

Easing into longitudinality

Assigning statistical and administrative information to firm identities over time

Business Statistics Data Warehouse
CBS

27 November 2025

Julie Cathrine Krabek Sørensen (jck@dst.dk)

Kalle Emil Holst Hansen (khs@dst.dk)



Introduction

- Aim: Supporting studies on questions such as ‘What is an enterprise?’, ‘How do enterprises change over time?’, and ‘How can enterprise performance be measured?’
- Advance: Facilitating work on enterprise identity identification (longitudinality) and statistical information (CVR data)
- Topics of the presentation
 - Demographic events in general
 - Tracking enterprise identity ID and CVR over time
 - Aligning enterprise identity ID with statistical information found at the CVR level
 - Model selection
 - Conclusion and questions

Firms as entities: four pillars

- Ownership
- Activity
- Geographical location

- Employees

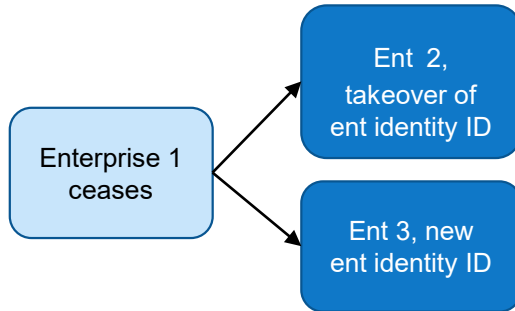
Relating the four pillars to the CVR and enterprise identity

- CVR entity (legal unit)
 - Defined as an administrative legal unit with legal obligations
 - Ownership (separate legal status)
 - Activity (related to total value creation)
 - Geographical location (CVR unit, regardless of legal units)
 - Employees (labour input producing the economic activity)

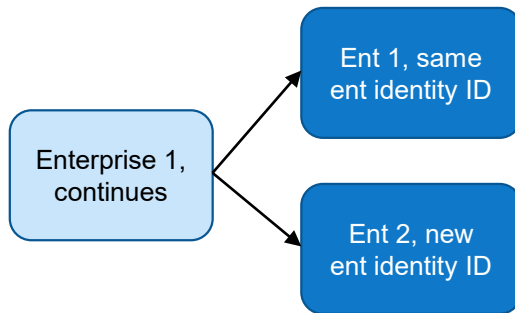
- Enterprise identity (purely a statistical unit)
 - Defined as the core (majority) set of employees, or local unit(s) related to the legal unit (CVR entity)
 - Employees (individuals employed at local unit(s) and employee changes)
 - Ownership (relationship between local unit(s) and CVR)
 - Activity (local unit production related to CVR activity)
 - Geographical location (local unit(s))

Demographic events

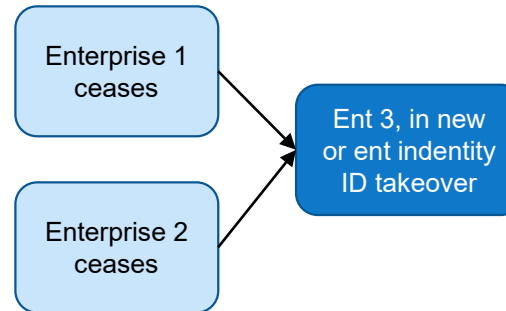
1. 'Split-ups'



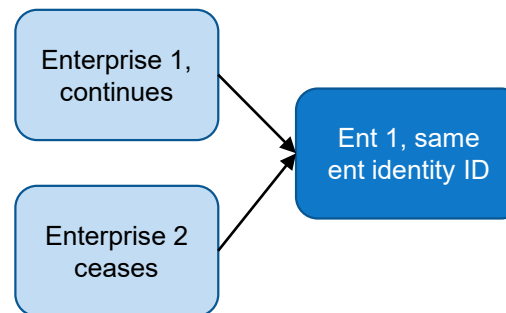
2. 'Spin-offs'



3. 'Mergers'



4. 'Takeovers'



Three scenarios when assessing demographic events of enterprises over time e.g. in a panel study:

Case 1:

$ID = ID$
 $CVR = CVR$

(with or without demographic events leading to no changes in the relationship between ID and CVR)

Case 2:

$CVR = CVR$
 $ID \neq ID$

(change in relation between ID and CVR)

