

Methodology - Insolvency (Shared Data Unit)

This is a text version of the methodology for this release:

https://docs.google.com/document/d/1toIXsxn0zMYHzY3YX8-ij4emWR2WI1IPDIczij_JCcw/edit?usp=sharing

The code used to process and analyse the data will be available on Github in R markdown format after embargo: <https://github.com/BBC-Data-Unit/shared-data-unit>

Sourcing:

We scraped data on insolvency notices from the London Gazette, the official public record, from January 2019 onwards.

This data is therefore from a different source to the government's national level Insolvency Statistics, which are based on monthly changes in the register and the Insolvency Service's own data on compulsory liquidations. The monthly statistics also show a drop in insolvencies year on year. <https://www.gov.uk/government/collections/monthly-insolvency-statistics>.

Another key difference is that the Insolvency Service does not count members voluntary liquidations as these are not included in the formal insolvency process. The Gazette data does include these, which may explain a difference in figures for March 2020.

NB: Data on Northern Ireland has been affected by a complete **closure of** the Courts and **Insolvency Service** offices in that region from April onwards. This means that there were no new compulsory liquidations, as these require a court order.

General process:

The Gazette notices included for analysis were: Appointment of liquidators, Appointment of Administrators, Appointment of Receivers Resolutions for Winding-Up and Winding-up Orders. Data was then deduplicated by company id to avoid double-counting, should a company have filed more than one of the above.

Gazette data was matched to a snapshot of Companies House by company number, using the Companies House API to resolve non-matches. Around 35 companies could not be matched, the majority of which were non-GB registrations. These were discounted.

We then looked up the location of the companies filing notices using the Companies House database. Location is therefore **by registered office address** by company.

Local authority areas were joined onto the data using the [2020 ONS geography lookup of postcode to Local Authority area](#). Non-matches were kept as 'unclassified' entries.

Region areas were added after data had been aggregated by local authority area. Regional averages were calculated using the median instead of the mode due to heavy distributional skew towards a few local authority areas.

Standard Industrial Classification (SIC):

For **industry analysis**, we matched the Standard Industrial Classification (4-digit siccode) in the Companies House columns of the data to broader industry description (Retail, Agriculture etc). We used SIC 2007 classification tables.

Some companies have multiple siccode descriptors. Data was 'reshaped' to allow this matching process to be on all siccodes given, instead of just the first one listed.

To avoid subsequent duplication issues, this more complex dataset was only used for analysis of impact by industry, and not for examining notices issued by time or location.

Because industry classification is broader than the 'original' 4-digit sic, it was possible for a company with multiple siccodes to have the same industry description twice. These double entries were removed to minimise distortion when looking at industry impact over time. Entries for different industry descriptors for the same company were preserved however, as one could not be said to be 'more valid' than the other.

Data that was compiled using the Companies House API (around 1000 entries of 39000), could only be retrieved with a single siccode instead of up to four. Given multiple siccodes are relatively uncommon, this is unlikely to have had much of an impact on findings.

New enterprises:

New companies were found by splitting the most recent snapshot available from Companies House by incorporation date.

Siccode matching was used as for insolvencies to find the breakdown by sector and also by granular industry description.

As for insolvencies, the non-siccode matched dataset was used to find new enterprises by area, matching local authority districts to postcodes on Companies House.