

LMDG Dataset Documentation

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

Henning Bunzel

1 ESR_SE_CVR_EVENTS, Relations between *SENR* and *CVRNR*, events data

DSECONAU Updated: 03-04-2022 Special ESR version to ECONAU. *AJOUR_DATO* is 03-04-2022. Last full year is 2021.

ECONAU Updated: 03-04-2022 Special ESR version to ECONAU, last year 2021. Delivered to LMDG project 702728.

1.1 Description

ESR datasets The *ESR* datasets is a special ESR sample made for *LMDG*. It was supposed to be spell datasets, but it is a mixture of yearly and spell 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. The current version has the latest *AJOUR_DATO* as 03-04-2022.

The following datasets contain the core units, see table 1.

The following datasets define relations between the core units, see table 2.

The following ECONAU datasets contain supplementary variables which can be merged on the relevant core dataset, see table 3.

Dataset ESR_SE_CVR_EVENTS The dataset used for input is 702728\RAW_DATA\VIEWS\ESR_UDV_CVR_SE_TOTAL which is an event dataset created 03-04-2022

Table 1 Core Units

ECON name	LMDG name	Contents	Type	Spell Period	Remarks
ESR.UDV_ARB	<i>ESR_ARB</i>	Workplace	Yearly	[2000-2020]	Earliest spell date is 01-01-1983
ESR.UDV_JUR	<i>ESR_JUR</i>	Early reporting units	Yearly		
ESR.UDV_CVR_SE.TOTAL	<i>ESR_SE.CVR_EVENTS</i>	Firm and reporting units	Events	[2000-1921]	Earliest SE spell date is 01-01-1900. Earliest Rel spell date is 24-11-1848
	<i>ESR_SE.CVR</i>	Relation between SENR and CVRNR	Yearly	[2000-1921]	Created from <i>ESR_SE.CVR_EVENTS</i>
ESR.UDV_OK	<i>ESR_OK</i>	Economic units	Yearly		
ESR.UDV_KONKURS.TOTAL		Economic units, bankruptsy			

Table 2 Relations between Core units

ECON name	LMDG name	Contents	Type	Spell Period	Remarks
ESR.UDV_ARB_P	<i>ESR_ARB_P</i>	Relation units Workplace-Production	Yearly	[2000-2020]	
ESR.UDV_OK_ARB	<i>ESR_OK_ARB</i>	Relation units Economic-Workplace	Yearly	[2000-2020]	
ESR.UDV_OK_CVR	<i>ESR_OK_CVR</i>	Relation Economic-Legal units	Yearly	[2000-2021]	

Table 3 Supplementary variables, Core units

ECON name	LMDG name	Contents	Variables	Type	Spell Period	Remarks
ESR.UDV_ARB.ADRESSE						
ESR.UDV_ARB.BRANCHE						
ESR.UDV_ARB_P						
ESR.UDV_CVR.ADRESSE						
ESR.UDV_CVR.BRANCHE						
ESR.UDV_CVR.VIRK.FORM						
ESR.UDV_OK.ADRESSE						
ESR.UDV_OK.VIRK.FORM						

as an extraction from *ESR* defining *SENR* spells for (*START_DATO*,*SLUT_DATO*), and a relation between *SENR* and *CVRNR* with the start and end date for the Rel Spells (*REL_START_DATO*, *REL_SLUT_DATO*).

The key for a *SENR* spell is (*SENR*, *START_DATO*).

A Rel spell maps a *SENR* to zero, one, or more *CVRNR* for a specific period and the key is (*SENR*, *CVRNR*, *REL_START_DATO*).

A *SENR* spell and a Rel spell is still active if end date is equal to '31DEC9999'd.