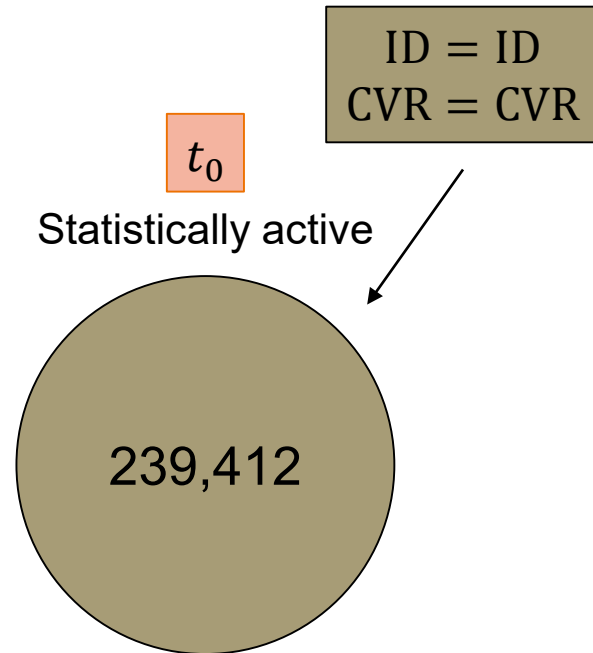
Case 3:

$ID = ID$
 $CVR \neq CVR$

(change in relation between ID and CVR)

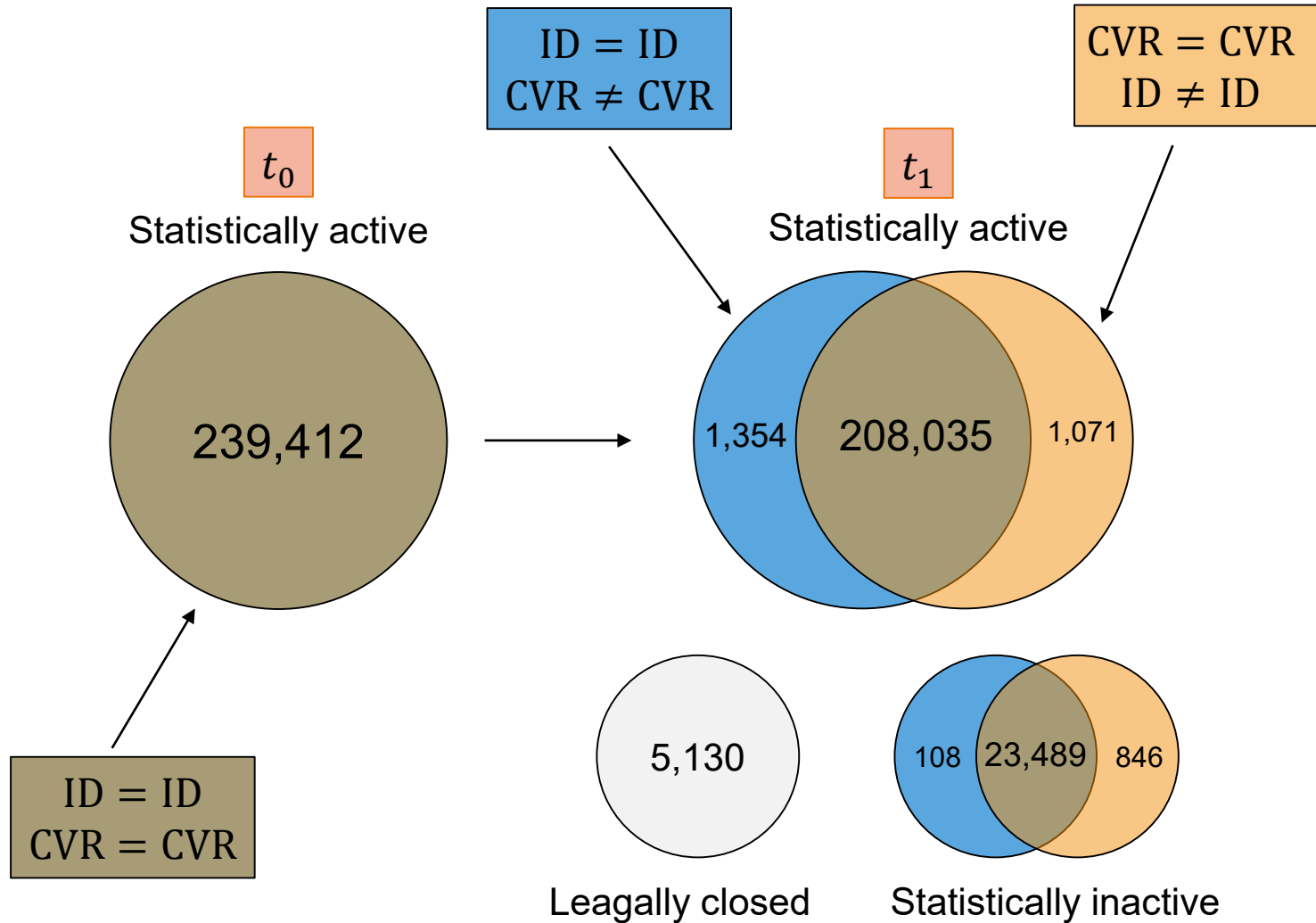
Following Enterprise ID and CVR over time

