Statistical investigation of needs proxies

Contract: C204/2008/09

Final report

for

Independent Commission

on Funding & Finance for Wales

LE Wales

LE Wales

LE Wales is an economics and policy consultancy based in Wales and is a division of London Economics. London Economics also has offices in London, Dublin, Budapest and Brussels, and associated offices in Paris and Valletta.

We advise clients in both the public and private sectors on economic and financial analysis, policy development and evaluation, business strategy, and regulatory and competition policy. We are able to use a wide variety of analytical techniques to assist our work, including cost-benefit analysis, multi-criteria analysis, policy simulation, scenario building, statistical analysis and mathematical modelling. We are also experienced in using a wide range of data collection techniques including literature reviews, survey questionnaires, interviews and focus groups.

Further information about LE Wales is available at www.le-wales.co.uk.

Acknowledgements

Members of the Holtham Commission and the Commission secretariat provided us with useful guidance and feedback throughout this research. The secretariat also provided us with much of the data that we used. Members of the research team included Siôn Jones, Pau Salsas and Derek Scully (all LE Wales), Peter Watt (Institute of Local Government Studies, University of Birmingham) and Professor David King (University of Strathclyde). Professor Andrew Chesher at University College London also provided helpful feedback on some of the methodological issues at an early stage in the research. Responsibility for the contents of this report remain with LE Wales.

Contents

Page

Executive Summary

Introduction

This research was undertaken by LE Wales during 2009 for the Independent Commission on Funding and Finance for Wales (the 'Holtham Commission').

The Commission, and others, have suggested that the current Barnett Formula for the allocation of public funds to the devolved administrations of Wales, Northern Ireland and Scotland should be replaced by a funding allocation system that reflects relative needs for public expenditure.

In the debate on the use of a needs-based formula the suggestion has been made that a simple formula may be desirable even though this could lead to some loss in the accuracy with which the formula tracks relative needs. The aim of this research is to contribute to an understanding of the impact of using a simple needs-based formula rather than more complex needs-based formulae. We focus on impact in the sense of the potential loss of accuracy in reflecting variations in need.

Methods

We do not have any direct measures of need and so our approach has been to take existing complex needs-based formulae and seeing how accurately we can match the expenditure allocations implied by these formulae with those implied by a more simple formula.

In order to meet our research objective we use the formulae that allocate public expenditure resources to local authorities, NHS Primary Care Trusts and schools in England. For the purposes of our analysis, this is the complex system of formula against which we are testing simpler formulae. We use a number of methods for producing simpler formulae. Our simpler formulae have relatively small numbers of needs indicators – typically between 1 and 28, though we focus on formulae at the smaller end of this scale.

This research does not seek to develop a formula that accurately tracks any view of our own of what the underlying distribution of needs is, nor is it seeking to estimate a formula that could be used to replace the Barnett Formula.

Results and conclusions

We were able to find simple formulae that provide very similar funding allocations to the current formula-based systems in England. In these cases the loss in accuracy from using simpler formulae in place of more complex formulae appears to be very limited.

It is not possible to draw direct inferences from these results about whether or not a simple needs-based replacement for the Barnett Formula would necessarily accurately reflect relevant variations in need across the devolved nations or over time. The analysis does however provide one input into an informed judgement on whether or not a simple, as opposed to a complex, needs-based formula could be a suitable replacement for the Barnett Formula.

1 Introduction

This research was undertaken by LE Wales during 2009 for the Independent Commission on Funding and Finance for Wales (the 'Holtham Commission')[1] under contract C204/2008/09.

The Commission's own terms of reference are to:

- look at the pros and cons of the present formula-based approach to the distribution of public expenditure resources to the Welsh Assembly Government; and
- identify possible alternative funding mechanisms, including the scope for the Welsh Assembly Government to have tax varying powers as well as greater powers to borrow.

The Commission, and others, have suggested that the current Barnett Formula for the allocation of public funds to the devolved administrations of Wales, Northern Ireland and Scotland should be replaced by a funding allocation system that reflects relative needs for public expenditure.

In the debate on the use of a needs-based formula, the suggestion has been made that a simple formula may be desirable even though this could lead to some loss in the accuracy with which the formula tracks relative needs.

The aim of the research presented here is to contribute to understanding of the impact of using a simple needs-based formula rather than a more complex needs-based formula (or system of formulae). We focus on impact in the sense of the potential loss of accuracy in reflecting variations in need.

The report is structured as follows.

- In Chapter 2 we discuss some of the background to this research;
- In Chapter 3 we set out our research objectives and summarise our main methods;
- In Chapter 4 we provide our main results and conclusions

A number of Annexes provide more detail on the methods and results of our different approaches to addressing the research question.

[1] Sometimes we refer to ICFFW.

2 Context

2.1 Allocation of public expenditure resources

At present there are several tiers of government in the UK – the central government; the devolved administrations in Wales, Northern Ireland and Scotland; and the local authorities. Under the current UK system, lower tiers of government have very limited revenue-raising powers and most revenue is raised by central government through taxation. The central government allocates public funds to lower tiers of government so that they can carry out their functions.

There are many ways in which these public funds can be allocated.[2] In the UK, a common approach is to use funding allocation formulae. Such formulae determine how much resource is allocated to a particular area based on the characteristics of that area. These characteristics are intended to reflect the need for public expenditure in the area. They vary with the formula, but may include factors such as the age structure and income of the population in that area. The Barnett formula, which allocates funds from central government to the devolved administrations in Wales, Northern Ireland and Scotland is an example.[3]

The devolved administrations also use formulae to allocate funds to local authorities. In Wales, for example, a range of service-specific formulae are used to allocate the main grant from the Welsh Assembly Government to Welsh local authorities. In England, central government allocates funds directly to local authorities using a number of service-specific formulae.

2.2 Recent developments

The Barnett formula has become the subject of much debate in recent times. The House of Lords Select Committee on the Barnet Formula[4] reported in July 2009 and the Holtham Commission published its First Report[5] in the same month. The

[2] A description of the many approaches used in practice can be found in our previous research for the Independent Commission on Funding and Finance for Wales – available at: http://wales.gov.uk/docs/icffw/report/090708barnettformulaen.pdf.

[3] A description and discussion of the Barnett formula can be found in previous research undertaken for the Independent Commission on Funding and Finance for Wales by Gillian Bristow (Cardiff University) – available at: http://wales.gov.uk/docs/icffw/report/090708literaturereviewen.pdf.

[4] House of Lords Select Committee on the Barnett Formula (2009) The Barnett formula, HL Paper 139, published 17 July 2009, London: The Stationary Office Limited.

[5] Independent Commission on Funding & Finance for Wales (2009) First Report – Funding devolved government in Wales: Barnett & beyond, July.

Commission on Scottish Devolution (the 'Calman Commission') also reported in June 2009.[6]

We do not focus any further here on the Calman Commission's report except to record that the Commission's view on the potential for a needs-based replacement for the Barnett Formula was that "*need is the only basis on which grant funding can be properly justified, and it should be need for the common welfare services that comprise the social Union.*" They suggested that the Barnett Formula should continue to be used until a needs assessment across the UK was conducted.

2.2.1 House of Lords Select Committee on the Barnett Formula

The main objective of the House of Lords Select Committee was "*to examine the purpose, methodology and application of the Barnett Formula as a means of determining funding for the devolved administrations of the united Kingdom, to assess the effectiveness of the calculation mechanism to meet its purpose and to consider alternative mechanisms*".

The Committee made a number of recommendations. One of them was that devolution funding should be based on relative need and that a new system should be based on the following principles:

- "*It should consider both the baseline and any increment in funds;*
- *It should be fair and seen to be fair;*
- *It should be comprehensible;*
- *It should respect territorial autonomy; and*
- *It should be stable and predictable (para. 88)*."

The Committee also argued that a new formula based on relative need should be derived using a 'top-down' approach – using a small number of aggregate statistics. They base this argument on the following considerations (para. 91):

- *"given the priority we accord to comprehensibility, a simple approach is a high priority. While it may reasonably be countered that the cost of simplicity is a certain rough justice, we would expect that cost to be lower at the national level than locally since differences between countries are much smaller than differences between localities within countries;*

[6] Commission on Scottish Devolution (2009) *Serving Scotland Better: Scotland and the United Kingdom in the 21st Century*, Final Report – June 2009.

- *we also favour a top-down approach because of the inherently top-down nature of the exercise itself. The opposite approach, that the formula should be built up from a detailed assessment of a full range of needs locality by locality, would mean trespassing on the domains of the devolved administrations; and*

- *finally, although we are not recommending the Big Lottery Fund formula should be adopted, its formula shows that such an approach is viable."*

2.2.2 Holtham Commission First Report

In their First report, the Holtham Commission recommended that, in the medium term, the funding arrangements for Wales should be based on relative needs. The Commission propose. six desirable characteristics of systems for financing sub-national government:

- autonomy;
- efficiency;
- Stability/predictability;
- simplicity/transparency;
- accountability; and
- equity.

The Commission favours a system based on relative needs on the basis that this is equitable. The Commission also note that:

"An ideal needs-based formula would be both simple (making it easy to implement, transparent and readily explicable to the non-specialist) and complete (i.e. it would capture all relevant aspects of need). In practice there is a trade-off between simplicity and completeness. When allocating an unhypothecated block grant there is a strong case for favouring a relatively simple formula since resources allocated on the basis of needs in one field may, in practice, be used in another." (para. 3.11)

2.3 The trade off between simplicity and accuracy

Both the House of Lords Select Committee on the Barnett Formula and the Holtham Commission propose the replacement of the Barnett Formula with a needs-based formula. Both also propose that the needs-based formula should be 'simple'.

The potential trade off between simplicity and accuracy (or completeness) is recognised by both organisations. Whilst a simple formula may be desirable because it makes it more comprehensible, a simple formula may also reflect variations in need less accurately than a more complex formula because the drivers of need are complex. Other commentators also point to this trade off. McLean, Lodge and Schmuecker (2008), for example, make the following point:

"*However, too fine-grained a needs assessment can be opaque, a criticism that is often heard in relation to the system of local authority grants in England, which is based on detailed measures of relative deprivation. This can obscure procedural fairness, so a coarser but more transparent needs assessment may be preferable.*"[7].

Understanding the importance of this trade off between simplicity and accuracy will play an important role in determining the size and structure of any future needs-based formula. The main role of this research is to contribute to the understanding of that trade off. In the words of the House of Lords Select Committee, we are trying to understand how rough is the "certain rough justice" implied by a simple formula.

[7] McLean, Lodge and Schmuecker (2008) *Fair Shares? Barnett and the politics of public expenditure*, IPPR, July.

3 Research objectives and method

3.1 The research objective

The aim of this research is to contribute to an understanding of the impact of using a simple needs-based formula rather than a more complex needs-based formula (or system of formulae) for the allocation of public funds. We focus on impact in the sense of the potential loss of accuracy in reflecting variations in need.

We do not have any direct measures of need and so our approach has been to take existing complex needs-based formulae and seeing how accurately we can match the expenditure allocations implied by these formulae with those implied by a simpler formula.

In order to meet our research objective, we use the formulae that allocate public expenditure resources to local authorities, health authorities and schools in England. For the purposes of our analysis, this is the complex system of formulae against which we are testing simpler formulae. This is useful because the range of services that these formulae reflect are a significant proportion, in public expenditure terms, of the range of services that are devolved to Wales.

We use a number of methods for producing simpler formulae to test against our complex system of formulae. Our simpler formulae consist of one formula (rather than a system of formulae) and have relatively small numbers of needs indicators. We test 'simple' formulae with different numbers of explanatory variables (typically between 1 and 28 explanatory variables) in order to understand how the accuracy of the simplified formulae changes with the number of needs indicators.

It is important to be clear that this research does not seek to develop a formula that accurately tracks any view of optimal needs distribution, nor is it seeking to estimate an exact formula that could be used to replace the Barnett Formula. Rather, our focus is on understanding how well different simple formulae perform in matching the public funding allocations implied by a more complex system of formulae.

Our analysis is illustrative in the sense that it relates to the specific circumstances of the complex system of formulae for allocating funds to various authorities within England. It is not possible to draw direct inferences from this about whether or not a simple needs-based replacement for the Barnett Formula would accurately reflect relevant variations in need across the devolved nations or over time. The analysis does however provide one input into an informed judgement on whether or not a simple, as opposed to a complex, needs-based formula could be a suitable replacement for the Barnett Formula.

3.2 Methods

The first step in our research was to develop a number of 'simple' formulae that might be likely to provide similar funding allocations to the more complex formula system. The next step was then to test the funding allocations implied by these simple formulae against the funding allocations implied by the current complex system of formulae in England.

3.2.1 Developing simple formulae

We developed and tested simple formulae in three different ways, calling these 'Approach 1', 'Approach 2', and 'Approach 3'. Full details on the methods and the results of each of these three approaches are provided in the Annexes.

Approach 1

For Approach 1 we used data on the formula-based allocation of funds to English local authorities as the complex formula baseline for the testing of simpler formulae. For Approaches 2 and 3 we used the same local authorities grant and also add the formula-based allocation of funds for health and schools.

The UK central government allocates the main grant, Formula Grant, to English local authorities using a complex system called the 'Four-Block Model'. The 'Relative Needs Formulae' that we use for our analysis feed into one of the four blocks.[8] There are 15 formulae in total using more than 140 needs indicators between them. Each formula represents a different local authority service and the needs indicators used in each formula are intended to reflect the relative need for spending on that particular service.

For Approach 1 we develop a number of different single formulae with the aim of testing how each of these 'simple' formula can replicate the funding allocation implied by the current system of 13 formulae[9] and over 140 needs indicators. The number of needs indicators in our simple formulae range from 3 to 28.

Our simple formulae are developed by taking the full set of needs indicators used in the English local government formulae and using various statistical methods for eliminating indicators that contribute relatively little to the explanation of the funding allocation implied by the English local government formulae. This process reduced the number of needs indicators and led to a number of simple formulae. These were then tested using the methods described in Section 3.2.2.

---

[8] A more detailed discussion of the four-block model is provided in **Error! Reference source not found.**.

[9] We excluded two of the 15 services (police; and fire and rescue) from our analysis because the geographic areas for which these services are provided are very different to the geographic areas over which other services are provided.

Approach 2

For Approach 2 we used a similar process to Approach 1, but this time our baseline system of complex formulae was different. In addition to the system of formulae used to allocate funds to English local authorities, we added the formulae used to allocate funds to English NHS Primary Care Trusts and English Local Education Authorities (LEA Direct School Grant).[10]

We then developed a number of simple formulae with the aim of matching the funding allocation implied by the combination of the formulae used for local authorities, NHS Primary Care Trusts and LEA Direct School Grant in England.[11]

As with Approach 1, our simple formulae are developed by taking the full set of needs indicators used in the English local government formulae and using various statistical methods for eliminating indicators that contribute relatively little to the explanation of the funding allocation implied by the combined formulae as described above. This process reduced the number of needs indicators and led to a number of simple formulae. These were then tested using the methods described in Section 3.2.2.

Approach 3

For Approach 3, we undertook the same exercise as Approach 2 but we used a different set of needs indicators for our modelling. For Approach 2 we took the full set of needs indicators used in the English local government formulae as our starting point. For Approach 3 we started from the needs indicators proposed in two recent reports relating to the needs-based allocation of expenditure to the devolved nations: the report of the House of Lords Select Committee on the Barnet Formula;[12] and the first report of the Holtham Commission.[13]

Both reports provide an illustration of the authors' views on which variables might potentially be good indicators of need for use in comparisons of need at the level of England and the devolved nations. Neither claim that the needs indicators they discuss provide either a definitive or a complete list. For our purposes, they provide a useful alternative starting point for our analysis.

---

[10] Health funding accounts for about 60% of the new funding aggregate and education funding accounts for a further 20% or so. Consequently these two components account for almost 80% of all funding.

[11] As the geographic areas covered by these grants are slightly different some modifications to the data needed to be made in order to be able to use a consistent set of geographical areas. This process is explained in more detail in **Error! Reference source not found.** and was undertaken with the assistance of the Commission secretariat.

[12] House of Lords Select Committee on the Barnett Formula (2009) The Barnett formula, HL Paper 139, published 17 July 2009, London: The Stationary Office Limited.

[13] Independent Commission on Funding & Finance for Wales (2009) First Report – Funding devolved government in Wales: Barnett & beyond, July.

Our simple formulae are developed by taking four alternative starting sets of needs indicators, based on those discussed in the two reports,[14] and using various statistical methods for eliminating indicators that contribute relatively little to the explanation of the funding allocation implied by the combined formulae as described above. This process reduced the number of needs indicators and led to a number of simple formulae. These were then tested using the methods described in Section 3.2.2.

3.2.2 Testing the 'simple' formulae

For all three approaches, we used the same methods for testing how well each of the simple formulae that we developed matched the funding allocation implied by the complex formulae baselines.

We used a number of different approaches to testing the performance of the simple formulae.

First, we measured what percentage of the variation in allocations implied by the complex formulae is explained by the simple formulae (using the statistical measures R^2 and adjusted R^2).

This measure of average performance does not show whether there are individual local areas that would lose or gain significantly from the replacement of the complex formula system with the simple formula. In order to measure and compare this effect we calculated, for each simple formula, the number and the percentage of local areas that would have changes in funding allocations of less than 5% and less than 10%. We also calculated the size (in percentage terms) of the five largest changes (positive or negative).

Finally, we tested how well our simple formulae predicted funding allocations by the complex formulae in a small number of local areas that were deliberately excluded from the local areas used in the above analysis. This process is known as out-of-sample prediction.

[14] The four sets of indicators are the core set (8 indicators) and the expanded set (12 indicators) discussed in the House of Lords Select Committee report, the 11 needs indicators discussed in Chapter 4 of the Holtham Commission First Report and a set of 21 indicators that combines these three sets of indicators. Note that we were not able to find suitable measures for all indicators and because some indicators were common to the first three sets, the number of indicators in the fourth set is less than the sum of the indicators in the other three sets.

4 Results and conclusions

4.1 Approach 1

When we tested the range of simple formulae derived using Approach 1 mixed results were obtained.

In general, formulae of all sizes performed well in explaining a high percentage of the variation in the data. Even some of the smaller formulae, with 3 needs indicators, explained over 90% of the variation across local areas in the resource allocation implied by the complex formulae. The better performing larger formulae (with 10-28 needs indicators) explained as much as 99% of this variation.

When we examined the potential impacts on resource allocations to individual local areas the general performance of the simple models was less good. The best of the larger models (with 10-28 needs indicators) led to no local areas having a change in resource allocation of more than 10% as a result of replacing the complex formulae with the simple formula. One formula, with 28 needs indicators, resulted in no local areas having a change in resource allocation of more than 3%. Nevertheless, the formulae with fewer needs indicators (1-3) performed significantly less well, with the better formulae showing some local areas having changes in resource allocation in the range 15% - 70%.

4.2 Approach 2

Overall, the simple formulae derived in Approach 2 performed significantly better than the simple formulae derived in Approach 1.

The four largest formulae (with 18-30 needs indicators) performed very well in explaining more than 98% of the variation in the data. They also performed very well on other criteria with 98% - 100% of local areas having a change in resource allocation of less than 5% as a result of replacing the complex formulae with the simple formula. In the out-of-sample test 89% - 95% of local areas had a change in resource allocation of less than 5%.

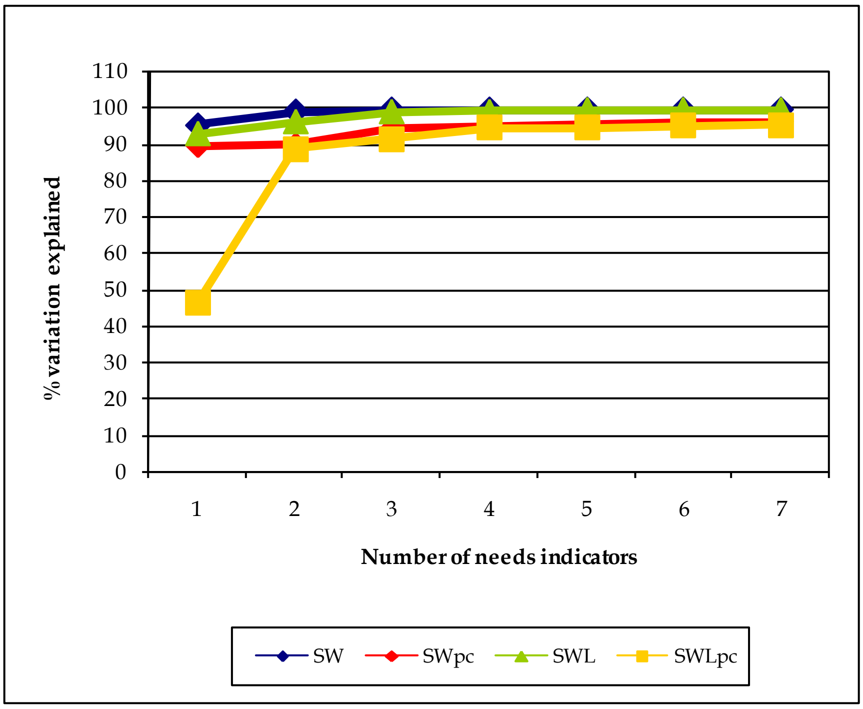
We also tested a number of smaller formulae (with 1 – 7 needs indicators). The performance of four types of formula, each with different needs indicators, is summarised in Figure 1 and Figure 2.

