How many UK companies are *really* active?

A novel data tool for better business insights

Arathi Bhaskar, Yannis Galanakis and Nik Wolf
Office for National Statistics and ESCoE

20 May 2026

Motivation

Large differences in coverage between publicly available business statistics

2.8m	**>6m**	**2.3m**
active enterprises	incorporated companies	VAT-registered businesses
ONS Business Demography	**Companies House**	**HMRC**

Conceptual differences in how legal registration, statistical coverage, and economic activity are measured

What do we do?

In a nutshell

How many UK companies are *really* active? What is their economic footprint?

1

Construct the LBD Companies House Spine

1. linking table between **LBD** and **CH** identifiers, and
2. reproduction of national business statistics

2

Measurement Audit of UK business population

Assess coverage and accuracy of the UK business population across administrative sources

3

Match Characteristics

What distinguishes companies that match vs. do *not* match between IDBR and Companies House?

Data: LBD Companies House Spine

Longitudinal Business Database (LBD)
Database of longitudinally linked business references

Complete Linkage

Uses UK Inter-Departmental Business Register
Links **all representations** of each business entity:
Administrative VAT • PAYE • Companies House Registration Number
Statistical and Observation: Enterprise Groups • Enterprise • Reporting Units • Local Units

Result: Single longitudinal record per observation unit with all administrative and statistical identities linked

Enhanced coverage

Longitudinal and timely coverage

Four core products

One of our core products is being used via the Secure Research Service and the UKDS

Companies House Spine

Unit of observation

Enterprise Group (wowref) **x**
Enterprise (entref) **x**
Company (CRN)

Frequency: quarter

Eligibility

- *IDBR*:
  (1) VAT > threshold;
  (2) PAYE
- *Companies House* incorporated businesses: CRN

EXCLUDES: sole-traders and freelancers
[unless they meet the VAT threshold]

Sources

1. LBD L-WE (links ENTREF to WOWREF)
2. IDBR-processed *linking* tables

Challenge

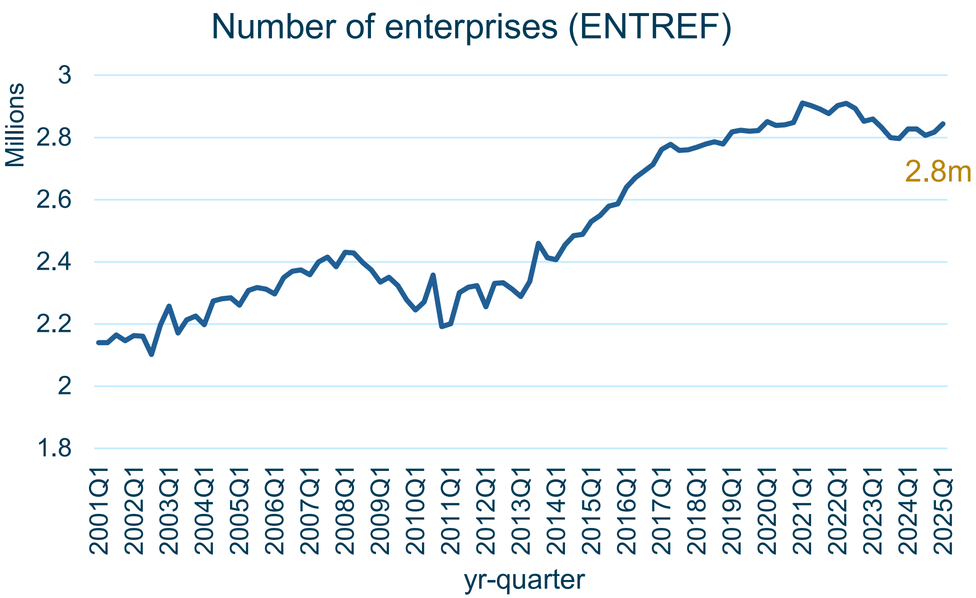
Coverage due to company activity criteria