- Accounts Statistics for Non-Agricultural Private Sector (FIRE) in 2019



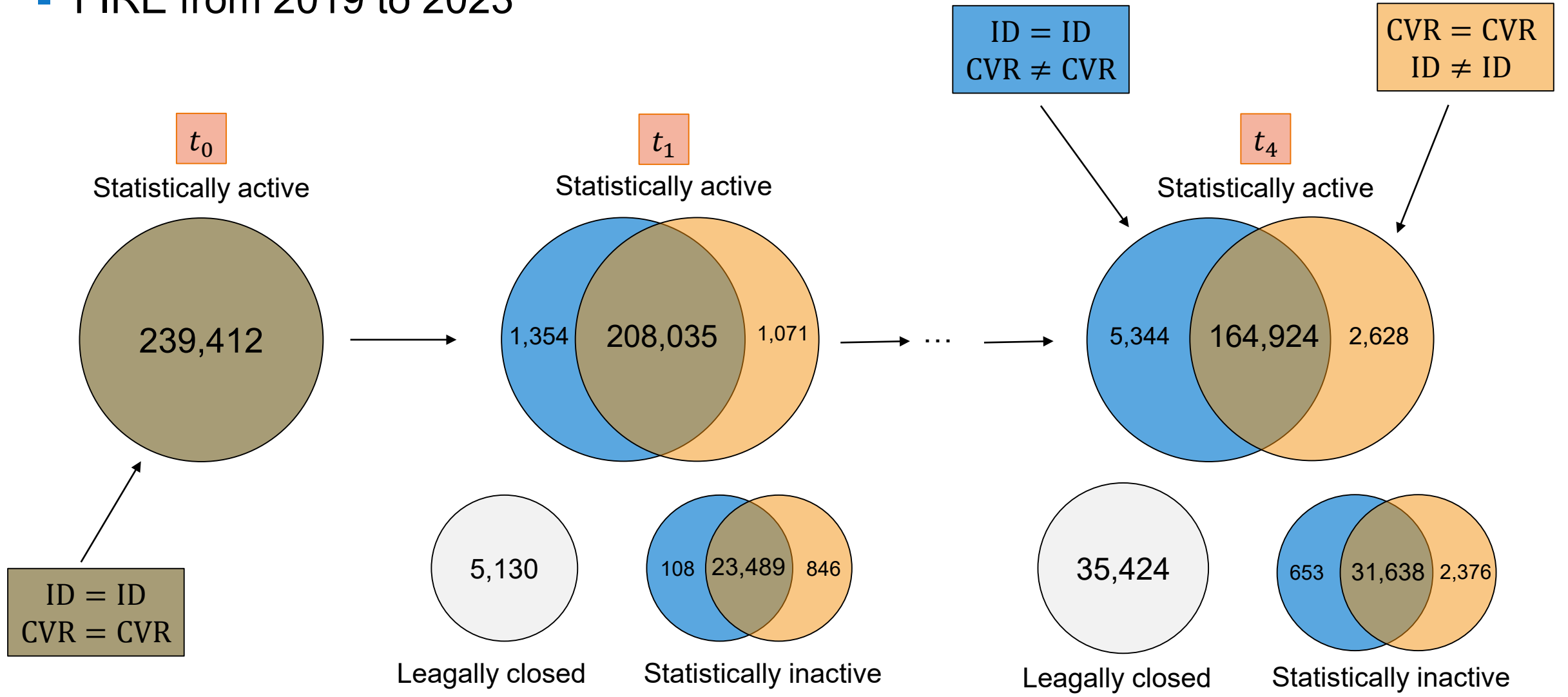
Following Enterprise ID and CVR over time