Figure 1 shows the four formula types (SW, SWpc, SWL and SWLpc) each represented by a different coloured line.[15] The vertical axis shows the percentage of

[15] 'pc' represents formulae where the variables are expressed in per capita terms and 'L' represents formulae where

the variation in the data explained by each formula and the horizontal axis shows the number of needs indicators in each formula. Hence each coloured line shows how the performance of each formula type changes with the number of needs indicators used in the formula. Overall the Figure shows that the performance of the formulae, in terms of the percentage of the variation explained, is strong in all cases except where there is just one needs indicator in formula SWLpc.

Figure 1: Percentage variation in the data explained by simple formulae with different numbers of needs indicators

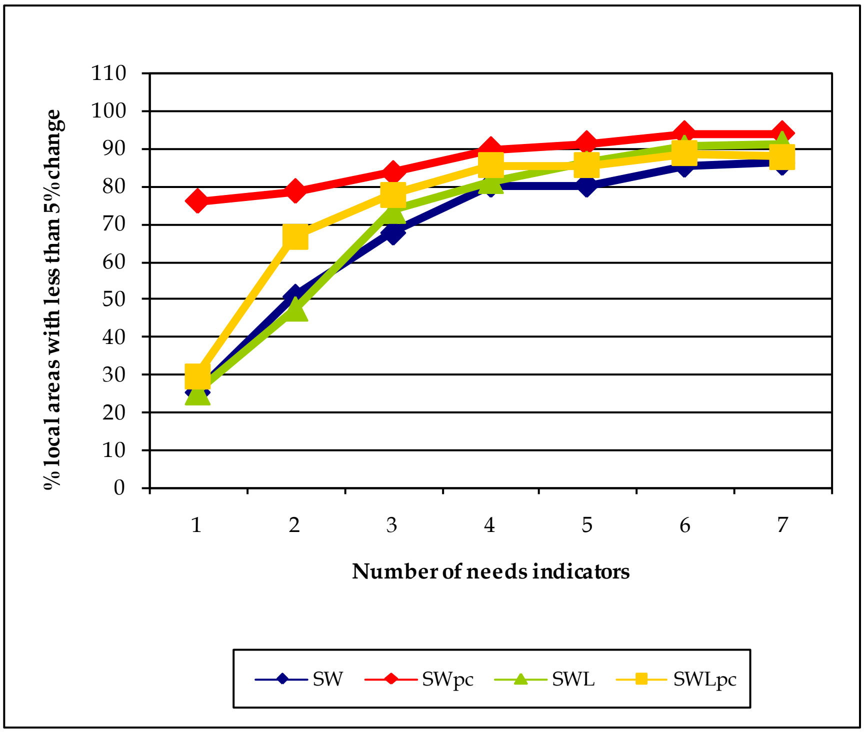


Source: LE Wales

Figure 2 has the same structure as Figure 1 but this time, for the same formulae, it shows the results of the test of the percentage of local areas that experience changes in funding allocation of less than 5% through application of the simple formula in place of the current complex system of formulae. This time the number of needs indicators in the formulae has a much stronger influence on the performance of the formulae against this test. With four and five needs indicators all four formula types show around 80% to 90% of local areas experiencing a change in funding allocation of less than 5%.

the variables are expressed in logarithms. Hence SWLpc is expressed in both logarithms and per capita terms whilst SW is expressed in levels only.

Figure 2: Percentage of local areas with a less than 5% change in funding allocation from introducing the simple formula



Source: LE Wales

The results for Approach 2 suggest that, for this particular data set, it is possible for a relatively simple formula to reasonably accurately replicate the funding allocations implied by the more complex system of formulae that we used as our baseline.

4.3 Approach 3

The results from Approaches 1 and 2 enable us to draw conclusions in respect of our research objective. Nevertheless, it was felt that it would be useful to expand on Approach 2 to see if it is possible to derive a simple formula that better reflects views on which needs indicators should be in a simple needs-based formula.

For the purposes of this research we took as our starting point the suggestions for needs indicators made by the House of Lords Select Committee on the Barnett Formula in their recent report and those made in the First Report of the Holtham Commission.

Four formulae were developed and tested initially. These included between 6 and 12 needs indicators. In general these formulae performed very well, explaining 93% -

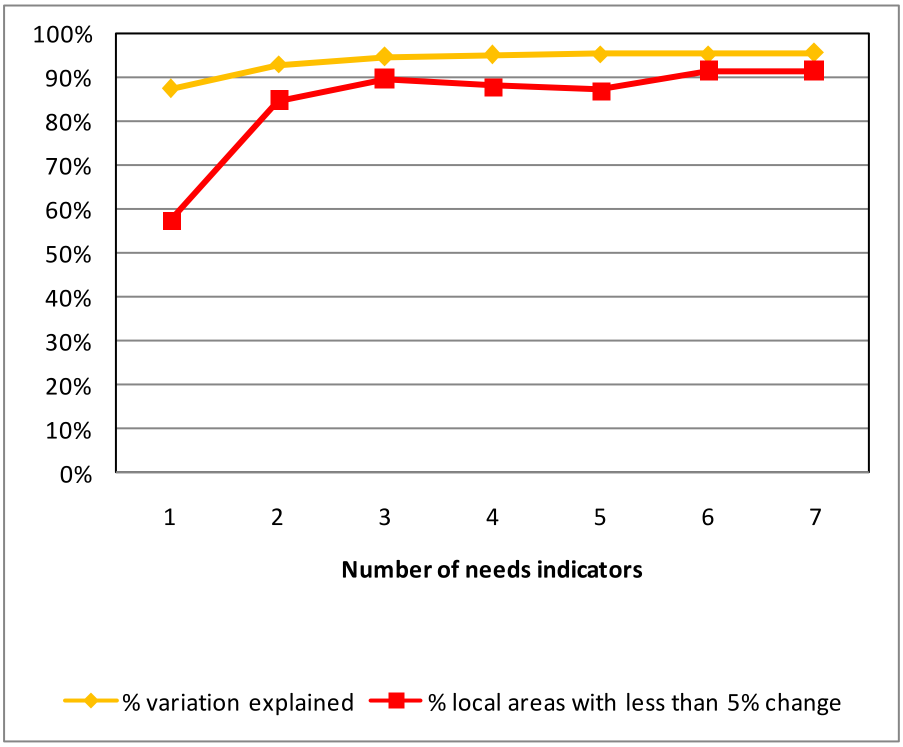
99% of overall variation and with 83% - 92% of local areas experiencing changes in funding allocation of less than 5% through application of the simple formula in place of the current complex system of formulae. Performance in the out-of-sample test was very similar.

We selected the best performer from amongst these formulae and tested the effect of reducing the number of needs indicators in the formula. The results are shown in Figure 3. The vertical axis shows percentage scores and the horizontal axis shows the number of needs indicators in each formula. The yellow line shows how the percentage of the variation explained by the formula[16] changes with the number of needs indicators included. The red line shows the percentage of local areas that experience changes in funding allocation of less than 5% through application of the simple formula in place of the current complex system of formulae.

This Figure shows the trade-off between the number of needs indicators included and the performance of the formula: performance improves significantly from 1 to 2 need indicators. Formulae with three or more needs indicators perform well on both of these criteria.

[16] Measured by R^2.

Figure 3: % variation explained and % local areas with less than 5% change by number of needs indicators



Source: LE Wales

Details on all of these models and the needs indicators that they include are provided in the Annexes but, for example, the formula with three needs indicators that is illustrated in Figure 3 includes the following needs indicators:

- % Population classified as non-white;
- % Working age population with a limiting long-term illness;
- % Children in the population whose parents are on benefit (IS/IB JSA).

For illustrative purposes, we also tested the performance of two alternative simple formulae. The first included the following three needs indicator variables:

- % children in the population whose parents are on benefit (IS/IB JSA);
- the inverse of an index of gross household disposable income per head; and
- % of the working age population who are unemployed.

This model explained 90% of overall variation; 70% of local areas experience changes in funding allocation of less than 5% through application of this simple formula in place of the current complex system of formulae; and 95% of local areas experienced changes in funding allocation of less than 10%.

The second alternative simple formula included the following four needs indicator variables:

- % children in the population whose parents are on benefit (IS/IB JSA);
- % working age population with a limiting long-term illness;
- % population classified as non-white; and
- the dependency ratio (population not of working age).

This model explained 95% of overall variation; 88% of local areas experience changes in funding allocation of less than 5% through application of this simple formula in place of the current complex system of formulae; and 98% of local areas experienced changes in funding allocation of less than 10%.

4.4 Overall conclusion

It is important to be clear that throughout our research we have not been trying to find a model that accurately reflects the distribution of expenditure needs, nor have we been trying to find a model that could act as a suitable replacement for the Barnett formula.

Instead, our aim has been to test whether it is possible for a simple formula to replicate reasonably accurately the funding allocation implied by a more complex system of funding allocation formulae. We have done that by comparing the funding allocations suggested by the current complex system of formulae in England with the funding allocations implied by a number of alternative 'simple' formulae that we have developed during the course of this research.

Using data on the allocation of funds to local government only, in Approach 1, we were unable to find a simple formula with a limited number of needs indicators that could be used to make a very similar funding allocation to the current local government formula system used in England. Using data on the allocation of funds to local government, health and education expenditures in aggregate (Approaches 2 and 3), however, we were able to find simple formulae that provide very similar funding allocations to the current formula-based systems in England. In these cases the loss in accuracy from using simpler rather than more complex formulae appears to be very limited using several statistical measures.

It is not possible to draw direct inferences from these results about whether or not a simple needs-based replacement for the Barnett Formula would necessarily accurately reflect relevant variations in need across the devolved nations or over time. The analysis does however provide one input into an informed judgement on whether or not a simple, as opposed to a complex, needs-based formula could be a suitable replacement for the Barnett Formula.

Annex 1 Approach 1 – Local government spend

A1.1 Introduction

This Annex details our method and results for Approach 1, which is an analysis of the distribution of funds to English local authority areas through the English local government funding formulae. The analysis is mainly explorative and is interesting because it develops some of the methods used further in Approaches 2 and 3 and which constitute the main results for this research.

This Annex has the following structure:

- In Section A1.2 we provide some background material on the allocation of funds to local authorities in England;
- In Section A1.3 we provide a description on the data used;
- In Section A1.4 we describe the results of our analysis for Approach 1.

It should be noted that the original analysis was based on models which use the Relative Needs Amount (RNA) as the dependent variable. Following the presentation of results and discussion with ICFFW it was agreed to test alternative models with variables expressed per-head and in logarithms as well as using Relative Need Formulae (RNF) as a dependent variable.

We believe that RNF is more appropriate as a dependent variable in this context because it aims to reflect need, whereas RNA is simply the authority's share of the relative need block control total. In consequence, the key results to consider in this Annex are those that use RNF, rather than RNA, as the dependent variable. These models are presented in section A1.4 as 'alternative specifications'. The analysis using RNA as the dependent variable is retained in this Annex for completeness.

In addition, there are a number of appendices to this Annex, which provide further detail on the data, the method and the results.

A1.2 Funding allocation in England

At the present time, allocation of funding to local government in England is based on the 'Four-Block Model' formula distribution system, introduced in 2006/07 as a replacement of the previous method of distribution (the Formula Spending Shares, FSS).

The allocation is based on the relative annual needs of each local authority for a total of three years. The relative needs are calculated as the aggregation of a number of services, and are provided at different levels of local government. This section provides some background to the English model of funding, and describes how are relative needs calculated for the different services provided and local units of funding.

Background

The current model allocates funding to the different regions according to four different concepts or blocks:

- The Relative Needs Block;
- The Relative Resource Block;
- The Central Allocation; and
- Damping.

Our analysis focuses on the Relative Needs Block, which is designed to determine the different authorities' relative "needs" (i.e. their needs above a minimum threshold) in different service areas using a Relative Need Formulae (RNF)[17]. The Central Allocation Block provides a basic fixed amount per authority, while the other blocks can be seen as adjustments of the central and relative blocks for needs[18].

The size of the Relative Needs Block is set by ministers as a percentage of the overall Formula Grant available, and for 2008/09 its size is £17,046,680,143, or 73.0% of the total available to receiving-authorities.

Relative Needs

Services provided

The needs are split into the following seven categories of needs (or service groups):

- Children's Services
- Adults' Personal Social Services
- Police
- Fire & Rescue

[17] The formulae used are similar to those used in the old system. However, the Relative Need Block is a cash block (as opposed to a measure of relative spend or "need" as is often used in the FSS).

[18] The Relative Resource Block is a negative block and removes funding from those authorities with a larger tax base relative to their population. The Damping block is set to scale back authorities above a certain floor.

- Highway Maintenance
- Environmental, Protection and Cultural Services (EPCS) and
- Capital Financing.

The relative needs of each authority are calculated from different *indicator variables* using the RNFs. In turn, indicator variables are calculated from *underlying data* obtained from different sources (through calculations described in the methodology of the Four-Block Model)[19].

For example, the RNF for the Youth and Community (Y&C) sub-group is calculated as the projected population aged 13 to 19 multiplied by the Y&C Basic amount, and multiplied by a number of factors to account for a deprivation top-up[20], ethnicity top-up, area cost adjustment, and a specific Y&C scaling factor. The final RNF value is expressed in units divided by 10,000,000,000.

Details of each service group and sub-groups are given in Appendix A1.6, together with a full list of indicators and underlying data for each service and subservice.

Units of funding

Funds are provided at different local government and local authority levels for the different needs. The majority of services are provided at the same local level (or unit), but there are special provisions for EPCS, Capital, Police, and Fire and Rescue needs. The units of funding for different services are the following:

- Principal unit: all services (except Police, and Fire and Rescue) are provided for each of the principal units (counties; unitaries; Metropolitan districts; and Greater London).
- Additional funding: EPCS and Capital needs are also determined at the district level (this is, within and independent of County needs).
- Exceptions: Police, and Fire and Rescue are provided at a different unit level.
    - For Police, the relevant authorities are: London–City and Greater London Authority; Metropolitan authorities; and Shire Police Authorities.

[19] For services related to children, adults and EPCS, relative needs are determined as the sum of different sub-group services.

[20] Deprivation top-up is an indicator variable that is calculated from underlying data as the proportion of Children of IS/IBJSA Claimants above the threshold multiplied by 57.7598.

- For Fire and Rescue the relevant authorities are categorised as: Counties with Fire Responsibility; Unitaries with Fire Responsibility; Combined Fire Authorities; and Metropolitan Fire Authorities.

A1.3 Data

The data for the analysis uses information from the English funding allocation model[21]. The dataset combines needs, indicator variables and underlying data of the five service groups with common units.

The different variables and observations used in the model have been defined as follows (a detailed description of the observations contained in the data is provided in Appendix A1.7).

Dependent variable

For Approach 1, we use the Relative Needs Amount (RNA), which is calculated as the authority's "need" (from the Relative Need Formulae) compared to that of other authorities providing a similar service. An authority's RNA is their share of the relative need block control total.

We need to note the following:

- Information related to Police and Fire and Rescue are not included in RNA (and in the analysis) because data are not consistently separated along the same geographic or administrative boundaries as the other service groups.

- For EPCS and Capital services the different variables provided at the district level have been added to the principal unit. Hence, for each county, the values of these services include the needs of each of the county's districts in addition to the needs at the county level.

Independent variables

For Approach 1, we use indicator variables and underlying data that constitute the service groups RNF. There is a total of 143 variables (93 underlying variables and 50 indicators and 28 additional variables used in the calculations of RNF and RNA values).

[21] We are grateful to the Commission secretariat for providing the data in raw form used in the calculations of the Relative Needs Formulae. We are also grateful to Jennifer Owens (SCT Technical Support Team, Treasury Department, Somerset County Council) for answering queries and providing additional details on the Four Block Model.

Underlying and indicator variables largely consist of demographic and socio-economic variables. For simplification purposes, we have grouped all the variables into four broad categories (and different subgroups) (Table 1).

Table 1: Independent variables: Groups and sub-groups

Group	Subgroup	No. of vars	Group	Subgroup	No. of vars
Environ ment	Weather	3	Transport	Road Length	5
	Additional environmental expenditure	3		Traffic flow	4
Financial	Area cost adjustment	6	Population	Population (general)	46
	Assumed debt	11		On benefit / disadvantaged	29
	Estimated supported capital expenditure	8		Ethnicity	11
	Debt repayment and interest	2		Sparsity / density	14
Total		33			109

Observations used - summary

Information on local government funding is provided for a total of 149 district/observations. This is: 34 non-metropolitan counties, 46 unitary districts, 36 districts for the 6 metropolitan counties, and 33 London Boroughs.

One of the principal objectives of the study is to assess the accuracy of prediction of the new proposed models. We do this by testing the accuracy in “out-of-sample” prediction, which analyses the performance of the model on a different sample. The usual procedure to perform out-of-sample prediction is to leave some observations aside which will not be used for the analysis and estimation of the parameters. In the assessment stage, the model prediction performance is tested against these observations to compare observed with predicted values.

Overall the dataset has:

- Observations: 126 for estimation, and 23 for out-of-sample prediction[22]
- Variables: 143 (93 underlying variables and 50 indicators).

A detailed description of the in-sample and out-of-sample observations is provided in Appendix A1.7.

[22] Data for out-of-sample prediction was obtained as a 15% random draw (by district classes).

A1.4 Results

This section presents the results of the statistical and econometric prediction exercise.

Our analysis follows a number of differentiated stages.

1) Reduction of number of available variables
2) Estimation of alternative regression models
3) Assessment of appropriateness of the model and goodness-of-fit
4) Assessment of model accuracy by testing performance in out-of-sample prediction
5) Parsimonious model
6) Alternative specifications
7) Parsimonious model: Alternative specifications

A1.4.1 Reduction of number of variables

This was done using three different methods:

- Principal components analysis;
- Stepwise backward selection; and
- Manual/judgemental selection.

The results for each method are outlined below, in turn.

Principal components analysis

Principal components (PC) analysis is a statistical technique for data reduction. Taking a number of variables (possibly correlated) it transforms these into a smaller number of uncorrelated linear combinations of the variables (called principal components) that explain most of the data's variation.

PC analysis was undertaken for the 143 indicators and underlying variables in the data set. Additionally, PC analysis was undertaken separately for the variables used to compute each of the 5 service need blocs. The principal axis method was used to extract the components, and this was followed by a varimax (orthogonal) rotation. The components that displayed eigenvalues greater than 1, were retained and the results verified by visual inspection on a screen plot.

The analysis including all variables yielded 11 components. For Adult, EPCS and Highway Maintenance we selected four components which were able to explain, respectively, 91%, 86% and 85% of total variation. For Children five components satisfied our selection criteria, explaining 92% of the total variation. For capital just the first component was selected and it was capable of explaining 93% of the total variation. Detailed principal components calculations are shown in Figure 4.

Figure 4: Principal components. Detailed results after varimax rotation

PC1: All indicators and variables

```
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =        11
                                                  Trace            =       138
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.9362

    --------------------------------------------------------------------------
       Component |       Variance    Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |       56.8953       42.4722              0.4123       0.4123
           Comp2 |       14.4232        1.5195              0.1045       0.5168
           Comp3 |       12.9037       1.68193              0.0935       0.6103
           Comp4 |       11.2217       3.44279              0.0813       0.6916
           Comp5 |       7.77893      .0308823              0.0564       0.7480
           Comp6 |       7.74805       2.25921              0.0561       0.8041
           Comp7 |       5.48884       .961394              0.0398       0.8439
           Comp8 |       4.52745       1.08402              0.0328       0.8767
           Comp9 |       3.44343       .630549              0.0250       0.9017
          Comp10 |       2.81288       .856217              0.0204       0.9221
          Comp11 |       1.95666             .              0.0142       0.9362
    --------------------------------------------------------------------------
```

PC2: Indicators and variables used for RNF Adults

```
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =         4
                                                  Trace            =        34
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.9099

    --------------------------------------------------------------------------
       Component |       Variance    Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |       15.7266        9.5002              0.4625       0.4625
           Comp2 |       6.22636      .0356805              0.1831       0.6457
           Comp3 |       6.19068       3.39915              0.1821       0.8278
           Comp4 |       2.79154             .              0.0821       0.9099
    ----------------------------------------------------------------------
```

PC2: Indicators and variables used for RNF Children

```
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =         5
                                                  Trace            =        46
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.9222

    --------------------------------------------------------------------------
       Component |       Variance    Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |       17.6558       7.72484              0.3838       0.3838
           Comp2 |       9.93097       1.57795              0.2159       0.5997
           Comp3 |       8.35302        4.8118              0.1816       0.7813
           Comp4 |       3.54122       .601012              0.0770       0.8583
           Comp5 |       2.94021             .              0.0639       0.9222
    --------------------------------------------------------------------------
```

PC2: Indicators and variables used for RNF Capital

```
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =         1
                                                  Trace            =        21
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.9268

    --------------------------------------------------------------------------
       Component |       Variance    Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |       19.4618             .              0.9268       0.9268
    --------------------------------------------------------------------------
```

```
PC2: Indicators and variables used for RNF Adults ECPS
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =         4
                                                  Trace            =        23
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.8557

    --------------------------------------------------------------------------
       Component |       Variance   Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |        8.20332       3.1454              0.3567       0.3567
           Comp2 |        5.05792      1.66127              0.2199       0.5766
           Comp3 |        3.39665      .373777              0.1477       0.7243
           Comp4 |        3.02287            .              0.1314       0.8557
    --------------------------------------------------------------------------

PC2: Indicators and variables used for RNF Highway Maintenance
Principal components/correlation                  Number of obs    =       126
                                                  Number of comp.  =         4
                                                  Trace            =        18
    Rotation: orthogonal varimax (Horst off)      Rho              =    0.8499

    --------------------------------------------------------------------------
       Component |       Variance   Difference          Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |        5.52504      1.96452              0.3069       0.3069
           Comp2 |        3.56051     .0156797              0.1978       0.5048
           Comp3 |        3.54483      .876888              0.1969       0.7017
           Comp4 |        2.66795            .              0.1482       0.8499
    --------------------------------------------------------------------------
```

Source: LE Wales.

