECD, 2012) introduced conceptual and practical rules to distinguish the entrepreneur (i.e. for owned enterprises statistics by gender).

From the demand-side: difficulty in identifying SE not for sole-proprietorships but for owners of the shares of the equity in partnerships and limited liability companies.

Definition: **Entrepreneurs** of are those persons (business owners) who seek to generate value, through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets (OECD, A Framework for Addressing and Measuring Entrepreneurship).
Administrative sources and the system of registers

- VAT register
- Business Income participation in partnerships
- Studi di settore
- Tax register (Remun.-770-form)
- Employer decl. on each employee
- Employer decl. on outworkers
- Self-employed persons in agric. handicraft and trade data
- Chamber of Commerce
  - enterprise ownership shares
  - partners and positions
  - shareholders
- Insurance against work related injuries
- Insurance positions workplace
- Links identification codes (fiscal code, br code)
- Fiscal data
- Social security data
- Chamber of Commerce
- LEED
- Frame for surveys
- BR-ASIA
- CIS
- Longitudinal analyses

Links identification codes (fiscal code, br code)
The statistical process

The Administrative database received are of an overall high quality for the administrative purpose. Linking, transformation processes and cleaning (editing and imputation processes) are then required to produce robust official statistics.

In order to integrate all the needed information the following link have to be carried on:

a. Links between job positions of different administrative sources;
b. Links between job positions and enterprises;
c. Links between job positions and individuals;
d. Links between job positions and enterprises at different time;
Let be $U$, $Z$ and $W$ data matrixes where:

- $U$: contains data on the $N$ individual characteristics. The row elements is denote by $[u_i]_{i \in \{1,2,\ldots,N\}}$. Here, we assume that are time-invariant data;

- $Z$: contains data on the $M$ firms characteristics. The row elements is denoted by $[z_{jt}]_{j \in \{1,2,\ldots,J\}, t \in \{t_{j1},\ldots,t_{jT_j}\}}$, where $T_j$ is the total number of periods of data available for firm $j$. The data are time dependent;

- $W$: contains data on job position i.e. for each employer who employed individual $i$ during period $t$. The row elements is denote by $[w_{it}]_{i \in \{1,2,\ldots,N\}, t \in \{t_{i1},\ldots,t_{iT_i}\}}$, where $T_i$ is the total number of periods of data available for individual $i$. 

The data $U$ and $W$ are linked by a person identifier (fiscal code).

The data $Z$ and $W$ are linked by a firm identifier (firm fiscal code). The link between the data $Z$ and $W$ can be conceptualized by the link function $j = J(i; t)$ which indicates the firms where worker $i$ was employed at date $t$. 

![Diagram of data structure](image)
The informative data structure for self employment:

(1) Structure of Links (enterprise-individuals) BASE: links between the administrative legal unit and the individuals assured by a system of identification codes (fiscal codes)

(2) information used as proxy to measure any flow of remuneration from the company going to its members or any evidence of working activity

Administrative Sources (1) – CCIAA and Tax Agency;
Administrative Sources (2) – Social Security, CCIAA, INAIL, Fiscal data (Vat code)
Register of Persons with positions in the enterprise (owned by the Chamber of Commerce)

- Yearly firm declarations related to persons covering a position inside partnerships or companies (i.e. member of the board of directors, administrator, simple partner). Covered position reflects the underlying legal regulation followed by the business.

- Input records amount at about 6.5 millions of records (links).

- The main information are:
  - Unique identity number: firm/ person fiscal codes;
  - Variables: validity dates of positions with respect to the reference year, type of position(s) covered during the year, share holding.
Administrative Data Sources – (1)

Shareholders’ Register *(owned by the Chamber of Commerce)*

- Yearly firm declarations related to shareholders belonging to companies and cooperatives that contribute to the enterprise’s capital because they hold a percentage of its shares.
- Input records amount at about 3,9 millions of links.

Fiscal declaration’ Register *(owned by the Tax Agency- quadro RH, modello Unico-PF)*

- Yearly declarations made by taxpayers on income deriving from the participation in a partnerships (i.e. association of professionals).
- Input records amount at about 2,6 millions of links. The source has been treated in order to derive the structure at firm level and attribute to each partner income (profit share) in proportion to the share held. Data have one year time lag.
The craftsmen and tradesmen archive (owned by Italian National Institute for Social Security -INPS)

- Yearly social contributions paid by the holders of craft or trade businesses. These people use mainly their own work or family workers for carrying out the business activity. The level of contribution they have to pay is calculated on the base of their overall produced income.

- Input records amount at about 4,2 millions of declarations. Information is available at the owner fiscal code level and, where present, for its relatives (family workers).
Declarations of social contribution for Outworkers (Parasubordinati) (owned by Italian National Institute for Social Security - INPS)

- Collaborators register:
  ✓ The main information are the gross wages paid by employers, type of workers (i.e. administrators of companies, associated with participation, members of cooperatives, etc.)

- Professionals register:
  ✓ Declarations made by free lance that do not have a proper pension fund or professional register.

Declarations of Insurance for Outworkers (owned by agency for the insurance against work-related injuries -INAIL)

✓ Yearly firm declarations, related to Outworkers carrying out a project, for insurance reasons due to the risk of work-injuries.
Criteria for the identification of self-employment:

Step 1 - deterministic classification rules

**Self-employed estimation**

*Based on particular combination of the vector’s component*

- a. he/she must be a member of the company (shareholder)
- b. with an office (or position)
- c. and working in it (with a mixed-income),

*a. and b. are easier to measure*

Problems may arise to measure the mixed income ....

