Ministry of Housing,
Communities &
Local Government

Overview of statistical techniques (NR TWG 18-01)

Technical Working Group

Local Policy Analysis

16 January 2018

Ministry of Housing, Communities & Local Government

Technical consultation on relative need

Open for 12 weeks until 12 March (Link)

- The consultation sets out the Government's approach to the review of relative needs and resources, and seeks wider views on what important factors should be included in a new funding formula.

- The consultation:

  - provides a summary of how relative needs assessments in England have worked in the past (chapter 2),
  - presents the idea of using a simple foundation formula to measure the relative needs of local authorities, based on a small number of common cost drivers (chapter 3),
  - considers a number of service areas where in addition, a more sophisticated approach to measuring relative needs may potentially be required (chapter 4),
  - outlines the statistical techniques that could be used to construct relative needs formulas (chapter 5), and
  - seeks views on the potential impact of the options outlined in the consultation on persons who share a protected characteristic (chapter 6).

- This presentation will focus specifically on the issues discussed in chapter 4 and 5.

Ministry of Housing, Communities & Local Government

Why use statistical techniques?

Regression against expenditure

Other statistical techniques

Ministry of Housing, Communities & Local Government

How do we develop a measure of relative need?

Cost drivers are identified and weightings given to each cost driver within a service area

A measure of total need is then estimated for each LA by multiplying ***that LA's*** cost drivers by the ***nationally-derived*** weights

The measure of total need is then scaled so the England total equals the given control total for that service area

Each LA then has a relative measure of need before other factors considered (ACA, resource, transitional arrangements etc.)

Ministry of Housing, Communities & Local Government

Key criteria for cost drivers

Relevant

- there should be evidence to demonstrate that the cost driver has a significant impact on the cost of providing services

Objective

- the cost driver should be measurable using robust, up-to-date data that is collected on a consistent basis across England. Local authorities should not be able to directly affect the indicator (i.e. the cost drivers should not create perverse incentives to 'game' the system)

Distinct

- the cost driver should explain significant variation in the 'need to spend' that is not covered by another cost driver

Stable

- the cost driver should not exhibit unpredictable or large changes year-on-year

Future proof

- the cost driver should be expected to drive the on-going costs of providing services (i.e. they should not be one-off events)

Ministry of Housing, Communities & Local Government

What are cost drivers?

- LAs and sector experts generally have a good idea about what factors drive need for services and service costs in different areas of local government spending.

- Identifying these factors and isolating national data that represents them is the first stage in determining relative need – we generally call these the cost drivers.

- However, we also need to understand the relative importance of each factor e.g. how *much* more important are the number of HGVs travelling along a section of road than snow days in determining how much it costs to maintain it?

- This is what we refer to as the weighting of a cost driver.

Ministry of Housing, Communities & Local Government

How do we determine weightings for cost drivers?

- But we also need to understand how the different cost drivers interact with each other and whether this changes the weighting we give to each of them.
- For example, both low levels of educational attainment and low income are found to be key cost drivers for Children's Social Care. However, we know that low levels of educational attainment and low income are often linked.
- In order not to 'double count' the effect of both of these cost drivers, we need to consider how they interact with each other – we need to know how much *additional* difference each cost driver describes.
- Statistical techniques can help to answer these questions.

Ministry of Housing,
Communities &
Local Government

Why use statistical techniques?

Regression against expenditure

Other statistical techniques

Regression

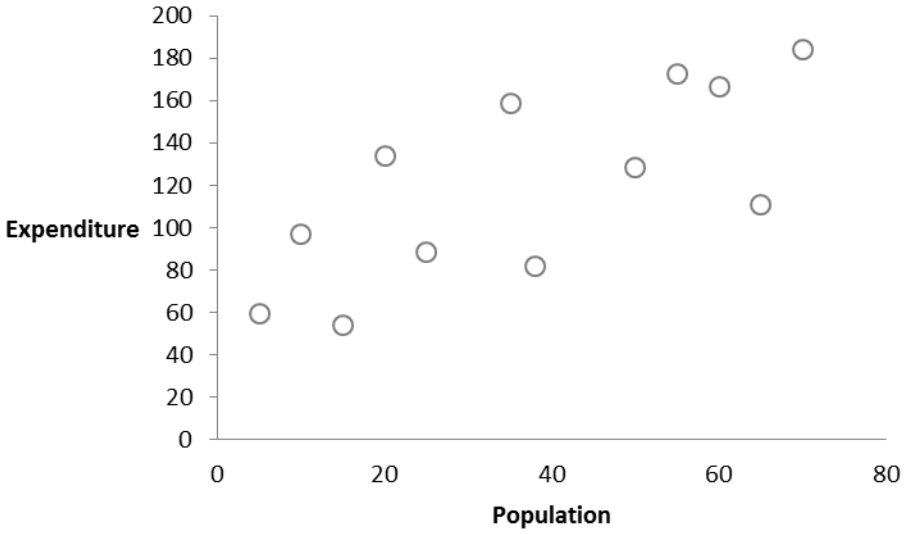
- Regressions enable us to assign a weight to each cost driver by determining the relationship between cost drivers of need e.g. population and deprivation, and a dependent variable (a measure of local authority need).