Each selected component was related to one of the original variables (this is usually referred as a "surrogate" in the literature) using the factor loading of the variables in each principal component.[23] The surrogate variable was selected from among those with the highest factor loadings.[24]

Hence, PC analysis selected two sets of variables:

- 11 variables were selected using the analysis for the whole variables dataset (PC1); and
- 18 variables were selected from the 5 different service need blocs (PC2).

The different variables selected are summarised in Table 11 (for PC1 and PC2) in Appendix A1.8 of this Annex.

Interestingly, the variables selected reflect the different groupings we had previously identified within the independent variables (Table 1). Both PC1 and PC2 identify one

[23] A variable's factor loading is the correlation coefficient between the variable and the factor.

[24] The factor loadings were generally low with factor loadings ranging from between 12.6% and 63.3% for PC1, and 20.2% and 59.8% for PC2.

or more variables from each of the four main groups. In addition, PC1 selects at least one variable from 9 of the 12 subgroups resulting from the grouping exercise, and PC2 chooses at least one variable from 8 of the subgroups.

Stepwise (backward selection)

Stepwise (SW) backward selection is a method of data reduction that excludes variables by subsequent iterations of a regression based on the statistical significance of the estimated coefficients[25]. A p-value of 0.01 was used as a critical value for statistical significance.

As before, SW selection was performed by including all the variables (excluding linear combinations) in a regression containing the total relative needs value, and for separate regressions for the 5 different service need blocs.

SW selected:

- 28 variables when including all variables and
- 22 for independent regression models for the 5 different services.

The different variables selected are summarised in Table 11 (for SW1 and SW2) in Appendix A1.8 of this Annex.

Judgemental selection

Finally, we used our own experience and judgement (OJ) criteria to select a subset of variables that could be expected to explain, theoretically, the relative levels of needs.

Our criteria were based on the previous grouping established for 143 potential independent variables, in Table 1. For each group we chose the variable that was best suited to explain the relative needs of a local authority. Our reasoning also used our own knowledge on the indicators and underlying variables used to compute the relative needs formulae.

Through OJ we selected 15 variables to be included in the regression model[26].

Further details of the variables included are contained in Table 11 in Appendix A1.8 of this Annex, as OJ1.

[25] Under a stepwise backward selection approach, the first regression contains all variables specified in the model. Following the first estimation variables whose estimated coefficients have greater p-values than that specified as being the minimum in the model, are dropped. This process is repeated until all of the variables in the final model estimated have estimated coefficients with p-values less than the minimum specified.

[26] The variables were chosen to reflect the different groupings: 2 financial, 11 population, and 2 transport.

A1.4.2 Estimation of alternative regression models

We estimated two sets of models. In a first step, we estimated the regression models using the set of variables obtained with different selection processes. This resulted in models PC1, PC2, SW1, SW2, OJ1.

We refined such models by reducing the number of variables using statistical significance and own judgement. This resulted in models PC3, PC4, SW3, SW4, OJ2.

The methods used for estimation were standard OLS, without any previous transformation of the data. The variables included in all of these models are shown in Table 11.

A1.4.3 Assessment of appropriateness of the model and goodness-of-fit

The goodness of fit of the models, in summarised form, are provided in Table 2, for original (PC1, PC2, SW1, SW2, OJ1) and refined models (PC3, PC4, SW3, SW4, OJ2).

For each model, we provide the number of observations (126 in all cases), the number of variables included by different methods (ranging from 7 to 28), and the R-squared and adjusted R-squared[27].

In general, all models have very high R-squared (for models with 18 or more variables the R-squared is never below 90%, which suggests a very good fit).

Nevertheless, a closer look at the models' residuals (this is, the difference between the relative needs amount and the values fitted by the model) shows that these are noticeably high in a number of cases. For example, looking at the five observations with higher errors (expressed as a percentage of needs) we can observe that in all models the largest error is greater than 40%, and in 6 of the 10 models the largest error is more than 100% of the observed needs. A table of residuals and the graph plots for the different models is reported in Appendix A1.9 of this Annex.

We also examined whether there were significant differences in the residuals' dispersion for each model and type of local authorities for which government funding is provided (counties, unitaries, metropolitan districts and London), and, although these are generally larger for unitaries, we found no systematic patterns in the remaining groups (counties, metropolitan districts, and London boroughs). This is shown in Figure 6 and Figure 7, in Appendix A1.9 of this Annex.

As an alternative measure, we have also calculated, in Table 2, the number of observations which fall within the boundaries of a pre-determined critical error

[27] R-squared and adjusted R-squared are common measures of goodness-of-fit used to assess the relationship between the outcome of the models and the original values. Typically, a low R-squared indicates poor fit, while and R-squared of close to 1 (or 100%) indicates very good fit.

band. The critical errors are subjective and have been chosen as 5% and 10%, so that for each model it can be shown the number of observations within each of those bands. Only model SW1 with 28 variables is able to predict 80% observations within a 5% error band. Finally, only the two SW models with more than 20 explanatory variables are only able to predict more than 90% of the observations within a 10% error band. The remaining models only predict between 53% and 85% of the observations within a 10% error band.

Table 2: Results of regression models

	Preliminary models					Refined models				
Model	**PC1**	**PC2**	**SW1**	**SW2**	**OJ1**	**PC3**	**PC4**	**SW3**	**SW4**	**OJ2**
Number of obs.	126	126	126	126	126	126	126	126	126	126
Number of vars.	11	18	28	22	15	9	11	8	7	10
R2 (%)	95.8	99.1	99.8	99.8	98.3	95.7	98.9	98.0	98.3	98.1
adj R2 (%)	95.4	98.9	99.8	99.8	98.0	95.4	98.8	97.9	98.2	98.0
largest errors (%): 1	107	78	68	40	173	109	112	176	92	158
2	68	68	20	21	51	82	80	39	44	55
3	53	33	13	19	42	57	59	38	40	39
4	49	31	13	14	31	52	44	38	36	27
5	46	29	11	13	29	52	36	31	35	25
Obs within 5% error band (%)	41 (33%)	74 (59%)	101 (80%)	95 (75%)	67 (53%)	43 (34%)	68 (54%)	47 (37%)	67 (53%)	62 (49%)
Obs within 10% error band (%)	69 (55%)	106 (84%)	121 (96%)	118 (94%)	98 (78%)	67 (53%)	100 (79%)	91 (72%)	98 (78%)	95 (75%)

A1.4.4 Assessment of model accuracy by testing performance in out-of-sample prediction

We now consider the performance of the models in out-of-sample prediction (this is, in predicting the needs for a subset of observations that have not been used at any stage of the analysis conducted so far).

Not surprisingly, the models do not perform well in predicting out-of-sample observations. The models are only able to predict less than half of out-of-sample observations within a 5% error band (Table 3).

Table 3: Prediction error of out-of-sample prediction											
		Preliminary models					Refined models				
Model		PC1	PC2	SW1	SW2	OJ1	PC3	PC4	SW3	SW4	OJ2
Number of out-of-sample obs.		23	23	23	23	23	23	23	23	23	23
largest errors (%):	1	89.9	50.7	25.3	46.5	26.2	93.3	76.8	44.6	42.9	23.8
	2	46.2	23.8	13.9	26.3	19.9	26.7	69.5	21.0	23.1	18.6
	3	32.6	17.4	11.3	8.1	16.8	24.8	18.6	19.2	13.4	16.6
	4	23.6	13.7	11.0	7.9	16.3	20.4	13.7	15.4	11.1	15.9
	5	22.8	12.6	10.8	7.5	15.1	19.6	13.0	14.9	9.5	15.7
Obs. within 5% error band (%)		2 (9%)	9 (39%)	10 (43%)	11 (48%)	6 (26%)	3 (13%)	11 (48%)	6 (26%)	11 (48%)	8 (35%)
Obs. within 10% error band (%)		8 (35%)	15 (65%)	18 (78%)	21 (91%)	12 (52%)	7 (30%)	16 (70%)	17 (74%)	19 (83%)	12 (52%)

A1.4.5 Parsimonious model (OJ3)

One of the objectives of the study was to test the performance of models with a reduced number of variables. We have seen how different models with different numbers of variables (ranging from 9 to 28) achieve very modest results.

Not surprisingly, any model with fewer variables will only be inferior. As an example we provide a new model that includes 3 variables only. The independent variables included:

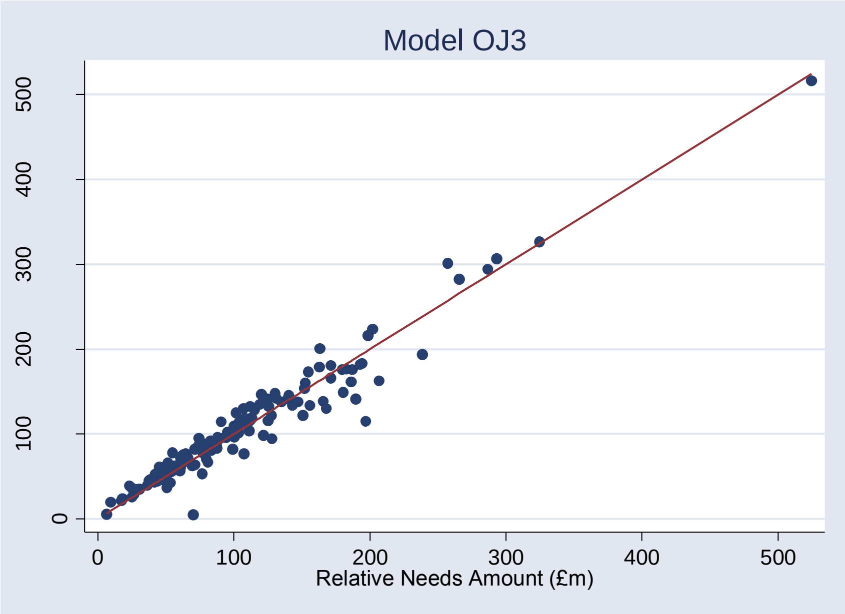
- The resident population (number);
- The number of IS/IB JSA claimants aged 18-64 years; and
- Area cost adjustment for adult personal social services, (the largest single expenditure block).

The model shows an R-squared of 94%. However, the prediction is again poor in terms of the residuals: the largest error is greater than 100% of the modelled needs (Table 4). Out-of-sample prediction is also limited, as it only predicts 6 observations (26%) within a 5% error band (and 14, or 61%, within a 10% error band).

Table 4: Results of regression model (OJ3)		
Model		**OJ3**
Number of obs.		126
Number of vars.		3
R2 (%)		93.9
adj R2 (%)		93.8
largest errors (%):	**1**	109
	2	82
	3	57
	4	52
	5	52
Prediction error	**Obs within 5% error band (%)**	37 (29%)
	Obs within 10% error band (%)	65 (52%)
Out-of-sample prediction errors	**Obs. within 5% error band (%)**	6 (26%)
	Obs. within 10% error band (%)	14 (61%)

A plot of the actual needs values and those predicted by the model is shown in Figure 5. The relationship between the observed and predicted values is given by the 45-degree line. The closer the observations are to the line, the better the model performed in predicting the actual values.

Figure 5: Actual and predicted RNA values – Model OJ3



Source: LE Wales.

A1.4.6 Alternative specifications

Following the presentation of results of Approach 1 to the ICFFW it was agreed to include RNF[28] as a dependent variable, and using alternative models for both RNA and RNF with variables expressed per-head[29] and in logarithms.

Independent variables drawn from the 2001 Census have been expressed in per capita terms by dividing them by the 2001 resident population (from the Census). Variables that were originally in per capita terms in the four block model have been left unchanged.

There were a number of zeroes in a few of the independent variables so instead of using the logarithm of the variable we used the logarithm of the variable plus one, i.e. for each variable *x* we used *ln(x+1)*.

The strategy for variable selection was the stepwise backward selection procedure (from general to specific) with a significant level threshold of 0.01. We included all

[28] The RNFs for the different services are expressed in percentage terms, adjusted for costs and multiplied by a scaling factor (and divided by 10^{10}). A total RNF variable (trnf) was calculated as the simple sum of the five different RNF (children; adults: highway maintenance: environmental protective and cultural services: and capital financing) and multiplied by 10^{10}.

[29] Variables to reflect needs per head have been constructed by dividing the needs variable for each of the local units by the 2009 expected population in that unit, and multiplying by 10^{6}.

the variables used for the calculation of the RNFs except: duplicated variables, indicator variables (i.e. transformations from underlying data), variables referred as a total (in the case of models per capita), and variables that were collinear.

Four step-wise models were estimated for RNA and RNF dependent variables:

- RNA per capita and independent variables per capita (SWpc1)
- RNA in logs and independent variables in logs (SWl1)
- RNF index and independent variables per capita (SWpc2)
- RNF in logs and independent variables in logs and per capita (SWl2)

All models use more than 10 variables selected with the step-wise procedure, and all models show very high R2 values, very close to 100% (Table 5).

Because we are interested in comparing the performance of the models in predicting needs for each local unit we make the following transformations[30]:

- For per-capita models we construct a needs predicted value ($\hat{y}$) that is a result of the model fit ($\hat{\hat{y}}$) times the population and divided by 10^6. This is $\hat{y} = \hat{\hat{y}} \times p^* / 10^6$.
- For the variables logged we construct a needs predicted value ($\hat{y}$) as the exponential of *y* less one 1. This is $\hat{y}$ = *exp(y)-1*.

The performance of the models is assessed by comparing the differences in original needs and predicted needs values for each local unit. Hence, we are interested in the analysis of the residuals, *r*, defined as $r = (y - \hat{y}) / y \times 100$.

Models in per capita, show quite high residuals, reaching 14% and 30% differences of original values, for RNA and RNF, respectively (residuals are shown in Table 5 for the different models against their dependent variables, RNA and total RNF). However, models in logarithms perform quite well: in both models residuals are never greater than a 10% off the original value (for both RNA and RNF).

It is even more noticeable that in the case of the RNF logged model (SWl2), the residuals are really low: all residuals are within a 3% error band, and only 3 are outside a 2% (Table 5).

[30] These are simply a "de-construction" or reverse process of the calculation of the per-head and logged variables.

Table 5: Results of alternative specifications					
Model		SWpc1	SWI1	SWpc2	SWI2
Number of obs.		126	126	126	126
Number of vars.		28	28	10	28
R2 (%)		99.9	99.7	99.1	99.9
adj R2 (%)		99.9	99.7	99.1	99.9
largest errors (%):	1	13.7	9.4	30.0	2.8
	2	8.7	9.0	28.9	2.5
	3	8.6	7.6	28.1	2.5
	4	7.7	7.3	26.5	2.0
	5	7.1	7.0	26.2	2.0
Obs within 5% error band (%)		118 (94%)	107 (85%)	68 (54%)	126 (100%)
Obs within 10% error band (%)		125 (99%)	125 (99%)	95 (75%)	126 (100%)

We now consider the performance of these same models in out-of-sample prediction. There is again a clear separation between model SWl2 and the rest. The former shows a very high predicting performance, whereas the rest of models predictions are limited or poor (Table 6).

Table 6: Prediction error of out-of-sample prediction					
Model		SWpc1	SWI1	SWpc2	SWI2
Number of obs.		126	126	126	126
Number of vars.		28	28	10	28
R2 (%)		99.9	99.7	99.1	99.9
adj R2 (%)		99.9	99.7	99.1	99.9
largest errors (%):	1	32.6	18.7	37.2	2.9
	2	7.1	15.4	25.9	2.5
	3	6.7	12.7	20.0	2.5
	4	6.6	8.7	18.8	2.5
	5	6.1	7.7	15.0	2.2
Obs within 5% error band (%)		18 (78%)	11 (52%)	6 (26%)	23 (100%)
Obs within 10% error band (%)		22 (96%)	20 (87%)	13 (57%)	23 (100%)

A1.4.7 Parsimonious models: alternative specifications

Starting from the model with best fit (SWl2) we tried with simplified models, using 1, 2 and 3 variables. Models Pl1 to Pl5 use the following variables:

- Pl1: Benefit recipients in logs (l1chl12)
- Pl2: Resident population in logs (l1hma7)
- Pl3: Benefit recipients and ACA, in logs (l1chl12 l1adl31)
- Pl4 Resident population in logs, and ACA, in logs (l1hma7, l1adl31)
- Pl5: Benefit recipients in logs, Resident population in logs, and ACA, in logs (l1chl12, l1hma7, l1adl31)

Results are presented in Table 7 (detailed results are presented in the Appendix A1.10). The main features are:

- R2 are lower and in the range of 76% to 93%
- The residuals for individual observations are much larger than in the previous SWl2 model.

Table 7: Results of parsimonious models: alternative specifications

Model		P1	P2	P3	P4	P5
Number of obs.		126	126	126	126	126
Number of vars.		1	1	2	2	3
R2 (%)		76.6	87.8	77.0	90.1	93.8
adj R2 (%)		76.4	87.7	76.6	90.1	93.7
largest errors (%):	1	73.1	75.7	67.1	64.2	71.3
	2	70.1	74.4	66.2	59.9	30.6
	3	67.9	49.9	63.2	54.0	28.9
	4	60.1	41.5	56.4	51.4	28.2
	5	53.1	40.7	51.5	50.3	23.9
Obs within 5% error band (%)		14 (11%)	35 (28%)	11 (9%)	35 (28%)	49 (39%)
Obs within 10% error band (%)		30 (24%)	63 (50%)	28 (22%)	65 (52%)	83 (66%)

Not surprisingly, the performance of the models in out-of-sample prediction is also poor (results omitted for brevity).

A1.5 Conclusions

A indicated in the introduction to this Annex, we believe that the important results derive from those models that use RNF as the dependent variable.

In general, RNF formulae of all sizes performed well in explaining a high percentage of the variation in the data (high R^2). Even some of the smaller formulae, with 2 and 3 needs indicators, had R^2 values over 90%. The better performing larger formulae (with 10-28 needs indicators) had R^2 values as high as 99%.

When we examined the potential impacts on resource allocations to individual local areas the general performance of the simple models was less good. One formula, with 28 needs indicators, resulted in no local areas having a change in resource allocation of more than 3%. Nevertheless, the formulae with fewer needs indicators (1, 2, 3 and 10) performed significantly less well, showing some local areas having changes in resource allocation in the range 30% - 70%.

A1.6 Appendix: Groups and sub-groups of funding

Local funding needs in England are divided into seven service groups (and sub-groupings). The relative needs of each local authority are based on a series of formulae that include underlying and indicator data specific to each service grouping (Table 8). The underlying and indicator variables included in the formulae are presented further below for each of the service groups.

