**More detailed jobs series, Spring 2022 update**

**Ammar Ljubijankić**

**Aim/purpose**

**Produce a continuous long-running series of employee job numbers across different hierarchical industry levels within London. The series is uploaded as a workbook to the** [London Datastore](https://data.london.gov.uk/dataset/london-s-sectors)**.**

**Data sources**

**Data on employee job counts is imported from Business Register and Employment Survey (BRES) data on the Secure Research System (SRS) accessed on Office for National Statistics (ONS) servers. This data is at a higher level of granularity than that** [published by the ONS](https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/businessregisterandemploymentsurvey)**, as it contains information at the firm level, including number of employees, geographical location and detailed industry classification on an annual basis. This data is collected at the industry class-level for London firms within the SRS before being exported and processed further.**

**Data at the highest industry section-level is collected from the published** [ONS Workforce Jobs](https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/workforcejobsbyindustryjobs02/current) **(WFJ) series. This data includes a continuous time series measure of the number of workforce jobs within London on a quarterly basis.**

**Method**

**The class-level data is initially collected for the years 2018 to 2020 within the SRS separately. Data for previous years were already available on the GLA servers as they had been collected and published previously**[[1]](#footnote-1)**. The ONS releases two versions of data for each year: a provisional version and a revised version once the provisional data has been corrected and verified. In this iteration, only the 2020 data was provisional.**

**The annual data is then subjected to a rigorous review process within the SRS to ensure that the ONS’s safe researcher standards are observed and there would be no privacy violations when exporting the data. This involves the following process:**

* **The raw BRES data is collected within the GLA Economics working folders and collapsed from firm-level to less granular levels within the Standard Industry Classification (SIC) system.**
  + **We use the four highest levels within SIC: 1-digit sections (highest level of industry aggregation); 2-digit divisions; 3-digit groups; and 4-digit classes.**
* **We keep information on the number of employee jobs per SIC level and the number of jobs in the largest firm within the relevant level. This information is loaded into Stata.**
* **We use a Stata script to perform a thorough evaluation process of the data to ensure that we only export data aggregated to the extent that no personal information can be extracted. This includes**[[2]](#footnote-2)**:**
  + **Ensuring that each SIC level contain at least ten employee jobs and ten firms, or otherwise suppressing the data point;**
  + **Ensuring that any individual firm does not include more than 40% of employees within the SIC level, or otherwise suppressing the data point;**
  + **Ensuring that no suppressed data points can be reconstructed through comparison with other industries at the same level.**
* **Once the suppression process is complete, the data is submitted to the ONS for review and approval before being exported outside of the SRS within Excel.**

**The exported data is then collected in the INPUT data folder in the Detailed jobs project folder. Using an R script, the data is processed as follows:**

* **Annual data is reshaped into a uniform long format, where each row represents one constituent level within the SIC hierarchy, and columns indicate the year.**
* **The section G is split into two constituent parts: “Retail” and “Wholesale and motor repair”. Some classes are aggregated to conform more closely to pre-2015 data, e.g., classes 1723 and 1729 are combined into 172\_39.**
* **At this point we also import the ONS WFJ data in order to align the two data series.** 
  + **Factors are calculated at the section-level by dividing WFJ data points with the respective BRES data point (by section and year).**
  + **These factors are then applied to all sub-levels within the same section to constrain the BRES data to the WFJ series, which is considered a more accurate estimate of the number of employee jobs.**
* **For publication, data is rounded to nearest 250 for section-level and nearest 100 for all other levels. Each level is exported to a separate sheet in a workbook. These sheets are copied manually to a separate workbook which uses Excel formulas to fill in ready formatted tables.**

Once the data was exported to Excel in a uniform format, it was subjected to manual checks, comparing it to previous publications and performing sense checks.

**Potential risks, uncertainty and assurance**

The analysis does not use any modelling nor is it reliant on assumptions for its estimates.[[3]](#footnote-3)

Hence the main risks associated with the release are: (i) exposure of private information and (ii) release of incorrect figures. The first risk is based

The first risk has been alleviated before the data is exported from the SRS. The script used to review data and perform the necessary suppression runs without further user interaction – it will therefore produce a correct result each time assuming the code is appropriate. This was confirmed when used the first time on the 2019 data as follows:

* The raw data was first subjected to a manual review where each data point was checked by the analysis lead within Excel using a previously established procedure, and individual cells were suppressed where needed.
* The script was then run on the same raw data and the outputs were compared. The tables produced by the script were found to correspond with the manual output.
* The script was then used to produce the output for the 2018 and 2020 data. Instead of conducting the full manual replication, we performed sense checks and spot checks to confirm the results still appeared correct.

All tables were then submitted to the ONS for an external review, which it passed before being approved.

The potential use of the published figures for further analyses or policy decisions is the source of the second risk. Inaccuracy in the figures may appear at any point between sourcing the data from BRES within the SRS and the final publication, either through human error or mistakes in the script codes. This could occur during the suppression Stata process within the SRS; the processing of data in R to reshape and constrain the data; or during the final formatting of the tables.

We believe they were not corrupted by the Stata process within the SRS as the figures were subject to sense checks and spot checks (by comparison to raw BRES data) within before being submitted for review to the ONS. The figures were not manipulated manually at any point within SRS, which should reduce the risk of human error.[[4]](#footnote-4)

The data was also likewise only processed using scripts once exported, to minimise random human error. We then compared the final outputs with previous published and unpublished figures to ensure that numbers had not changed substantially or unexpectedly.

It is important to note that both BRES and WFJ data are subject to revisions and corrections by the ONS, such that we do not expect data collected on different years to match perfectly. This is especially true for provisional datasets. For instance, the BRES data for 2016 was deemed provisional at the time of GLA Economics published the previous iteration of the “Detailed jobs series” in 2018. There were therefore significant differences between the new section-level figures and those published in 2018. However, comparing the new figures with those produced in an unpublished version of the analysis from 2020 did not show any significant differences.[[5]](#footnote-5)

Due to suppression, there may also be differences in which data points are retained across years, particularly for lower SIC levels such as groups and classes.

The reviewing and checking process was conducted by the analysis author (Ammar Ljubijankić) and the supervisory economist (Christopher Rocks).

1. The “More detailed jobs series” was first released in 2014 and most recently in 2018. [↑](#footnote-ref-1)
2. The script includes more detailed information on how the data is reviewed and suppression performed. [↑](#footnote-ref-2)
3. Previous versions of the release involved more complex methods of aligning new and old data to produce a continuous time series. However, as this has already been reviewed and released, it is assumed that there are no risks associated with it continued use. See the [2014 series methodology paper](https://data.london.gov.uk/download/london-s-sectors/5cd17d76-50cc-42f4-b672-6f9c9ec45866/Methodology-London's%20sectors%20-%20more%20detailed%20jobs%20data.pdf). [↑](#footnote-ref-3)
4. Note that the ONS does not check the data for accuracy, only for compliance with SRS terms of use. [↑](#footnote-ref-4)
5. As all figures are rounded, small differences are to be expected from release to release, especially where figures are not rounded to the same level. While the employment counts are often above 100,000 at the section level, the difference in counts between releases were generally below 1,000 on the few occasions they did not match exactly. [↑](#footnote-ref-5)