Sequence for the jobs in *esr_se_cvr_events_2022*

- The job *_1_esr_se_cvr_events* in directory *esr_se_cvr_events_2022\ dan*.
 - It changes still-active-date to 31DEC9999.
 - It selects the date part of the date-time value in variable with type date.
 - It determines the range of dates.
 - It renames a number of variables.
 - It has 2,297,080 observations.
- The job *_2_esr_se_cvr_update* in subdirectory *dan* updates the *ESR_SE_CVR_EVENTS*.

- It changes illegal *CVRNR* (cryptized *cvrnr* blank values) to " ". All observations have a *CVRNR* values, but 323,722 observations have a value '*****9509' which most likely is a cryptized " " value. They are changed to " ", see the observations in dataset *data.ILLEGALE.CVRNR*.
- *_3_esr_se_cvr_events_cvr_miss*. Some observations with a missing *CVRNR* have a following spell with a *CVRNR*, see datasets *data.SENR_MISSING_CVRNR_x*. These 2 spells are concatenated with this *CVRNR*. For these observations *ED-ITS="9"*. The dataset now has 2,270,249 observations.
- The job *_4_standard_miss_program_se_cvr_events* in subdirectory *dan*. Check the result for missing values.
- The job *_1_esr_se_cvr_events_old_new* in subdirectory *test* compare 2018 dataset ved 2022 dataset. It is not run for this update because the structure is different and th quality is better
- The job *_2_dups_senr_rel_cvrnr_stat_dato* in subdirectory *test* test for duplicates of keys.
- The job *_3_rel_spell_not_in_se_spell* in subdirectory *test*.
- The job *_4_spell_datoer_senere_end_ajour_dato* in subdirectory *test*.
- The job *_4_esr_se_cvr_events_year* in subdirectory *dan* creates the datasets *DATA\YEARS* and *DATA\YEARS_TRUNC* with yearly start and end dates. It is assumed that a Rel spell must be contained with the SE spell. For many observations this is not the case and the job deletes and edit observations observations to make the dates consistent. See the table ??
- The job *_5_se_cvr_events_year_overlaps* in subdirectory *test* uses the dataset *YEARS* dataset to test if Rel spells are contained within SE spells and if Rel spells overlaps, and if the key (*SENR, CVRNR, REL_START_AAR*) is unique..
- The job *_desc_esr_se_cvr_events* in subdirectory *desc* provides descriptions of the dataset.

For some observations *START_DATO* is missing. Observations with missing start date all have *SLUT_DATO* before 31-12-1999. Some have a *CVRNR* but most have a missing *CVRNR*.

At any point in time a *SENR* can only be assigned to one *CVRNR*, but a *CVRNR* can be assigned many *SENR*. The *SENR* is assigned by the tax or custom authority when the unit is created. An economic unit can always change *CVRNR*.

323,722 *SENR* numbers are not assigned a *CVRNR*. It is typically an activity without employment, but in this dataset some *SENR* are missing because they were created before *CVR* was established in 1999. In the latest years about 200-250 observations have missing *CVRNR*. This is less than the number of units without employment, which according to DS is about 8,000. In the years earlier than 1995 the number of missing values for *CVRNR* is 500-15,000.

We use *CVRNR* to match other data sets, such as *FIRM*, *FIRE*, *BFL*, *CONESR* and in these datasets all units have employment.

All Rel spell observations must have valid values for *SENR*, *CVRNR*, *REL_START_DATO*, and *REL_SLUT_DATO* and the dataset *ESR_SE_CVR_EVENTS* has missing *CVRNR*

values, hence the Relation Spells data *YEARS* is created by selecting observations with valid values for *CVRNR*.

Remarks on the 2022 dataset *ESR_SE_CVR_EVENTS* .

