

LMDG Spell Data

**Dale T. Mortensen Centre, Aarhus University —
Version 15/05/2025**

Henning Bunzel and Mads Hejlesen

Department of Economics and Business Economics, Aarhus University, e-mail: hbunzel@econ.au.dk, e-mail: mhejlesen@econ.au.dk

Contents

LMDG Spell Data	1
Henning Bunzel and Mads Hejlesen	
1 Structure and data sources for LMDG spell data	5
1 Introduction	5
2 Data source	6
2.1 Firm identifiers	6
2.2 Workplace identifier	7
3 Employment spells	9
4 Structure of the spell data	9
4.1 SPELL RECORD	10
4.2 EMPLOYMENT SPELL FIELDS	10
4.3 SUBSPELL FIELDS IN A MASTER SPELL	10
4.4 SUBTYPE SPELL RECORD	10
4.5 Final spell datasets	11
2 Firm and Workplace Identifiers	13
5 Identifiers	13
6 New ECONAU datasets	13
6.1 LMDG CVRNR	14
6.2 LMDG ENTERPRISE REGISTER	14
6.3 LMDG WORKPLACE IDENTIFIER	14
6.4 LMDG OVER_TIME ENTERPRISE IDENTIFIER	15
7 LMDG Firm and Workplace Identifiers	15
8 Matching datasets with FIRM identifiers	15
3 Types of spells	17
9 Subtype spells	18
9.1 YEAR SUBTYPE SPELL	18
9.2 BENEFITS, LABOR MARKET PROGRAMS, UNEMPLOYMENT	18

	9.3	ESTABLISHMENT	18
4		Comparison of spell data and other labour market statistics	21
	10	Employment	21
	11	Employment and firms	21
	12	Unemployment	21
	13	Out of labour force	21
	14	Dream	21
5		Dynamics over the business cycle	23
	15	Termination of spells	23
	16	New spells	23
	17	Flows of jobs	23
	18	Flows of workers	23
6		A descriptive model for Ins and Outs of the labour market	25

Chapter 1

Structure and data sources for LMDG spell data

1 Introduction

Individual spell data, i.e. a longitudinal dataset of transitions between different status on the labour market, is required for the analysis of many important aspects of labour market theory.

The major part of the personal register data in Statistics Denmark is yearly data, for example IDA only has the labor market state in the last week of November and no information about the states in between.

Using a number of data sets it is possible to generate a longitudinal data for persons' transitions between different **states** in the labor market.

LMDG spell data is such a longitudinal dataset from 1984. For the E,U, and N states it is possible to create spells back to 1980. The major states in the labor market are

- *E*, Employed
- *U*, Unemployed
- *N*, Not in the labor Market

If possible, a spell will always contain variables for *pay* (i.e. *wage*, *salary*, *income*, *remuneration*, *social benefit*) and *hours* associated with the pay. A spell will typically contain the following core variables and some specific variables.

- *PNR*, An identifier of the person.
- *SPELL_ID*, A sequence number for this spell type for this person.
- *START_DATE*, The day on which the spell starts.
- *END_DATE*, The day on which the spell ends.
- *STATE*, A code for the person's state in this spell. Each spell type has a number of states.
- *PAY*, The person's pay for this period and state.
- *HOURS*, Hours of effort associated with the pay.
- *SPELL_FIRM_ID*,
- *SPELL_ARB_ID*,

For some analyses it is necessary to subdivide the major status, f.ex. unemployed into real unemployed and participating in a Labour Market Programme.

More detailed information about a spell is available in *sub_spells* which are linked with *subspell keys* to the major spell. A *sub_spell* itself may have a *sub_spell*.

Also, for some spells there exist repeated data, for example monthly income. The spell will contain the total or average income and the detailed income data will be in subspells.

The data becomes better over time and today the important datasets report events dated to a *calendar day*. The *LMDG spell data* uses calendar day for *start* and *end* of a spell, even if the source data has a lower frequency for example monthly.

The source datasets have different frequencies, like week, month, and year, hence many assumptions are needed in order to construct spells with day as frequency. The content and definitions of each source change 2-3 times over the period. The dating of early spells has more errors because the raw source datasets have lower frequencies and many missing value in the early years.