Table 8: RNF, groups sub-groups and indicator variables

Service Group	Service sub-group	Indicator variables
Children's Services	Youth & Community; Local Authority Central Education Functions; and Children's Social Care.	Projected Population Aged 13 to 19 in 2008 (number) Children of Income Support/Income Based Job Seeker's Allowance Claimants Above Threshold (proportion) Youth and Community Deprivation Top-Up (index) Secondary Low Achieving Ethnic Groups Above Threshold (proportion) Youth and Community Ethnicity Top-Up (index) Pupils Aged 3 to 18 (number) Resident Pupils Aged 3 to 18 (number) Pupils Deprivation Top-Up (index) Ward Sparsity (index) Sparsity Top-Up (index) Resident Pupils Deprivation Top-Up (index) CEF Fixed Cost Amount (number) Projected Population Aged 0-17 in 2008 (number) Children's Social Care Deprivation Top-Up (index) Foster Cost Adjustment (index) Area Cost Adjustment for Education Area Cost Adjustment for Children and Younger Adults
Adults' Personal Social Services	Social Services for Older People; and Social Services for Younger Adults.	Older People PSS Age Top-Up (index) Older People PSS Deprivation Top-Up (index) Low Income Adjustment (index) Sparsity Adjustment for People Aged 65 and Over (index) Area Cost Adjustment for Older People's PSS Projected Population Aged 18 to 64 in 2008 (number) Younger Adults PSS Deprivation Top-Up (index) Area Cost Adjustment for Children and Younger Adults
Police		Projected Population in 2008 (number) Police Crime Top-Up 1 (index) Police Crime Top-Up 2 (index) Police Crime Top-Up 3 (index) Police Crime Top-Up 4 (index) Police Crime Top-Up 5 (index) Police Crime Top-Up 6 (index) Police Crime Top-Up 7 (index) Police Incidents Top-Up (index) Police Fear of Crime Top-Up (index) Police Traffic Top-Up (index) Police Sparsity Top-Up (index) Projected Daytime Population in 2008 (number) Area Cost Adjustment for Police
Fire & Rescue		Projected Population in 2008 (number) Fire and Rescue Coastline Top-Up (index) Risk Index (index) Fire and Rescue Deprivation Top-Up (index)

		High Risk Top-Up (index) Property and Societal Risk (index) Property and Societal Risk Top-Up (index) Community Fire Safety (index) Community Fire Safety Top-Up (index) Area Cost Adjustment for Fire and Rescue
Highway Maintenance		Weighted Road Lengths (index) Traffic Flow (index) Daytime Population per Km (number) Usage Top-Up (index) Winter Maintenance Top-Up (index) Area Cost Adjustment for Highway Maintenance
Environmental, Protection and Cultural Services (EPCS) and	Services provided predominantly by non-metropolitan district councils in non-metropolitan areas (District Services); Services provided predominantly by county councils in non-metropolitan areas (County Services); Fixed Costs - minimum needed to run day to day operations; Flood Defence - adjusted net current expenditure for flood defence plus levies payable to the Internal Drainage Board; Continuing Environmental Agency Levies – adjusted average amount payable to English Regional Flood Defence Committees; Coast Protection – adjusted average expenditure on coast protection.	Projected Population in 2008 [2009] [2010] (number) Population Sparsity (index) Flood Defence Expenditure (number) Environment Agency (England) Levy (number) Coast Protection Expenditure (number) District Services EPCS Density Top-Up District Services EPCS Sparsity Top-Up District Services EPCS Additional Population Top-Up District Services EPCS Deprivation Top-Up County Services EPCS Density Top-Up County Services EPCS Additional Population Top-Up County Services EPCS Deprivation Top-Up Area Cost Adjustment for EPCS
Capital Financing.		Assumed Mid 2008 Debt (£m) 2008/09 Debt Repayment and Interest Charges (£m)

Children's services

Underlying Data

1. Children of Income Support/Income Based Job Seeker's Allowance Claimants (number)
2. Resident Population Under 18 (number)
3. Children of Income Support/Income Based Job Seeker's Allowance Claimants (proportion)
4. Pupils of Secondary School Age in Low Achieving Ethnic Groups (number)
5. Pupils in Secondary School with an Ethnic Group Recorded (number)
6. Pupils of Secondary School Age in Low Achieving Ethnic Groups (proportion)
7. Sparsity <= 0.5 Residents per Hectare (proportion)
8. Sparsity 0.5 – 4 Residents per Hectare (proportion)
9. Children Without Good Health (number)
10. Population Aged 0 to 17 (number)
11. Children Without Good Health (proportion)
12. Income Support/Income Based Job Seeker's Allowance Claimants Aged 18 to 64 Years (number)
13. Resident Population Aged 18 to 64 (number)
14. Income Support/Income Based Job Seeker's Allowance Claimants Aged 18 to 64 Years (proportion)
15. Children in Black Ethnic Groups (number)
16. Population Aged 1 to 15 (number)
17. Children in Black Ethnic Groups (proportion)
18. People in Other Ethnic Groups (number)
19. Resident Population (number)
20. People in Other Ethnic Groups (proportion)
21. People in Mixed Ethnic Groups (number)
22. People in Mixed Ethnic Groups (proportion)
23. People Aged 16 to 74 Whose Highest Qualification Attained Was Level 1 or 2 (number)
24. Resident Population Aged 16 to 74 (number)
25. People Aged 16 to 74 Whose Highest Qualification Attained Was Level 1 or 2 (proportion)

26. People Aged 16 to 74 Whose Highest Qualification Attained Was Level 4 or 5 (number)
27. People Aged 16 to 74 Whose Highest Qualification Attained Was Level 4 or 5 (proportion)
28. Females Aged 16 to 74 Looking After Home and/or Family (number)
29. Resident Females Aged 16 to 74 (number)
30. Females Aged 16 to 74 Looking After Home and/or Family (proportion)

Indicator Data

31. Projected Population Aged 13 to 19 in 2008 [2009] [2010] (number)
32. Children of Income Support/Income Based Job Seeker's Allowance Claimants Above Threshold (proportion)
33. Youth and Community Deprivation Top-Up (index)
34. Secondary Low Achieving Ethnic Groups Above Threshold (proportion)
35. Youth and Community Ethnicity Top-Up (index)
36. Pupils Aged 3 to 18 (number)
37. Resident Pupils Aged 3 to 18 (number)
38. Pupils Deprivation Top-Up (index)
39. Ward Sparsity (index)
40. Sparsity Top-Up (index)
41. Resident Pupils Deprivation Top-Up (index)
42. CEF Fixed Cost Amount (number)
43. Projected Population Aged 0-17 in 2008 [2009] [2010] (number)
44. Children's Social Care Deprivation Top-Up (index)
45. Foster Cost Adjustment (index)
46. Area Cost Adjustment for Education
47. Area Cost Adjustment for Children and Younger Adults

Relative Need

48. RNF Youth and Community
49. RNF LEA Central Functions
50. RNF Children's Social Care
51. RNF Total Children's Services
52. ACA Element of Youth and Community
53. ACA Element of Local Authority Central Education Functions
54. ACA Element of Children's Services
55. Area Cost Adjustment (included in column (51))

Adults personal social services

Underlying Data

1. Household and Supported Residents Aged 65 and Over (number)
2. Household and Supported Residents Aged 90 and Over (number)
3. Projected Household and Supported Residents Aged 65 and Over in 2008 [2009] [2010] (number)
4. Older People Receiving Attendance Allowance (number)
5. Resident Population Aged 65 Plus (number)
6. Older People Receiving Attendance Allowance (proportion)
7. Older People in Rented Accommodation (number)
8. People Aged 65 or Over (number)
9. Older People in Rented Accommodation (proportion)
10. Older People Living in One Person Households (number)
11. Older People Living in One Person Households (proportion)
12. Older People Receiving Income Support/Income Based Job Seeker's Allowance/Guarantee Element of Pension Credit (number)
13. Older People Receiving Income Support/Income Based Job Seeker's Allowance/Guarantee Element of Pension Credit (proportion)
14. Sparsity <= 0.08 Residents Aged 65 and Over per Hectare (proportion)
15. Sparsity 0.08 – 0.64 Residents Aged 65 and Over per Hectare (proportion)
16. People Aged 18 to 64 Receiving Disability Living Allowance (number)
17. Resident Population Aged 18 to 64 (number)
18. People Aged 18 to 64 Receiving Disability Living Allowance (proportion)
19. People Aged 18 to 64 Who Are Long Term Unemployed or Have Never Worked (number)

20. Population Aged 18 to 64 (number)
21. People Aged 18 to 64 Who Are Long Term Unemployed or Have Never Worked (proportion)
22. People Aged 18 to 64 Who Work in Routine or Semi Routine Occupations (number)
23. People Aged 18 to 64 Who Work in Routine or Semi Routine Occupations (proportion)
24. Households With No Family (number)
25. Households (number)
26. Households With No Family (proportion)

Indicator Data

27. Older People PSS Age Top-Up (index)
28. Older People PSS Deprivation Top-Up (index)
29. Low Income Adjustment (index)
30. Sparsity Adjustment for People Aged 65 and Over (index)
31. Area Cost Adjustment for Older People's PSS
32. Projected Population Aged 18 to 64 in 2008 [2009] [2010] (number)
33. Younger Adults PSS Deprivation Top-Up (index)
34. Area Cost Adjustment for Children and Younger Adults

Relative Need

35. RNF Social Services for Older People
36. RNF Social Services for Younger Adults
37. RNF Total Adults' Personal Social Services
38. ACA Element of Elderly Services
39. ACA Element of Younger Adults Services
40. Area Cost Adjustment (included in column (39))

Police

Underlying Data

1. Daytime Net Inflow (number)
2. Total Resident Population (number)
3. Daytime Net Inflow per Resident Population (proportion)
4. Bars (number)
5. Hectare (number)
6. Log of Bars per 100 Hectares (proportion)
7. Income Support/Income Based Job Seeker's Allowance/Guarantee Element of Pension Credit Claimants (number)
8. Resident Population (number)
9. Income Support/Income Based Job Seeker's Allowance/Guarantee Element of Pension Credit Claimants (proportion)
10. Single Parent Household (number)
11. Households (number)
12. Single Parent Households (proportion)
13. Population Density (index)
14. Long-Term Unemployment-Related Benefit Claimants (number)
15. Resident Males Aged 18 to 64 and Females Aged 18 to 59 (number)
16. Long-Term Unemployment-Related Benefit Claimants (proportion)
17. Residents in Routine/Semi Routine Occupations or Never Worked/Long Term Unemployed (number)
18. Residents in Routine/Semi Routine Occupations or Never Worked/Long Term Unemployed (proportion)
19. Hard Pressed (proportion)
20. Student Households (number)
21. Student Housing (proportion)
22. Young Male Unemployment Related Benefit Claimants (number)
23. Young Male Unemployment Related Benefit Claimants (proportion)
24. Overcrowded Households (number)
25. Log of Overcrowded Households (proportion)
26. Sparsity <= 0.5 Residents per Hectare (proportion)
27. Sparsity 0.5 – 4 Residents per Hectare (proportion)
28. Population Sparsity (proportion)

29. Log of Population Sparsity (proportion)
30. Wealthy Achievers (proportion)
31. Residents in Terraced Accommodation (number)
32. Households (number)
33. Residents in Terraced Accommodation (proportion)

Indicator Data

34. Projected Population in 2008 [2009] [2010] (number)
35. Police Crime Top-Up 1 (index)
36. Police Crime Top-Up 2 (index)
37. Police Crime Top-Up 3 (index)
38. Police Crime Top-Up 4 (index)
39. Police Crime Top-Up 5 (index)
40. Police Crime Top-Up 6 (index)
41. Police Crime Top-Up 7 (index)
42. Police Incidents Top-Up (index)
43. Police Fear of Crime Top-Up (index)
44. Police Traffic Top-Up (index)
45. Police Sparsity Top-Up (index)
46. Projected Daytime Population in 2008 [2009] [2010] (number)
47. Area Cost Adjustment for Police

Relative Need

48. RNF Police
49. Area Cost Adjustment (included in column (48))

Fire & Rescue

Underlying Data

1. Coastline (number)
2. Resident Population (number)
3. Coastline (proportion)
4. Children of Income Support/Income Based Job Seeker's Allowance Claimants (number)
5. Resident Population Under 18 (number)
6. Children of IS/IB JSA Claimants (proportion)
7. Households Not Containing A Couple With No Children (proportion)
8. People in Rented Accommodation (proportion)
9. Absences in Pupils of Primary School Age (number)
10. Average Number of Rooms per Household Resident (number)
11. ACORN Types 50 and 53 (proportion)
12. COMAH Sites (number)
13. COMAH Sites (proportion)

Indicator Data

14. Projected Population in 2008 [2009] [2010] (number)
15. Fire and Rescue Coastline Top-Up (index)
16. Risk Index (index)
17. Fire and Rescue Deprivation Top-Up (index)
18. High Risk Top-Up (index)
19. Property and Societal Risk (index)
20. Property and Societal Risk Top-Up (index)
21. Community Fire Safety (index)
22. Community Fire Safety Top-Up (index)
23. Area Cost Adjustment for Fire and Rescue

Relative Need

24. RNF Fire and Rescue
25. Area Cost Adjustment (included in column (24))

Highway maintenance

Underlying Data

1. Principal Built-Up Roads (km)
2. Principal Non Built-Up Roads (km)
3. Other Built-Up Roads (km)
4. Other Non Built-Up Roads (km)
5. Traffic Flow of All Vehicles (millions)
6. Traffic Flow of HGVs, Buses and Coaches (millions)
7. Resident Population (number)
8. Net In-Commuters (number)
9. Annual Number of Night Visitors (number)
10. Annual Number of Day Visitors (number)
11. Days with Snow Lying (number)
12. Predicted Gritting Days (number)

Indicator Data

13. Weighted Road Lengths (index)
14. Traffic Flow (index)
15. Daytime Population per Km (number)
16. Usage Top-Up (index)
17. Winter Maintenance Top-Up (index)
18. Area Cost Adjustment for Highway Maintenance

Relative Need

19. RNF Highway Maintenance
20. Area Cost Adjustment (included in column (19))

Environmental, Protective and Cultural Services

Underlying Data

1. Population Density (proportion)
2. Sparsity <= 0.5 Residents per Hectare (proportion)
3. Sparsity 0.5 – 4 Residents per Hectare (proportion)
4. Net In-Commuters (number)
5. Day Visitors (number)
6. Incapacity Benefit and Severe Disablement Allowance (number)
7. Income Support/Income Based Job Seeker's Allowance/Guarantee Element of Pension Credit Claimants (number)
8. Older People on Income Support/Income Based Job Seeker's Allowance/ Guarantee Element of Pension Credit (number)
9. Unemployment Related Benefit Claimants (number)
10. Country of Birth Residents (number)

Indicator Data

11. Projected Population in 2008 [2009] [2010] (number)
12. Population Sparsity (index)
13. Flood Defence Expenditure (number)
14. Environment Agency (England) Levy (number)
15. Coast Protection Expenditure (number)
16. District Services EPCS Density Top-Up
17. District Services EPCS Sparsity Top-Up
18. District Services EPCS Additional Population Top-Up
19. District Services EPCS Deprivation Top-Up
20. County Services EPCS Density Top-Up
21. County Services EPCS Additional Population Top-Up
22. County Services EPCS Deprivation Top-Up
23. Area Cost Adjustment for EPCS

Relative Need

24. RNF District Level EPCS
25. RNF County Level EPCS
26. RNF Fixed Costs
27. RNF Flood Defence
28. RNF Continuing Environment Agency Levies
29. RNF Coast Protection
30. RNF Total EPCS
31. ACA Element of Districts RNF
32. ACA Element of County RNF
33. Area Cost Adjustment (included in column (30))

Capital Financing

Underlying Data

1. Assumed Debt at Start of Year 2004-05 (£m)
2. Estimated 2004/05 Supported Capital Expenditure (Revenue) (£m)
3. Assumed Debt at Start of Year 2005-06 (£m)
4. Estimated 2005/06 Supported Capital Expenditure (Revenue) (£m)
5. Assumed Debt at Start of Year 2006-07 (£m)
6. Estimated 2006/07 Supported Capital Expenditure (Revenue) (£m)
7. Assumed Debt at Start of Year 2007-08 (£m)
8. Estimated 2007/08 Supported Capital Expenditure (Revenue) (£m)
9. Assumed Debt at Start of Year 2008-09 (£m)
10. Estimated 2008/09 Supported Capital Expenditure (Revenue) (£m)
11. Assumed Debt at Start of Year 2009-10 (£m)
12. Estimated 2009/10 Supported Capital Expenditure (Revenue) (£m)
13. Assumed Debt at Start of Year 2010-11 (£m)
14. Estimated 2010/11 Supported Capital Expenditure (Revenue) (£m)
15. Assumed Debt at Start of Year 2011-12 (£m)

Indicator Data

16. Assumed Mid 2008 Debt (£m)
17. 2008/09 Debt Repayment and Interest Charges (£m)

Relative Need

18. RNF Capital Financing

A1.7 Appendix: Observations contained in the data

For the purposes of local government outside Greater London, in 1974 England was divided into counties (Metropolitan and Non-Metropolitan). Counties were further divided into districts, each with its own district council. The reorganisation of local government (in the 1990s) introduced Unitary Authorities[31] (or Unitaries), this are

[31] These are generally large towns and cities that are deemed capable of functioning independently of other local government involvement.

counties which are directly responsible for all aspects of local government and operate alongside the remaining counties and districts[32].

Local government funding is organised differently in each of these divisions.

- Counties are organised on a two-tier approach to local government. Services at the county level are administered by a County Council, while services at the district level are provided by District Councils. Therefore, the overall needs of a County are calculated accounting for both levels of services (county and district level).

- Services for Unitaries are all administered by a single local council. Shire Unitaries largely came about in the 1990s and are used

- Metropolitan Districts administer local services independently, typically within urban areas. At the present time (following the abolition of Metropolitan Counties as administrative entities) Metropolitan Districts have effectively become Unitaries.

- London is divided into 32 London boroughs (and the City of London), each of these have their own borough council and discharge local services. Each of these can effectively be viewed as being akin to a Shire Unitary.

The list of English districts are shown in Table 9 for counties, metropolitan and non-metropolitan counties (including unitaries) and Greater London authority.

For Police and Fire and Rescue services the geographic and administrative boundaries of these services do not typically correspond with the local authority boundaries.

Table 9: Summary of English local units

Main division	Districts		
Non Metropolitan	**Counties**		
	Bedfordshire	Essex	Nottinghamshire
	Buckinghamshire	Gloucestershire	Oxfordshire
	Cambridgeshire	Hampshire	Shropshire
	Cheshire	Hertfordshire	Somerset
	Cornwall	Kent	Staffordshire
	Cumbria	Lancashire	Suffolk
	Derbyshire	Leicestershire	Surrey
	Devon	Lincolnshire	Warwickshire
	Dorset	Norfolk	West Sussex
	Durham	North Yorkshire	Wiltshire
	East Sussex	Northamptonshire	Worcestershire
		Northumberland	
	Unitaries		
	Bath & North East Somerset	Leicester	Slough
	Blackburn with Darwen	Luton	South Gloucestershire

[32] Wales was completely divided into Unitary Authorities while Scotland was divided into 32 Council Areas.