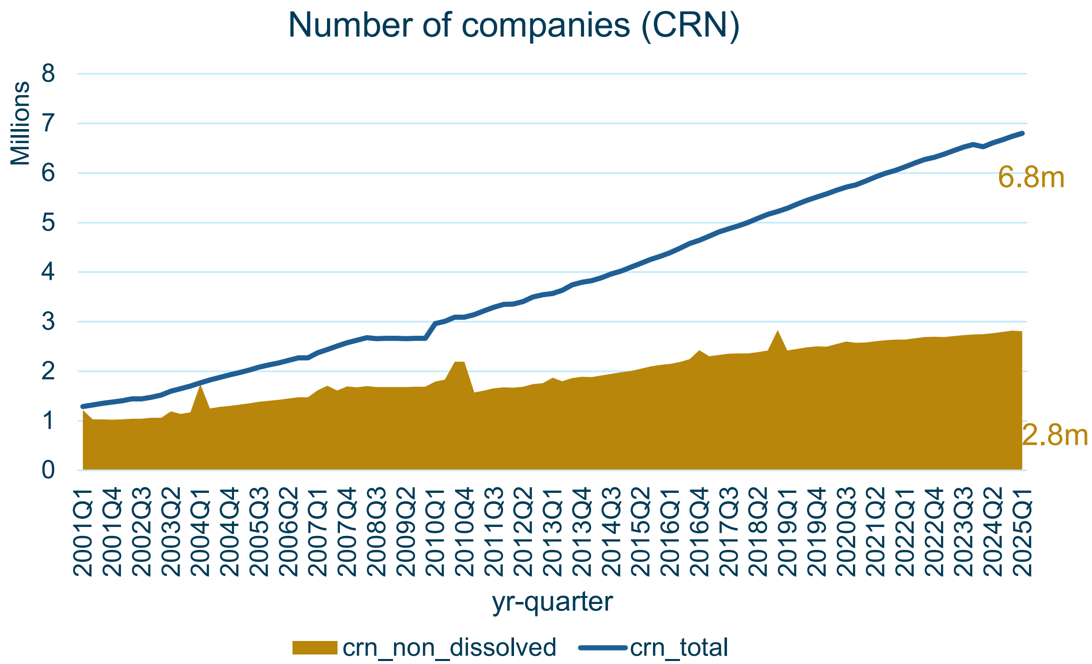
IDBR-active ≠ CH-active

Insights from the LBD Companies House Spine

Office for National Statistics

Counts of enterprises and companies

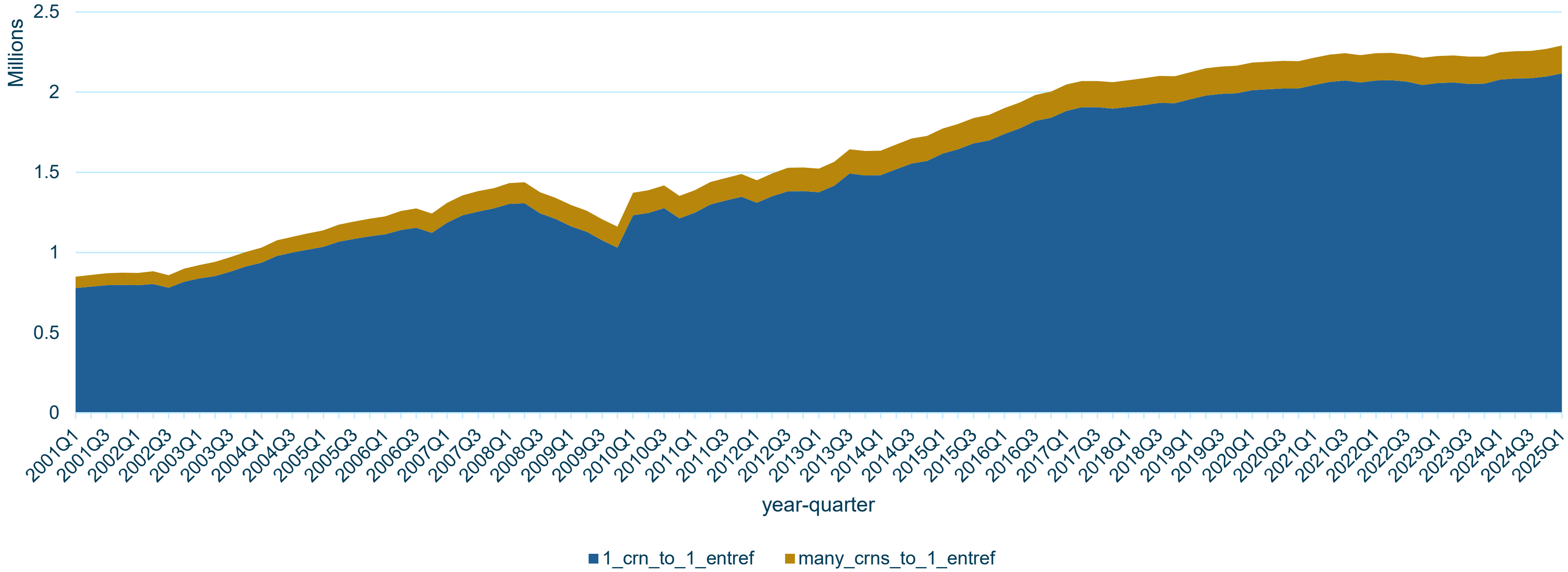




Source: Longitudinal Business Database, ONS

Notes: Enterprise figures exclude inactive or deleted enterprises. It further excludes enterprises and companies that are created for administrative purposes by the IDBR team. Right-hand side panel shows only companies that are matched to an enterprise. It excludes companies that are created for administrative purposes.

