



Title: **Needs & Redistribution Technical Working Group**

Paper: **NR TWG 16/20 – Different statistical techniques for assessing need (DCLG)**

Date: **Thursday 13th October 2016 (12:30pm)**

Venue: **DCLG - 2 Marsham Street - Fry Building - 4th Floor - Rooms 4A & 4B**

POLICY DEVELOPMENT: NOT A STATEMENT OF GOVERNMENT POLICY

There are a number of analytical techniques and approaches that could be used to derive estimates of local authority need.

1. Regressions against past expenditure

Regression analysis against past expenditure has been used at least since the introduction of council tax to derive the weights attached to different indicators within the needs formulae.

Past expenditure for each local authority is divided by the **client group**, generally the appropriate measure of population relating to the consumers of the service, and deflated by the **area cost adjustment**. This then forms the **dependent variable** in the regression. Regression analysis is used to derive the weights of appropriate **independent variables** that explain the greatest proportion of differences between authorities.

Pros

- Past expenditure data is readily available
- Uses a relatively simple statistical technique
- Importantly, the approach allocates money based on *predicted* expenditure, rather than past expenditure per se. Predicted expenditure is determined by the relationship between past expenditure and need characteristics.
- This can be used to forecast future needs (using updated independent variable data at LA level) assuming the relationships remain constant

Cons

- Regression analysis does not identify causal relationships. Therefore unless the panel of independent data used in the regression has an *a priori* causal link with the dependent variable (for example it may have been identified in academic literature) the link may be entirely coincidental.

- Past expenditure is being used as a proxy measure for need to spend. However past expenditure is heavily dependent on the resources available to local authorities, which are heavily influenced by central government either by way of grants or through legislative requirements. This means that this may not be a true proxy for comparing needs between local authorities.
- Using independent variables that are too **correlated** can artificially raise the measure defining the explanatory power of the model. To overcome this DCLG has previously checked the correlation matrix for the independent variables, ensuring that the variables entered into the regression are not too correlated.

2. Small area modelling

This technique has previously been used to define the Adult Social Care needs models and is a more sophisticated version of regression against past expenditure.

Expenditure data can be extracted from local authority systems at small area level (such as Lower Super Output Areas, LSOAs). For adult social care, a wide range of potentially relevant need, wealth and supply data are easily available at LSOA and other small area levels. LSOAs are designed to be relatively homogenous. Modelling can also be used to estimate survey results for geographies that ordinarily contain too small a sample.

Pros

- Uses a readily available statistical technique
- Uses variation within rather than just across local authorities, resulting in a more precise formula and larger sample sizes. Unlike local authority level regression, this captures where local authorities choose to spend more within their area, and how this links with need characteristics. Each of the 152 local authorities with social care responsibilities is divided into approximately 210 LSOAs on average.
- Importantly, the approach allocates money based on *predicted* expenditure, rather than past expenditure per se. Predicted expenditure is determined by the relationship between past expenditure and need characteristics at the small area level.
- This can be used to forecast future needs (using updated independent variable data at LA level) assuming the relationships remain constant

Cons

- Small area expenditure or utilisation data is present in local authority IT systems but is not routinely reported, so extraction takes time and money. Small area Census and DWP data (which are important components of the existing adult social care RNF) is however readily available.
- Regression analysis does not identify causal relationships. Therefore unless the panel of independent data used in the regression has an *a priori* causal link with the dependent variable (for example it may have been identified in academic literature; in adult social care the eligibility criteria imply that certain relationships should exist) the link may be entirely coincidental.
- Using independent variables that are too correlated can artificially raise the measure defining the explanatory power of the model.

3. Regressions against outcome variables

The Department for Education is currently proposing to use indicator data that has been shown to influence outcomes in their Dedicated Schools Funding model. The consultation paper does not explain how they intend to weight these measures in order to create the new grant distribution.

It would be possible to use the outcome measure itself, for example the proportion of children receiving at least 5 A*-C GSEs within an area, as the dependent variable within a regression. The indicator data would be formed from the agreed data shown to influence outcomes. The overall level of funding available to the authority will limit their ability to spend on services. It would be possible to include the overall level of funding available to each local authority as an additional independent variable within the regression. The regression equation would then become:

$$\hat{y} = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$$

where:

\hat{y}	is the estimated outcome
α	is a constant
$\beta_1 \dots \beta_n$	are the weights applied to the outcome variables
x_1	is the weight allocated to level of expenditure e.g. £ per head
$x_2 \dots x_n$	are the outcome variables

Pros

- Uses a relatively simple statistical technique
- This can be used to forecast future needs **assuming the relationships between the outcome and the explanatory variables remain constant**

Cons

- Identifying a suitable outcome measure, for which data exists, for all services may not be possible.
- **Regression analysis does not identify causal relationships.** Therefore unless the panel of independent data used in the regression has an *a priori* causal link with the dependent variable (for example it may have been identified in academic literature) the link may be entirely coincidental.
- Using independent variables that are too correlated can artificially raise the measure defining the explanatory power of the model.

