

ECONAU Data Set Documentation

Dale T. Mortensen Centre, Aarhus University —
Version 21/01/2022

Henning Bunzel

1 IDAN, IDA employments/jobs

DSECONAU Updated: 12072018 2008-2016, Update yearly data sets.

ECONAU Updated: 14-03-2019.

1.1 Remarks on Update

NYT 4-kvartal 2018 2007-2016. Alvorlig fejl. **NYT 2-kvartal 2018** 2007-2016. Alvorlig fejl. *IDAS* and *IDAN* have several variables defining the data the following year. They are not included in the latest year. Therefore, we must always have current and previous year at each update for *IDAS* and *IDAN*.

Changes in Version 2016:

1. In 2008 the sources change. Several variables are discontinued, and new variables are introduced.
 - The source of data on employee jobs changes from *CON* to *BFL*. *BFL* is monthly and has a more exact dating of a job and number of hours worked. *BFL* It has a basic wage and an extended wage. The basic wage corresponds to the wage in 1980 to 2007, but there are minor differences. In 1980 to 2007 there is not a number of hours worked, but it can be imputed from job wage and hourly wage.

- The source of non-employee activities changes from *LC* to *OF*. *OF* has a net and gross unemployment. The net unemployment corresponds to the unemployment in 1980 to 2007.
- SOCIO and DISCO codes are now available from *RAS*.

Errors and Questions:

1. 1.092.954 jobs have a *PNR* which does not exist *PERSONER*. In 2013 version of IDAN it was only 43.916 observations, and all observations were in 2004 with *TYPE* equal to N,E,3. Now it is in 2004, 2008-2016 and include primary jobs. Is this an error? *RAS* population is end november for 2008-2016, and not end of year. Is this also the case for the IDA population? or are jobs now included even if the person does not live in Denmark End of November or End of Year? This is the case in *RAS*, therefor the observations are not deleted. *ECONAU* has created a variable *I_BEFOLKNINGEN_KODE* as in *RAS*.
2. *JOB_ADGANGSADRESSE* 16.151.707 observations have an encrypted missing value.
3. *JOB_ADGANGSADRESSE*. It is not described in FSE documentation, but it is assumed that KOM and OPGIKOM are address of the establishment where the person works. In *RAS* they are prefixed with JOB and BOPAEL. In *RAS* both BOPAEL addresses exist for 1980 to 2016. *JOB_ADGANGSADRESSE*, *Access street code in municipality (street, house number, house letter) for a job* exists for 1980 to 2016, while *JOB_ENHEDSADRESSE*, *Job unit address* only exists for 2009-2016. It exists in *ESR*.
4. *TYPE_2008* is not in TIMES dokumentation - link from FSE datasets does not work.
5. The following variables was in previous 2008-2013 data sets, but are not included in 2008-2016. Will they be included in the next IDAN version? *TILKNYT*, *JOBKAT*, *STILL*.
6. *ARB_HOVED_BRA_DB07* is not available for 2000-2007. It is available in *RAS*, will it be available from 2000 in the next IDA version?
7. *JOB_STOETTE_BESK_KODE* is not available for 1980-2007. It is available in *RAS* for 1994-2007, will it be available from 1994 in the next IDAN version?
8. SOCIO code is only available from 2008, it exists from 1996 in *RAS*, will it be available from 1996 in the next IDAN version?.

ECONAU modifications of the data set:

1. 90.355 observations have an illegal *PNR*. These observations are deleted.
2. *PERSADR* is a packed version of *KOM* and *JOB_ENHEDSADRESSE*. We need the variable split up in two variables. Otherwise it cannot be matched to any other data set. About 3 mio observations have a encrypted missing value code.
3. *ARBNR* 14.748 observations in the period 2008-2013 have an illegal value. The illegal values accoure in all years, and for all job types.
4. *CVRNR* 6.214 observations in the period 2008-2013 has missing values for all types of employees. Some of the observations have a valid *ARBNR*.