For each source we construct a *raw spell dataset* with the same contents for all years, maybe with missing values for the early years.

The MergeSpells algorithm determines how to merge these spells together to a *spell datasets* without *overlapping periods*. If two datasets are involved in an overlap one of the datasets is often more reliable than the other data source and therefore its start- and end dates are used. A number of *sub_spells* are constructed containing data on any conflicting periods, or for an aggregated period like year.

2 Data source

2.1 Firm identifiers

The two major firm identifiers are *SENR*, *A firm account number for tax and custom payments* and *CVRNR*, *Encrypted legal unit identifier. Data break 1994*

SENR was introduced in 1985 as an identifier for fiscal payments from firms to customs and tax authorities.

CVRNR was introduced in 1999 as a unique identifier of legal units. Firms may also choose to have several *CVRNR* and to change a *CVRNR*. Therefore, *CVRNR* is not a unique identifier of an *economic unit*.

OKNR is an identifier of an economic unit. One *OKNR* may be owner of one or more *CVRNR*. One *CVRNR* is marked as being carrying data about an *economic unit*.

A *CVRNR* may own one or more *SENR*. If the legal unit only has one *SENR* then the *CVRNR* is often equal to the *SENR*.

The DS-DRDS project "Firm Warehouse" is making *OK_OT_ID* an over-time identifier for the economic unit from 2000.

The spell project has constructed a similar over-time identifier using worker-flows among *IDAS* workplaces. These identifiers can be constructed from 2000. They are compared to *OK_OT_ID*.

Until 2018, several datasets did not include *SENR* before 1999, but only *ARBGNR* or *JURNR*.

JURNR is an *ESR* unique identifier of legal units and *ARBGNR* is an *IDA* unique identifier of owners of workplaces. *JURNR* and *ARBGNR* are similar. For privately owned legal units they are equal to *SENR*. For non-private units different identifiers are used, see the description in *ARBGNR*. For privately owned legal units *ARBGNR* is a *SENR*, for others it is similar to *JURNR*.

Before 2018 *ARBGNR* was equal to *ARBGNR9* i.e. *ARBGNR8* with a one digit prefix identifying the system reporting taxable income. From 2018 *ARBGNR* is equal to *ARBGNR8* in *FSE* datasets, but all *ECONAU* datasets may not yet have been updated. If the dataset has been updated *ARBGNR* is renamed to *SENR*.

From 2018 all relevant *FSE* datasets contain *SENR*, hence *ARBGNR* and *JURNR* are no longer needed.

The *SENR* can be used to map *CVRNR* on these datasets. The mapping is done using dataset *ESR_SE_CVR*. *ESR_SE_CVR* is also used to verify *SENR* and *CVRNR* in all new data. All raw spell dataset use this mapped *CVRNR*.

2.2 Workplace identifier

Until 2008 a physical workplace is identified by (*AAR*, *ARBGNR*, *ARB_STK*) in *CONESR* and *RAS* or (*AAR*, *ARBGNR*, *DSKOD*) in *IDA*. *ARB_STK* and *DSKOD* are similar but not identical. A fictitious workplace is identified by specific *ARB_STK* code values.

From 2008 a workplace is identified by *ARB_NR* and *ARB_SUBTYPE* and an indicator *FIKTIV_ARB_KODE* define the workplace to be a physical (workplace) or a fictitious workplace.

IDA has constructed an over-time physical workplace identifier *LBNR*. *LBNR* is 00.00.00.00 for fictitious workplaces, hence they are not uniquely identified by *LBNR*, but must be identified by *ARBGNR*. If a *ARBGNR* has only one workplace, then *ARB_STK* has a missing value, but (*AAR*, *ARBGNR*, *ARB_STK*) is still a unique key. An establishment is a physical workplace.

IDAS only has physical workplaces but *CONESR*, *BFL*, *IDAN*, and *RAS* have both physical and fictitious workplaces. *CONESR* and *BFL* only have jobs occupied by employed workers.

All the workplace keys are collected in spell dataset *SPELL_ARB*.