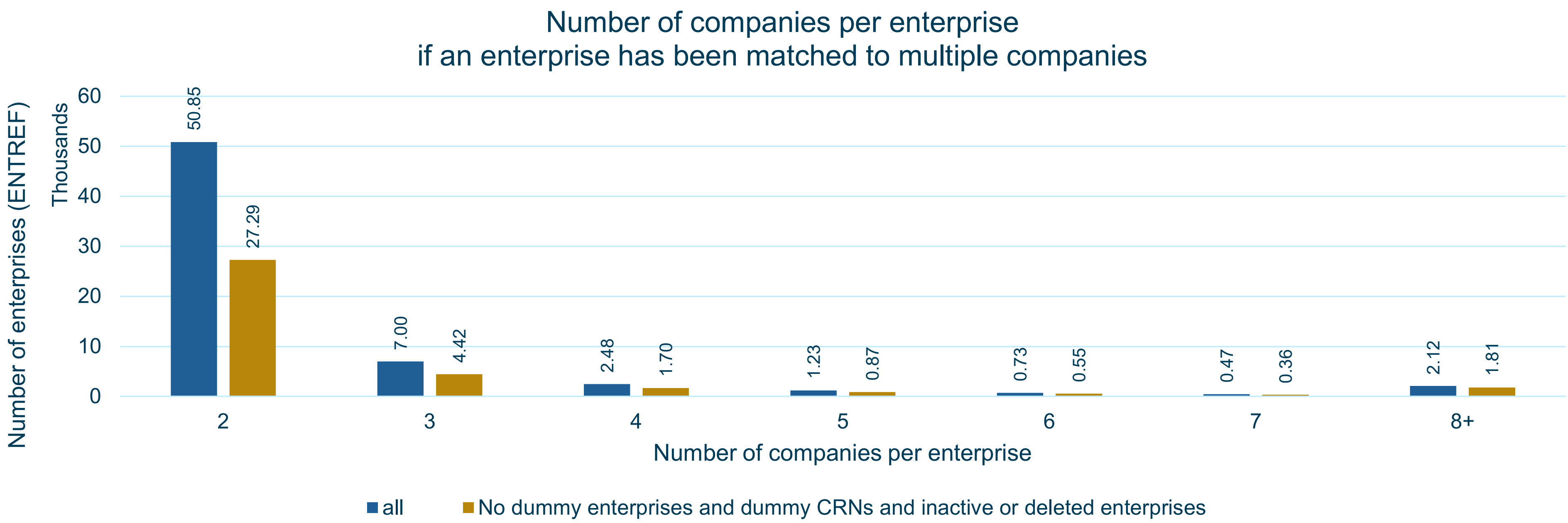
Number of company enterprise matches



Source: Longitudinal Business Database, ONS

Notes: This figure excludes inactive or deleted enterprises. It further excludes dummy enterprises and dummy CRNs that are created for administrative purposes by the IDBR team

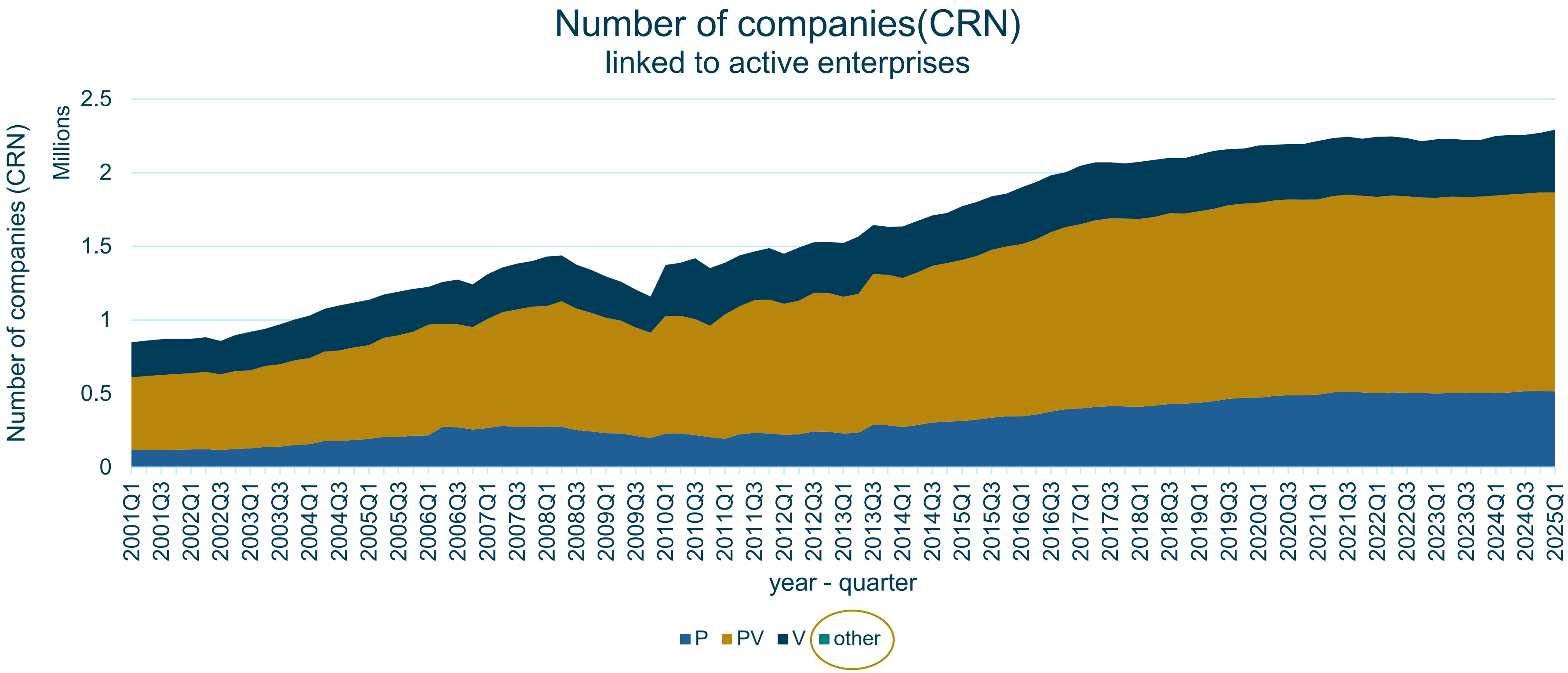
Distribution of multi-company enterprise matches



Source: Longitudinal Business Database, ONS

How do companies find their way to a matched enterprise?

P: PAYE, V: VAT, other: E, EP, EPV, EV



Source: Longitudinal Business Database, ONS

Notes: This figure excludes inactive or deleted enterprises.