	Blackpool	Medway	Southampton
	Bournemouth	Middlesbrough	Southend-on-Sea
	Bracknell Forest	Milton Keynes	Stockton-on-Tees
	Brighton & Hove	North East Lincolnshire	Stoke-on-Trent
	Bristol	North Lincolnshire	Swindon
	Darlington	North Somerset	Telford and the Wrekin
	Derby	Nottingham	Thurrock
	East Riding of Yorkshire	Peterborough	Torbay
	Halton	Plymouth	Warrington
	Hartlepool	Poole	West Berkshire
	Herefordshire	Portsmouth	Windsor and Maidenhead
	Isle of Wight Council	Reading	Wokingham
	Kingston upon Hull	Redcar and Cleveland	York
		Rutland	
Metropolitan Counties	**Greater Manchester**	**South Yorkshire**	**West Midlands**
	Bolton	Barnsley	Birmingham
	Bury	Doncaster	Coventry
	Manchester	Rotherham	Dudley
	Oldham	Sheffield	Sandwell
	Rochdale	**Tyne and Wear**	Solihull
	Salford	Gateshead	Walsall
	Stockport	Newcastle upon Tyne	Wolverhampton
	Tameside	North Tyneside	**West Yorkshire**
	Trafford	South Tyneside	Bradford
	Wigan	Sunderland	Calderdale
	Merseyside		Kirklees
	Knowsley		Leeds
	Liverpool		Wakefield
	Sefton		
	St Helens		
	Wirral		
Greater London	**Inner**	**Outer**	
	Camden	Barking and Dagenham	Havering
	City of London	Barnet	Hillingdon
	Greenwich	Bexley	Hounslow
	Hackney	Brent	Kingston upon Thames
	Hammersmith and Fulham	Bromley	Merton
	Islington	Croydon	Newham
	Kensington and Chelsea	Ealing	Redbridge
	Lambeth	Enfield	Richmond upon Thames
	Lewisham	Haringey	Sutton
	Southwark	Harrow	Waltham Forest
	Tower Hamlets		
	Wandsworth		
	Westminster		

Table 10: Out of sample units

Main division	Districts	
Non Metropolitan	**Counties**	
	Cambridgeshire	West Sussex
	Cheshire	Wiltshire
	Derbyshire	
	Unitaries	
	Halton	Southend-on-Sea
	Medway	Windsor and Maidenhead
	North Somerset	Wokingham
	Portsmouth	
Metropolitan Counties	**Greater Manchester**	**West Midlands**
	Wigan	Dudley
	Merseyside	**West Yorkshire**
	Liverpool	Kirklees
	Wirral	Leeds
		Wakefield
Greater London	**Inner**	**Outer**
	Hackney	Hounslow
	Southwark	Waltham Forest

A1.8 Appendix: Selection of variables

Table 11: Selection of variables using PC, SW, and OJ

Group	Variable type	Variable	PC1 & PC3	PC2 & PC4	SW1 & SW3	SW2 & SW4	OJ1 & OJ2
Environmental	Exp (Flood, Coast)	Environment Agency (England) Levy	X		X*	X*	
Environmental	Weather	Winter Maintenance Top-Up in 2009/10	X	X			
Financial	Area Cost Adjustment	ACA for Children and Younger Adults		X*			
Financial	Area Cost Adjustment	ACA for Education			X		
Financial	Area Cost Adjustment	ACA for Older People's PSS	X	X		X	
Financial	Assumed Debt	Assumed Debt at Start of Year 2007/08			X		
Financial	Assumed Debt	Assumed Debt at Start of Year 2009/10					X
Financial	Assumed Debt	Assumed Debt at Start of Year 2011/12				X	
Financial	Assumed Debt	Assumed Mid 2008/09 Debt	X				
Financial	Assumed Debt	Assumed Mid 2010/11 Debt		X			
Financial	Est. Supported Cap Exp	Est. 2009/10 Supported Cap Ex (Revenue)					X
Financial	Est. Supported Cap Exp	Est. 2010/11 Supported Cap Ex (Revenue)			X		
Population	on Benefit / Disadvantaged	Children in Black Ethnic Groups			X*	X	
Population	on Benefit / Disadvantaged	Children of IS/IB JSA Claimants			X	X*	X
Population	on Benefit / Disadvantaged	Children of IS/IB JSA Claimants Above Threshold		X*			
Population	on Benefit / Disadvantaged	County Services EPCS Deprivation Top-Up in 2009/10			X	X	
Population	on Benefit / Disadvantaged	District Services EPCS Deprivation Top-Up in 2009/10		X			
Population	on Benefit / Disadvantaged	Foster Cost Adjustment in 2009		X			
Population	on Benefit / Disadvantaged	Incapacity Benefit and Severe Disablement Allowance			X*		X
Population	on Benefit / Disadvantaged	IS/IB JSA Claimants Aged 18 to 64 Years			X*		
Population	on Benefit / Disadvantaged	IS/IB JSA / Guarantee Element of Pension Credit					X
Population	on Benefit / Disadvantaged	Older People Pss Deprivation Top-Up 2009/10		X	X*		
Population	on Benefit / Disadvantaged	Older People Receiving attendance Allowance					
Population	on Benefit / Disadvantaged	Older People IS/IB JSA/Guarantee of Pension Credit				X*	X*
Population	on Benefit / Disadvantaged	People 18-64 Receiving Disability Living Allowance			X*		X
Population	on Benefit / Disadvantaged	People 18-64 Who Are LT Unemployed or Never Worked					X*
Population	on Benefit / Disadvantaged	People in Mixed Ethnic Groups	X				
Population	on Benefit / Disadvantaged	Pupils Sec School Age in Low Achieving Ethnic Groups	X				X
Population	on Benefit / Disadvantaged	Younger Adults PSS Deprivation Top-Up 2009/10	X			X*	
Population	Population	Country of Birth of Residents			X*		
Population	Population	County Services EPCS Additional Pop Top-Up in 2009/10	X		X*	X*	
Population	Population	Daytime Population Per Km		X			
Population	Population	District Services EPCS Additional Pop Top-Up in 2009/10		X			
Population	Population	Females 16 to 74 Looking After Home and/or Family	X*				
Population	Population	Household and Supported Residents Aged 90+			X*		
Population	Population	Households with No Family				X*	
Population	Population	Older People in Rented Accommodation			X	X*	
Population	Population	Older People in Rented Accommodation			X*		
Population	Population	Older People Living in One Person Households				X*	
Population	Population	Older People Living in One Person Households				X*	
Population	Population	Older People Pss Age Top-Up 2009/10			X*		
Population	Population	People 16-74 Highest Qual. attained was Level 1 or 2			X*		
Population	Population	People 16-74 Highest Qual. attained was Level 4 or 5			X*		
Population	Population	People 16-74 Highest Qual. attained was Level 4 or 5 (%)				X*	
Population	Population	People 18-64 Work semi- or Routine Occupations			X*		
Population	Population	People Aged 65 Plus					X
Population	Population	Population Aged 18 to 64					X*
Population	Population	Proj Household and Sup. Residents Aged 65+ in 2009				X*	
Population	Population	Projected Population Aged 0 to 17 in 2009		X	X	X*	
Population	Population	Projected Population Aged 18 to 64 in 2009		X			
Population	Population	Pupils Aged 3 to 18			X*		
Population	Population	Resident Population			X*	X*	
Population	Population	Resident Population Aged 18 to 64				X	
Population	Population	Resident Population Under 18			X*		X
Population	Sparsity / Density	District Services EPCS Density Top-Up in 2009/10				X	
Population	Sparsity / Density	District Services EPCS Sparsity Top-Up in 2009/10			X*	X*	
Population	Sparsity / Density	Population Density		X*	X		X
Population	Sparsity / Density	Population Sparsity		X*			
Population	Sparsity / Density	Sparsity Adjustment For People Aged 65 and Over		X*			
Population	Sparsity / Density	Sparsity Top-Up in 2009					
Population	Sparsity / Density	Ward Sparsity	X	X			
Transport	Road Length	Road Lengths Other Built-Up Roads				X*	
Transport	Road Length	Road Lengths Other Non Built-Up Roads				X*	
Transport	Road Length	Weighted Road Lengths		X*			X*
Transport	Traffic Flow	Traffic Flow	X*	X*			X*
Total – Preliminary Models			**11**	**18**	**28**	**22**	**15**
Total – Refined Models			**9**	**11**	**8**	**7**	**10**

Note: X denotes variables selected by each technique. Variables identified by * were dropped in the refined models (PC3, PC4, SW3, SW4 and OJ2).

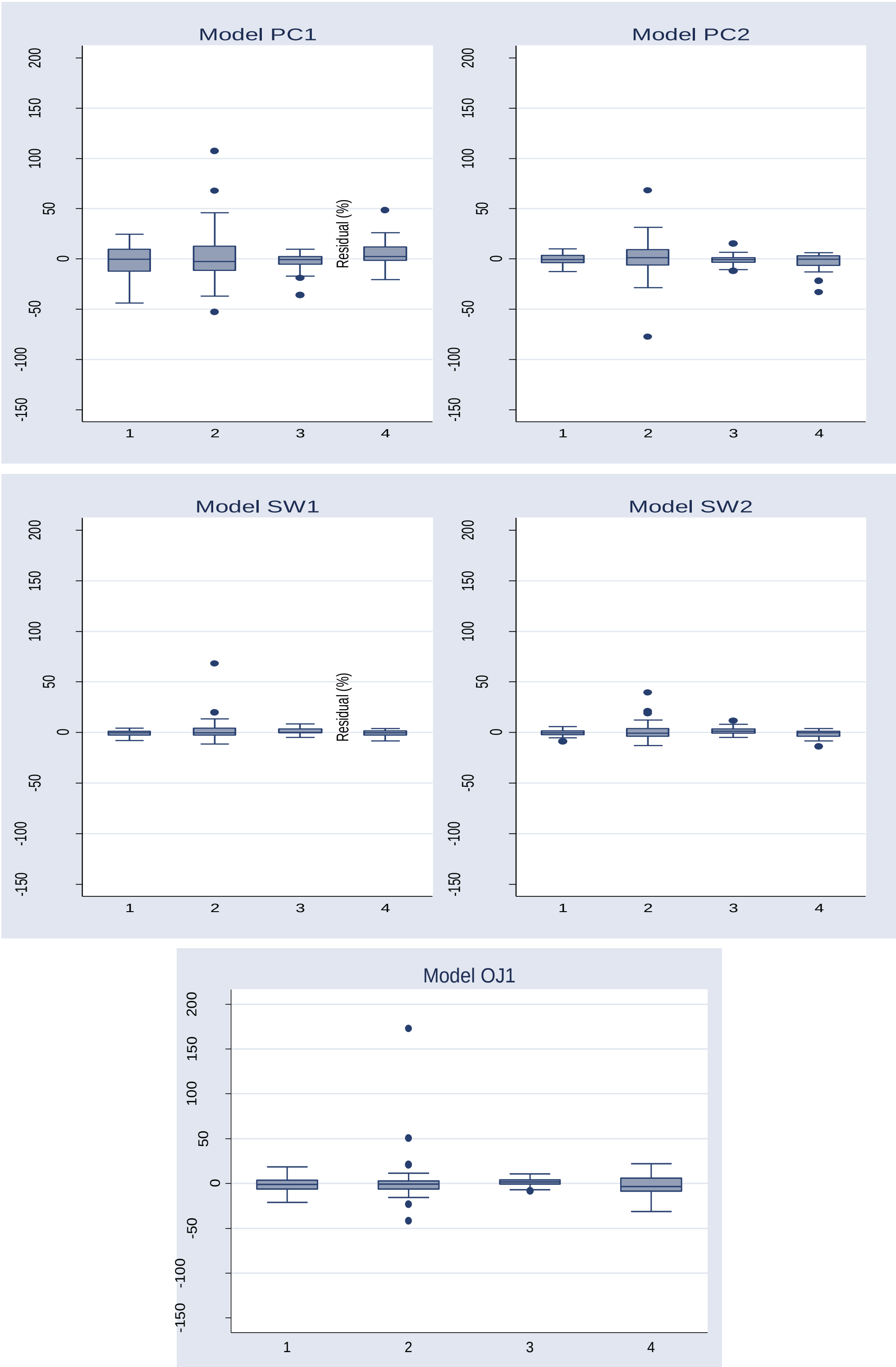
A1.9 Appendix: Analysis of residuals

Table 12: Residuals (in %)

	Shire/County	Area	PC1	PC2	SW1	SW2	OJ1	PC3	PC4	SW3	SW4	OJ2
Non-Metropolitan	**Counties**	Cornwall	13	3	4	2	3	12	2	12	7	4
		Cumbria	11	6	3	4	7	11	7	5	2	8
		Gloucestershire	-6	0	0	-1	-4	-7	-1	-6	-5	-7
		Hertfordshire	0	7	3	1	13	0	8	5	10	12
		Lincolnshire	9	5	1	1	4	9	5	8	6	4
		Norfolk	13	6	2	1	1	13	7	4	3	2
		Northamptonshire	-5	2	-1	-3	-2	-2	2	-3	3	-3
		Northumberland	5	2	0	-3	5	5	1	-5	0	8
		Oxfordshire	-36	10	-5	-5	-7	-36	6	-12	-1	-7
		Suffolk	-3	3	-2	-2	-3	-3	3	2	1	-4
		Surrey	-25	10	3	3	19	-26	8	9	13	21
		Warwickshire	5	-1	-5	-3	-6	5	-1	-3	-4	-7
		Bedfordshire	-13	-4	-7	-9	-21	-12	-5	-16	-4	-25
		Buckinghamshire	-44	-8	1	6	3	-46	-8	4	5	2
		Devon	3	1	-2	-2	-6	2	3	1	-1	-4
		Dorset	-14	-4	-8	0	1	-14	2	16	3	-2
		Durham	25	-6	-2	0	9	24	-8	-6	1	10
		East Sussex	18	-4	-3	-2	-3	16	2	1	-2	-9
		Essex	13	3	1	2	2	13	3	2	5	1
		Hampshire	-32	-13	-1	-1	-10	-32	-13	-15	-11	-9
		Kent	-1	-5	0	-1	-1	-2	-4	-4	-4	-3
		Lancashire	12	-8	1	-1	-3	12	-9	-3	-7	-4
		Leicestershire	-18	-1	2	1	-16	-15	0	-7	-9	-11
		North Yorkshire	-8	-3	1	4	2	-7	0	3	-1	5
		Nottinghamshire	10	-3	-3	-2	-4	9	-3	-3	-2	-4
		Shropshire	-12	-7	2	2	4	-12	-6	8	-5	5
		Somerset	-2	3	-3	-5	6	-2	4	12	7	7
		Staffordshire	1	-7	1	1	-8	1	-9	-4	-7	-4
		Worcestershire	4	-2	-5	4	-15	4	-4	-10	-9	-15
	Unitaries	Isle of Wight Council	-27	8	2	-1	21	-33	4	22	7	11
		Bath & North East Somerset	-5	4	5	0	-1	-8	9	0	2	-6
		Blackburn with Darwen	-28	-5	-2	-4	-1	-34	-9	10	5	-6
		Blackpool	-28	4	-6	-10	-1	-27	5	9	-1	-8
		Bournemouth	19	11	4	-1	-9	20	8	1	-2	-7
		Bracknell Forest	107	-29	-7	-13	3	109	-59	-38	-36	5
		Brighton & Hove	9	3	1	1	-23	8	4	-25	-22	-25
		Bristol	10	2	-1	-1	-3	11	-1	-9	0	-1
		Darlington	-2	9	13	12	-6	0	23	3	0	-8
		Derby	-3	7	0	5	-7	-3	12	3	2	-9
		East Riding of Yorkshire	-7	-2	-3	-3	0	-5	1	11	-6	2
		Hartlepool	-53	-6	-4	-9	-2	-52	-8	-2	0	0
		Herefordshire	-29	-14	0	-3	21	-28	-14	20	-13	20
		Kingston upon Hull	22	0	-1	-4	-7	23	0	-18	-2	-4
		Leicester	7	-6	0	-1	1	7	-6	-9	-10	4
		Luton	-12	-5	-1	-5	-16	-14	2	-8	-2	-17
		Middlesbrough	-13	-8	3	5	-5	-11	-2	-4	3	1
		Milton Keynes	-37	-8	-1	2	11	-36	-18	0	4	6
		North East Lincolnshire	27	9	-2	-6	0	22	14	-3	-2	-2
		North Lincolnshire	-19	10	-5	-9	10	-23	-3	15	5	8
		Nottingham	0	1	-1	-5	-10	0	4	-8	6	-10
		Peterborough	-9	4	-4	2	3	-2	-2	14	20	10
		Plymouth	4	4	1	8	-10	8	4	-2	5	-10
		Poole	30	31	20	21	-3	29	33	20	27	-3
		Reading	-6	-7	2	1	-10	-7	-8	-28	-9	-10
		Redcar and Cleveland	-11	1	8	-2	3	-14	2	0	-2	2
		Rutland	68	-78	68	40	173	82	-112	176	-92	158
		Slough	-7	-17	-7	-6	-3	-13	-29	-12	-2	-6
		South Gloucestershire	-33	-10	-11	-8	1	-29	-13	4	-8	1
		Southampton	20	19	7	3	-15	19	15	-18	7	-16
		Stockton-on-Tees	11	1	3	-4	-2	8	0	-9	-2	0
		Stoke-on-Trent	-11	4	-3	-1	2	-11	6	0	3	-3
		Swindon	7	1	-6	-2	-1	9	-3	-8	-3	-3
		Telford and the Wrekin	13	12	0	2	5	12	7	14	21	6
		Thurrock	5	-1	7	6	6	13	-3	12	12	16
		Torbay	-8	16	5	4	5	-10	11	22	10	-3
		Warrington	11	16	-1	10	0	13	2	4	9	0
		West Berkshire	38	-14	-7	-1	51	39	-44	17	-1	55
		York	46	68	13	19	-42	57	80	-26	-14	-39
Metropolitan	**Greater Manchester**	Bolton	10	0	0	2	9	10	-1	8	5	9
		Bury	3	4	9	12	3	4	6	10	6	0
		Manchester	-4	-3	0	-1	2	-4	-5	-4	3	1
		Oldham	-10	1	3	3	4	-12	6	8	5	-1
		Rochdale	-2	1	1	-3	2	-2	1	1	4	-1
		Salford	-5	-4	2	1	5	-3	-5	0	5	5
		Stockport	5	0	7	8	4	6	2	6	-1	3
		Tameside	0	1	3	5	11	1	-1	7	5	7
		Trafford	-4	-1	1	8	-1	-3	0	10	6	-2
	Merseyside	Knowsley	-36	-11	-3	-5	-7	-31	-4	3	7	-3
		Sefton	9	-1	5	4	2	8	0	7	-3	-3
		St Helens	-19	-3	0	2	3	-19	-2	5	2	0
	South Yorkshire	Barnsley	-6	5	1	-1	7	-7	5	5	9	4
		Doncaster	2	-3	2	1	4	2	0	5	3	3
		Rotherham	-3	7	-1	3	0	-2	4	6	10	0
		Sheffield	-1	3	-2	-2	-4	0	2	-5	-2	-2

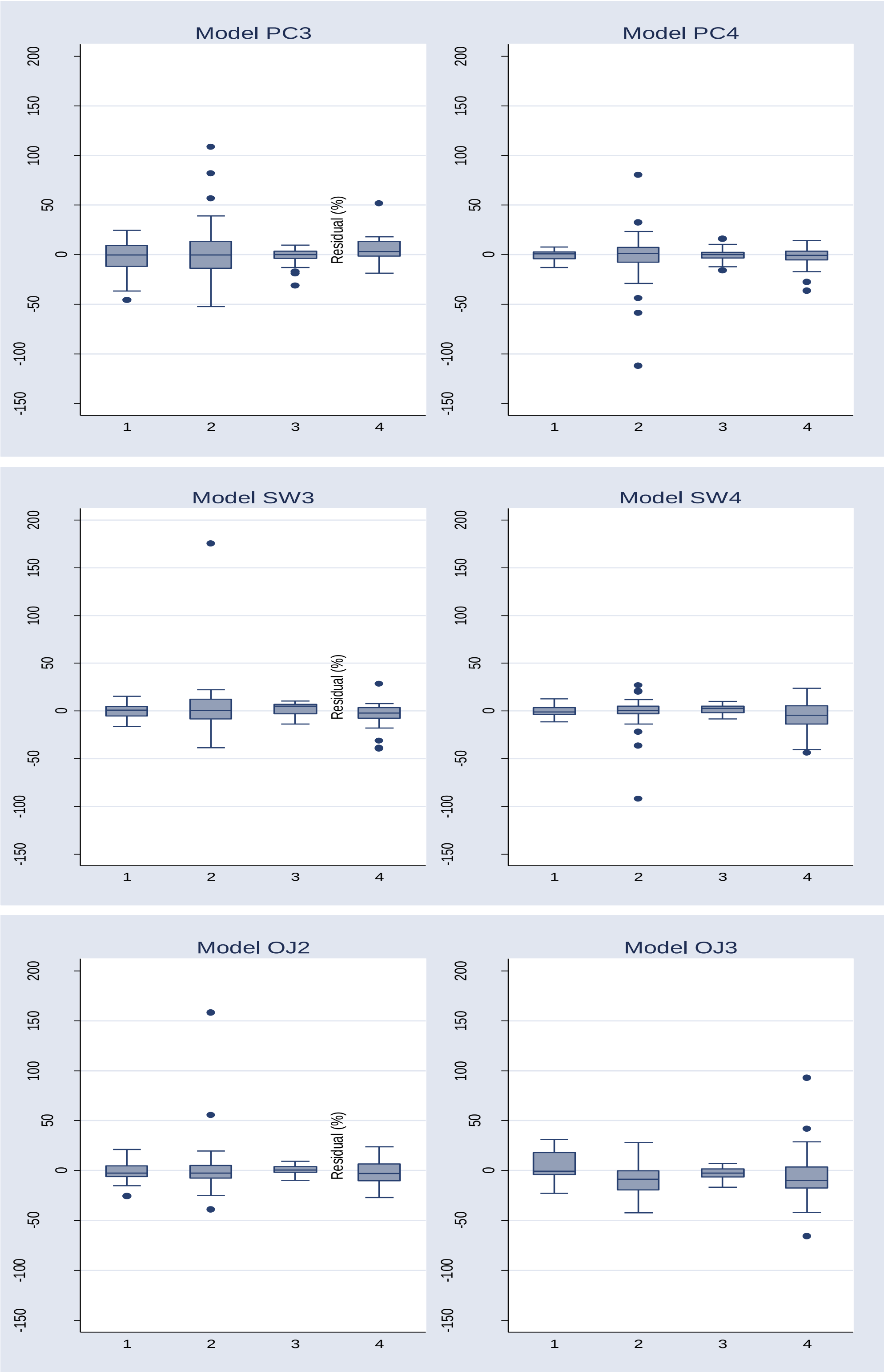
Cont'd.	Shire/County	Area	PC1	PC2	SW1	SW2	OJ1	PC3	PC4	SW3	SW4	OJ2
	Tyne and Wear	Gateshead	-6	-4	0	-3	4	-3	-4	-4	2	8
		Newcastle upon Tyne	0	-4	0	-2	-2	0	-6	-11	-2	1
		North Tyneside	0	0	-5	-2	1	-1	-6	-8	-3	-1
		South Tyneside	-10	-12	-3	-4	-1	-13	-16	-14	-3	0
		Sunderland	2	-8	-2	0	1	3	-12	-8	-4	4
	West Midlands	Birmingham	4	3	0	1	2	5	3	5	0	2
		Coventry	2	-4	-1	3	-3	3	-2	-3	-8	-3
		Sandwell	3	0	0	0	4	3	0	5	4	5
		Solihull	-1	-5	6	6	-3	1	10	10	-4	-5
		Walsall	4	-2	4	2	5	4	2	7	3	7
		Wolverhampton	-6	-5	2	1	4	-7	-3	1	-4	6
	West Yorkshire	Bradford	-17	1	0	0	-8	-17	1	5	0	-10
		Calderdale	9	15	5	5	1	9	16	2	9	-2
London	**Inner London (inc City)**	Camden	8	3	2	1	7	7	3	8	6	7
		City of London	-21	-1	-1	-6	-8	-19	-7	-8	-35	-18
		Greenwich	1	3	2	3	16	1	5	8	14	14
		Hammersmith and Fulham	1	-9	-5	0	-13	4	-1	-14	-13	-10
		Islington	1	-10	0	0	0	1	-11	-6	-1	-1
		Kensington and Chelsea	2	3	4	1	-10	1	12	-5	-17	-8
		Lambeth	12	6	2	0	11	13	6	-2	3	14
		Lewisham	4	4	1	3	13	5	4	3	10	13
		Tower Hamlets	8	3	1	2	-1	6	3	4	10	-3
		Wandsworth	-6	6	-4	-1	-9	-2	7	-5	-1	-4
		Westminster	0	0	0	1	22	-1	2	28	24	24
	Outer London	Barking and Dagenham	21	-2	-8	-4	-4	17	-2	-4	7	-10
		Barnet	18	-2	-5	-6	6	14	-6	-2	-10	7
		Bexley	17	4	1	1	-4	18	14	5	2	-6
		Brent	2	1	-1	-3	-1	2	0	-8	-14	3
		Bromley	18	-3	-4	-7	1	14	-1	3	-5	0
		Croydon	-7	4	3	-7	-3	-7	7	0	-1	-3
		Ealing	-1	0	2	4	-6	1	-2	-11	-10	-1
		Enfield	-1	1	1	2	-5	-2	3	-3	-9	-6
		Haringey	-18	-7	-2	-1	-15	-18	-6	-18	-17	-12
		Harrow	19	-13	-1	0	8	15	-17	-6	-19	11
		Havering	49	2	-2	-8	-5	52	7	8	5	-1
		Hillingdon	4	-9	-5	-4	16	3	-10	8	8	16
		Kingston upon Thames	11	-22	1	-3	-31	7	-28	-39	-40	-27
		Merton	-8	-7	2	-2	-22	-7	-3	-31	-27	-17
		Newham	-12	5	0	1	-8	-13	4	-1	1	-9
		Redbridge	-12	-2	-3	2	-15	-15	-3	2	-11	-20
		Richmond upon Thames	26	-33	-7	-14	-29	18	-36	-38	-44	-22
		Sutton	9	-5	0	-3	-2	11	-4	-1	-5	-1