Table 1 Datasets contributing to SPELL_ARB

Dataset	Period	Variables											
		AAR	Month	ARBGNR	CVRNR	SENR	OKNR	ARB-STK	ARB-NR	LBNR	FIKTV_ARB-KODE	FUNKTION-KODE	ADGANG-ADRESSE
ESR_ARB	1999-	+			+		+						
IDAN	1980-2007	+		+	mapped			+		mapped			
RAS	2008-	+			+		mapped		+		mapped		
IDAS	1980-2007	+						+		+	mapped		
BFL	2008-	+	+		+			+		+	+		+
CONESR													

3 Employment spells

The variable *SOC_STATUS.KODE* defines for all persons the association with the labour market according to the ILO definition. *LMDG_SPELL* data only includes persons 15 years or older and living in Denmark. Persons are considered under education if

The population is grouped as ALL persons

- employed
- employed workers
- self-employed, assistant spouses
- unemployed in publicly supported jobs or training
- temporary leave, publicly supported
- unemployed
- out of the labour force

- publicly supported
- students
- pensioners
- others

Raw spell datasets are constructed for each group. *CONESR* and *BFL* only include employed workers. *IDAN* includes all.

For *employed workers* the data source is *CONESR* before 2008 and *BFL* from 2008. For *self employed persons and assistant spouses* the source is *ESR*, *RAS*, *IDAN*, and *IND*.

For persons on leave from the labour market or working in supported jobs or training the data sources are

For *unemployed persons* the data source is

For *persons out of labour market* the data source is

4 Structure of the spell data

The datasets are organized in a number of spell and subspell datasets, The important types of spells are

1. PERSONS, a spell for each person with address in Denmark
2. UNEMPLOYMENT, a spell for each person registered as unemployed or receiving benefits which require that you are searching for jobs
3. EMPLOYMENT, a spell of employment for an employee, employer, self-employed, assisting spouse working for a self-employed. During any period a person may have several employments. One employment is selected as the main employment.
4. LABOR MARKET PROGRAMS, a spell for persons in a labor market program
5. NOT IN THE LABOR MARKET, a spell for persons under education or receiving pensions and benefits which do not require that you are searching for jobs

4.1 SPELL RECORD

A Spell Record has the following fields

- pnr, an identifier of the person
- seqno, sequence number for this spell type for this person
- from-date, day on which the spell starts
- to-date, day on which the spell ends
- state, a code identifying the state
- end_november, 1 if the spell includes end of November as defined by IDA
- left_censored, 1 if censored, else 0
- right_censored, 1 if censored, else 0
- fields specific for this spell type

4.2 EMPLOYMENT SPELL FIELDS

If it is an E, Employment spell, it has a set of fields which identifies the firm, workplace, wage and contains imputed values for hours worked and hourly wage

- CVRNR, a firm identifier
- ARB_STK, ARB_NR, a code identifying a workplace for this enterprise
- LBNR, IDA Over_Time identifier of an establishment
- JOB_LOEN_BELOEB, wage for entire spell
- HOURS_WORKED, hours worked entire spell
- HOURLY_WAGE, imputed hourly wage for the spell

Other spells like Pensions, Benefits, and Labor Market Programs have Income for the entire spell.

4.3 SUBSPELL FIELDS IN A MASTER SPELL

If this spell type has subsPELLs it will have the following fields for each subtype

- no_subSPELLs
- aggregated fields specific for this spell subtype
- sub_spell identifier

4.4 SUBTYPE SPELL RECORD

A subtype spell record has the following fields

- pnr, an identifier of the person
- spell_number, sequence number for the master spell
- sub_spell number, sequence number for this subtype spell for this master spell for this person
- from-date, day on which the spell starts
- to-date, day on which the spell ends
- state, a code identifying the state
- fields specific for this subspell type

4.5 Final spell datasets

Chapter 2

Firm and Workplace Identifiers

5 Identifiers

A firm is described by a large number of factors.