- FIRE from 2019 to 2020



Following Enterprise ID and CVR over time

- FIRE from 2019 to 2023



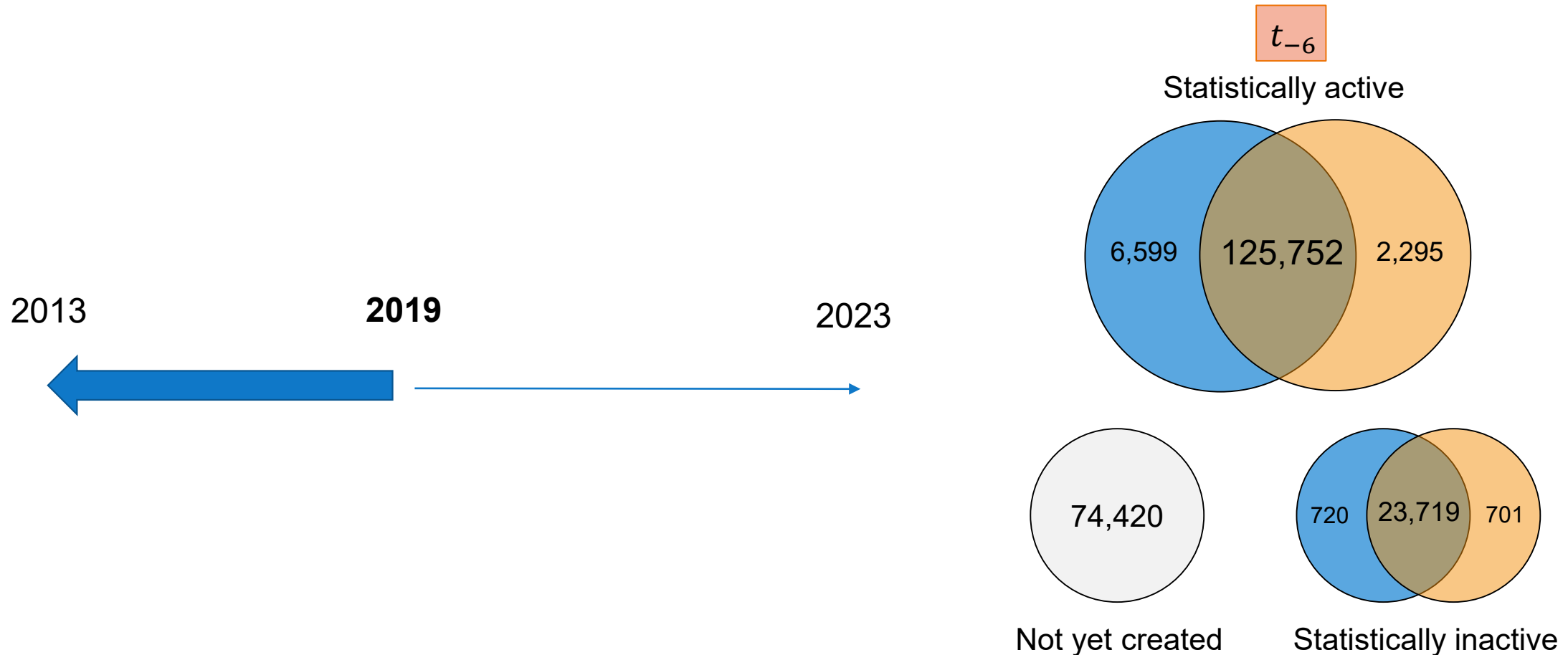
Statistical Information

- FIRE from 2019 to 2023, statistically active

	ID = ID CVR ≠ CVR	ID = ID CVR = CVR	CVR = CVR ID ≠ ID
No. enterprises (t_4)	5,344	164,924	2,628
Full-time employees (t_4)	33,624	1,249,858	28,287
Full-time employees (t_0)	26,703	1,231,304	35,544