5. *JOBLON*, 69.407 observations for employees have a zero value. This is an error. A positive value exists in *CON*, *RAS*, and *BFL* for those checked. Errors exist in both periods 1980 to 2007, 2008-2013.
6. In the period 1980 to 2007 77.476 persons do not have a prime job. In 2008-2013 311.608 persons do not have a prime job. Comparing with data in *RAS* it looks like the observations should not have been included in *IDAN*.
7. *ANSDAGE*, 7.191.090 observations with value 0 have been changed to missing, see Remarks below.
8. *JOB_KAT*, 2.577.808 observations with value 0 has been changed to missing. *JOB_KAT* has code value 9 for "not available";
9. *JOB_ADGANGSADRESSE*, 16.151.707 observations have an encrypted missing value. It has been replaced to missing.
10. *ARB_STK*, 115.444 observations have a value "000" or "0000". The FSE documentation specifies that "0000" is not a legal value, therefore it is replaced to missing;
11. *JOB_LOEN_BELOEB_BRED*, 1.886.323 observations with value 0 have been changed to missing.
12. *JOB_LOEN_BELOEB_SMAL*, 4.397.624 observations with value 0 have been changed to missing.
13. *ARB_NR*, 96604 observations have blank characters as part of the value. According to *ESR* data set *ARB*, these values are illegal. They are changed to missing.
14. *CVRNR*, 96.604 observations have an illegal *cvmnr*. They are changed to missing.

Remarks:

1. There are 110743 duplicate records sorted on (*AAR*, *PNR*, *TYPE*, *CVRNR*) , or sorted on (*AAR*, *PNR*, *TYPE*), why?
2. *JOBLON* has no missing values in 2004, 2012 and 2013, is the variable just set to 0?
3. *ARBGNR* has 188.365 missing values in 1980 falling to 130.000 in 2007. *CVRNR* has app 150.000 missing values in 2008-2013.
4. *ANSAAR*. Break in 2007-2008. In 2007, 1.212.276 missing values, in 2008, 2.941.486 missing values, number of observations has increased by 746.560. *ANSDAGE*, number of missing values has just increased by 27.000, but the mean falls from 277 to 184, more zeroes? The same pattern for *ARBLED*. A break in mean value for all other series, too.
5. *ARBSTK* has app 1.5 mio missing values 1980 to 2007, but *ARBNR* has only app 155.000. *ARBSTK* is missing, "", for firm with only 1 workplace?
6. *ARBGNR8* for 1980 to 2007 has been renamed to *CVRNR*, even though it is NOT a *CVRNR* for the years 1980- 2000, it may be a *JURNR* or *SENR*.
7. *JOB_KAT* has both missing ' ' and '0' values. In code values '9' is defined as missing
8. RENAMED: *ARBGNR8* to *CVRNR* for the years 1980 to 2007, *ARBNR* to *ARB_NR*, *ARBSTK* to *ARB_STK*, *JOBATP* to *JOB_ATP_BELOEB*, *JOBLON* to *JOB_LOEN_BELOEB*, *JOBKAT* to *JOB_KAT*, *TYPE* to *JOB_TYPE*.

1.2 Description

IDAN contains data about activities on the labour market since 1980.

The population is all persons with a permanent address in Denmark end of the year.

JOB.TYPE In *IDAN* the activities are grouped by *JOB.TYPE*.

- H - November primary job.
- A - Employer
- S - Self-employed
- M - Assisting spouse
- B - November secondary job
- N - Other November job, exist from 2004.
- 3 - Most important non-November job, exists 1980 to 2007.
- E - Other Non-November job, exists from 2004.

Hence an employee has an End-of-November Job and maybe a secondary End-of-November job, and other jobs. The types of jobs included in *IDAN* vary over time.

- *Primary Activity* (H, A, S, M) All active persons are assigned one End-of-November activity, either as employee or employer/self-employed.
- B, N. Other November jobs.
- 3, E. Non-November jobs. These jobs exist in *IDAN* from 1994.
- E. Most important non-November job.

Data available for different activity types Different data are available for the different activity types, see Table 1. In summary:

- H, Employee, and A, Employer. All variables in *IDAN* have a value.
- S, Self-employed and M, Assisting spouse. No job data, like firm and workplace, hours and salary. Income exists in *IND* and main yearly activity in *AKM* like *SOCIO*, *DISCO*, *BRANCHE*.

Summary xx shows number of job types over the years.

Data Break 2008 There is a major data break in 2008 as new data sources are introduced.

- In 1980 to 2007 *CON* is major source of employee earnings. It is a yearly report from employer. The employment spell is specified as entire year, a specified sub-period, a number of unspecified employments. From 2008 the source is *BFL*, which is a monthly report with start-date and end-date of the employment. It looks like 1980 to 2007 contains too many all-year employment. The result is a drop in number of jobs in 2008.
- In 1980 to 2007 the source of non-employee earnings are several data sets. From 2008 these data sources have been combined in *OF* and data from the different data sets have been cross-validated and overlapping spells have been removed.