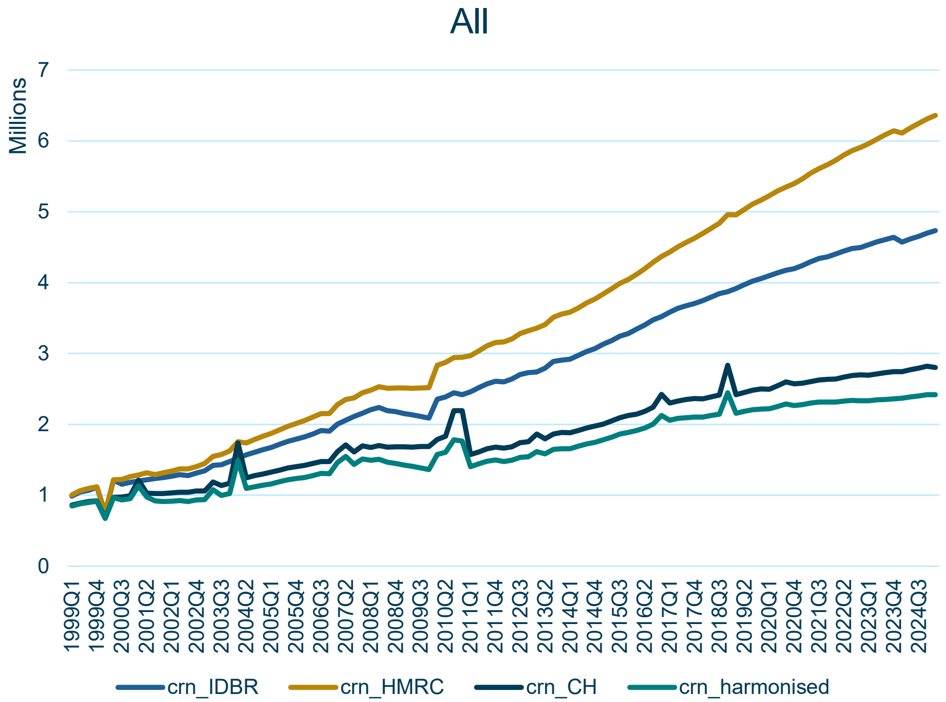
Company Activity Criteria

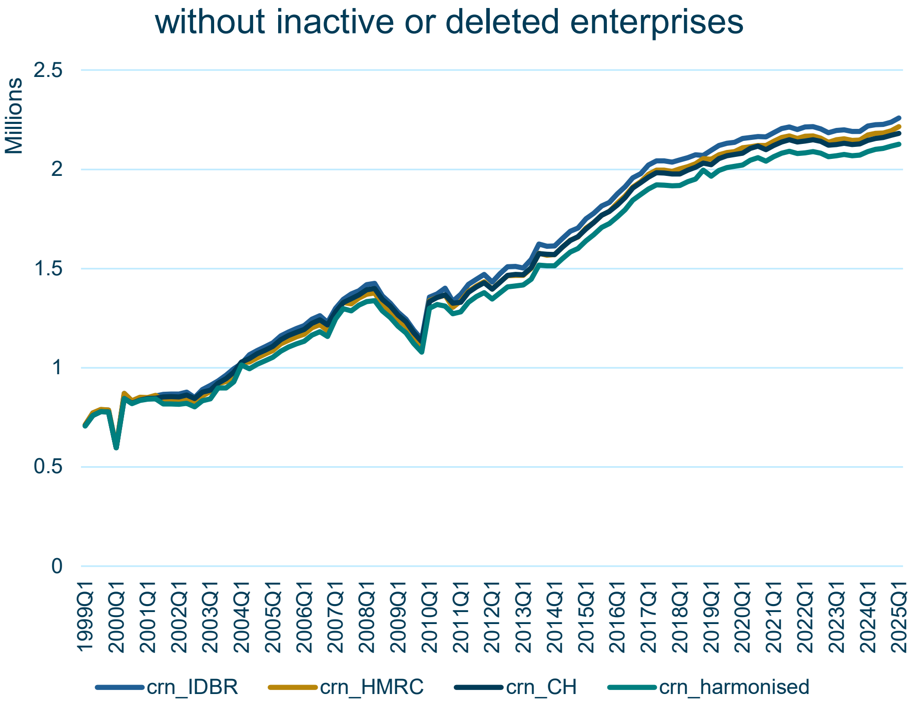
3+1 'Activity' flag variables:

CH active	*Company inactive if it has been dissolved*
HMRC active	*Company active if business active or receiving income; inactive if dormant or non-trading*
IDBR active	*Company active if linked to active VATs and/or PAYEs*

