



Department of Economics and Business – University of Aarhus

New Firm and Workplace Identifiers in the LMDG Data

Henning Bunzel

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 labour market state in the last week of November and no information about the states in between.
- ★ Using a number of datasets it is possible to generate a longitudinal dataset for persons' transitions between different **states** in the labour market.

Introduction

- ★ **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 labour market are
 - E, Employed
 - U, Unemployed
 - N, Not in the labour 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.

Introduction

- ★ For each source we construct a RAWSPELLS 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 SPELLS 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 SUBSPELLS is constructed containing data on any conflicting periods, or for an aggregated period like year.
- ★ Other SUBSPELLS are constructed for special *states* E, U, and N.
- ★ In the new version of LMDG spells we have one version with FIRM as units and one version with WORKPLACES as units.

ERS

- ★ Up to now the firm identifier has been ARBGNR (1984-2007) and SENR (2007-2018). Both are tax/duty reporting units.
- ★ Using different data sources these identifiers have been mapped to CVRNR which was the identifier used in f.ex. FIRM and FIRE datasets.
- ★ Someone, not documented, has replaced ARBGNR with SENR in dataset CONESR_RAS,
- ★ ESR datasets are now available and it is assumed that all SENR in other datasets should be included in this dataset. But that is *NOT* the case.
- ★ Using these SENR in the source datasets it is expected that the LMDG spells will have less E-E transitions due to just change of firm identifiers.

Firm Identifiers

- ★ The two major firm identifiers are SENR and CVRNR.
- ★ 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 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 own one or more CVRNR. One CVRNR is marked as being carrying data about an ECONOMIC UNIT. Which data? Is it the CVRNR used in FIRM and FIRE?
- ★ 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 "Firm Warehouse" has constructed a variable OK_OT_ID which is an over-time identifier for the economic units from 2000.

OK Units

- ★ 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,
 - Financial structure, patents, customer section, production
- ★ An *Enterprise* is an organisational unit producing goods or services which has a certain degree of autonomy in decision making.
 - ★ For tax, liability and organisational reasons an economic unit may choose to organise its activities like sales, and different production in two or more legal units, use more than one fiscal registration.
 - ★ *Group*, if legal units have the same management they are considered one economic unit.

These are the EU Statistical Manual definitions. How are they implemented in Statistics Denmark? Unkown! One CVRNR has been

OK Units

defined as the one having data about the enterprise. What data? Datawarehouse has an over_time identifier for OK for 2000-2022. Can it be extended back to 1984? The experian group data is very inconsistent. And DS data cannot be compared to external data.

Workplace Identifier

- ★ Until 2008 a physical workplace is identified by AAR, ARBGNR, ARB_STK in datasets CONESR, RAS and by 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 defines the workplace to be a physical or a fictitious workplace.
- ★ IDA has constructed an over-time physical workplace identifier LBNR. LBNR is 00000000 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.
- ★ IDAS only has physical workplaces but CONESR, BFL, IDAN, and RAS has both physical and fictitious workplaces. CONESR and BFL only have jobs occupied by employed workers.

Workplace Identifier

- ★ Data on Privately Owned Firms and holidays must be obtained from other sources. Yearly data.
- ★ ARB_NR is an ESR version of CVR Register variable PRODNR. ARB_NR and LBNR are both over-time identifiers, but how is ARB_NR defined? The project is trying to compare them.

ESR datasets

- ★ The ESR dataset is a special ESR sample made for LMDG.
- ★ It was supposed to be event datasets, but is a mixture of yearly and event datasets. The yearly datasets are constructed from different versions of ESR (AJOUR_DATO) made for statistical datasets.
- ★ The design of the ESR database and the creation of versions are not documented.
- ★ ESR_ARB, workplace
- ★ ESR_JUR, early reporting units
- ★ ESR_SE_CVR_EVENTS, firm and reporting units
- ★ ESR_SE_CVR, relation between SENR and CVRNR
- ★ ESR_OK_ARB, relation units
- ★ ESR_OK_CVR, relation economic-legal units

ESR datasets

- ★ An Rel spell maps a SENR to zero, one, or more CVRNR for a specific period and the key is (SENR, CVRNR, REL_START_DATO)
- ★ There exist start- and end dates for both SENR and Rel spells. It is assumed that Rel spells are contained within an SE spell, but this not always the case. Therefore, some editing of the dates has taken place

Over-time Identifiers

- ★ Statistics Denmark has created an over-time firm and enterprise identifier for the period 2000-2022. The main criterium is the maintenance of workers. They have chosen some rules which probably differ from LBNR in IDA.
- ★ Spell has constructed an over-time identifier for a firm based on LBNR for the period 1984-2021.
- ★ The two identifiers will be compared and merged into one identifier used for a new version of E-E spells. The other version is based on OKNR/CVRNR.

FIDA

- ★ FIDA used to move a group of workers from one firm to another because reporting of wages was not reported by correct firm app. 80,000 persons were moved according to an internal memo.
- ★ From 20?? construction of FIDA was changed and the content does not correspond to the Data Warehouse Documentation.
- ★ FSE has informed us that PNR-CVRNR is correct in IDAN.
- ★ The Spell project tries to confirm it. It must be the case that a group of workers with key (CVRNR, ARB_NR) in BFL cannot be found in IDAN with this key but with another key.
- ★ But the data is very noisy

FIDA

Number of matches in datasets IDAN-BFL

Key	Match	In BFL, not in IDAN	Not in BFL, in IDAN
PNR	4,615,824	319,910	119,775
CVRNR	379,030	35,677	279,234
ARB_NR	474,930	48,609	300,523
PNR, CVRNR, ARB_NR	20,035,732	3,548,997	834,221

- ★ Privately owned firms (owner, assisting spouse) do not exist in BFL.
- ★ Non wage earners do not exist in BFL and IDAN?
- ★ Only workers with permanent address in Denmark ultimo November are included in IDAN.

Some results with the new identifiers

Number of observations edited in SPELL_FIRM_ID

Code	Condition	No of observations	Action
0	Total	25,287,909	
1	SE spell dates missing	897,154	assigned Rel 2,463,893spell
2	START_AAR < 1980	4,098,055	START_AAR is assigned 1980
3	SLUT_AAR > LAST_YEAR	11,810,300	SLUT_AAR is assigned LAST_YEAR
4	REL_SLUT_AAR > SLUT_AAR	22,268,985	REL_SLUT_AAR
5	REL_START_AAR	4,013,766	REL_START_AAR is assigned 1980
6	REL_SLUT_AAR > SLUT_AAR	9,837,485	REL_SLUT_AAR is assigned SLUT_AAR
7	REL_START_AAR < START_AAR	25,646	REL_START_AAR is assigned START_AAR
8	CVRNR is missing for first spell	657,796	Is assigned CVRNR from next spell
9	CVNR missing	409	CVRNR is assigned SENR
10	Wrong spell dates in ESR	2,463,893	Years added for each year in raw employment
11	SENR does not exist in ESR	25,572	CVRNR assigned SENR and years added for each year in raw employment

Some results with the new identifiers

Number of matches SPELL_FIRM_ID – FIRM			
Key	Match	Not in FIRM	Not in SPELL_FIRM_ID
PNR, CVNR	6,370,543	5,342,218	466,083
CVRNR	1,041,775	270,832	125,074

Some results with the new identifiers

Further information can be found on the following web pages which are in progress: [LMDG documents](#)