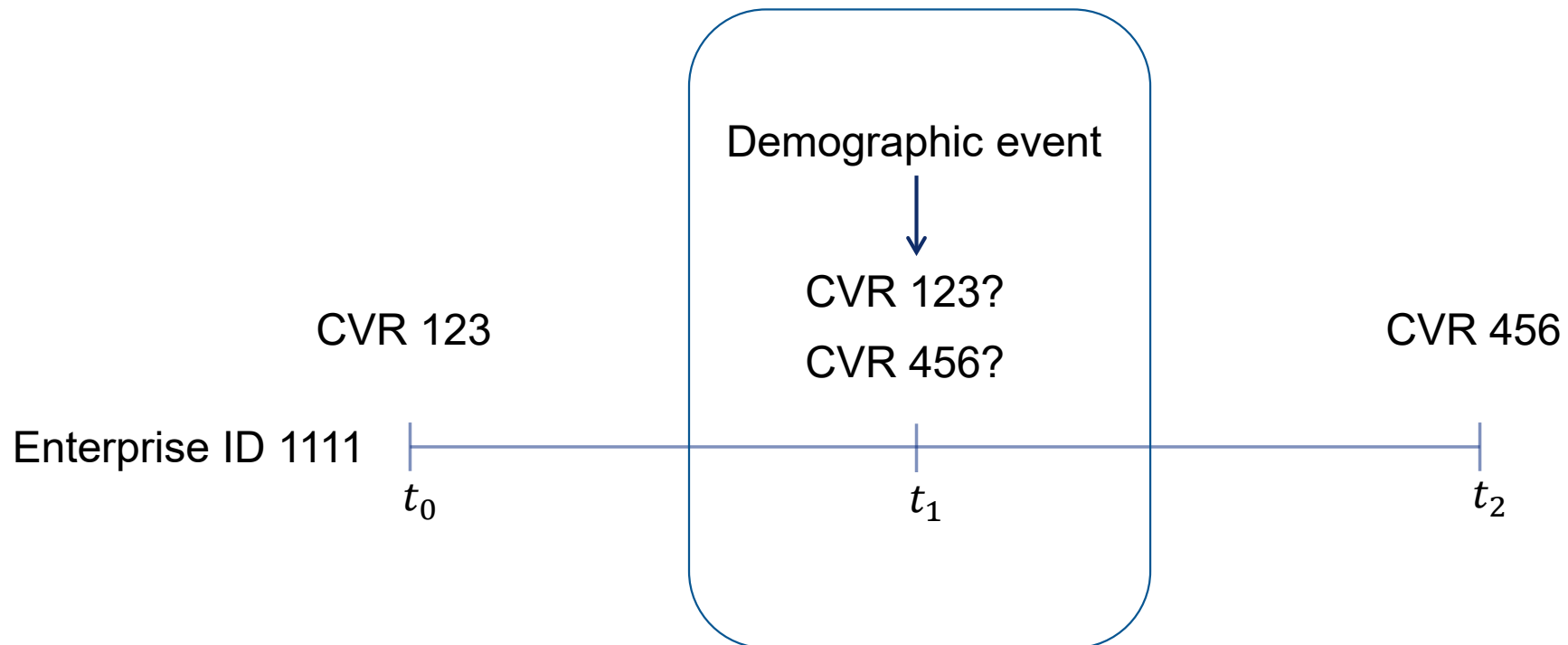
The long perspective

- Applies equally going back in time: FIRE 2019 population



Reconciling and aligning enterprise identity (ID) and statistical information (CVR) – Unit Selection