Harmonised active = *company is active across all three*

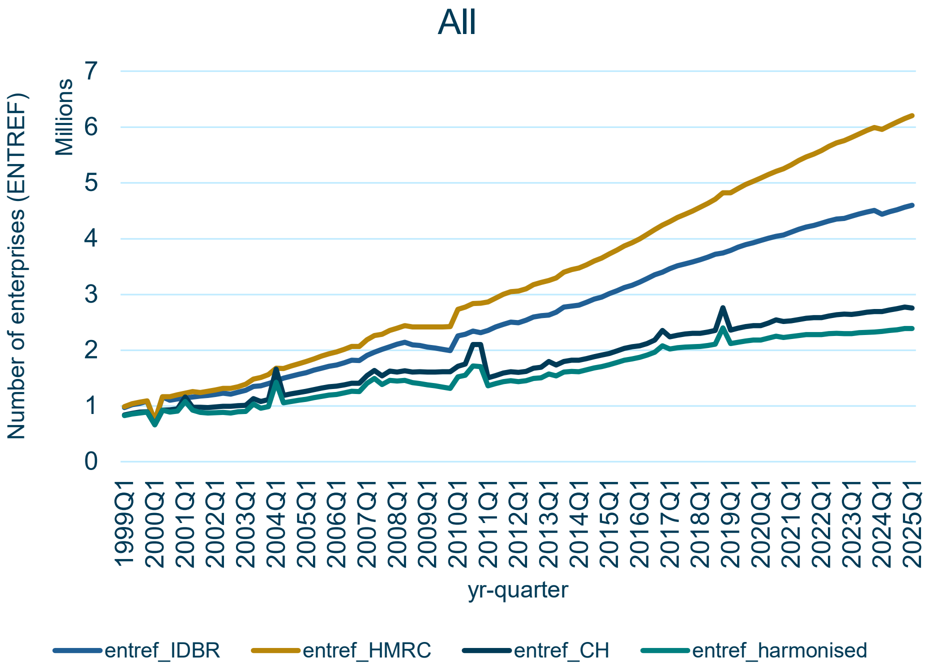
Activity variables -- companies

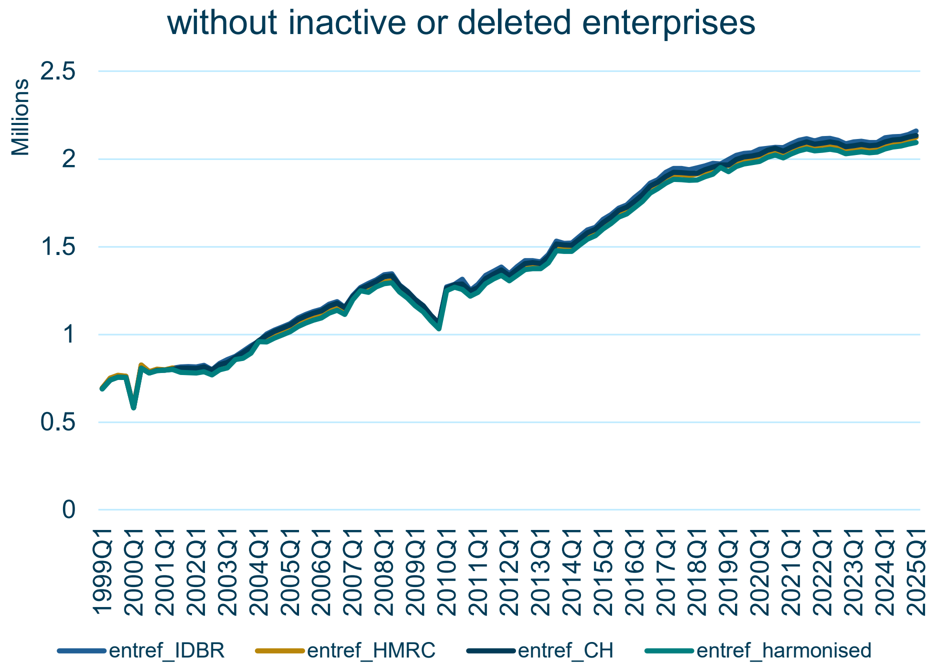




Source: Longitudinal Business Database, ONS

Activity variables -- enterprises





Source: Longitudinal Business Database, ONS

Limitations

CH spine limitations

LBD includes units that meet the IDBR eligibility criteria

Smaller entities that don't grow much or hire are not covered – their dynamics might differ

Coverage

RUREF/LUREF – CRN link

limited match for multi-unit ENTREFs

Linkages

It is based on CH additional SIC classification and accounts

HMRC_active

SRS-restricted because of the CRNs

Data Access

Successful registered companies

Characteristics of the matched population

Office for National Statistics

Methodology

What are the observed characteristics of the companies (CRNs) that we match on IDBR?

$$\log(\text{CRNs matched to IDBR})_{irst} = \beta_0 + \beta_1 * Harmonised\ Active_{irst} + \beta_2 M_{irst} + \sum_j \gamma_j T_{j,irst} + \sum_k \delta_k E_{k,irst} + \alpha_{rst} + \varepsilon_{irst}$$

where i indexes the matched CRN-ENTREF unit, r region, s 2-digit industry, and t year-quarter

- Harmonised active is the binary variable when the matched CRN – IDBR pair is active in IDBR, CH and HMRC.
- M: is the type of match
- $T_{j,irst}$: enterprise turnover band dummies
- $E_{k,irst}$: enterprise employment band dummies

Larger enterprises are associated with higher numbers of matched CRNs

	Dep. Var.: Log(Number of matched CRN-ENREF)	
	(1)	
	full sample	
Harmonised active	1.343	***
	(.002)	
Type of match		
1-CRN-to-1-ENTREF	omitted	
many-CRNs-to-1-ENTREF	-0.056	***
	(0.002)	
Enterprise turnover bands		
10k--100k	-0.458	***
	(.002)	
100k--1m	-0.269	***
	(.003)	
1m—10m	0.387	***
	(.004)	
10m--25m	0.364	***
	(0.005)	
25m—500m	0.659	***
	(0.007)	
>500m	-0.393	***
	(.086)	
Enterprise employment band		
10--49	-0.4556	***
	(.003)	
50--249	-0.504	***
	(0.003)	
>=250	0.086	***
	(.003)	
N	3,748,629	
R^2	0.268	

Heterogenous size effects

- Smaller enterprises have more 1-CRN-to-1-ENTREF matches
- Larger enterprises have more many-CRN-to-1-ENTREF matches

	Dep. Var.: Log(Number of matched CRN-ENTREF)			
	(2) 1-CRN-to-1-ENTREF		(3) many-CRNs-to-1-ENTREF	
CH active	1.984	***	0.393	***
	(0.004)		(0.001)	
IDBR active	2.282	***	0.974	***
	(0.006)		(0.001)	
HMRC active	1.984	***	0.388	***
	(0.003)		(0.001)	
Enterprise turnover bands				
10k--100k	-1.160	***	-0.094	***
	(0.004)		(0.002)	
100k—1m	-1.986	***	0.344	***
	(0.005)		(0.003)	
1m—10m	-2.6	***	0.885	***
	(0.008)		(0.004)	
10m--25m	-3.199	***	0.662	***
	(0.020)		(0.005)	
25m—500m	-3.239	***	0.927	***
	(0.026)		(0.006)	
>500m	-3.537	***	-0.1801	**
	(0.108)		(0.071)	
Enterprise employment band				
10-49	-0.736	***	-0.294	***
	(0.004)		(0.003)	
50-249	-1.224	***	-0.1537	***
	(0.005)		(0.003)	
>250	-1.426	***	0.4802	***
	(0.006)		(0.003)	
N	1,121,119		2,627,510	
R2	0.499		0.387	

Next Steps

Where are we heading now?