- The Extraction date has value 03-04-2022.
- The dataset has 2,297,080 observations.
- 323,722 observations have missing value for *CVRNR*.
- 52,702 observations have missing values of *START_DATO* and *SLUT_DATO*. They are distributed over the entire period.
- The key is (*SENR*, *START_DATO*). It has 56,440 duplicates. The *SENR* are assigned more than 1 *CVRNR* over time.
- The dataset has 2,268,657 unique *SENR*.
- The dataset has 1,883,904 unique *CVRNR*.
- 1,882,885 observations have *SENR* equal to *CVRNR*.
- 1,436,135 *SENR* numbers are still active.
- The range of *START_DATO* is [01-01-1900, 25-03-2022].
- Rel spells have the key (*SENR*, *CVRNR*, *REL_START_DATO*).) The key is unique.
- The range of *SLUT_DATO* for terminated *SENR* spells is [18-1-1901, 31-12-2022].
- 2,297,075 observations have a terminated SE spell, but the REL spell is still active.
- 26,767 observations have SE spell start date later than REL spell start date.
- 77 observations have SE spell start date later than REL spell end date.
- 77 observations have *SLUT_DATO* later than extraction date *AJOUR_DATO*. They all have a date later in 2022 and all *REL_SLUT_DATO* are 31-12-9999. This not a problem as the last year used in spells is 2021.

Remarks on the 2022 Rel spell datasets *YEARS* and *YEARS_TRUNC* .

- The Rel spell dataset has 1,947,326 observations .
- See table ?? to see number of observations deleted and edited in the creatin of dataset *YEARS* from *ESR_SE_CVR_EVENTS*.
- In the Rel spell dataset all observations have valid values for *SENR*, *CVRNR*, *REL_START_DATO* and *REL_SLUT_DATO*.
- The datasets are truncated to [1980,2021]
- In the *YEAR_TRUNC* dataset the *REL_SLUT_AAR* is set to *SLUT_AAR*.
- The datasets are used to create *ESR_SE_CVR*. The expected number of observations in *YEAR* is 32,346,279, and the expected number of observations in *YEAR_TRUNC* is 21,347,392.
- The key for Rel spells (*SENR*, *CVRNR*, *REL_START_DATO*) is unique.
- The key for Yearly Rel spells (*SENR*, *CVRNR*, *REL_START_AAR*) is unique.
- The key for Yearly Rel spells (*SENR*, *START_AAR*) has 43 duplicates. There are several (*SENR*, *CVRNR*) spells with gaps within a year.
- The Rel set has 1,946,001 unique *SENR*.
- The Rel set has 1,947,326 unique *CVRNR*.

- 1,373,593 Rel spells are active. The all have a closed SE spell. Maybe this is not an error but by design.
- The range of *REL_START_DATO* is [24-11-1848, 01-09]
- The range of *REL_SLUT_DATO* for terminated relations is [27-11-2001, 12-08-2021].
- 1,829,739 *CVRNR* have just 1 *SENR* assigned. The highest number of *SENR* assigned to a *CVRNR* is 473, see Table 4
- The Rel spells do not overlap.
- 26,766 observations have *START_DATO* later than *REL_START_DATO*.
- 7,679 observations have *SLUT_DATO* before *REL_SLUT_DATO*.
- 77 observations have Rel spell before SE spell.
- 0 observations have REL spell start date later than *AJOUR_DATO*.
- The above 3 problems with Rel spells should not effect the match of yearly economic datasets with Rel spell, but there will be Rel spells which maybe cannot be matched to economic datasets

Table 4 Frequency of *CVRNR* assigned 1,2,..., *SENR*

No of <i>SENR</i> assigned	Frequency of <i>CVRNR</i>
1	151,301
2	183,495
3	160,005
4	149,925
5	112,211
< 5	1,108,479

The LMDG dataset has been modified:

1. In the raw 2022 dataset all dates are day-time. They are all converted to day format yymmdd10.
2. 323,722 observations have the *CVRNR* value "xxxxxxx509" replaced with a missing value.

1.2 Remarks on Update

Errors and Questions, Update of LMDG version

- In the new dataset all dates are day-time. They are all converted to day format yymmdd10.
- In 2018 data a new *CVRNR* xxxxxxx509 is associated with 771,898 *SENR*. In 2022 data 323,722 observations have this *CVRNR*. They are changed to " ". This *CVRNR* is NOT a valid *CVRNR*, but most likely a wrongly encrypted blank or

illegal value. It has also showed up as a new *CVRNR* value in the current *RAS* version! It is important to get this error fixed.