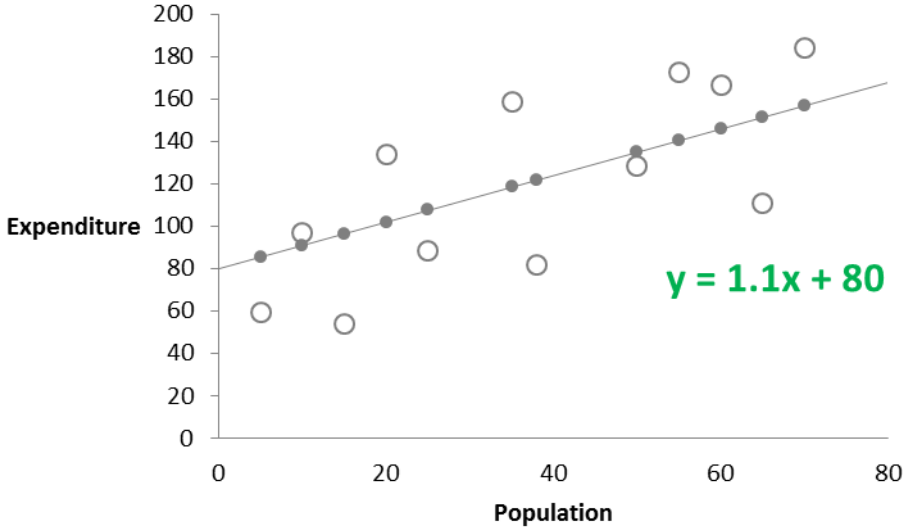
- Regression models will tell us how *much* more funding an LA should receive depending on how it performs against each of the cost drivers included. This means we can use the weights generated to allocate funding based on specific characteristics of the LA.

- The current dependent variable used in relative need assessments is local authority past expenditure. Outcome data could also be used as a dependent variable if a suitable variable is found.

- It is generally accepted that regressions based on larger datasets and smaller geographies are likely to be more analytically robust.

Example with one cost driver





- Expenditure data from several local authorities is plotted against their population (white dots).
- A line of best fit expresses an average pattern of expenditure given all England data (the grey line).
- *In this case, on average each authority spends about £80 plus an extra £1.10 for each member of their population.*

Ministry of Housing, Communities & Local Government

Example with one cost driver

- An LA's estimate of need would then be their data plugged into this national average formula (the grey dots on the grey line).

So an authority with 50 people would have an estimated relative need of:

*£80 + £1.10 * 50 people = £135*

- Were this to exceed the national ceilings, the amount for each authority would be scaled back.

- In reality we would use multiple cost drivers across several areas of spending to reflect the complexity of a needs assessment. The regression technique allows us to reduce the risk of 'double-counting' the same impact when cost drivers are closely related (as it looks at the additional variation each cost driver explains).

Ministry of Housing, Communities & Local Government

Pros and cons of expenditure-based regression

Pros:

- Up-to-date past expenditure data readily available
- Simple statistical technique
- Allocations are made on *predicted* expenditure
- Can be used to forecast needs