Next steps

Data improvements:

- Improve enterprise age definition
- Improve HMRC_active
- Investigate unexpected CRN link to enterprise

Analysis:

- Patterns of companies that do not grow enough to have an economic foot-print

CH Spine Potential

link to any CRN-based dataset

e.g. Corporation Tax/FAME, further with ONS or DBT surveys

Conclusion

Conclusions

1. **Created the LBD Companies House Spine**
   Dual role: (1) linking table between any CRN-based dataset and IDBR-based surveys; (2) reproducing national business statistics with the right filtering

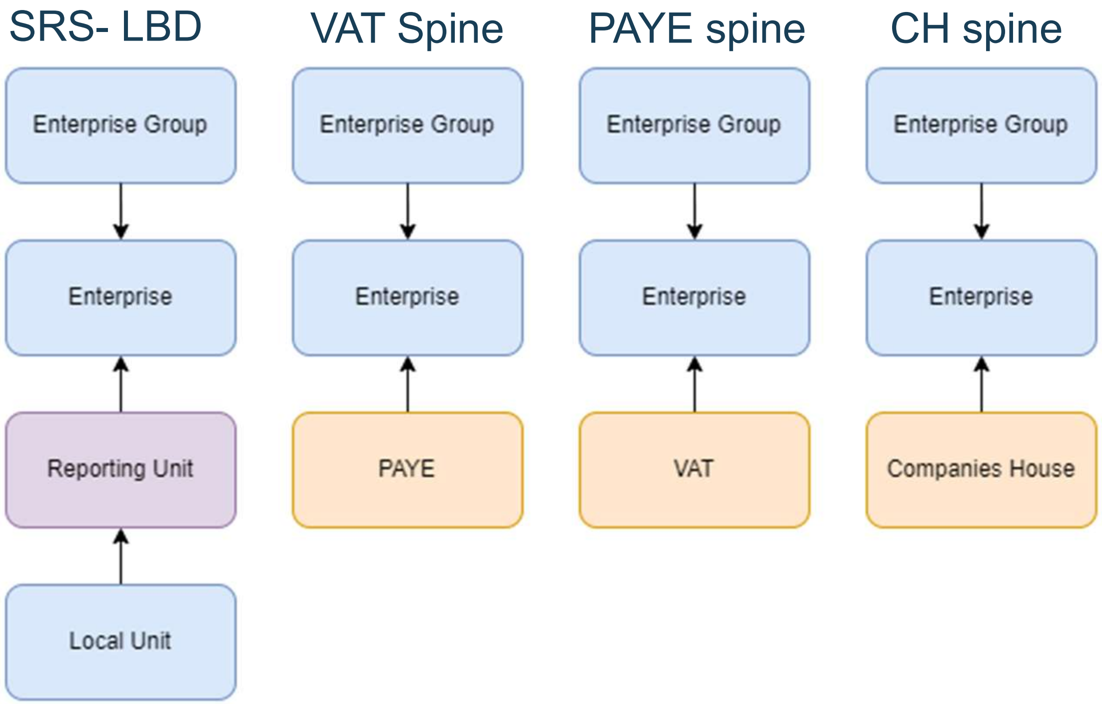
2. **Harmonised business activity across admin sources**
   Consistent activity classification across IDBR, Companies House, and HMRC

3. **Larger enterprises are more likely to match with a company**
   ...but with more complex structures (many CRNs-to-1-ENTREF)

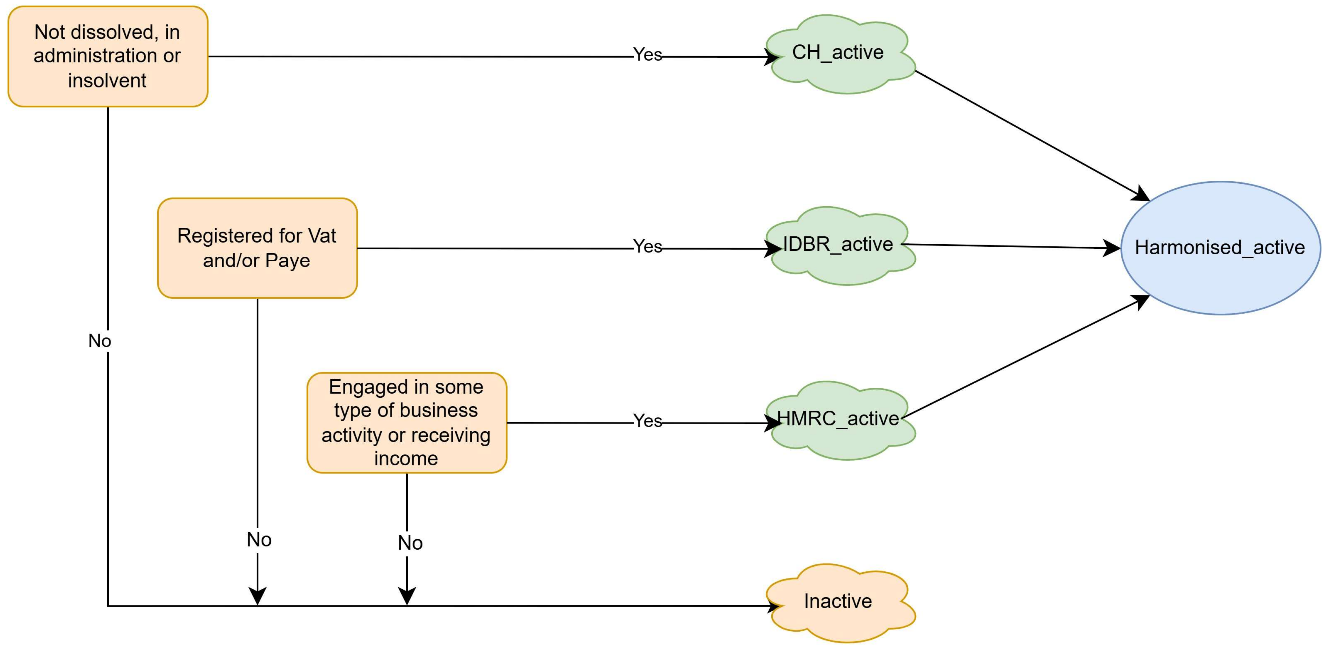
4. **Smaller enterprises tend to have simple matches**
   1-CRN-to-1-ENTREF mapping dominates among smaller firms