Figure 6: Residual plots – preliminary models (1)



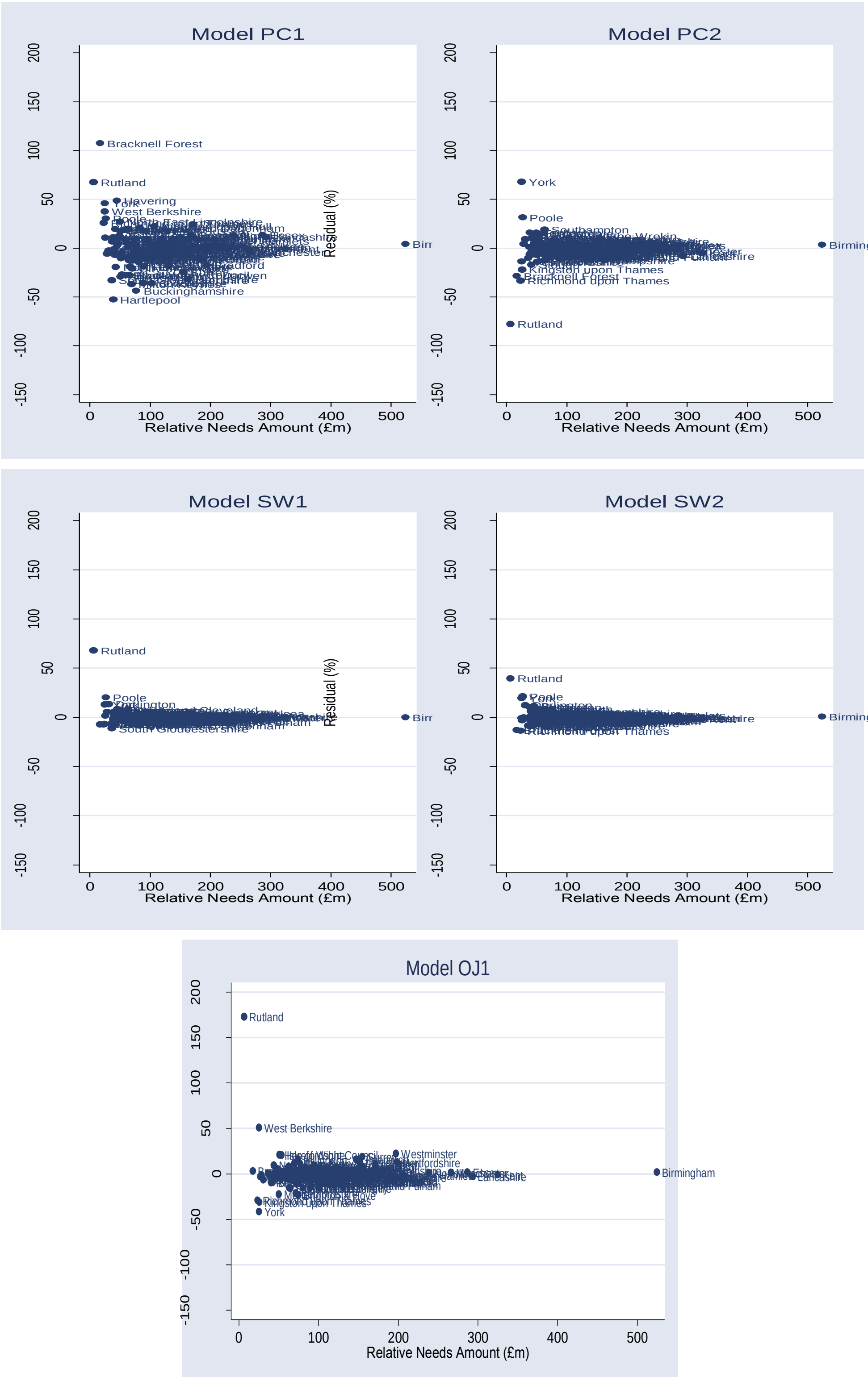
Note: 1 – counties; 2 – unitaries 3 – metropolitan districts, group 4 – London Boroughs.
Source: LE Wales.

Figure 7: Residual plots – refined models (1)



Note: 1 – counties; 2 – unitaries 3 – metropolitan districts, group 4 – London Boroughs.
Source: LE Wales.

Figure 8: Residual plots – preliminary models (2)



Source: LE Wales.

Figure 9: Residual plots – refined models (2)

Source: LE Wales.

A1.10 Appendix: Alternative specifications

Figure 10: Parsimonious models: additional specifications

```
. * Parsimonious Models (Dependent Variable: ltrnf)

. * Pl1
.          reg l1trnf l1chl12

      Source |       SS       df       MS              Number of obs =     126
-------------+------------------------------           F(  1,   124) =  404.97
       Model |  36.4082992     1  36.4082992           Prob > F      =  0.0000
    Residual |  11.1481022   124   .08990405           R-squared     =  0.7656
-------------+------------------------------           Adj R-squared =  0.7637
       Total |  47.5564014   125  .380451211           Root MSE      =  .29984

------------------------------------------------------------------------------
      l1trnf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     l1chl12 |   .7281681   .0361844    20.12   0.000     .6565491    .7997871
       _cons |   10.43326   .3463045    30.13   0.000     9.747827    11.11869
------------------------------------------------------------------------------

. * Pl2
.          reg l1trnf l1hma7

      Source |       SS       df       MS              Number of obs =     126
-------------+------------------------------           F(  1,   124) =  889.08
       Model |  41.7355638     1  41.7355638           Prob > F      =  0.0000
    Residual |  5.82083756   124  .046942238           R-squared     =  0.8776
-------------+------------------------------           Adj R-squared =  0.8766
       Total |  47.5564014   125  .380451211           Root MSE      =  .21666

------------------------------------------------------------------------------
      l1trnf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      l1hma7 |   .8303012   .0278461    29.82   0.000      .775186    .8854165
       _cons |   7.007113   .3484634    20.11   0.000     6.317406     7.69682
------------------------------------------------------------------------------

. * Pl3
.          reg l1trnf l1chl12 l1adl31

      Source |       SS       df       MS              Number of obs =     126
-------------+------------------------------           F(  2,   123) =  205.61
       Model |  36.6070162     2  18.3035081           Prob > F      =  0.0000
    Residual |  10.9493852   123  .089019392           R-squared     =  0.7698
-------------+------------------------------           Adj R-squared =  0.7660
       Total |  47.5564014   125  .380451211           Root MSE      =  .29836

------------------------------------------------------------------------------
      l1trnf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     l1chl12 |   .7399884   .0368648    20.07   0.000     .6670167      .81296
     l1adl31 |   1.054866   .7060281     1.49   0.138    -.3426737    2.452405
       _cons |   9.561262   .6777725    14.11   0.000     8.219653    10.90287
------------------------------------------------------------------------------

. * Pl4
.          reg l1trnf l1hma7 l1adl31

      Source |       SS       df       MS              Number of obs =     126
-------------+------------------------------           F(  2,   123) =  596.88
       Model |  43.1141092     2  21.5570546           Prob > F      =  0.0000
    Residual |  4.44229219   123  .036116197           R-squared     =  0.9066
-------------+------------------------------           Adj R-squared =  0.9051
       Total |  47.5564014   125  .380451211           Root MSE      =  .19004
```

```
------------------------------------------------------------------------------
      l1trnf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      l1hma7 |   .8788154   .0256562    34.25   0.000     .8280306    .9296003
     l1adl31 |    2.85043   .4613716     6.18   0.000     1.937173    3.763687
       _cons |   4.349428   .5277047     8.24   0.000     3.304869    5.393987
------------------------------------------------------------------------------


. * P15
.         reg l1trnf l1chl12 l1hma7 l1adl31

      Source |       SS       df       MS              Number of obs =     126
-------------+------------------------------           F(  3,   122) =  617.01
       Model |  44.6158209     3  14.8719403           Prob > F      =  0.0000
    Residual |   2.9405805   122  .024103119           R-squared     =  0.9382
-------------+------------------------------           Adj R-squared =  0.9366
       Total |  47.5564014   125  .380451211           Root MSE      =  .15525

------------------------------------------------------------------------------
      l1trnf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     l1chl12 |   .2578885    .032672     7.89   0.000     .1932111    .3225659
      l1hma7 |   .6507194   .0356982    18.23   0.000     .5800512    .7213876
     l1adl31 |   2.654938   .3777218     7.03   0.000       1.9072    3.402676
       _cons |   4.879335   .4362946    11.18   0.000     4.015646    5.743023
------------------------------------------------------------------------------
```

Source: LE Wales.

Annex 2 Approach 2 – Local govt, health & education spend

A2.1 Introduction

This Annex details our method and results for Approach 2.

The Annex is structured as follows:

- In Section A2.2 we provide a description on the data used and how these relate to the data used in Approach 1;
- In Section A2.3 we describe the results of our analysis for Approach 2.

A2.2 Data

The data for the analysis uses combined allocations of public expenditure resources. The same variables from Approach 1 are used as independent variables[33]. This is explained in detail further below.

Dependent variable

For Approach 2, we use the combined targeted allocations as the dependent variable. This variable is provided in per capita terms for 137 different combined local government areas in England for 2010/11[34] and is calculated as the sum of the relevant expenditure allocations estimated for[35]:

- Primary Care Trust NHS;
- LEA Direct School Grant; and,
- Local Government formula grant.[36]

[33] A detailed description of the dataset used in Approach 1 has been provided in the first draft report.

[34] Data was also made available on the actual per capita allocation for 2010/11. Further information provided by the Welsh Assembly Government who prepared the data indicated that the targeted allocations should be the focus of the analysis.

[35] Additional figures are also available as an adjusted sum, which proportionally adjusts Council Tax deductions (per head) in each Authority such that they are proportional to the tax base across all areas. As agreed with the Welsh Assembly Government, the adjusted sum figures are not used for the analysis.

[36] NHS and Local Government estimates are prior to damping adjustments.

Independent variables

As in Approach 1, we use indicator variables and underlying data that constitute the service groups in the relative needs formulae, which consist of demographic and socio-economic variables (a summary of these can be seen in Table 1 shown for Approach 1). There are 99 variables (93 underlying and 6 area cost adjustment variables), which are provided for each of the 149 districts.

Combining datasets

As noted, the units of analysis of the dependent and independent variables are different:

- Data on spend is provided in per capita terms for 137 different combined Local Government Areas (LGAs) in England, whereas

- Underlying and variables for area cost adjustment are provided for each of the 149 LGAs (districts) in England.

To make the data comparable, 21 districts in the data set for underlying and area cost adjustment variables were aggregated into 9 combined LGAs (this resulted into 137 combined LGAs, 149-21+9). The correspondence between combined LGAs and the 21 districts was provided by the ICFFW and it is shown in Table 13.

Table 13: Correspondence: Combined local government areas – districts

Class	Combined local government areas	Districts
Shire Unitary	Bournemouth & Poole	Bournemouth; Poole
	Essex, Southend on Sea & Thurrock	Essex; Southend on Sea; Thurrock
	West Berkshire, Reading & Wokingham	West Berkshire; Reading; Wokingham
	Leicester & Rutland	Leicester; Rutland
	Bracknell Forest, Windsor & Maidenhead, Slough	Bracknell Forest; Windsor & Maidenhead; Slough
	North Yorkshire & York	North Yorkshire; York
	Halton & St. Helens	Halton; St. Helens
London	Merton & Sutton	Merton; Sutton
	City & Hackney	City of London; Hackney

The different variables were combined according to different criteria:

- Variables in totals were simply summed across the relevant areas (e.g. chl19 – Resident Population, number);
- Proportional variables were averaged using population weights (e.g. chl11 – Children Without Good Health, proportion);
- Scaling factors were averaged using population weights (e.g. chl46 – Area Cost Adjustment for Education);
- Variables related to density and sparsity were excluded, as it was not possible to construct an aggregate.[37]

The final dataset contains 92 variables as it excludes 7 density/scarsity variables.

Finally, to allow for different model specifications variables were expressed in per capita terms and in logarithms. Independent variables drawn from the 2001 Census have been expressed in per capita terms by dividing them by the 2001 resident population (from the Census), to keep consistency with the indicators used. Dependent variables relating to expenditures were converted to per capita terms by

[37] The Density/Sparsity variables dropped were: chl7, chl8, adl14, adl15, epc1, epc2 and epc3.

dividing by population in 2007[38]. There were a number of zeroes in the data for some of the independent variables so instead of using the logarithm of the variable we used the logarithm of the variable plus one for the models in logarithms, i.e. for each variable *x* we used *ln(x+1)*.

Observations used

The final dataset for Approach 2 analysis contains 137 combined local government areas/observations: 32 non-metropolitan counties, 39 unitary districts, 35 districts for the 6 metropolitan counties, and 31 London Boroughs.

As with Approach 1, the accuracy of prediction of the new proposed models is tested using "out-of-sample" prediction.[39] One should note that because the number of observations has changed, the out-of-sample will be different from that in Approach 1. Nevertheless, conclusions from both approaches (in terms of prediction performance) are comparable.

Overall the dataset has:

- Observations: 118 for estimation, and 19 for out-of-sample prediction[40]
- Variables: 92 (86 underlying variables and 6 area cost adjustment variables).

The out-of-sample observations are provided in Table 14.

[38] The population data for 2007 was provided by the ICFFW Secretariat alongside the expenditure data and was taken from previous allocation models/formulae.

[39] Further details of testing using "out-of-sample" predictions can be found in the discussion of Approach 1.

[40] Data for out-of-sample prediction was obtained as a 15% random draw (by class/type of area).

Table 14: Out of sample units – Approach 2		
Main division	**Districts**	
Non Metropolitan	**Counties** Cambridgeshire Devon Dorset	 Durham Gloucestershire
	Unitaries Blackpool Leicester & Rutland North Lincolnshire	 Plymouth Stockton-on-Tees
Metropolitan Counties	**Greater Manchester** Wigan **Merseyside** Sefton	**West Midlands** Birmingham Coventry **South Yorkshire** Barnsley
Greater London	**Inner** Hammersmith & Fulham Tower Hamlets	**Outer** Barnet Merton & Sutton

A2.3 Results

As in Approach 1, our analysis followed four differentiated stages.

1) Reduction of number of variables
2) Estimation of alternative regression models
3) Assessment of appropriateness of the model and goodness-of-fit
4) Assessment of model accuracy by testing performance in out-of-sample prediction
5) Parsimonious models

A2.3.1 Reduction of number of variables

This was done using stepwise backward selection and manual/judgemental selection. The results for each method are outlined below, in turn.

Stepwise (backward selection)

Stepwise (SW) backward selection method was used for reduction of the number of variables.[41] A p-value of 0.01 was used as a critical value for statistical significance.

Four different models were estimated with the dependent variable in total and per capita terms, expressed in levels and in logs.

[41] For more information on Stepwise Selection, refer to the description of the procedure in Approach 1.

- Total allocation (with underlying and cost adjustment variables included as regressors) is modelled in:
  - Levels (model SW)
  - Logs (model SWL)
- Allocations per capita (only underlying and cost adjustment variables expressed in per capita terms are included as regressors) is modelled in:
  - Levels (model SWpc)
  - Logs (model SWLpc)

The total number of independent variables included in models SW and SWL was 81. For models SWpc and SWLpc, 52 independent variables were included. The SW selected the following variables: 30 and 18 variables, in levels and logs respectively, for the models for total allocation; and 21 and 19 variables for models of allocations per capita, in levels and logs respectively. The detailed list of variables selected by each model are presented in Table 15.

Models selected with SW criteria showed reasonably good fit (see Section A2.3.3, Table 5) but used a large number of variables. We estimated additional models with fewer variables.

Table 15: Selection of variables using SW analysis –Target Allocations

Group	Variable type	Variable	SW	SWpc	SWL	SWLpc
Financial	Assumed Debt	Assumed Debt At Start of Year 2008/09	X	X	X	X
Financial	Assumed Debt	Assumed Debt at Start of Year 2011/12	X	X		X
Financial	Assumed Debt	Assumed Debt at Start of Year 2010/11	X		X	
Financial	Est. Supported Cap Exp	Est 2008/09 Supported Capital Expenditure (Revenue)		X	X*	X
Financial	Est. Supported Cap Exp	Est 2005/06 Supported Capital Expenditure (Revenue)	X	X		
Financial	Est. Supported Cap Exp	Est 2009/10 Supported Capital Expenditure (Revenue)		X		X*
Financial	Est. Supported Cap Exp	Est 2010/11 Supported Capital Expenditure (Revenue)				X
Population	on Benefit / Disadvantaged	People Aged 18-64 Receiving Disability Living Allowance (%)	X	X		X*
Population	on Benefit / Disadvantaged	Sec School Pupils in Low Achieving Ethnic Groups		X*	X*	X*,**
Population	on Benefit / Disadvantaged	Children in Black Ethnic Groups	X*		X	
Population	on Benefit / Disadvantaged	People in Mixed Ethnic Groups	X	X*		
Population	on Benefit / Disadvantaged	Incapacity Benefit and Severe Disablement Allowance	X**			X**
Population	on Benefit / Disadvantaged	IS/IB JSA/ Guarantee Element of Pension Credit	X			X
Population	on Benefit / Disadvantaged	IS/IBJSA Claimants Aged 18-64		X*,**	X*,**	
Population	on Benefit / Disadvantaged	Older People IS/IBJSA/Guarantee Element Pension Credit(%)	X			X**
Population	on Benefit / Disadvantaged	People Aged 18-64 Receiving Disability Living Allowance	X*		X*,**	
Population	on Benefit / Disadvantaged	People in Other Ethnic Groups	X		X*	
Population	on Benefit / Disadvantaged	People Aged 18-64 Long Term Unemployed or Never Worked	X**			
Population	on Benefit / Disadvantaged	Children of IS/IBJSA Claimants				X*
Population	on Benefit / Disadvantaged	Older People Receiving Attendance Allowance			X	
Population	on Benefit / Disadvantaged	Older People Receiving Attendance Allowance (%)		X		
Population	on Benefit / Disadvantaged	Older People IS/IBJSA/Guarantee Element of Pension Credit	X*			
Population	on Benefit / Disadvantaged	People in Other Ethnic Groups			X	
Population	on Benefit / Disadvantaged	Pupils in Sec Schools with an Ethnic Group Recorded			X	
Population	on Benefit / Disadvantaged	Pupils of Sec Schools Age in Low Achieving Ethnic Group	X			
Population	on Benefit / Disadvantaged	Unemployment Related Benefit Claimants	X			
Population	Population	Net In-Commuters	X*	X		X
Population	Population	Older People Living in One Person Households		X	X	X*
Population	Population	People Aged 16-74 Highest Qualification Attained 4 or 5 (%)		X*,**	X*	X*
Population	Population	Population Aged 0-15	X	X		X*
Population	Population	Population Aged 0-17	X	X		X*
Population	Population	Resident Population Aged 18-64		X	X	X*
Population	Population	Annual Number of Day Visitors	X	X		
Population	Population	Households	X		X	
Population	Population	Older People in Rented Accommodation	X		X	
Population	Population	Resident Population	X**		X*,**	
Population	Population	Country of Birth of Residents		X		
Population	Population	Household and Supported Residents Aged 90+		X		
Population	Population	Households with No Family	X			
Population	Population	Older People in Rented Accommodation (%)		X		
Population	Population	People Aged 16-74 Highest Qualification Attained 4 or 5	X**			
Population	Population	People Aged 18-64 Work in Semi- or Routine Occupations			X	
Population	Population	People Aged 65+	X			
Population	Population	Resident Population under 18	X*			
Population	Population	Households with No Family (%)				X*
Transport	Road Length	Road Lengths-Other Built-Up Roads	X	X		X
Transport	Traffic Flow	Traffic Flow of All Vehicles	X			
Transport	Traffic Flow	Traffic Flow of HGVs, Buses & Coaches	X			
Total			**30**	**21**	**18**	**19**
Total (SW-reduced)*			**5**	**4**	**7**	**10**
Total (Own Judgement)**			**4**	**2**	**3**	**3**

SW-reduced models

We firstly used a smaller cut-off point for SW selection of variables to be in the model. Hence, we excluded variables with a p-value larger than $p=0.0001$. This is an atypical cut-off point but in this context is justified because the interest of the exercise is to compare predictive performance of models with a fewer number of variables.