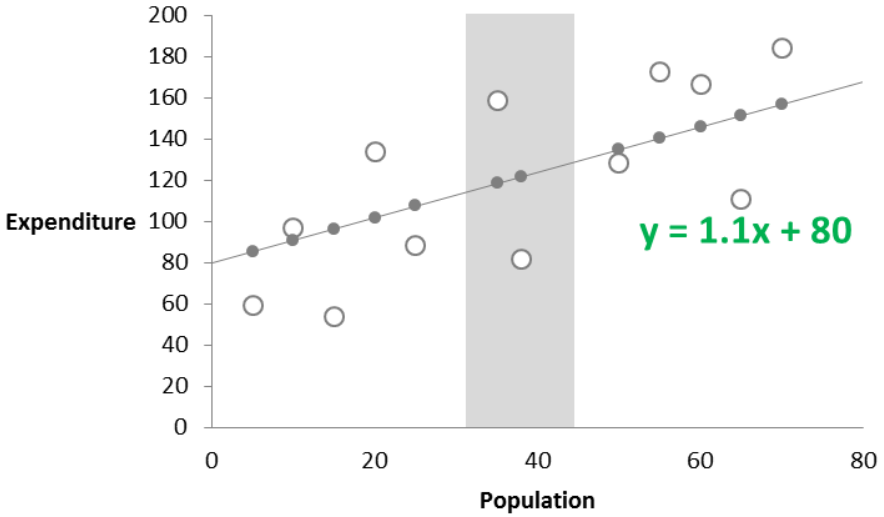
Cons:

- Does not identify **causal** relationships
- Past expenditure may not be an accurate proxy for comparing need
- Variables cannot be too correlated with one another
- Use of RO forms to produce expenditure based regression

Ministry of Housing, Communities & Local Government

Rewarding inefficiency?

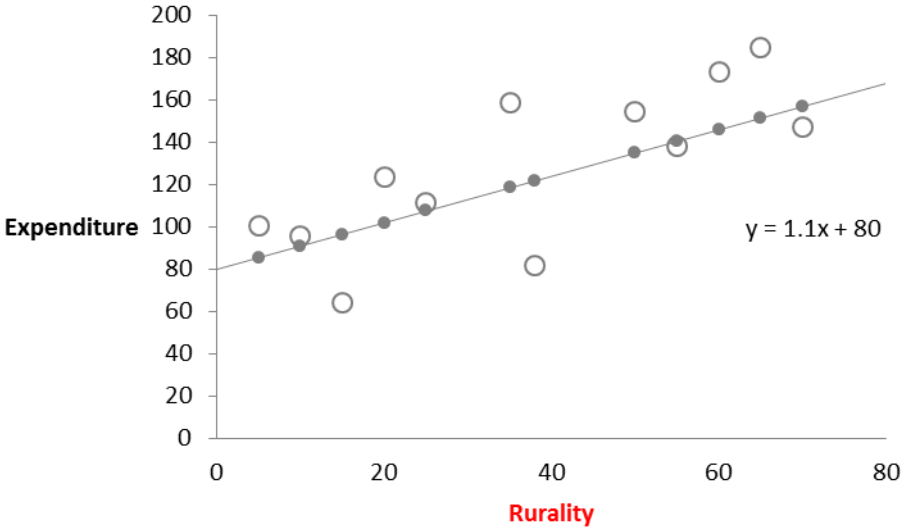
- The idea that regressions reward inefficiency is a misconception
- Regression looks at “average England” patterns of expenditure
- Any LA whose expenditure is above average (controlling for their cost factors) will be **brought down** to the average
- Likewise any LA whose expenditure is below average will be **brought up** to a level commensurate with the patterns across England as a whole
- This means that there is no penalty for efficiency – in fact it could be argued the opposite is true

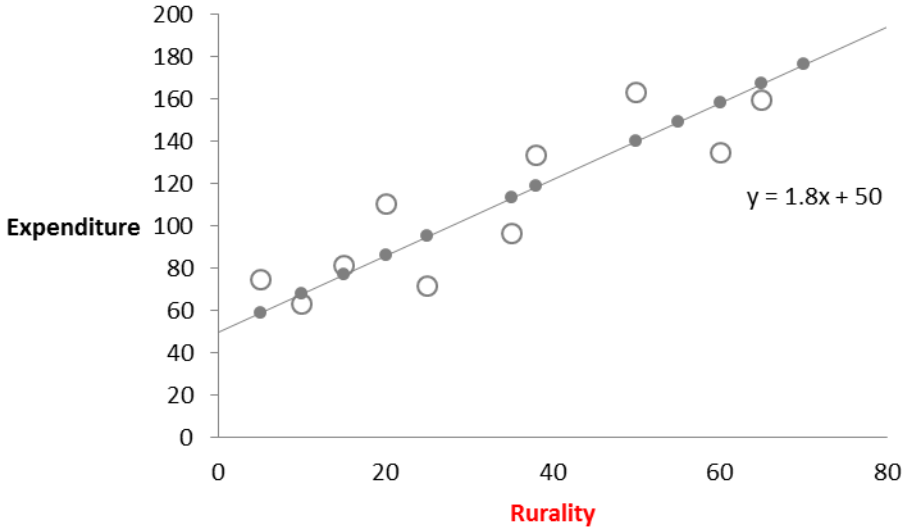


The allocations (grey dots) for two authorities with similar levels of cost drivers will be similar, even if their expenditure levels were very different.