Appendix

LBD – The core spines



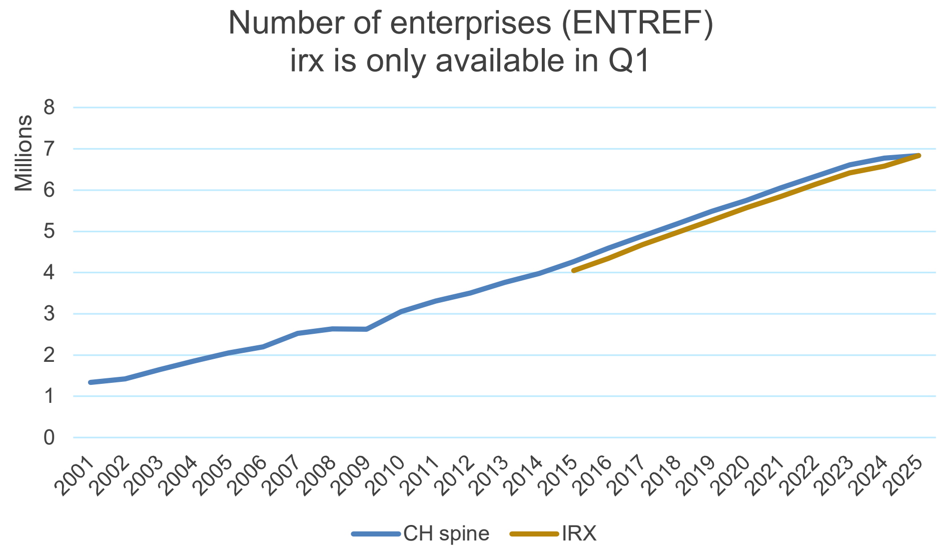
Company Activity Criteria

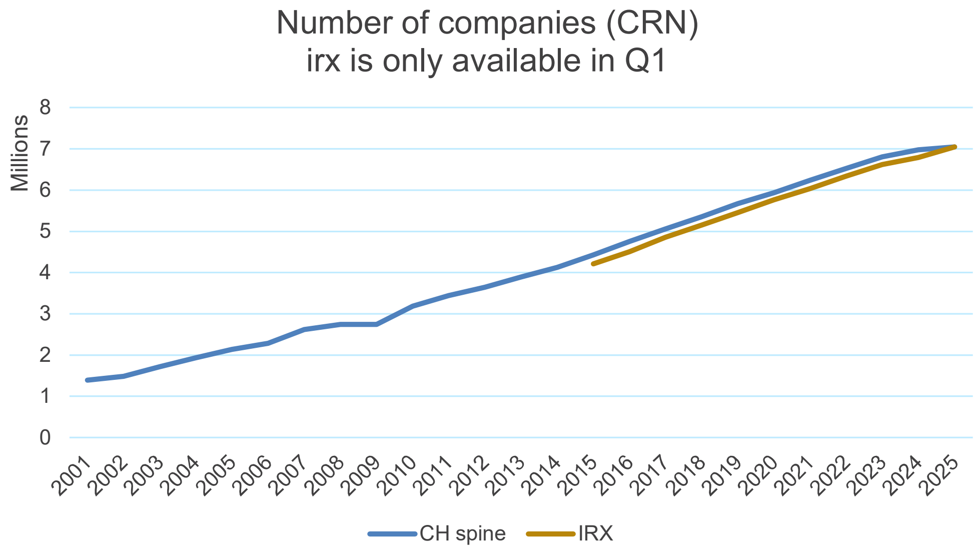


External validity

Compare with other data products

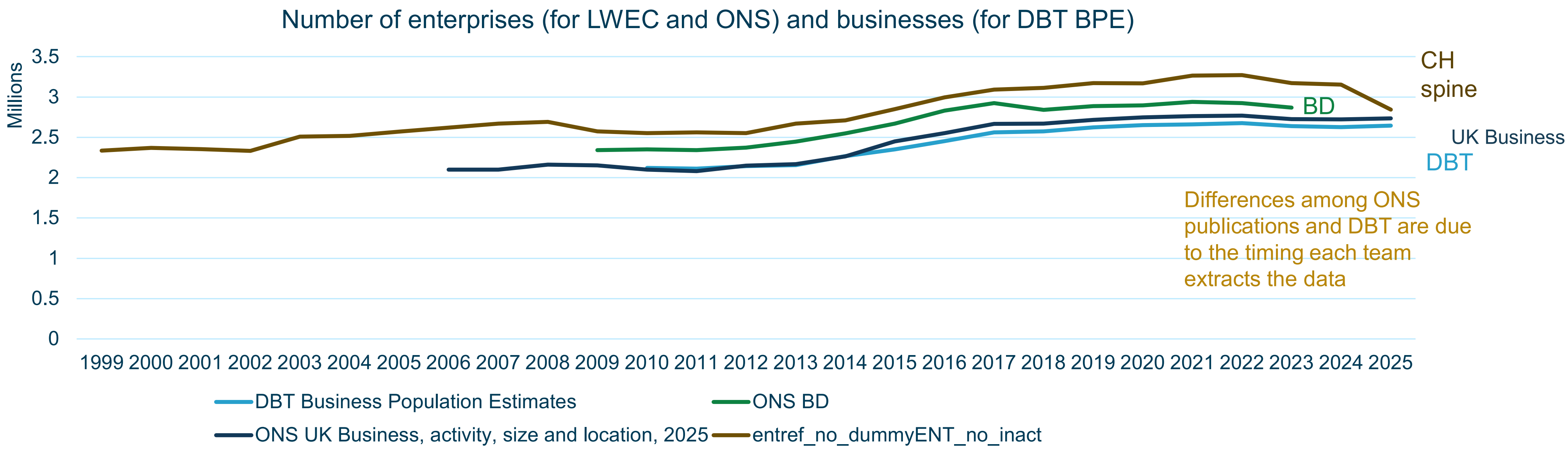
1. CH spine as a "matching tool": Does it do what IDBR does? YES!