- *DST* has informed that many *SENR* do not have start- and end-date. These existed before *CVR* was created in 1999. It has not been possible to obtain these dates. If the *SENR* has had the same owner then it is possible to assign a *CVRNR*. 200-250 current *SENR* are *VAT* units (foreign companies or Danish units without employment) Should it not be 8,000?.
- In the 2022 dataset a *SENR* spell is added. We assume that the Rel spell must be contained in the *SENR* spell. This is not the case, see above.
- All 31DEC4712 values for *SLUT_DATO* are changed to 31DEC9999.
- Rel spell dates have no missing values. 52702 SE spell dates have missing values. The rel start of these observations are distributed over all years, as late as 2022.
- Many observations with *CVRNR* missing still have rel spell dates.
- The key (*SENR*, *CVRNR*, *REL_START_DATO*) is unique.
- 26,766 observations have *START_DATO* later than *REL_START_DATO*.
- 7,679 observations have *SLUT_DATO* before *REL_SLUT_DATO*.
- 77 observations have Rel spell before SE spell.
- 1,375,593 observations have a closed SE spell but Rel spell is still active. Maybe this is not an error but by design.

ECONAU Variables in ESR_SE_CVR_EVENTS

Henning Bunzel

1 AJOUR_DATO, Date for sampling from ESR.

Dato for udtræk fra ESR.

1.1 *Beskrivelse*

AJOUR_DATO er den historiske dato, med hvilken *ESR* ønskes betragtet.

Det betyder, at kun rækker ajourført på eller før *AJOUR_DATO*, og som er ukorrigerede eller korrigerede efter *AJOUR_DATO* indgår i betragtede rækker.

1.2 *Description*

AJOUR_DATO is the historical date to be used in *ESR*.

It means that only rows which have been updated at or before *AJOUR_DATO*, and which are not corrected or corrected after *AJOUR_DATO* are part of the rows considere.

2 CVRNR, Encrypted legal unit identifier. Data break 1994

CVRNR, krypteret nøgle for juridisk enhed. Databrud 1994

FSE Name: ARBGNR, ARBGNR8, JURNR, CVR_NR

2.1 Beskrivelse

CVRNR er et entydigt identifikationsnummer, som alle juridiske enheder tildeles i *CVR*, *Centrale Virksomheds Register*. Nummeret skal bruges til at identificere virksomheden entydigt over for alle offentlige myndigheder, og det bruges også i meget vid udstrækning til at identificere virksomheden over for andre virksomheder.

En økonomisk enhed kan have flek *CVRNR* eller skifte *CVRNR*. De fleste økonomiske enheder har samme *CVRNR* i hele levetiden. For bestemte virksomhedstyper kan *CVRNR* genbruges.

I ECONAU data er *JURNR* omdøbt til *CVRNR* selvom de ikke er et *CVRNR*. Perioden afhænger af datasættet. I nyere version af data sæt has FSE converteret *JURNR* og *SENR* til *CVRNR*.

I ECONAU findes *CVRNR* i følgende datasæt:

- *AKM* 1985-2002
- *BFL* alle år fra 2008
- *CONESR* 1980-2007
- *LOEN* alle år fra 2002
- *MOMS* 1990-2000
- *RAS* fra 2006
- *ESR_SE.CVR* alle år fra 1989
- *UHDI*, *UHDM* alle år fra 1993

I ECONAU findes *CVRNR* i følgende datasæt:

- *BFL* alle år fra 2008
- *MOMS* 2001-2009
- *FIKS* alle år fra 2001
- *RAS* fra 2008
- *ESR_SE.CVR* alle år fra 1989
- *UHDI*, *UHDM*, alle år fra 1993