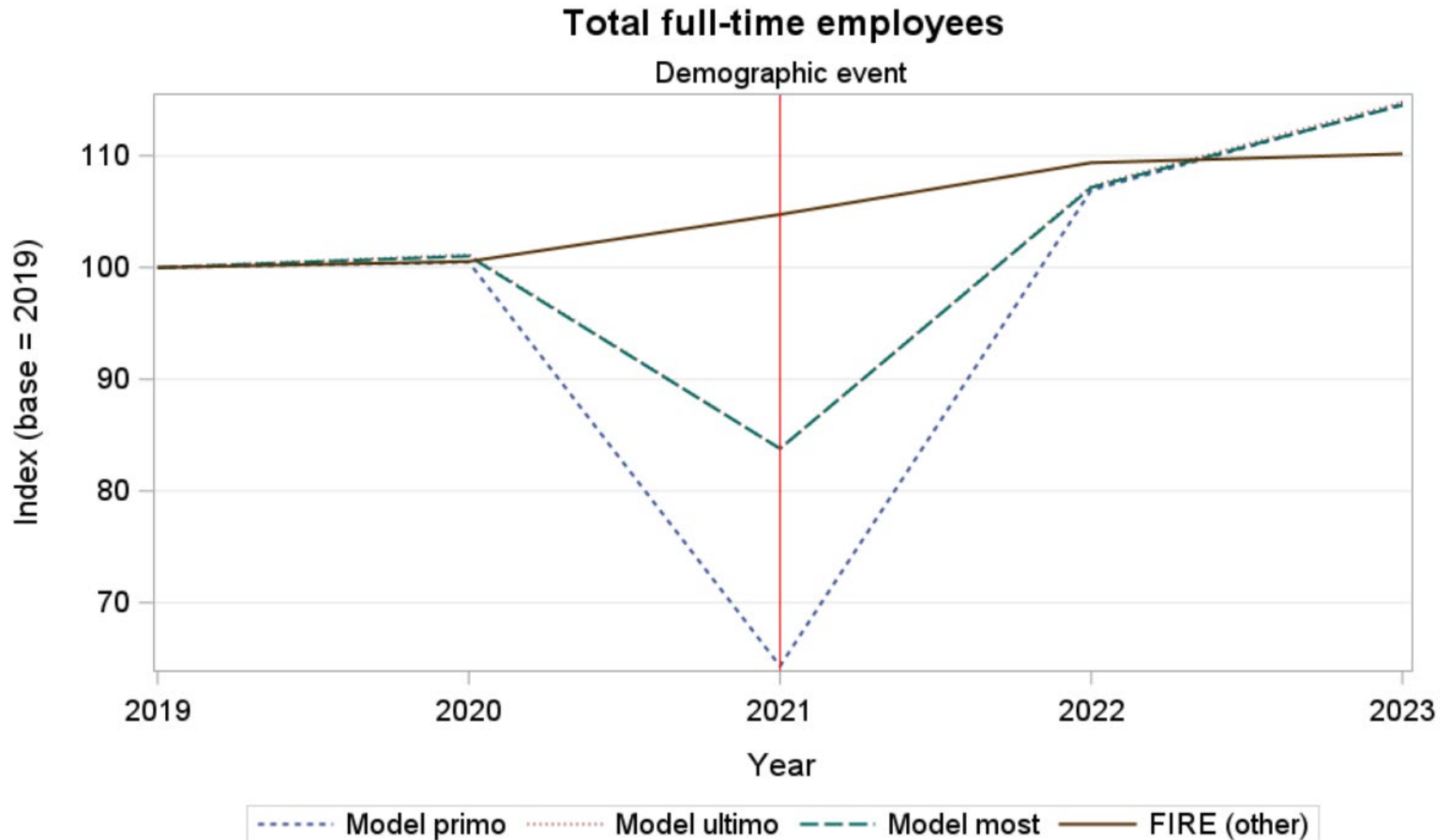
- Issue: Enterprise identity ID have no statistical information
- Question: Which CVR (and its statistical information) should be assigned to an enterprise identity ID in a year with demographic events?



Models Tested

- Three selection models tested to assess the effect of different ways to determine the relationship between enterprise ID and CVR number in a reference year
 - Primo (as recorded 1 January)
 - Combination enterprise identity ID + CVR(1/1)
 - Ultimo (as recorded 31 December)
 - Combination enterprise identity ID + CVR(31/12)
 - Most of the year (days)
 - Combination enterprise identity ID + CVR(longest relationship in a reference year)