Table 1 Table 1

Variable	Job Type							
	3	A	S	M	H	N ^j	E ^j	B
ANSAAR	Only missing values	Values all years ^e	Only missing values	Only missing values	Values all years ^g	Only missing values	Only missing values	Only missing values
ANSDAGE	Values all years ^e	Only missing values	Only missing values	Values all years ^f	Values all years ^e	Values all years ^e	Values all years ^e	Values all years ^e
ANSXFREM ^k	Only missing values	Only missing values	Values all years ^b	Values all years ^f	Only missing values	Only missing values	Only missing values	Only missing values
ANSXTILB ^k	Only missing values	Values all years	Only missing values	Values all years	Values all years ^e	Only missing values	Only missing values	Values all years ^e
ARB.ENHEDSADRESSE.IDAN ^a	Only missing values	Values all years	Values all years	Values all years	Values all years	Only missing values	Only missing values	Only missing values
ARBLEID	Only missing values	Only missing values	Only missing values	Only missing values	Values all years ^b	Only missing values	Only missing values	Only missing values
ARB.NR	Values 2008-2013	Values 2008-2013	Only missing values	Only missing values	Values 2008-2013	Values 2008-2013	Values 2008-2013	Values 2008-2013
ARB.STK	Values 1980 to 2007 ^e	Only missing values	Only missing values	Only missing values	Values 1980 to 2007 ⁱ	Values 2004-2007 ^e	Values 2004-2007 ^e	Values 1980 to 2007 ^e
CVRNR	Values all years	Values all years	Only missing values	Only missing values	Values all years	Values all years	Values all years	Values all years
JOB.KAT	Only missing values	Only missing values ^c	Only missing values	Only missing values	Values all years	Only missing values	Only missing values	Only missing values
JOB.LOEN.BELOEB	Values all years	Only missing values	Only missing values	Only missing values	Values all years ^e	Values 2004-2013	Values 2004-2013	Values all years
LBNR	Values 2004-2013	Only 0's	Only 0's	Only 0's	Values all years	Values 2004-2013	Values 2004-2013	Values all years
LOENM.FREM	Only missing values	Values all years	Only missing values	Only missing values	Values all years	Only missing values	Only missing values	Only missing values
LOENM.TILB	Only missing values	Values all years	Only missing values	Only missing values	Values all years	Only missing values	Only missing values	Only missing values
STILL	Only missing values	Values all years ^d	Only missing values	Only missing values	Values all years ^e	Only missing values	Only missing values	Values all years
TILKNYT	Only missing values	Values all years	Only missing values	Only missing values	Values all years ^e	Only missing values	Only missing values	Only missing values
JOB.TIME.LOEN.SMAL	Only missing values	Only missing values	Only missing values	Only missing values	Values all years ^e	Only missing values	Only missing values	Values all years ^e

^a The address cannot be matched to other data sets. ^b Many missing values from 2004. ^c *JOB.KAT* has a few values in 1980. Probably an error. ^d All values are set to 5.

^e A few missing values. ^f Many missing values. ^g A few missing values 1980-2003. ^h Many missing values because they are not unemployed during a year. ⁱ Values 1980 to 2007. Most values are missing. ^j This job type only exists from 2004. ^k According to documentation values should be available fro H,B, A, M, S which is not the case for ANSXFREM

- A person is in the labour force, but unemployed, if the persons is full time unemployed in the End-of-November week.
- In 1980 to 2007 a person was defined as an employee if the total, yearly salary exceed app. 10.000 DKR. After 2008 a person is defined as an employee if the salary exceeds 4 hours of regular low-paid employment. To be employed end of november either the job spell must include this period or the ATP-payments are equal to the contributions of a yearly, full-time employment.
- In summary the changes result in fewer employee jobs. The largest decline is for the group of age 18-29. For the group below 18 there is an increase because their yearly salary is below 10.000 DKR.
- The definition of yearly income was changed in 2013.
 - The income of self-employed was changed in 2013. Interest and capital payments related to the firm are included in the income of self-employed. The change change the number of persons grouped as employee, self-employed and out-of-labour force. App. 15.000 persons were moved from self-employed to early retirement (efterløn)
 - The imputed income from owning ones residence is reduced to the macro value used in national accounts . Before the change it was 4% of the value of the residence.
 - Danish national church tax is no longer part of the income/tax figures
 - Employer paid sickness payments are moved from public transfers to earnings
 - Green compensation payment is included as income from 2010
 - Transfer payments are split into private pensions and public transfer payments
- Change in SOCIO in 2013. Persons, age above 15 years, with a salary below 10.000 DKR and less than 950 hours of work are classified as "Persons above 15 years in education". App 115.000 persons have been moved to this group for employee.
- The 2013 income and socio variables have been extended back to 1987, income of self-employed has a break in 1994, because *HONNY*, does not exist before 1994.
- *CON* and *BFL* are available in January in the year following the year of the income. Income of self-employed is not available until the final tax-income register in November. For a number of self-employed persons data is not available until later than November. In such cases they may be groupd as employee, hence we observe a number of wrong sequences: self-employed, employee, self-employed.