- Owner. The owner defines wage and hiring policy, marketing policies, investment and growth.
- CEO and board. In some firm types the administration may be determined by CEO, or CEO, board, and owner in cooperation,

An *enterprise* is an organisational unit producing goods or services which has a certain degree of autonomy in decision making. An enterprise can carry out one or more economic activities, produce one or more goods and have one or more *establishments* where the activities take place. An enterprise may consist of one or more *legal units* and each legal unit may have one or more *fiscal or administrative identifiers*

For tax, liability and organisational reasons an economic unit may choose to organise its activities like sales, transportation and different production in two or more legal units, use more than one fiscal registration. But if these legal units have the same management they are considered one economic unit. But one or more economic units may form an *enterprise group* if they are bounded together by legal and/or financial restrictions. These restrictions may be across several countries.

6 New ECONAU datasets

The construction of spell data, but also other applications of the data, require a modified version of the enterprise and workplace identifiers, named lmdg-firm-identifiers.

6.1 LMDG CVRNR

The Statistical Denmark Dataset uses different universes of firm identifiers. SENR, Fiscal Identifier, is used for tax, income report, value added reports, internal and external trade. In 2000 CVRNR, a legal firm identifier, was introduced, and for most legal firm units with just one SENR, this SENR is used as CVRNR. An enterprise is the smallest set of legal firms which makes economic decisions and is in most datasets identified with one of the CVRNRs.

Hence, an enterprise may have many SENRs and CVRNRs and they may change over time.

In some datasets using SENR Statistics Denmark has replaced the SENR by CVRNR in some, or all years, but the actual mapping file is not available for each dataset. Statistics Denmark has one mapping file SE_CVR, which is undocumented.

The firm identifier, CVRNR, being used in IDA employee datasets does not use the same universe as the CVRNR being used in the Business datasets. Statistics Denmark supplies the FIDA dataset to help with this matching. This dataset has the same non-match problems as the other datasets, but documented. *FIDA* is discontinued in 2008. The current version identifies *CVRNR* for workers in *EBS*.

When trying to match different datasets on SENR and CVRNR, including the SE_CVR file, we get many no matches, mainly for small enterprises. Some datasets do include a variable describing the matches and for these datasets some documentation of the possible reason for a non-match is available.

6.2 LMDG ENTERPRISE REGISTER

We will add an *econau cvrn* to all data sets and create match variables describing the match with the different data sets. For some datasets and records the *econau cvrn* will differ from the *cvnr* of the dataset. Some of this information is available in The Statistics Denmark Enterprise Register, which is not in general available for research, but we are asking for some variables from this dataset.

This register will identify the *cvnr* forming the universe of enterprises in the Business datasets.

6.3 LMDG WORKPLACE IDENTIFIER

The IDA datasets have an Over_Time_Establishment identifier, *lbnr*, for all establishments. Some employees are not working in an establishment, like travelling salesmen, and they have *lbnr* value 0, but they are still connected to an enterprise. They have a fictitious workplace. We change this *lbnr* value to a unique value such that all employees have a unique *lbnr*.

6.4 LMDG OVER TIME ENTERPRISE IDENTIFIER

It is well known that establishments, legal firm units, enterprises, and enterprise groups change SENR and CVRNR over time. Hence, it is not correct to use the cvnr of an enterprise over time, but it is done all the time.

Economic theory does not have an operational definition of an enterprise/firm. Therefore, we will define an *Over_Time_Enterprise_Identifier* as an identifier where more than 50% of the workforce is the same in the two years. In the long run we hope to add a lot of firm variables to these data sets, like CEO, Board, Financial structure.

We try this definition based on flows between enterprises and based on flows of establishments. If we use establishments it will also require that the workforce is the same in this establishment in the two years and that they are producing the same goods.

7 LMDG Firm and Workplace Identifiers

The enterprise dynamics project which generates *Over_Time_Enterprise_Identifier* also generates a change in identifier, which in the spell data might be treated as the same enterprise.

If cvnr-x transfers more than 50% of the workers to cvnr-y next year, then cvnr-x will be closed, and the employees working in a new cvnr-z. In the spell dataset we may want to treat cvnr-x and cvnr-z as the same enterprise.

In 2005 many cities were merged to less, larger cities. In spell data we may want to treat Aarhus before 2006 as the same employer as Aarhus 2006 and later.

8 Matching datasets with FIRM identifiers

It seems to be very difficult to merge enterprise statistics across FSE data sets. There are several reasons for the missing matches.