Ministry of Housing, Communities & Local Government

Historically favoured characteristics?





However, to an extent it can be seen that historical favouring of some characteristics might be perpetuated.

- Say a former government decided to give extra funding to Local Authorities who all share a particular characteristic
- This might mean these areas have higher levels of expenditure due to something other than true need
- This would make it look like this characteristic causes them to need more funding

This is a criticism of using expenditure as a proxy for need rather than a criticism of regression. It should be noted that not all spending is defined by settlement allocations – this can mitigate the effect of historical allocations.

Ministry of Housing, Communities & Local Government

Why use statistical techniques?

Regression against expenditure

Other statistical techniques

Ministry of Housing, Communities & Local Government

Outcome based regression models

- Instead of using past expenditure, we could use **outcomes** to measure need
- In theory, outcome variables - for example the proportion of children receiving at least 5 A*-C GSEs within an area - could be the dependent variable within a regression. In addition, for many services, it may be appropriate to include the level of expenditure as an additional independent variable if this is considered to influence the outcome.

Pros:

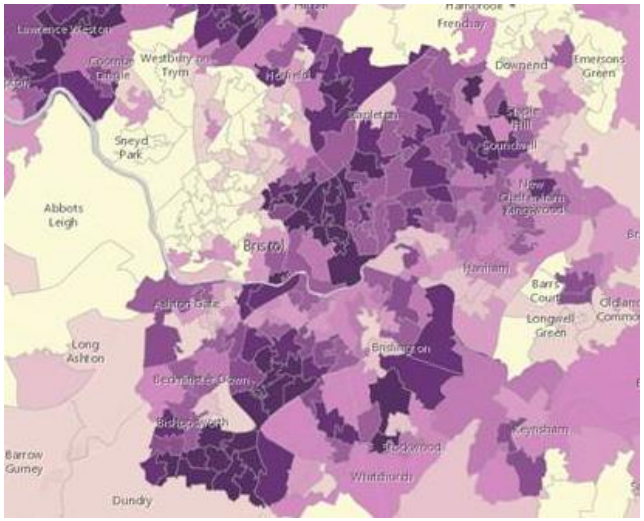
- Doesn't use past expenditure as the dependent variable
- Can use outcomes to forecast future needs **assuming the relationships between the outcome and the explanatory variables remain constant**

Cons:

- Identifying a suitable outcome measure, for which data exists, for all services may not be possible. Do all services have easily measureable outcomes?
- Many such variables could create perverse incentives, e.g. if collecting more rubbish meant more money
- Like expenditure-based regression, does not identify **causal** relationships. Still regression.

