LABORatorio R. Revelli
Centre for Employment Studies

WHIP

Work Histories Italian Panel

WHIP

Work Histories Italian Panel

Summary

WHIP is a database of individual work histories, based on Inps (National Social Security) administrative.

The reference population is made up by all the people – Italian and foreign – who have worked in Italy even only for a part of their working career. A large representative sample has been extracted from this population: in the standard file the sampling coefficient is about 1:90, for a dynamic population of about 740,000 people.

For each of these people the main episodes of their working careers are observed. The complete list of observations includes: private employee working contracts, atypical contracts, self-employment activities as artisan, trader and some activities as freelancer, retirement spells, as well as non-working spells in which the individual received social benefits, like unemployment subsidies or mobility benefits. The workers for whom activity is not observed in WHIP are those who worked in the public sector or as freelancers (lawyers or notaries) – who have an autonomous security fund. The WHIP section concerning employee contracts is a Linked Employer Employee Database: in addition to the data about the contract, thanks to a linkage with the Inps Firm Observatory, data concerning the firm in which the worker is employed is also available. The observed period goes from 1985 to 1999, but within a few months data up to 2002 will be released.

WHIP allows exhaustive empirical analysis about labour economics and industrial economics, and in particular about themes concerning employment and labour mobility, wages, dynamics and firm demography – themes which can usefully interact.

Given the longitudinal nature of the sample, and the quantity of different observable events in the working careers of individuals, in general, exhaustive analysis of the duration and transition among events are possible.

What is WHIP

WHIP is a database of individual work histories, based on Inps (National Social Security) administrative.

The reference population is made up by all the people – Italian and foreign – who have worked in Italy even only for only a part of their working career. A large representative sample has been extracted from this population: in the standard file the sampling coefficient is about 1:90, for a dynamic population of about 740,000 people.

For each of these people the main episodes of their working careers are observed. The complete list of observations includes: private employee working contracts, atypical contracts, self-employment activities as artisan, trader and some activities as freelancer, retirement spells, as well as non-working spells in which the individual received social benefits, like unemployment subsidies or mobility benefits. The workers for whom activity is not observed in WHIP are those who worked in the public sector or as freelancers (lawyers or notaries) – who have an autonomous security fund. The WHIP section concerning employee contracts is a Linked Employer Employee Database: in addition to the data about the contract, thanks to a linkage with the Inps Firm Observatory, data concerning the firm in which the worker is employed is also available. The observed period goes from 1985 to 1999, but within a few months data up to 2002 will be released.

What types of analysis does WHIP allow

WHIP allows exhaustive empirical analysis about labour economics and industrial economics, and in particular about themes concerning employment and labour mobility, wages, dynamics and firm demography – themes which can usefully interact. Here is a brief list of basic themes for which WHIP could be a source of information:

- studies concerning flexibility of the labour market and worker mobility;
- wage distribution, inequalities and wage mobility;
- evaluation analysis of active labour policies like apprenticeships or training and work contracts (CFL);
- studies about the diffusion and impact of new kinds of contracts like temporary work or "lavoro interinale";
- labour and retirement choices of the elderly;
- creation and destruction of job positions and firm birth and death;

Given the longitudinal nature of the sample, and the quantity of different observable events in the working careers of individuals, in general, exhaustive analysis of the duration and transition among events are possible.

What is the source data

WHIP comes from administrative source data. In fact the data on which it is founded is a sample collection extracted from the National Institute of Social Security's (Inps) administrative archives, and have been provided to LABORatorio Revelli according to an agreement signed between Inps and the

University of Torino. The administrative archives from which data have been extracted are:

- unified individual register
- employees' contributions
- artisans' contributions
- traders' contributions
- contributions of the IV gestione separata (atypical workers)
- retirement benefits
- other social security benefits (unemployment subsidies etc.)
- business register
- monthly contributions of firms

The global supply is rather huge – the quantity is around 6 million records provided, for 1.400 million observations (record per variable) – and is growing with new extractions: in addition to annual updating, in the new agreement which is being signed, extractions from the summary of accured contributions and Isae (Indicator of the Equivalent Economical Situation) archives will be provided. Inps provides statistical observatories on most part of these archives, the results of which are published online (see on the Institute website www.inps.it), where it is possible to consult the methodological notes, useful as documentation for WHIP input archives

How WHIP was constructed

The data which LABORatorio Revelli receives from Inps are extracted from archives which respond to the managing purposes of the National Institute for Social Security. WHIP building procedure may be seen as a black box which processes the archives provided by Inps and gives as result archives that allow for statistical-economical analysis. If we open the black box we find a block of programs (about a hundred) which carry out four kinds of fundamental operations (and other optional operations) in a completely automatic way:

- the normalization, cleaning and re-codification of most variables contained in the source archives;
- firm longitudinal identification;
- working contracts longitudinal identification;
- strengthening of the contribution record in the annual data of working contracts.

