***Summary data on local authority procurement, trimmed to remove small payments and payments to individuals***

***(file: trimlaprocur.dta)***

Each row is a summary of the total amount paid to a given provider of service (or recipient) in a given financial year. This reduces the scale of the data to c. 1.8 Mn observations, representing 560 000 distinct organisations.

To aid analysis we have also attached ONS codes for local authorities, rather than their names; and we have also included charity commission or company numbers where we were able to identify them.

|  |  |
| --- | --- |
| Variable | Description |
| Funder | Local authority – identified by ONS code, e.g. “E08000025” (Birmingham) |
| Recipient  | Name of organisation receiving funds |
| Recipient\_charity\_number | Charity Commission registration number, if matched (blank otherwise) |
| Recipient\_company\_number  | Companies House registration number, if matched (blank otherwise) |
| Recipient\_organisation\_type | Type of organisation, based on most plausible match: types include Charity, company, educational establishment, health provider (various: GP surgeries, pharmacies, but a range of other providers also included), public sector (principally named public authorities, e.g. parish councils, district or county councils); sports organisations (from lists of Community Amateur Sports Clubs); and unallocated (organisations which did not match to any of the above categories) |
| Year | Calendar year of payment |
| Amount | Value of transaction |
| Amband | Banded amount of transaction £1000-9999; £10000-99999; £100k+) |

Classification of organisation type is inherently challenging and approximately one quarter of organisations were unclassifiable given the time and resources we had available. However they only account for some 5% of the total value of transactions.

These data were compiled from information that is **publicly available** – all we have done is put it together in a relatively accessible form. We assume that anyone using it has as a primary purpose the identification of broad patterns in the data rather than the identification of named individuals, of whom there are many in this dataset.

Individuals can be identified through the source material which we have used. However we assume that for analytical purposes most users will be interested in the relationships between local authorities and external providers of services, such as companies and charities. To limit the extent of inclusion of information about individuals in this dataset, we have excluded smaller payments and also payments classified as having been made to individuals.

This dataset contains information on the following entities:

Charity 19813 5.1 5.1

Company 208153 53.5 58.6

Education 5111 1.3 59.9

Health 6452 1.7 61.6

Public-Sector 5415 1.4 63.0

Sports 475 0.1 63.1

Unallocated 143551 36.9 100.00

The least problematic categories are charities and companies, where we are very confident that we have successfully matched names recorded in the procurement data. Over 98% of organisations classed as charities have a charity registration number; almost all those organisations identified as companies have a Companies House number, or their names contain the string “limited” or “Ltd”, denoting company status.

Investigation of a sample of records shows that the matches against lists of education providers, and health providers, almost wholly identify organisations, such as NHS Trusts, hospitals, GP Practices, or schools; the same is true also for the “Public sector” category, where we were matching lists from the procurement data against lists of known public authorities, and the “sports” category (lists of sports clubs).

This leaves a large “unallocated” category, accounting for 37% of observations (143000 entities). From an investigation of a sample of these, it is clear that well over 95% of records contain the names of companies, charities, or other organisations that have not been successfully matched to lists of such entities. This is almost invariably because the way in which their name is recorded differs in non-trivial ways from the way it is captured on the list of a regulator. Resources available in the project did not permit us to refine these matches in any way.

We have recoded the data into size bands for payments of £0 – 100, £100-999, etc., and dropped anything categorised as a payment to an individual as well as any payments below a certain threshold – in this case £1000.

An alternative is to drop entries on purely financial grounds; for example, a threshold for payments of at least £100k in any one year would reduce the dataset to 149000 observations.

With these restrictions in mind, we have dropped entries that relate to individuals, and everything below a threshold of £1000. This results in a dataset of 1.35Mn observations. This is the file “trimlaprocur.dta”.