4. Multi-level modelling

This technique has previously been used to define the Children's Social Care needs model.

This technique allows hierarchical data to be analysed at multiple levels, for example at both the ward level within authorities and at the authority level in the case of the Children's Social Care formula. The dependent data must be held at the lowest level i.e. in this case parish level. The independent variables can be defined at any level.

Pros

- In the case of past expenditure, the effect of different policy choices and different funding levels between individual authorities can be discarded to give an overall estimate of the weights and factors driving need to spend.
- This methodology outperforms classical regressions in its predictive accuracy when looking at hierarchical data.

Cons

- In order to perform the technique, the dependent data is required at the lowest level. Generally this data is not routinely collected and therefore a special data collection is required; **this extends the time taken to derive the model**. It is unlikely that this technique could be used to derive formulae in time for use in 2019-20.
- The statistical technique is more complicated than simple linear regression models.

5. Principal Component Analysis

This technique was used by Home Office to derive the new police formula consulted on in July 2015. It was also used to generate the Risk Index within the Fire Relative Needs Formula.

For example, if we start with a large list of indicators (or local authority cost drivers), then some of those indicators may be related, for example a large variety of indicators are used to measure deprivation in the 2013-14 Relative Needs Formulae. Principal component analysis is used to identify a relatively small number of factors that can be used to represent relationships amongst this set. These factors can help to identify underlying concepts which are not directly measurable e.g. deprivation.

The factors are created using linear combinations of the original indicators. The first principal component is the weighted combination of all the original indicators that account for the most variance in the sample; the second the next largest variance etc. Each of the principal components are uncorrelated with each other. Normally only a few indicators (those which explain the greatest variance) are chosen to represent the factor, and are used in the form of an index.

The indices can form one of the dependent variables within a regression. An alternative approach was used by the Home Office, who intended to use the weights produced by the principal component analysis to weight the indicators in their formula.

Pros

- Doesn't require a dependent variable.
- A number of assumptions underlying regression analysis can be violated without affecting results.

Cons

- The results are particularly sensitive to outliers and missing data.

6. Future Needs

Under the new 100% business rates system the quantum of funding available will be defined by the amount of business rates that can be collected. The new system will therefore need to look at relative need to spend.

This group previously suggested that national level expenditure of services may be a good proxy for relative need to spend on those services, and that this might be used to determine control totals for the different services. One way of allocating funding between authorities for a particular service is to forecast the drivers of that service. For example, the number of children in need is likely to be a cost driver for children's services. The *Children in Need Census* provides historical data on the number of children in need during a year. Detailed data is collected by gender, age, ethnicity and primary need at assessment. By definition, future data does not exist – the data therefore has to be either estimated or forecast. There are a number of statistical techniques that can be used to do this.

6a. Crude estimation

The simplest way to estimate the number of children in need going forward is to look at the proportion of children in need compared to all children aged 0 to 17 in a particular area and to apply these proportions to the latest population projection data.

Pros

- The methodology is very simple.

Cons

- It does not take into account any historical trends in past data.

6b. Time series analysis

An alternative way of estimating the number of children in need would be to look at previous trends in the data and project these forward. Data on children in need is only available on an annual basis and therefore it is not necessary to check the data for seasonal variations.

Pros

- The method takes into account historical trends in the data.

Cons

- Each time series will need to be analysed independently. There is therefore a limited number of such time series that can be analysed within the resources available. Increasing the level of granularity of the data will increase the number of time series that will need to be run.

Annex A: Glossary of Terms

Area cost adjustment

The scaling factor applied to need formulae to reflect higher costs – mostly pay – in some council areas.

Client group

This is the client group which potentially consume the service, such as projections of population, numbers of school children, numbers of elderly people or lengths of road.

Correlation

The extent of association between two or more variables. Correlation is positive or direct when two variables move in the same direction and negative or inverse when they move in opposite directions.

Indicator data

In this context, information used in the calculation of the needs formulae, such as the proportion of people aged 65 and over receiving attendance allowance or proportion of people receiving unemployment-related benefits.

Regression

This is a statistical approach to forecasting change in a dependent variable (past expenditure, for example) on the basis of change in one or more independent variables (client group and other indicators, for example). Regression analysis can be used in fitting either a curve or a line to data points.

In a cause and effect relationship, the **independent variable** is the cause, and the **dependent variable** is the effect. Least squares linear regression is a method for predicting the value of a dependent variable Y, based on the value of an independent variable X.

Linear regression finds the straight line, in a manner such that the differences in the distances of data points from the line are minimized. Relationships depicted in a regression analysis are, however, associative only, and any causal inference is purely subjective.

Revenue Outturn (RO) Data

Final audited data supplied by Local Government on their annual expenditure on different types of services.