Reelt aktive firmaer I *CVR* der danner grundlag for *ESR*, registreres alle firmaer, der er registreringspligtige i henhold til skatte- og afgiftslovene eller i henhold til selskabslovgivningen, uanset aktivitetsniveauet. Det betyder, at selv den mindste aktivitet, selvom den ikke er egentlig erhvervsmæssig, resulterer i en registrering af et firma. I firmastatistikken medtages kun reelt aktive firmaer. Et firma anses for at

være reelt aktivt når det har ansatte, der leverer, hvad der samlet svarer til mindst et halvt fuldtidsårsværk.

2.2 Description

The *CVRNR* is a unique identifier given to all legal units in *CVR*, *Central Firm Register*. The identifier is used to identify the *firm* uniquely to all public authorities and it is also widely used to identify the *firm* to other *firms*.

The *CVRNR* is unchanged during the entire life cycle of the legal unit. *CVRNR* for closed *firms* may be reused for certain types of *firms*.

In ECONAU data *JURNR* is renamed to *CVRNR* even though they are not a *CVRNR*. The period depends on the dataset. In recent version of datasets FSE has mapped *JURNR* and *SENr* to *CVRNR*.

I ECONAU *SENr* is found in the following datasets:

- *AKM* 1985-2002
- *BFL all years* from 2008
- *CONESR* 1980-2007
- *LOEN all years* from 2002
- *MOMS* 1990-2000
- *RAS all years* from 2006
- *ESR_SE_CVR all years* from 1989
- *UHDl, UHDM all years* from 1993

I ECONAU *CVRNR* is found in the following datasets:

- *BFL all years* from 2008
- *LOEN all years* from 1999
- *MOMS* 2001-2009
- *FIKS all years* from 2001
- *RAS all years* from 2008
- *ESR_SE_CVR all years* from 1989
- *UHDl, UHDM, all years* from 1993

Really active firms In *CVR* which is the basis of *ESR*, all firms are registered which are subjects to registration according to the taxes laws or to the company legislation, regardless of the level of activity. It means that even the smallest activity results in a registration of a firm, even though it is not really professional. The firms statistics only includes really active firms. A firm is regarded as being really active when the employees produce what corresponds to at least half of full-time man-years.

3 SENR, A firm account number for tax and custom payments

Virksomhedskontonummer til skat- og toldbetalinger

3.1 Beskrivelse

Et *SENR* er et ottecifret nummer, der blev indført i foråret 1985 af det daværende Statsskattedirektorat i forbindelse med etableringen af *SE*, *Stamregister for erhvervsdrivende*.

Nummeret blev anvendt som identifikation af juridiske enheder som for eksempel virksomheder og selvstændige erhvervsdrivende i forbindelse med afregning af kildeskat. Senere hen blev det også brugt til afregningen og registreringen af moms, AM-bidrag og A-skat, tilhørsforhold til barselsudligningsordning, periodeoplysninger og evt. produktionsenhedsnummer mv. Den indberetningspligtige har adgang til oplysninger i *eIR*, *Danmarks Statistiks E-Indkomst Register* som denne selv har indberettet, eller som er indberettet på den indberetningspligtiges vegne.

CVR, *Centrale Virksomheds Register* blev sat i kraft i oktober 1999. Generelt skal *CVRNR* anvendes til identifikation af virksomheden, idet det alene er *CVRNR*, der udgør virksomhedens entydige identifikation.

Nogle virksomheder har flere *SENR* hos *SKAT*. Der er ingen begrænsninger, og det er alene op til virksomheden, hvad den finder mest praktisk. Til eksempel kan der anvendes et til afregning af moms, og et andet til afregning af A-skat m.v.

I langt de fleste tilfælde har virksomheder kun et *SENR*, som så er identisk med *CVRNR*.

I nogle indberetninger bruger flere *CVR* enheder et fælles *SENR* afregningsnummer, såkaldt fællesafregning.