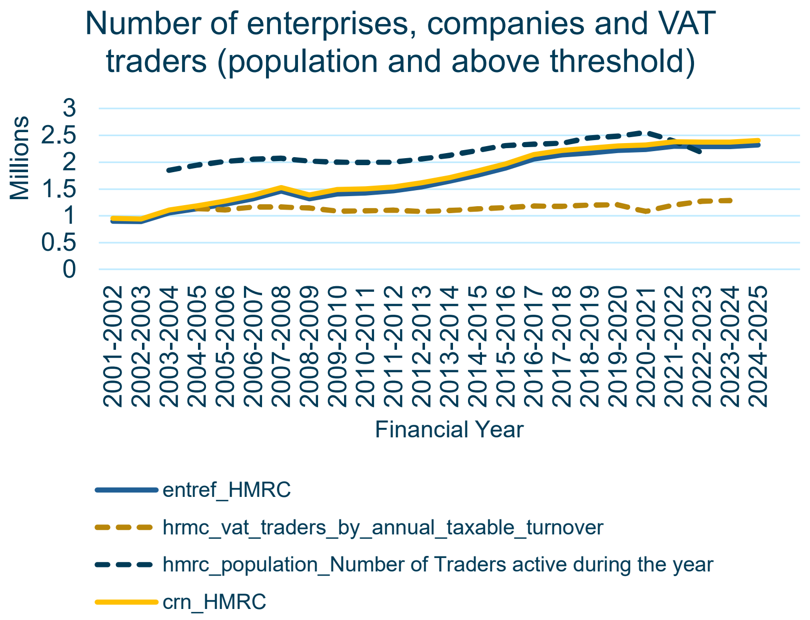
2. CH spine for official statistics

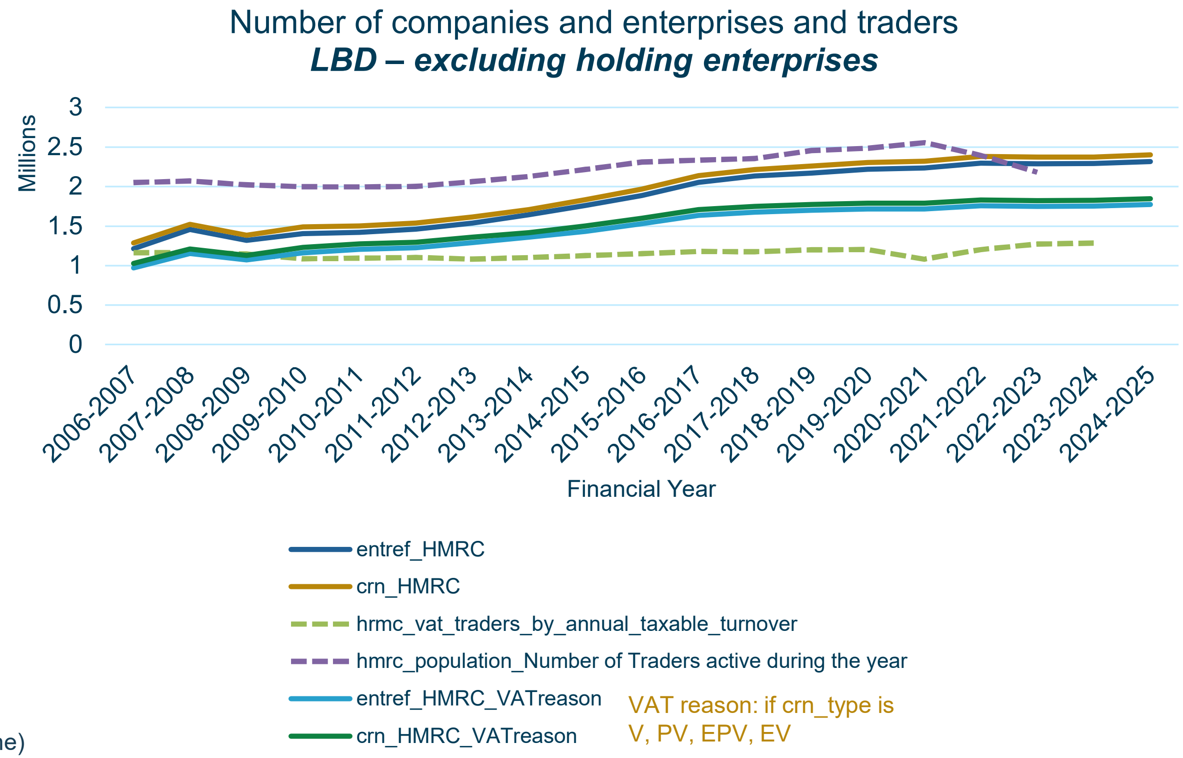
Can we produce the same series as other products do? Yes, with the right filtering...



Notes: This figure excludes inactive or deleted enterprises for the CH spine timeseries.

`HMRC_active` is our weakest activity marker. How far off are we?





Sources: HMRC VAT annual statistics 2023 to 2024 (dashed lines) and LBD (any other line)