This approach reduced the number of variables to 5 and 7 variables, in levels and logs respectively, for the models for total allocation. The models of per capita allocation were reduced to 4 and 10 variables, in the level and log models respectively.

Judgemental selection

Secondly, we used our own experience and judgement (OJ) criteria to select a subset of the variables chosen in the SW analysis that could be expected to explain, theoretically, the level of total allocation.

For models of total target allocation 5 variables were included in both models (SW and SWL). Following an SW selection: 4 variables were retained in the SW model; and 3 variables for SWL.

A2.3.2 Estimation of alternative regression models

We estimated four sets of models. SW, SWL, SWpc and SWLpc.

We refined such models by reducing the number of variables using statistical significance and own judgement. This resulted in models RSW, RSWL, RSWpc and RSWLpc for SW-reduced models, and OJ, OJL, OJpc and OJLpc for models with variables selected using own judgement.

The methods used for estimation were standard OLS. Data were transformed in logs and in per capita terms as explained in previous paragraphs.

A2.3.3 Assessment of appropriateness of the models and goodness-of-fit

SW models

The results of all four models indicate that the variables selected by the stepwise procedure perform well in predicting target allocation in all of the English LGAs. The adjusted R^2 of both models containing total allocation (SW and SWL) is 99.9%. The adjusted R^2 for the per capita regressions are slightly lower at 98.4% and 98.1% for SWpc and SWLpc respectively (Table 16).

Table 16: Results of regression models

Model		SW	SWpc	SWL	SWLpc
Number of obs.		118	118	118	118
Number of vars.		30	21	18	19
R2 (%)		99.9%	98.7%	99.9%	98.4%
adj R2 (%)		99.9%	98.4%	99.9%	98.1%
largest errors (%):	1	6	4	6	5
	2	5	4	5	5
	3	5	3	4	5
	4	4	3	4	4
	5	4	3	4	4
Obs within 5% error band (%)		116 (98%)	118 (100%)	117 (99%)	117 (99%)
Obs within 10% error band (%)		118 (100%)	118 (100%)	118 (100%)	118 (100%)

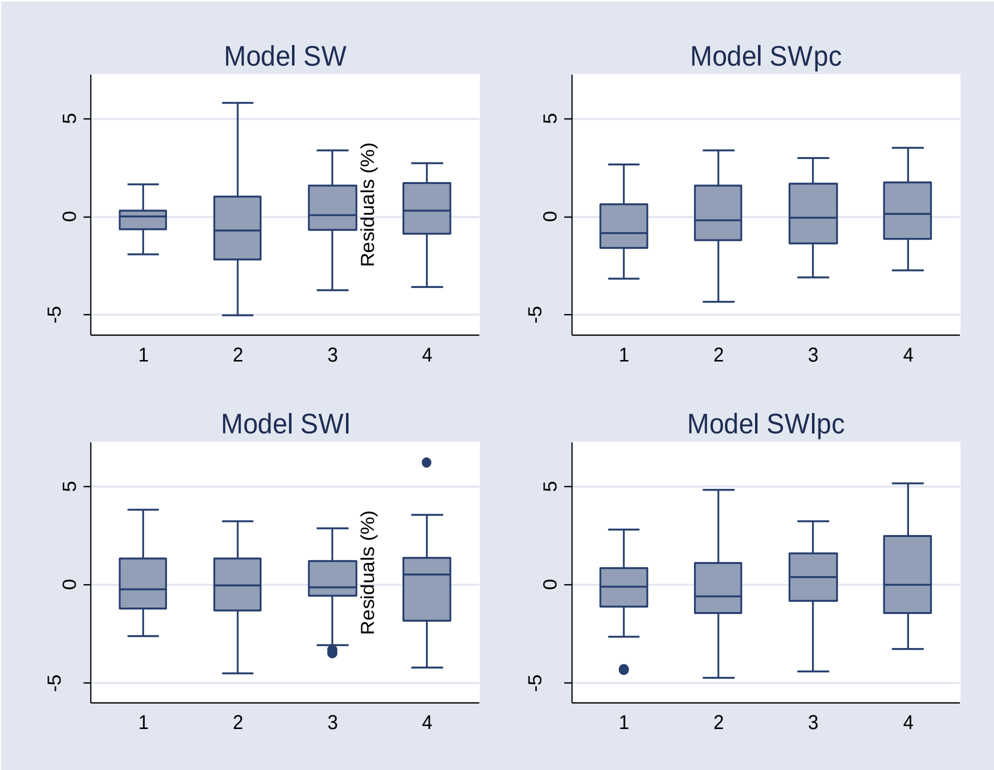
To assess the goodness-of-fit of the models we are interested in comparing the performance of the models in predicting the allocations of public expenditure for each LGA. Therefore, we transform the results of the per capita and log regressions to present the predicted values in total allocation terms.[42]

The performance of the models is then assessed by comparing the differences in the observed allocation values in the dataset and the predicted values for each local unit. Hence, we are interested in the analysis of the residuals, r, defined as $r = (y - \hat{y}) / y \times 100$.

Overall, each of the models indicate a very high goodness-of-fit with at least 98% of all predicted values found to be within ±5% of the observed values (an absolute residual of less than 5%). Table 16 presents a summary of the regression results. A detailed presentation of the regression results and the residuals is contained in Figure 11 and Figure 12.

[42] These are simply a "de-construction" or reverse process of the calculation of the per-head and logged variables. For more details on this de-construction, please see the Additional Specifications undertaken as part of Approach 1.

Figure 11: Plot of residuals – by area type



Note: 1 – counties; 2 – unitaries 3 – metropolitan districts, group 4 – London Boroughs.
Source: LE Wales.

Figure 12: Plot of residuals (authority's allocations)

Source: LE Wales.

SW -reduced models

New reduced SW models have been estimated using a different p-value for the variable selection criterion (Table 17). This reduction has a relatively small impact on the appropriateness of the models, with all models retaining an R^2 above 95%. Reducing the number of variables has impacted more significantly on the model errors, Nevertheless, all predicted values remain within a ±10% band of the observed values.

Table 17: Results of regression models (SW reduced models)

Model		RSW	RSWpc	RSWL	RSWLpc
Number of obs.		118	118	118	118
Number of vars.		5	4	7	10
R2 (%)		99.8%	95.2%	99.7%	97.3%
adj R2 (%)		99.8%	95.0%	99.7%	97.0%
largest errors (%):	1	9	9	8	7
	2	8	9	8	7
	3	8	8	7	6
	4	8	8	7	6
	5	7	7	6	6
Obs within 5% error band (%)		99 (84%)	106 (90%)	108 (92%)	110 (93%)
Obs within 10% error band (%)		118 (100%)	118 (100%)	118 (100%)	118 (100%)

Own Judgement

The results of this approach are broadly similar to those presented in relation to the SW-reduced models. This is not surprising as many of the models use the same variables (Table 15). The number of variables retained here are fewer although this does not appear to affect the R^2 of either model including total allocation (SW, SWL). The two per capita regressions are affected but to differing degrees. The regression in levels reports an R^2 of 91.2%, while model OJSWLpc no longer appears to be an appropriate model for this analysis (Table 18).

One might note the reduction in the number of variables appears to have a detrimental effect on accuracy of the models' predictions, this is something investigated further in Section A2.3.5.

Table 18: Results of regression models (own judgement models)					
Model		**OJSW**	**OJSWpc**	**OJSWL**	**OJSWLpc**
Number of obs.		118	118	118	118
Number of vars.		4	2	3	3
R2 (%)		99.6%	91.2%	99.1%	62.0%
adj R2 (%)		99.5%	91.0%	99.1%	61.0%
largest errors (%):	**1**	14	13	18	30
	2	13	11	18	29
	3	12	10	13	28
	4	11	10	13	22
	5	11	10	12	19
Obs within 5% error band (%)		80 (68%)	98 (83%)	88 (75%)	51 (43%)
Obs within 10% error band (%)		111 (94%)	114 (97%)	109 (92%)	78 (66%)

A2.3.4 Assessment of model accuracy by testing performance in out-of-sample prediction

Considering the performance of the models (SW, SWpc, SWL, SWLpc) in out-of-sample prediction, (predicting the needs for a subset of observations that have not been used at any stage of the analysis conducted so far), the models again appear to perform well. All models report an R^2 of greater than 98% and at least 89% of the models' predicted values are within a ±5% band of the observed values. The results are presented in Table 19

Table 19: Out-of –sample regression results					
Model		**SW**	**SWpc**	**SWL**	**SWLpc**
Number of obs.		19	19	19	19
Number of vars.		30	21	18	19
R2 (%)		99.9%	98.7%	99.9%	98.4%
adj R2 (%)		99.9%	98.4%	99.9%	98.1%
largest errors (%):	**1**	7	6	11	6
	2	6	5	7	5
	3	4	5	5	5
	4	4	5	4	4
	5	4	4	3	4
Obs within 5% error band (%)		17 (89%)	17 (89%)	17 (89%)	18 (95%)
Obs within 10% error band (%)		19 (100%)	19 (100%)	18 (95%)	19 (100%)

Overall the models can be considered to perform well in predicting the observed level of allocation for each LGA. These results are more encouraging than those previously seen in relation to Approach 1. One significant change between the two approaches is the change in the dependent variable. Previous models used allocations of LGA resources (as specified by the English local government needs formulae) whereas the ones presented here use an aggregate of health, education and LG allocations. Considering health expenditure makes up approximately 60% of this total, any improvement in the performance of these models may be more reflective of a close correlation between health spending and population, than improvements in the predictive power of the models in addressing regional needs, per se.

A2.3.5 Parsimonious models

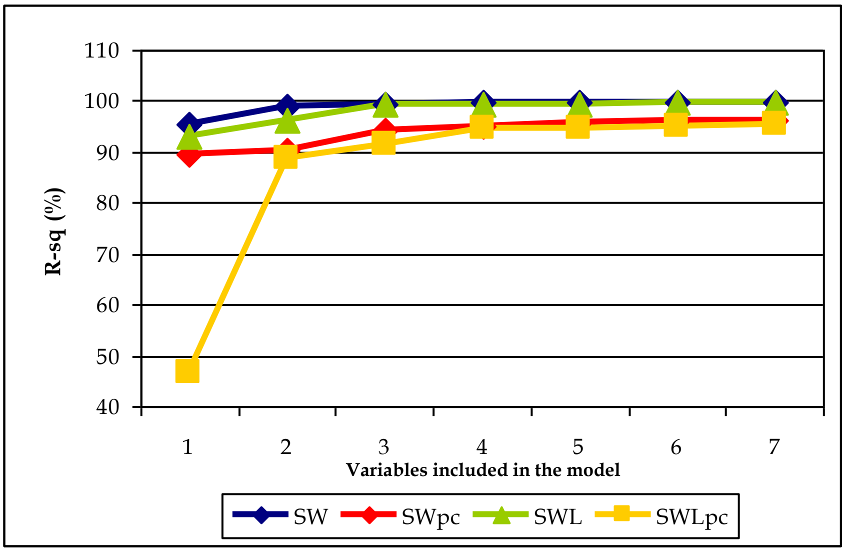
The objective of the study is to assess the goodness-of-fit of alternative models using a reduced number of variables. We have re-estimated the four models using the seven variables with the greatest statistical significance from each of the SW reduced regressions (plus a constant) The variables selected are contained in Table 20, where the row number in the first column corresponds to the number of variables contained in each iteration of each of the models (the variables included correspond to the variable in that row plus any variables in following rows). Therefore, regression (5) includes variables of rows (1-5) and regression (1) only one variable[43].

[43] Following a SW procedure, the variable with the smallest t-statistic is dropped in each iteration.

Table 20: Variables included in parsimonious models				
#	SW	SWpc	SWL	SWLpc
7	Assumed Debt At Start of Year 2008/09	Annual Number of Day Visitors	Est 2008/09 Supported Capital Expenditure (Revenue)	Older People Living in One Person Households
6	Households with No Family	Country of Birth of Residents	People in Other Ethnic Groups	Population Aged 0-15
5	Pupils of Sec Schools Age in Low Achieving Ethnic Group	People Aged 18-64 Receiving Disability Living Allowance	People Aged 16-74 Highest Qualification Attained 4 or 5 (%)	Population Aged 0-17
4	People Aged 18-64 Receiving Disability Living Allowance	People Aged 16-74 Highest Qualification Attained 4 or 5 (%)	IS/IBJSA Claimants Aged 18-64	Children of IS/IBJSA Claimants
3	Children in Black Ethnic Groups	Population Aged 0-15	Sec School Pupils in Low Achieving Ethnic Groups	People Aged 16-74 Highest Qualification Attained 4 or 5 (%)
2	Older People on IS/IBJSA/Guarantee Element of Pension Credit	People in Mixed Ethnic Groups	People Aged 18-64 Receiving Disability Living Allowance	Sec School Pupils in Low Achieving Ethnic Groups
1	Resident Population under 18	IS/IBJSA Claimants Aged 18-64	Resident Population	People Aged 18-64 Receiving Disability Living Allowance

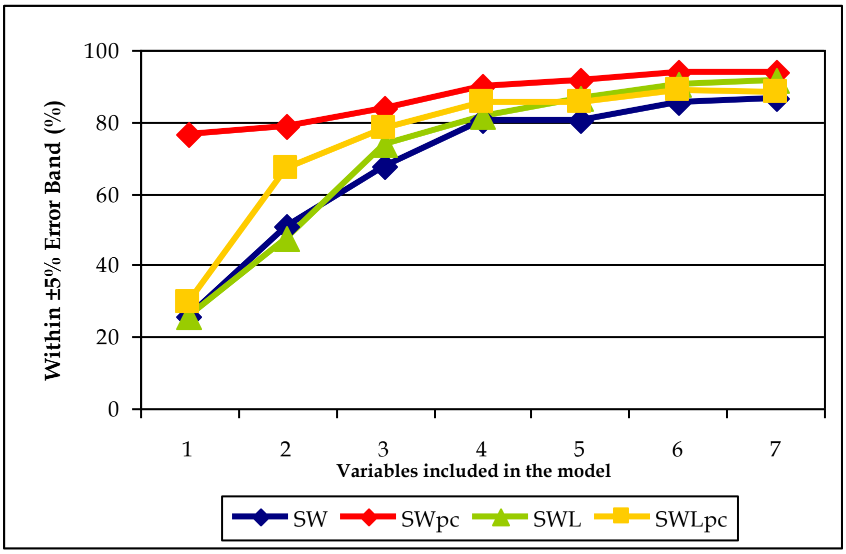
The different R^2 obtained are presented for the different models, in Figure 13. The predicted allocation values found to be within a ±5% band of the observed values are shown in Figure 14. The results of this exercise show the trade-off between the number of variables included and goodness-of-fit of each of the models: model SWLpc improves significantly from 1 to 2 variables, and in fact all models predict with an accuracy of 90% or better with 3 variables (measured in terms of R^2). The analysis on the errors is different but in general all models predict 80% (or better) of the observations within a ±5% error band.

Figure 13: Effect on R^2 from changing the number of variables



Source: LE Wales.

Figure 14: Effect on model error from changing the number of variables



Source: LE Wales.

Annex 3 Approach 3 – Alternative explanatory variables

A3.1 Introduction

This analysis considers the performance of a number of models that use external[44] publicly available data to predict the overall level of expenditure allocations of LGAs in England. The results presented in this report constitute what we call 'Approach 3'.

The report is structured as follows:

- In Section A3.2 we provide a summary of two recent reports investigating the issue of local authority funding and the Barnett Formula. These reports suggest a number of variables considered important to any calculation of relative need;
- In Section A3.3 we outline our approach and external data used for the analysis; and,
- In Section A3.4 we describe the results of our analysis.

There are a number of appendices which provide further detail on the data, the method and the results.

A3.2 Background

The Barnett Formula determines a block grant provided by the UK Government to the devolved administrations. This block grant is then used by the devolved administrations to decide the funding policy for its activities, as the legislation confers them freedom to allocate resources in line with its own policies.

The Barnett Formula has been used for the last 30 years and is applied annually to determine the block grant payable to each of the devolved administrations. For each year, the total block grant paid from the previous year (baseline) is used plus an increment which accounts for: changes in planned spending by departments in England; comparisons of the services provided by the devolved administrations; and the population proportion of each devolved administration.

[44] This refers to data that was not, per se, included in the original database used for the purpose of calculating the RNF or RNA. The data for this approach is taken from the most recent publicly available sources.

The allocation of funding to local government in England, and the Barnett formula, are currently being reviewed with the aim of identifying possible alternative funding mechanisms for Wales. Two separate studies are currently being undertaken:

- In December 2008, the House of Lords appointed the Select Committee on the Barnett Formula (SCBF) to examine, "the purpose, methodology and application of the Barnett Formula as a means of determining funding for the devolved administrations of the United Kingdom.[45]"
- The Independent Commission on Funding and Finance for Wales (ICFFW) was set up to look at the advantages and disadvantages of the present formula based approach to the distribution of public expenditure resources to the Welsh Assembly Government; and identify possible alternative funding mechanisms.

In this section we provide a brief summary of the first reports of the SCBF and the ICFFW. The reports are important because they provide recommendations on the indicators to be included in the calculation of relative expenditure need in Wales.

Select Committee on the Barnett Formula

In their first report, the SCBF noted two significant criticisms of the Barnett Formula. The first relates to the inability of the formula to account for population changes over time. The second refers to the equity of the formula and its failure to account for the increased divergence over time of the public services provided by the devolved administrations. The differences in public service provision are a function of devolution, as different devolved administrations pursue different policies that are not replicated throughout the UK.

In addressing the advantages of the Barnett Formula, the SCBF highlighted its simplicity, stability, and the absence of ring-fencing as important features that should be maintained in the future. The SCBF also found compelling that funding is based on relative needs characterised by[46]:

- A top-down approach using a small number of aggregate regional statistics;
- A small number of specific measures of needs (to be restricted to national statistics);
- A combined single measure for each of the devolved administrations in the UK using weights consistent with the level of UK public expenditure.

[45] The study excluded considering the distribution of funds within the different regions of the UK.

[46] The rationale for advocating these key features are: equal treatment of all people with a certain need across all areas; use of national statistics so that the indicators themselves are beyond reproach; and indicators chosen not affected by policy choices/changes that may be available to or enacted by devolved administrations.

Finally, the report suggests using 8 variables (plus 4 supplementary) for determining the baseline of relative needs. All of the variables are chosen to reflect the total number of people (not rates or proportions) and are described in Appendix A3.5 of this Annex.

Independent Commission on Funding and Finance for Wales

The report addresses the advantages and disadvantages of the present formula-based approach for the distribution of public expenditure resources to the Welsh Assembly government. It concludes that there is an urgent requirement to reform the funding arrangements for Wales, as left alone public services in Wales will become increasingly underfunded over the next decade.

As a metric for determining funding over the medium term in Wales, the ICFFW is of the view that this should be based on relative needs. The relative needs of different areas are to be determined by a series of variables to be used as indicators of 3 broad categories: demographics, cost, and deprivation.