Information from administrative sources are used as proxy for the mixed-income:

**Any flow of remuneration from the company going to its members**

- Remuneration from declarations at social security (“quasi-subordinate”)
- Any payment of contribution made by the person as Craftsmen and Trader, as professionals at social security or for any insurance at work
The informative data structure for identifying self employment

- Administrator (Y/N)
- Shareholder (Y/N)
- age
- sex
- place of birth
- education

- signals of remuneration by source
- estimation
- metadata

Self employed A
Shareholder B
Admin. only C
Self employed

LEED
Operational rules identifying a person as self-employed (SE) are specific according to the enterprise legal form:

**Sole proprietorships:**
- SE correspond to the owner and – if present - contributing family workers
- Represent a relevant component of the business structure

*Result: self-employment always present when the enterprise (i.e. the sole-proprietorship) is active (trivial)*
Partnerships:

**Association of professionals**: SE correspond to each linked person recognized as member (having or not a VAT code)

**Other forms (Partnerships)**

1. SE corresponds to a member that fulfils the basic requirements (a,b,c)

2. to overcome problems of missing data
   - Record linkage techniques (denomination and names)
   - Imputation by selection of the “best representative” having defined a rank function

3. Contributing family workers can be present

Result: self-employment always present with the exception if “all members are declared as employees”
The identification process – deterministic

Corporations (and cooperatives):

1. SE is identified differently for enterprise (E) and enterprise groups (EG)

2. SE corresponds to a member that fulfils the basic requirements (a,b,c) (relevance: to be a shareholder)

3. (EG): exclusions from the identification
   • participated only by juridical persons
   • branches
   • public control
   • complex groups
   • head group being a individual, partnership

4. (E): to overcome problems of missing data in the Shareholders’ Register “stochastic method” is developed (only for corporations)
The identification process

Step 2 - stochastic approach

Purpose: to implement stochastic models for estimation of zero self-employment in the sub-population of units with under-coverage in the administrative source on shareholders.

The model: to estimate the number of shareholders inside the sub-pop with zero self-employment (from deterministic initial rules) with undercoverage in Chamber of commerce source

Sources of data for the benchmark
- the fiscal data “Studi di settore (SDS)” (frame A on employment)
- the SBS SME (small and medium enterprises) survey

the chosen model is the zero inflated Poisson (ZIP)
Different typologies of workers (Job positions) employed by firms:

a. Employees
b. Self-employed/Contributing family workers
c. Outworkers- external component
d. Temporary workers - external component

- jobs characteristics by typology of workers
  - Job qualification
  - Contractual working time (part time/full time)
  - Type of contract (temporary or fixed time /permanent)
  - Maternity leave
  - Illness
• Individual (worker) demographic characteristics
  ✓ **Sex**
  ✓ **Age**
  ✓ **Place of Birth**
  ✓ **Residence**

• Enterprise characteristics
  ✓ **Industry sector (NACE 4 digits)**
  ✓ **Size (number of employee)**
  ✓ **Size (turnover)**
  ✓ **Geographical location (local unit)**
1. LEED data assesses the presence of the worker–employer link (job) at a point in time;

2. The longitudinal nature of the linked data on both the individual and business sides allow to track the flows of workers across detailed industries, across detailed demographic groups, and even for specific regional economic.

1. Measure on counts of jobs at a point-in-time (Census day), or means of jobs existing for an interval of time (year), by dimensions both from the individual and business sides;

2. Measure on the transitions of workers with respect to different individual characteristics or business characteristics. The origin–destination information permits measurement of status changes or earnings changes of the workers.
Measures and analyses

- Cross-sectional: Structure and composition of employment (in the business sector) - level of employment

- Longitudinal:
  - Tabulate job and workers flows by a variety of workers and firms characteristics
  - Origin and destination of transition for workers with different characteristics (age, gender, etc.) and relative earnings changes
  - To trace the job history positions of employed over time
  - Re-allocation of workers across industries (flows between origin-destination matrix by sectors of activity)
Measures and analyses

The profile of the entrepreneur

- Own-account workers ≅ self-employed without employees
- Employers ≅ self-employed with employees

By gender, place of birth, age

Indicators:
Share of self-employed over the total n. of person employed

BY GENDER
BY legal status
Measures and analyses

BY GENDER

Share of own-account workers with women self-employed (proxy of women-owned) over own-account workers

Share employers with women self-employed (proxy of women-owned) over total employers

Weighed by number of employees, turnover
By sectors of activity of the enterprise where they work
By size of enterprise
By age of enterprise
Measures and analyses

Matrix of information:

<table>
<thead>
<tr>
<th>Business Register</th>
<th>Business Demography</th>
<th>LEED - Employment</th>
<th>LEED- Self-employment Base structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sole proprietorships</td>
<td>Employer Enterprise Birth (EEB)</td>
<td>with Employees</td>
<td>whose self-employment is Male</td>
</tr>
<tr>
<td>Partnerships</td>
<td></td>
<td></td>
<td>alone</td>
</tr>
<tr>
<td>Corporations</td>
<td>Other Active Enterprises</td>
<td>with NO Employees</td>
<td>Female</td>
</tr>
<tr>
<td>Cooperatives</td>
<td></td>
<td></td>
<td>with contributing family workers</td>
</tr>
</tbody>
</table>

....working in progress

derived indicators on firm’s ownership structure:

✓ N. of owners, of which females
✓ N. of administrators, of which females
✓ Sum of owned shares (as %), of which females

The gender-run business is assigned by ranking individuals inside the firm according to each above simple indicators
Thanks for your attention!