- The universe of enterprises may be different, for example all enterprises or only enterprises with a minimum activity
- The enterprises existing on different dates or all enterprises which have existed during a specific period
- The units report using different identifiers
 - SENR a fiscal identifier
 - CVRNR an identifier for a legal unit, firm.
- An enterprise may consist of several legal units, may use more than one fiscal identifier, and some enterprises reports on the same fiscal identifier.

- Different fiscal identifiers are used when reporting different statistics, like Withheld Income Taxes, VAT, Export and Import Transactions, and Enterprise Taxes
- Both fiscal and legal identifiers can change over time even though the enterprise is the same when regarded as an economic unit, an enterprise.

Chapter 3

Types of spells

For most analyses the employment spell is the most important spell and it is also the one which is most difficult to construct. Therefore, we give a short description of the process.

The source is the taxable income reports, which all firms must submit to the Tax Authorities. This report also includes the amount paid to ATP pension for the employee. This amount is a step function of hours worked. IDA has used this data to impute number of hours worked and an hourly wage, but only for the end-of-November job. Rune Vejlin and Christian Lund have published a report of a revised algorithm which provides a better estimate of the number of hours worked and hourly wage. The algorithm is used to compute these variables for each employment spell. A simplified version of their algorithm is applied for the employment spell data.

During a year an employee may have many jobs, either with the same firm or with different firms. We need to choose which job is the primary job for each spell. For the period 1980-2007 we know if the employee has been employed in the firm all year, in a specified interval, or a number of times. In which case we do not know the start and end date of each spell. From 2008 all firms must report income, withheld taxes, paid ATP, and employment period each month.

Data shows that employees receive a number of 'random' amounts, often receive two full payments in the same month from two different firms.

The old spell data has two major problems: Many short employment spells where the employee returns to the same firm and too many spells start January 1st each year.

We decided to use number of hours worked over the next 3 months to select the current primary job. If the spell starts in the middle of a month we check if the primary job in the previous month continues until this new spell starts.

We choose the algorithm by looking at the data for the period 2008-2011 and compute different descriptive statistics. Then we applied the method to the period 1980-2007 using MIA data for 1997-2007 to improve the start-end dates on the income reports. MIA data is a monthly indicator for each person employed in a firm.

And it seems to work!

We still get too many spells starting January 1st, especially in the early years, each time the data source changes, and the year we had the municipality reform. When the new spells are recomputed using the `Over_Time_Enterprise_Identifier` we expect that this problem will be reduced a lot.

9 Subtype spells

The old spell data sets are very disaggregated, ie. have many states. Typically, researchers start by reducing the number of states to 3 and there exists a yearly spell data set - which I do not understand how to use.

The new spells are, aggregated but have the concept of a subtype spell. Standard jobs exist for replacing a master spell by all the subtype spells.

This method can also be used to document or debug the source spells that have been used to generate a master spell.

9.1 YEAR SUBTYPE SPELL

Many covariates are yearly and must be merged on from yearly files.

A YEAR SUBTYPE SPELL is generated for each year included in a spell. These spells will include core yearly variables, like income, highest completed education, years of experience, fraction of workdays and calendar for the year included in the spell. The year in the subtype spell can be used to merge yearly data sets on the subtype spells.

The master spell can contain aggregated values of the subtype spells.

9.2 BENEFITS, LABOR MARKET PROGRAMS, UNEMPLOYMENT

The old spell data has many states for each of these spell sources. In the new spell they are defined as subtype spells.

9.3 ESTABLISHMENT

It turns out that all source data sets for employment spells, CON, RAS, and BFL, have a number of persons with more than one record because he has been working in several establishments, either at the same time or consecutive. The master spell has

the LBNR of the most important establishment, and the wage and hours variables are aggregated.

Chapter 4

Comparison of spell data and other labour market statistics

10 Employment

11 Employment and firms

12 Unemployment

13 Out of labour force

14 DREAM

Chapter 5

Dynamics over the business cycle

15 Termination of spells

16 New spells

17 Flows of jobs

18 Flows of workers

Chapter 6

A descriptive model for Ins and Outs of the labour market