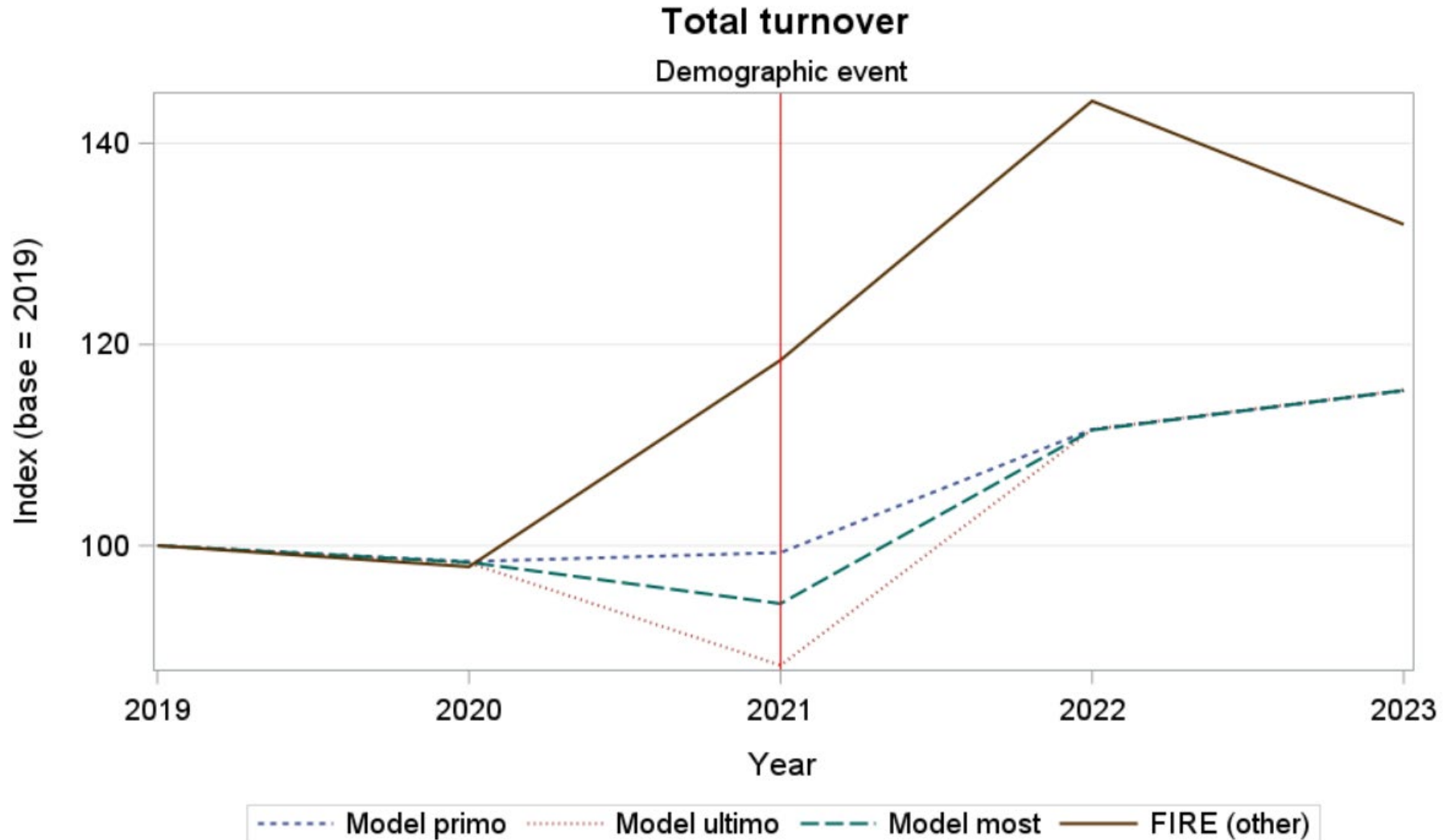
Demographic events affect the development of enterprises, especially in the years around the event