ESR er Danmarks Statistik register for virksomheder. Datasættet *ESR_SE_CVR* viser, hvilket *CVRNR* et *SENR* er tilknyttet på et givet tidspunkt.

I mange datasæt er *SENR* benyttet i indberetninger, men i de senere år har *DST* tilføjet *CVRNR* eller erstattet *SENR* med *CVRNR*. Det kan medføre flere observationer med samme nøgle (*AAR*, *CVRNR*).

I *ECONAU* findes *SENR* i følgende datasæt:

- *AKM* 1985-2002
- *BFL* fra 2008, alle år
- *CONESR* 1980-2007, alle år
- *LOEN* fra 2002
- *MOMS* 1990-2000
- *RAS* fra 2006
- *ESR_SE_CVR* fra 1989, alle år
- *UHDI*, *UHDM* fra 1993, alle år

I *ECONAU* findes *CVRNR* i følgende datasæt:

- *BFL* fra 2008, alle år
- *LOEN* fra 1999, alle år
- *MOMS* 2001-2009
- *FIKS* fra 2001, alle år
- *RAS* fra 2008
- *ESR_SE_CVR* fra 1989,
- *UHDI*, *UHDM*, fra 1993 alle år

3.2 Description

A *SENR* is an 8-digit number which was introduced in Spring 1985 by the then State Tax Directorate in connection with the creation of the *SE, Master register for traders*.

The number was used as identification of legal units as for instance firms and self-employed in connection with withholding of tax. Later on it was also used for registration and payment of VAT, Labour market contributions and income taxes, relation to contributions to maternity leave payments, periodical information and maybe production unit number etc. The reporting unit has access to information in *eIR, Danmarks Statistiks E-Indkomst Register* which has been reported by itself or on behalf of the reporting unit.

CVR, *Central Firm Register* was initialized in October 1999. Generally *CVRNR* is an identifier of a legal unit. *CVRNR* is a unique identifier of an economic unit, named firm or enterprise. Later, *OKNR, Identifier for Economic Units* was introduced as an identifier of the economic unit. An economic unit may have several *CVRNR*. In many datasets one of the *CVRNR* is selected as identifying the economic unit. Often it is the firm that decides to have several *CVRNR*, and several *SENR* associated each *CVRNR*. At any time one *SENR* can only be associated one *CVRNR* and some firms have more *SENR* at *SKAT*. The registration of *SENR* can be done for different purposes, for example one number can be used for withholding of VAT, and another for withholding of A-taxes mv.

In the vast majority of cases firms have only one *SENR* which then is identical to *CVRNR*.

In some data several *CVRNR* units use a common *SENR*, a so-called common settlement.

ESR is the Statistics Denmark register for firms. *ESR_SE_CVR* shows which *CVRNR* a *SENR* is associated at a given time.

In many datasets *SENR* is used in reports but in recent years *DST* has added *CVRNR* or replaced *SENR* with *CVRNR*. It may cause more observations with the same key (*AAR*, *CVRNR*.)

I ECONAU *SENR* is found in the following datasets:

- *AKM* 1985-2002
- *BFL* from 2008, **all years**
- *CONESR* 1980-2007, **all years**

- *LOEN* from 2002
- *MOMS* 1990-2000
- *RAS* from 2006
- *ESR_SE_CVR* from 1989, **all years**
- *UHDI, UHDM* from 1993 **all years**

I ECONAU *CVRNR* is found in the following datasets:

- *BFL* from 2008, **all years**
- *LOEN* from 1999, **all years**
- *MOMS* 2001-2009
- *FIKS* from 2001, **all years**
- *RAS* from 2008
- *ESR_SE_CVR* from 1989, **all years**
- *UHDI, UHDM*, from 1993 **all years**

SE_CVR_REL_FRA_DATO not found SE_CVR_REL_FRA_DATO not found SE_FRA_DATO
not found SE_FRA_DATO not found