The following modifications have been made to the *ECONAU* version of *IDAN*:

- *PNR*. The observations with a *PNR* which does not exist in *PERSONER* are deleted. They are all dated to 2004.
- *CVRNR* and *ANSXFREM* with illegal values are set to missing.
- *ARBNR* with illegal values is set to '999999999999'.
- *LBNR*. After 2008 exist both '0' and '0000000000' values. '0' is changed to '00000000000'.
- *JOB_LOEN_BELOEB* and *TIMELON* have 0 values replaced by missing values.

- *ARBSTK* has values '000' and '0000' set to missing.
- *ANSAAR* has value '0' set to missing.
- *ANSDAGE* has value 0 set to missing..
- *ARBLED* has value 0 set to missing.
- *JOB_KAT* has value '0' set to missing. The variable has value '9' for missing

Fictitious workplaces. Fictitious workplaces were introduced i 1991.

For the primary employee job there are jobs where no establishment is assigned to *ARBGNR*. For the period 1980-1990 it is approximately 96,080 jobs in 1980 falling to 15,041 in 1990 out of approximately 2,200,000 jobs. For *ARBGNR* with an establishment some jobs could not be assigned to an establishment. In the period 1980-1990 they were assigned to the largest establishment, from 1991 they were assigned a fictitious workplace. Both missing establishments and fictitious workplaces have a *LBNR* with a 0 value. In 1991 the number of observations with *LBNR* equal to 0000000000 jumped to 1 30,545 increasing to 171,718 in 2007. Due to a data break in 2000 it fell to a level of approx. 125,000.

For job type B observations with *LBNR* = 0000000000 jump from 5,000 to 14,000 out of 225,000 jobs in 1991.

Job type B has no jump in number of observations with *LBNR* = 0000000000 in 1991.

Job types N and E are introduced in 2004.

For the period 1980 to 2007 *CON* has 1 yearly income tax report for each employee at each establishment. Employees not assigned to an establishment are assigned to the *IDA* tax reporting unit, *CVRNR*. An establishment is identified by *LBNR* or (*CVRNR*, *ARB_STK*.) But the key has (*AAR*, *PNR*, *CVRNR*, *LBNR*) has 46,612 duplicates, and using (*CVRNR*, *ARB_STK*) instead results in 326,377 duplicates.

For the years from 2008 income tax reports are monthly, hence an employer can have different *JOB_TYPE* for the same establishment.

Primary Jobs. The documentation states that all persons in *IDAN* have a primary job at the end of November, i.e. a job of type H, A, S, or M, and zero or more other jobss as employees, i.e. a job of type B, 3, E, N. However, this is not the case.

In the period 1980 to 2007 *IDAN* has 5,106, 444 persons of whom only 5,28,968 have a primary job. But in the years 2002-2007 all persons have a primary job.

In the period 2008-2013 *IDAN* has 3,695,549 ersons of whom only 3,383,941 have a primary job.

81 % of persons without a primary job have job type 3. Most of these observations have a missing value of *STILL*, otherwise the code values 30, 31, 39. It is most likely that the observations should not have been included in *IDAN*.

Key for IDAN jobs. For the period 1980 to 2007 he source for jobs is *CON*, *Employer's annual report on employee earnings and withheld tax*. There is one yearly income tax report for each employee working some time during theyear at an establishment. The establishment is identified by (*AAR*, *CVRNR*, *ARB_STK*.)

For the period from 2008 the source for jobs is *BFL*, *Employment for employees, since 2008*. There is one monthly income tax report for each employee working

at an establishment during the month. The establishment is identified by (*CVRNR*, *ARB_NR*.) In *BFL* the FSE variables are (*AJO_CVR_NR_FRA_PROD_JOB*, *AJO_ARBNR_SENR*.)

 If *AJO_FIKTIV_ARB_KODE* is different from 0, *LBNR* is set to 0, but a value of *AJO_ARBNR_SENR* is defined.