From the point of view of ease of use, the main passages are the first and the fourth, since they allow the final user to avoid having to untangle the codifications used in the contribution declarations which Inps receives – that are often particularly far from interesting subjects for economists – and which allows one to work on the archives organized per year in a coherent and unambiguous way. From a more substantial point of view, the information produced, the more relevant passages are the second and the third ones, because they produce one of the more important added values of WHIP.

Firm longitudinal identification and firm transformation identification is fundamental to have reliable firm data (and consequently to identify labour contracts), and is necessary since WHIP has a longitudinal dimension which is superior to the number of years (five) for which Inps builds the record of the Firm Observatory, from which the sample object of supply is extracted. The problem is that between one edition and another the firm identification codes and the composition of the firm itself can change, due to a split up, merger or other corporate transformation, therefore it is necessity to identify longitudinally the firms among the different releases of the Observatory.

Longitudinal identification of labour contracts, for what concerns labour market studies, is probably the greatest added value produced by WHIP. The data, records are organized by contribution declarations produced for a worker, since, simplifying a little, the amount of the contribution represents the most interesting variable for the Institute, which uses it for the computation of retirement benefits and other benefits which are in favour of the insured person. However, a one to one relationship does not exist among the declared contributions events and the labour contracts established by the person (the relationship is many to many). Moreover, the date of beginning of the contract is not explicitly registered, and the date of cessation – even if present – lacks quality, both in terms of missing values and correctness. The steps of the procedure devoted to the longitudinal identification of labour contracts adjust for these problems and produce exactly this: the dates of the beginning and the end of labour contracts established by individuals.

Future developments

Presently, we are planning different important developments of the database. Some of them are linked to the Inps data supply and consist of:

- the updating of the database to include 2002 data;
- the extraction from the Isee (Indicator of the Equivalent Economical Situation) archive, from where it will be possible to have information about the composition of the household and about the economic situation of the itself;
- the creation of the complete contributions statement for all the individuals of the sample, through which it will be possible to have information about pre-1985 working histories and about contribution seniority

Other developments are designed to enrich the archives already received. One of the first projects, close to completion, consists of the cleansing and re-coding of the variables of contract code and contractual level of the employee, and linking the information available in the corresponding collective contracts (contractual minimum wage, advancement of contingency etc.). There are also two projects which are now in a preliminary stage; they will be launched in the next months and we are open to external collaboration proposals about them. One of them concerns further development of the firm transformations identification procedure, whose targets are at present medium-large firms. The target of the second one is the imputation of an eligible flag, which through the interaction of the seniority contribution observable in WHIP,

and the retirement regulations, state whether or not the worker was entitled to access to retirement benefits.

Diffusion policies

Two policies have been implemented to make WHIP available, according to two different versions of the same database.

- File Standard edition: the main data diffusion policy was realized though the construction of a standard file of WHIP, composed by anonymous micro data according to art. 1 of law 675/1996 on the safeguarding of privacy. After the processing of Inps data, the data contained in the WHIP standard file can't be associated with identified or identifiable subjects, and are therefore classified as anonymous data. This allows their diffusion and guarantees their full availability for researchers. The files are available from the page download.

- File Full edition: Universities, Public and Private Research Institutes, Scientific Societies and individual researchers in the field can ask for access to the complete version of WHIP, based on the presentation of a research plan, in compliance with the ethical norms. For both file standard and file full edition see: http://www.laboratoriorevelli.it/whip/

What documentation is available on line

Site: http://www.laboratoriorevelli.it/whip/

WHIP documentation is organized into three sections.

The metadata, namely the 'data about data', contains a description of all archives and variables which compose WHIP. The description includes the official name and label assigned to the variables; when the data are not self-explanatory or exhaustive, a more thorough description is also included, as well as the codes used. Moreover, there is a brief indication of the quality of the data, and when there are missing values or corrections there is a table to quantify the missing values and/or the corrections required.

A reports section, where technical reports concerning WHIP will be gathered, is forthcoming. It is an open section, which will accept contributions from users who have results related to data issues, or comparisons between WHIP and other databases.

Information concerning WHIP and its use, which cannot be filed in the meta data or technical reports, will be collected in a section of Frequently Asked Questions.