

1 Integrated database for labour market research

One important research register constructed by *Statistics Denmark* is *IDA*, *Integrated Danish Labour Market data base*.

The units in *IDA* are *establishments*, *jobs*, *persons*, and *firms*. There exists a data set for each unit *IDAS*, *IDA establishments*, *IDAN*, *IDA employments/jobs*, *IDAP*, *IDA, Persons*, *labour market attachment and experience*, and *IDFI*, *IDA Income tax reporting units*. *IDA* has data since 1980. There is a major break in 2008.

IDA has constructed an establishment identifier, *LBNR*, which is unique over time based on a specific set of conditions which must be fulfilled for an establishment to be the same economic establishment over time, see below.

IDA provides longitudinal data on *workers* identified by *PNR* and *establishments* identified by *LBNR*. Employees can be merged using *PNR* and with establishments using *LBNR* in *IDAS*.

Persons The population of *persons* are all persons living in Denmark at the end of the year.

An employee is identified by *PNR*, which is unique over time.

The population of the *CPR*, *Central Population Register* is all persons living in Denmark at the beginning of the year. *ECONAU* has constructed the data sets *PERSONER*, *FAMILIE* and *HUSSTANDE* which are lagged one year in order to correspond to the *IDA* population. They are using *C family* for the period 1980-2006 and *E family* from 2007 and onwards.

Establishments The population of *establishments* are all establishments which during the year have employees paying income tax. *IDA* also includes *establishments* which only exist part of the year and hence may have no employees ultimo November. An *establishment* is identified by the key (*ARBGNR*, *ARB_STK*) for the period 1980-2007, (*CVRNR*, *ARB_NR*) for the period 2008- , or the constructed variable (*LBNR*).

An *establishment* is a well defined physical entity located at a particular address, and producing one or a few goods.

Jobs A *job* is identified by the person, *PNR*, and an establishment, *LBNR*, see below. A firm must report income and withhold income tax for all workers on each establishment. A person can have several jobs at the same time over the year. The job does not have information on start date and end date, but the number of days employed, the income, and an imputed hourly wage. The jobs are grouped by *JOB.TYPE*. An important job type is the *primary job* end of November.

It is possible to find job spells in the *SPELL* data set.

IDA Firms An *IDA FIRM* is identified by the unit paying income taxes. In the period 1980-2007 the *IDA FIRM* is identified by *ARBGNR*, which may be a *JURNR*, *SENR* or a *CVRNR*. From 2008 it is identified by *CVRNR*. But the firm identifier used in *IDA* may not correspond to the economic *firm identifier*. From 1995 and onwards the economic *firm identifier* can be determined by matching *FIDA* using *PNR* and *JOB.TYPE* from *IDAN*.

IDA was developed in 1988-1990 with support from "Statens Samfundsviden-skabelige Forskningråd" by Ruth Emerek, Aalborg University, Per Vejrup-Hansen, Copenhagen Business School, and Søren Leth-Sørensen, Statistics Denmark

There has been a change in data sources.

From 2008 the sources are *EBST, Occupational employment, RAS, Labour force, compiled from registers., OF, Social benefits to supplement labour income, IND, Income*. The new *AMRUN, Labour Market Accounts of persons' attachment to the labour market* is used to form *EBST* and *RAS*.

The new sources imply a break in *employment, hours, hourly wage, and unemployment* in 2008.

In *IDA* an establishment is identified by *ARBGNR* together with *ARB_STK* for the years 1980-2007, and *CVRNR* together with *ARB_NR* after 2008. For 1980-2007 it is possible to use *CVRNR*, it is equal to *ARBGNR* without the first digit. It may be a *JURNR, SENR, or CVRNR*.

ARBGNR and *CVRNR* identify the owner of the establishment. These identifiers change if the establishment gets a new owner, and even an existing owner may decide to change the *CVRNR* for different reasons. Even if the owner changes, a specific analysis may still want to consider the establishment to be the same.

The purpose of *IDA* was to study the flow of workers between *establishments* and the dynamics of *establishments*. How an over-time identifier for the *establishments* is defined will influence these flows.

IDA has physical establishment and fictitious establishment, or in common workplace. Some workers work in or out from home and not at a physical establishment, and they are assigned a fictitious establishment associated with the *IDA* firm. *LBNR* is "0000000000" for fictitious establishment. A fictitious establishment is identified by *CVRNR, LBNR*.

Determining *LBNR* In *IDA* an *establishment* is the same unit, if the owner or the labour force is unchanged, hence less emphasis is put on location. No information is available about physical capital and technologies.

The *IDA* over-time consistent identifier for *establishments, LBNR* is based on a set of rules.

An establishment is considered the same if one of the following rules in Table 1 is fulfilled:

Table 1 Rules for treating an establishment as the same economic entity as last year

Code ¹	Rule	Remarks
1	Unchanged owner AND industrial sector	Owner is defined by <i>ARBGNR, CVRNR</i> . Industrial sector is defined by a 5-digit DB code
2	Unchanged owner AND labour force	Owner is defined by <i>ARBGNR, CVRNR</i> . Labour force is defined as at least 30% of workers are part of the labour force in either year
3	Unchanged labour force AND unchanged address OR industrial sector	Labour force is defined as at least 30% of workers are part of the labour force in both years. Address is defined by (municipal code, street code). Industrial sector is defined by a 5-digit DB code
9	Not the same workplace	No relation. <i>establishment</i> has been created the year before or closed.

¹ Code values for *REGEL_FREM* and *REGEL_TILB*

The determination of *LBNR* is done through a sequence.

- determine the establishments which have the same owner in two consecutive years
- determine the establishments which have the same labour force in two subsequent years.
- define multiple relationships, i.e. an establishment which has been split into two or more establishments, or two or more establishments which have been merged into one establishment.
- determine for the split and merges if it took place within the same owner or between different owners.

Information about split and merges are not used in determining *LBNR* but is available in *ID_FREM* and *ID_TILB*.

It is important to be able to distinguish between normal mobility and split and merges. Especially for small establishments this is difficult.

Notes

1. There may exist a single relation or a multiple relation between establishments in the two years.
In a single relation there is one *establishment* in *current-year* and in the next year.
In a multirelation two or more *establishments* are merged to one *establishment* or one *establishment* is split into two *establishments*.
How unchanged owner is defined in multirelations needs to be established.
2. If several rules are fulfilled it seems that the first is selected.
3. In 2004 there is a jump in code values 1 and 3 because 15,000 *establishments* in the municipal sector changed *ARBGNR* from the old *JUR_NR* to *CVRNR* and some were allocated to the private sector