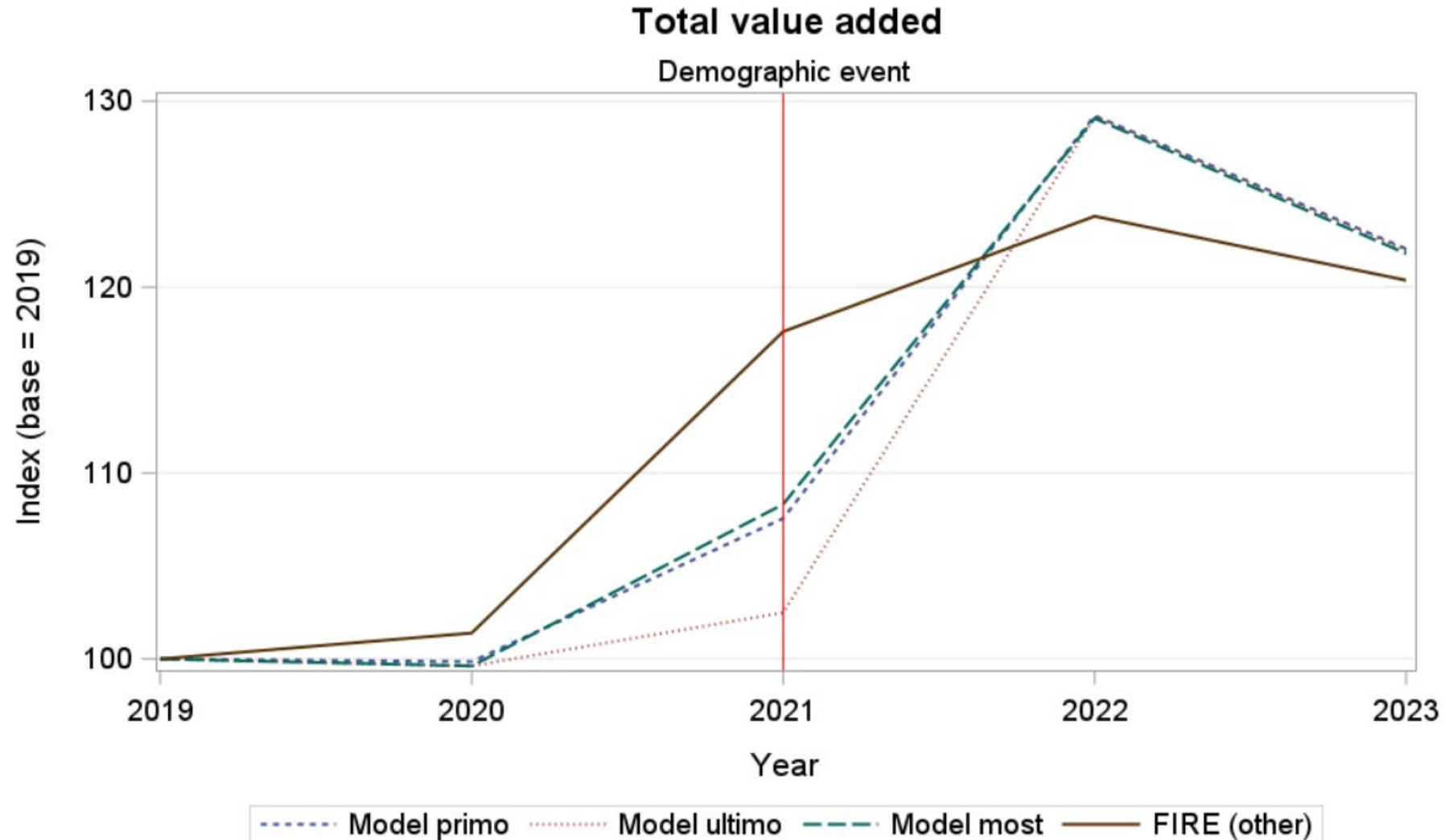
1.242 Enterprise IDs in all the models every year (change in relation)

158.256 Enterprise IDs in all the models every year (no change in relation)

Demographic events affect the development of enterprises, especially in the years around the event



Demographic events affect the development of enterprises, especially in the years around the event



Public available syntax (and dataset)?

- Primo (as recorded 1 January)
- Ultimo (as recorded 31 December)
- Most of the year (days)
- Demo (Indication of demographic event at enterprise identity ID level)

Year	ok_ot_id_primo	ok_ot_id_ultimo	ok_ot_id_most	cvr_nr_primo	cvr_nr_ultimo	cvr_nr_most	demo	
2020	A	A	A	111	111	111	0	ID = ID, CVR = CVR
2020	B	B	B	222	222	222	1	
2020	C	C	C	333	444	444	2	ID = ID, CVR ≠ CVR
2020	D	Z	D	555	555	555	2	CVR = CVR, ID ≠ ID
2020	E	E	E	777	777	777	0	
2020	F	F	F	888	888	888	0	ID = ID, CVR = CVR
2020	G	G	G	999	999	999	0	

Reference year

Used to track across years

Used to assign statistical information

Demo Event?

Conclusions

- Using enterprise identity IDs (not legal units – CVR) allows us to follow more enterprises (microdata units) over long timespans
 - + Keeps more employment data
 - + Holds the potential to eliminate ‘false positives’ from the data (CVR = CVR, but ID ≠ ID)
- The new syntax makes available three ways of assigning statistical information to enterprises (ID) in years with demographic
 - + Using the ‘most days’ selection model seems to create the least variance compared to the FIRE (other) population
 - but is that good thing?
 - + It produces a dataset that ‘untangles’ the relationship between enterprise identity ID and CVR in long registers and puts it into a yearly data structure
 - Simplifies complicated data structures (relational data)
 - Makes choices for the user to avoid, for instance, double-counting statistical information

Questions and comments

What are your thoughts on the approach?

Any suggestion for improvements?

Would making such syntax available to microdata users be of interest?