Ministry of Housing, Communities & Local Government

Multi-level modelling

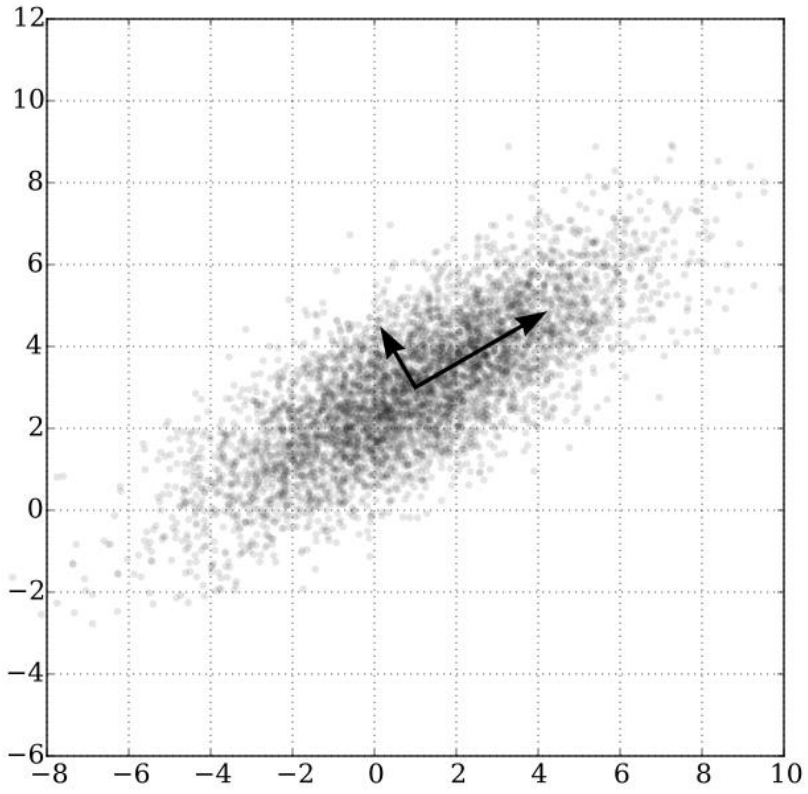


Variations in IMD within LSOAs in one local authority area

- Useful for data with a hierarchical or clustered structure - the map to the left shows how there can be significant variance within an LA)
- LA spend within LSOAs is a better proxy of need as it is more dependent on the service demand from that area and therefore less influenced by historical central Government spending patterns
- Multilevel models use small area estimation, but aim to account for variance, both within an LA and between LAs (accepting that both characteristics and place are key drivers in determining need)
- Multi-level modelling is more complex than simple linear regression models. Requires data at small geographies, which would require a bespoke and costly data collection exercise and therefore the time and cost of delivery may not be proportionate for all service areas.

Ministry of Housing, Communities & Local Government

Principal Component Analysis



Principal component analysis (**PCA**) uses a transformation to convert a set of correlated variables into a set of linearly uncorrelated variables called **principal components**. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible).

Advantages

- May not need a dependent variable e.g. past spending
- Components relate to broader concepts e.g. poverty instead of individual cost drivers e.g. population

Disadvantages

- Hard to interpret – not very transparent
- Strong evidence of a relationship needed before variable is included in the model
- Strong qualitative element needed to understand the concepts of the principle components.

Unit cost

- Need = Unit cost x Measurement unit
- E.g. the number of school and the number of pupils determine the number of head teachers, teachers and administrators required at a local authority level
- Unit cost is determined by dividing a model cost by the total number of teachers to arrive at a unit cost
- Need is then determined by multiplying unit cost by the number of measurement units (based on outputs from the previous year)

Pros

- Simple to understand and transparent in terms of calculation

Cons

- Works best in services with low variability
- Doesn't take account of all cost drivers e.g. assumes disposal of a unit of rubbish costs the same in any LA. This is particularly problematic for services such as Children's Services which have a variable unit cost across the country.

Ministry of Housing, Communities & Local Government

Options for analytical techniques

While this reflects general pros and cons of each technique, should we have more than one RNF different analytical techniques may be more or less appropriate to different service areas – this may mean a mixed approach of analytical techniques may be most appropriate

Regression (against past spending)

- ✓ Analytically robust
- ✓ Forms the basis of most of the current RNFs
- ✓ Data is readily available therefore implementation will be possible by 2019/20
- ✓ Relatively simple and transparent
- ✘ Uses past spending as a dependent variable (thus can be considered to perpetuate previous funding decisions)

Multi-level Modelling / Small Area Modelling

- ✓ Analytically most robust
- ✓ Used for revised DH Adult Social Care formula
- ✓ Controls for past spending as dependent variable
- ✘ Still uses past spending as a dependent variable
- ✘ Can be complex to understand
- ✘ Expensive and time-consuming data collection process means data may not be ready for 2019/20

Principal Component Analysis

- ✓ Does not use past spending as a dependent variable
- ✓ Data is readily available and implementation would be possible by 2019/20
- ✘ Complex to understand
- ✘ Query over analytical robustness (higher level of subjective interpretation involved)
- ✘ Considered by other Departments but rejected due to concerns about whether it is fit-for-purpose

Outcome-based regression

- ✓ Analytically robust
- ✓ Does not use past spending as a dependent variable
- ✓ Relatively simple and transparent
- ✘ Only suited to particular services where outcome variables can be identified and measured
- ✘ May require additional data collection and therefore possibility of delaying implementation beyond 2019/20