Summary of suggested variables

Table 21 presents a summary of the variables indicated by both the SCBF and ICFFW Reports. Although the ICFFW report does not explicitly reference variables to be used in such an analysis, the report does present variables within each category that have been taken as relevant proxies for the intended need. The variables are grouped into classes of variables as suggested in the ICFFW Report (demographics, cost, and deprivation).

Table 21: Suggested variables for needs based formulae			
Cat.	**Suggested Variables**	**Definition**	**Source**
Demographics	Aged under 5	Number of Children aged under 5	SCBF
	Aged 5 to 16	Number of Children aged 5 to 16	SCBF
	Aged 65+	Number of People aged 65+	SCBF
	Aged between 65 and 74	Number of People aged between 65 & 74	SCBF
	Aged 75+	Number of People aged 75+	SCBF
	Dependency ratio	% of the population aged 0-16 and 65+	ICFFW
	Ethnicity	% of the population belonging to a black or ethnic minority group	ICFFW
Cost	Sparsity	% population living outside settlements of 10,000 or more	ICFFW
	Labour costs	Public sector pay levels	ICFFW
	Other costs	Capital, rental & utility costs	ICFFW
Deprivation	Gross Value Added	Per capita measure of economic output or contribution to the overall economy	ICFFW
	Household income	Per capita household income	SCBF
	Adults with no qualifications	Number of Adults with no NVQ qualifications	ICFFW
	Employment rate	% of the working age population (16-64) in employment	ICFFW
	Unemployment rate	Number of people of working age without a job and actively seeking one	SCBF
	Not in employment	Number of people of working age who are not in employment	SCBF
	Mortality rate	Number of deaths, standardised for age profile of the population	SCBF
	Limiting long-term illness	% of the population with a limiting long term illness	ICFFW /SCBF
	Working age population with limiting LT illness	% of the population, of working age, with a limiting long term illness	ICFFW
	Work limiting disability	Number of people of working age with a work-limiting disability	SCBF
	Benefit claimants	% of working age population claiming social security benefits	ICFFW
	Child poverty	Number of children belonging to households with an income below 60% of the UK median	SCBF
	Adult poverty	Number of adults belonging to households with an income below 60% of the UK median	SCBF
	Index of multiple deprivation	Index providing a weighted score across a number of different measures of deprivation	ICFFW

A3.3 Data

As in Approach 2, we use the combined targeted allocations for 2010/11 as the dependent variable. The variables indicated in both the ICFFW and SCBF reports, or relevant proxies of those variables, are used as the independent variables.

Dependent variable

We use the estimate of the actual expenditure allocation as the dependent variable. This is the same dependent variable used in the Approach 2 analysis and is provided in per capita terms for 137 different combined local government areas in England for 2010/11.[47]

Independent variables

Data have been obtained from a number of publicly available sources, as well as directly from the Offices of the Welsh Assembly Government. The following data sources have been used:

- Annual Population Survey (APS) 2008;
- Department of Work and Pensions (DWP) 2008, data on benefit claimants;
- Census 2001;
- Office of National Statistics (ONS), data on Gross Value Added (GVA), 2006, and Gross Household Disposable Income (GHDI), 2007; and,
- Where relevant, underlying data from Approach 1 & 2 analysis.

The variables suggested by SCBF and ICFFW for measuring needs are presented in Table 22 and Table 23, with the different variables and their names we identified from different data sources. It should be noted we could not find a suitable variable for two of the indicators (Labour and Other Costs and Index of Multiple Deprivation) and these have been excluded in the regression analysis. Finally, all variables use the most recent data where possible.

[47] The variable is the sum of the relevant per capita expenditure allocations estimated for: Primary Care Trust NHS; LEA Direct School Grant; and, Local Government formula grant.

Table 22: Variables suggested by SCBF

Cat.	Sub-category	Variable	Name	Source
Demography	Age 1	Under 5	pop01_u5	Census 2001
	Age 2	5 – 15	pop01_5_19	Census 2001
	Age 3	65+	pop01_65	Census 2001
	Age 4*	65 – 74 & 75+	pop01_65_74 pop01_75	Census 2001
Cost	N/A	N/A	N/A	N/A
Deprivation	Household Income	Gross Household Disposable Income, per capita (inverse)	ghdi07pc_inv	ONS
	Labour Market	Unemployed	waunemp08	APS 2008
	Child Poverty	Children of parents on benefit	chl1	Approach 1&2
	Health	Mortality Rate (standardised)	smortr	1999-2003 (ICFFW)
		Population of working age with Limiting Long-Term Illness	wa_llti01	Census 2001
	Adult Poverty*	Working Age on Benefit	waben	DWP 2008
	Labour Market*	People not in Employment	wanemp08	APS 2008
	Health*	Population with Limiting Long-Term Illness	llti_tot01	Census 2001

Note: * denotes additional suggested variables.

Table 23: Variables suggested by ICFFW				
Cat.	**Sub-category**	**Variable**	**Name**	**Source**
Demography	Dependency Ratio	Dependency ratio	dep_ratio07	APS 2008
	Ethnicity	% Non-Whites in the Population	nwhite08_pct	APS 2008
Cost	Sparsity	% Population living in settlements >10,000 inhabitants	spars01	Census 2001
	Labour & Other Costs	N/A	N/A	N/A
Deprivation	Gross Value Added	GVA, per capita	GVA06pc_inv	ONS
	Skills & Labour Market Performance	% People of working age with No Qualifications	wanq08_pct	APS 2008
		% people of working age in Employment	waemp08_pct	APS 2008
	Health	% Population with Limiting Long-Term Illness	llti01_pct	Census 2001
		% Population of working age with Limiting Long-Term Illness	wallti01_pct	Census 2001
	Benefits	% Working age Population receiving Benefit(s)	wa_adlb2_pct	DWP 2008
	Index of Multiple Deprivation	N/A	N/A	N/A

Combining datasets

It is important to note that, in many cases the geographic area of the variables contained in the publicly available data differ from that of the dependent variable.

- Data on spend is provided in per capita terms for 137 different combined Local Government Areas (LGAs) in England, whereas
- Data available from the ONS (APS and Census) and DWP are provided for 143 LGAs (districts) in England.

To make the data comparable, 11 districts in ONS/DWP data were aggregated into the 5 LGAs geographical split of the dependent variable. This resulted in 137 LGAs, (143-11+5). The correspondence between combined LGAs and the 11 districts was provided by the Welsh Assembly Government and it is shown in Table 24.[48]

[48] The Welsh Assembly Government also provided a mapping of NUTS 3 areas to the LGAs. This was used to map the GVA and GHDI data from the ONS.

Table 24: Correspondence: Combined local government areas – districts		
Class	Combined local government areas	Districts
Shire County	Cumbria	East Cumbria; West Cumbria
	Derbyshire	East Derbyshire; South & West Derbyshire
	Nottinghamshire	North Nottinghamshire; South Nottinghamshire
Shire Unitary	Essex, Southend on Sea & Thurrock	Essex; Southend on Sea; Thurrock
	North Yorkshire & York	North Yorkshire; York

For all of the variables, the raw data was obtained in levels and as such could be easily combined (summed) for each area.

Finally, to allow for different model specifications, relevant variables have been expressed in per capita terms. Where per capita variables have been constructed, care was taken to ensure these were constructed relative to the appropriate population base. For example, independent variables drawn from the 2001 Census have been expressed in per capita terms by dividing them by the 2001 Census resident population to keep consistency with the indicators used.

Observations used

The final dataset for this analysis contains 137 combined local government areas/observations: 32 non-metropolitan counties, 39 unitary districts, 35 districts for the 6 metropolitan counties, and 31 London Boroughs.

As with Approach 1 and 2, the accuracy of prediction of the new proposed models is tested using "out-of-sample" prediction.[49] One should note that the out-of-sample observations are the same as those in Approach 2 so results from both approaches are comparable.

Overall the dataset has:

- Observations: 118 for estimation, and 19 for out-of-sample prediction.[50] A list of the out-of-sample observations is provided in Table 25.

Variables: 21, a summary of these has already been presented in Table 22 and Table 23. A more detailed description is available in Appendix A3.5 to this Annex.

---

[49] Further details of testing using "out-of-sample" predictions can be found in the discussion of Approach 1.

[50] Data for out-of-sample prediction was obtained as a 15% random draw (by class/type of area).

Table 25: Out of sample units – Approach 3		
Main division	**Districts**	
Non Metropolitan	**Counties**	
	Cambridgeshire	Durham
	Devon	Gloucestershire
	Dorset	
	Unitaries	
	Blackpool	Plymouth
	Leicester & Rutland	Stockton-on-Tees
	North Lincolnshire	
Metropolitan Counties	**Greater Manchester**	**West Midlands**
	Wigan	Birmingham
	Merseyside	Coventry
	Sefton	**South Yorkshire**
		Barnsley
Greater London	**Inner**	**Outer**
	Hammersmith & Fulham	Barnet
	Tower Hamlets	Merton & Sutton

A3.4 Results

As with our previous analysis for Approach 1 and Approach 2, our analysis followed differentiated stages:

- Estimation of regression models;
- Assessment of appropriateness of the model and goodness-of-fit; and,
- Assessment of model accuracy by testing performance in out-of-sample prediction.

A3.4.1 Estimation of regression models

We estimated 3 models using the suggested variables (described earlier in Table 22 and Table 23). The method used for estimation was standard OLS and the dependent variable is the expected level of Actual Expenditure (Target Expenditure). Where appropriate, the data have been transformed into per capita terms, or proportions (%), consistent with the approach outlined in section A3.3.

It should be noted that SCBF approach uses total targeted expenditure as the dependent variable, while the ICFFW analysis is based on target expenditure per capita. For the analysis of the SCBF approach two models are presented, the first (Model 1a) contains the first set of 8 suggested independent variables, Model 1b also includes the additional suggested variables.

Furthermore, to test the relative performance of these variables, a model (Model 3) has been estimated including all of the variables suggested in the SCBF and ICFFW reports, but excluding overlapping variables such as demographic variables related to population, and people in employment (as the model already uses an unemployment variable). This regression employs a stepwise backward variable selection criterion, preserving only variables that are found to be statistically significant at the 10% level. For this regression, all variables are converted to per capita (or %) terms and the dependent variable is per capita targeted expenditure.

Model 3 contains the following variables:

- Working age population who are unemployed (%);
- Gross household disposable per capita income (inverse);
- Dependency ratio (population not of working age), (%);
- Population classified as non-white (%);
- Population (working age) with a limiting long-term illness (%);
- Children in the population whose parents are on benefit (%).

A3.4.2 Assessment of appropriateness of the models and goodness-of-fit

The performance of the models in predicting the target level of expenditure of the LGAs can be seen to be mixed (Table 26). Model 1a and 1b appear to provide the best fit with an R^2 value of 99.8%, respectively; Model 2 provides the lowest R^2 value (93%); while Model 3 reports a value in between these two, approximately 97%.

To assess the goodness-of-fit of the models we are interested in comparing the performance of the models in predicting the total target expenditure for each LGA. Therefore, we transform the results of the per capita regressions to present the predicted values in total target expenditure terms.[51]

The performance of the models is then assessed by comparing the differences in the observed target expenditure values in the dataset and the predicted expenditure values for each local unit. Hence, we are interested in the analysis of the residuals, r, defined as $r = (y - \hat{y}) / y \times 100$.

Overall, the results are again mixed. Considering the predicted target expenditure values found to be within ±5% of the observed values (an absolute residual of less than 5%) Model 2 and particularly Model 3 provide a better fit than either of the Model 1 regressions. This largely reverses the performance ordering of the models based simply on R^2. At the ±10% level, 98% of all observations are contained within

[51] These are simply a "de-construction" or reverse process of the calculation of the per-head and logged variables. For more details on this de-construction, please see the Additional Specifications undertaken as part of Approach 1.

this band for Models 1a, 1b and 2. Almost all of the predicted expenditure values are found to be within ±10% of the predicted values from Model 3.

Table 26 presents a summary of the regression results, including residuals. A detailed presentation of the regression results and the residuals is contained in Appendices A3.6 and A3.7 to this Annex.

Table 26: Results of regression models

Model		Model 1a	Model 1b	Model 2	Model 3
Number of obs.		118	118	118	118
Number of vars.		8	12	8	6
R2 (%)		99.8%	99.8%	93.0%	95.6%
adj R2 (%)		99.8%	99.8%	92.7%	95.3%
largest errors (%):	1	16	16	13	10
	2	15	15	12	9
	3	8	10	10	8
	4	8	8	9	7
	5	8	7	9	6
Obs within 5% error band (%)		98 (83%)	101 (86%)	101 (86%)	108 (92%)
Obs within 10% error band (%)		116 (98%)	116 (98%)	116 (98%)	117 (99%)

A3.4.3 Assessment of model accuracy by testing performance in out-of-sample prediction

Considering the performance of the models (1a, 1b, 2, and 3) in out-of-sample prediction, (predicting the needs for a subset of observations that have not been used at any stage of the analysis conducted so far), the performance of the models is again mixed. The results are consistent with those already observed in relation to the in-sample predictions, with Model 3 performing best in terms of goodness-of-fit. All of the out-of-sample observed expenditure values were found to be within ±5% of the values predicted by the model. The results are presented in Table 27.

Table 27: Out-of–sample regression results

Model		Model 1a	Model 1b	Model 2	Model 3
Number of obs.		19	19	19	19
Number of vars.		8	12	8	6
R2 (%)		99.8%	99.8%	93.0%	95.6%
adj R2 (%)		99.8%	99.8%	92.5%	95.3%
largest errors (%):	1	13	13	14	5
	2	8	8	9	4
	3	5	5	5	4
	4	5	5	5	3
	5	4	5	4	3
Obs within 5% error band (%)		16 (84%)	17 (89%)	17 (89%)	19 (100%)
Obs within 10% error band (%)		18 (95%)	18 (95%)	18 (95%)	19 (100%)

Overall, the performance of the models in predicting the observed level of target expenditure for each LGA is mixed. Despite high R^2 values, Models 1a and 1b perform relatively poorly in terms of goodness-of-fit when the expenditure values predicted by the models were compared to the observed levels. This result is likely to be due to the relationship between the total level of targeted expenditure (dependent variable in these regressions) and the inclusion of population variables as explanatory variables in the regressions. Despite presenting significantly lower R^2 values, Model 2 was found to provide a similar goodness-of-fit to the observed data as Models 1a and 1b. As with Model 2, Model 3 uses per capita targeted expenditure as the dependent variable and this model is found to provide the best goodness-of-fit of all the models considered.

In relation to the results previously found in the Approach 2 analysis, the models considered here do not perform as well as the fully specified models under that approach. Relative to the reduced models considered under Approach 2, (with a similar number of explanatory variables) Model 3 can be seen to perform as well as its counterparts, using economically intuitive publicly available data.

A3.4.4 Parsimonious models

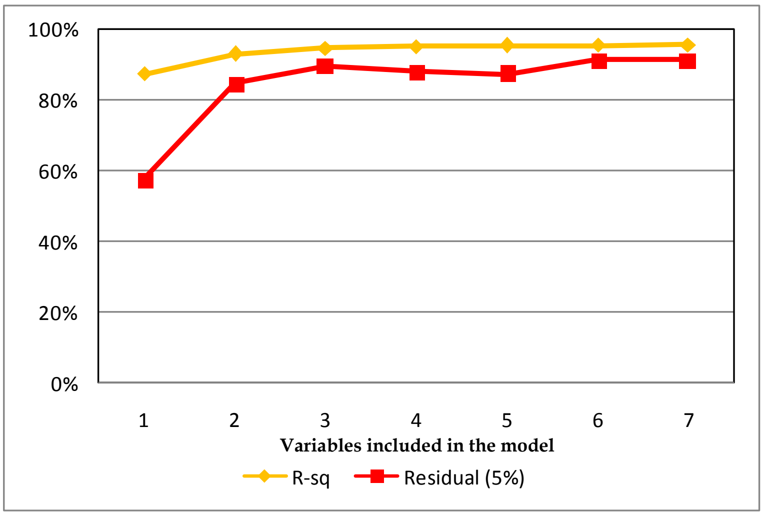
Adopting an analogous approach to that presented in Approach 2, we now examine alternative models using a reduced number of variables. We have started with the seven variables in Model 3 (plus a constant), and reduced the number of variables by one in successive iterations, based on statistical significance. The variables, and the sequence in which they were dropped are shown in Table 28, where the row number in the first column corresponds to the number of variables contained in respective iteration of the model (the variables included correspond to the variable in that row plus any variables in the following rows). Therefore, regression (5) includes variables of rows (5-7) and regression (1) only one variable[52].

Table 28: Variables included in reduced models

#	Variable Label	Description
7	spars01	Population in Settlements >10,000 (%)
6	waunemp08_pct	Working age population who are unemployed (%)
5	ghdi07pc_inv	Gross household disposable per capita income (England 2007=100) (inverse)
4	dep_ratio07	Dependency ratio (population not of working age), (%)
3	nwhite08_pct	Population classified as non-white (%)
2	wallti01_pct	Population (working age) with a limiting long-term illness (%)
1	chl1_pct	Children in the population whose parents are on benefit (IS/IB JSA) (%)

The results of this additional analysis are presented in Figure 15. From this one can see the trade-off between the number of variables included and goodness-of-fit of each of the models: the model improves significantly from 1 to 2 variables, in terms of the number of observations within a ±10% error band. All models report an accuracy of better than 90% (when measured in terms of R^2).

Figure 15: R^2 and model residuals and number of variables



Source: LE Wales.

[52] Following a SW procedure, the variable with the smallest t-statistic is dropped in each iteration.

The full regression results of these models are presented in Appendix A3.10 to this Annex. We have noted that the sign of the variable related to disposable income has an unexpected negative sign. This is because of the high correlation between the variables included in the models. The descriptive analyses of the variables show that gross household disposable income is highly correlated with the variable related to population with limiting long-term illness. The analysis also shows a very noticeable correlation of 0.93 between expenditure per head and the variable related to children with parents on benefit (see Appendix A3.8 to this Annex).

Another way of showing the importance of individual variables is by running univariate regressions. Again, this shows the importance of the variable related to children with parents on benefit, which is able to account for almost 88% of variation (see Appendix A3.8 to this Annex).

To explain expenditure based on a fewer number of variables using variables that can have an economic meaning which is easy to understand, one may want to explore alternative specifications. As an example of potential alternative specifications, we provide a model for expenditure per head with just three explanatory variables related to: children with parents on benefit (`chl1_pct`), the inverse of gross household disposable income per head (`ghdi07pc_inv`), and unemployment (`waunemp08_~t`). The model has good statistical properties and an R^2 of 90%[53] (Figure 16).

Figure 16: Alternative Model 1

```
.          reg exp_ph chl1_pct ghdi07pc_inv waunemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  3,    114) =  338.71
       Model |  20945562.6     3  6981854.2            Prob > F      =  0.0000
    Residual |  2349912.65   114  20613.2688           R-squared     =  0.8991
-------------+------------------------------           Adj R-squared =  0.8965
       Total |  23295475.3   117  199106.626           Root MSE      =  143.57

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   159.3744    7.59559    20.98   0.000     144.3276    174.4212
ghdi07pc_inv |    3559906    1199904     2.97   0.004      1182906     5936905
waunemp08_~t |   42.45731   13.37977     3.17   0.002      15.9521    68.96252
       _cons |   1543.658    82.2265    18.77   0.000     1380.768    1706.548
------------------------------------------------------------------------------
```

[53] This model predicts 70% of observations within ±5%, and 95% observations within ±10%. The largest errors are (in absolute terms): 18, 12, 11, 10 and 10.

Alternatively, one could use the previous Model 3, without the variable on gross disposable income and without the variable for unemployment (shown to be not statistically significant). The model is shown in Figure 17 with an R^2 of 95%[54].

Figure 17: Alternative Model 2

```
.         reg exp_ph chl1_pct wallti01_pct nwhite08_pct dep_ratio07 if
sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  4,    113) =  554.44
       Model |  22166060.1     4  5541515.03           Prob > F      =  0.0000
    Residual |  1129415.16   113  9994.82441           R-squared     =  0.9515
-------------+------------------------------           Adj R-squared =  0.9498
       Total |  23295475.3   117  199106.626           Root MSE      =  99.974

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   121.9765   9.841939    12.39   0.000     102.4778    141.4751
wallti01_pct |   61.29298   6.597028     9.29   0.000     48.22308    74.36289
nwhite08_pct |   7.290896   1.146711     6.36   0.000     5.019054    9.562737
 dep_ratio07 |    10.4751   3.764867     2.78   0.006     3.016223    17.93398
       _cons |   928.7929   134.8786     6.89   0.000     661.5741    1196.012
------------------------------------------------------------------------------
```

A3.4.5 Aggregation

The analysis reported above results in a large amount of aggregation with respect to service types compared to the funding allocation formulae that are currently used.[55] In their report, the House of Lords Select Committee on the Barnett Formula favoured a top-down approach, using a small number of aggregate statistics. Characterising this as a 'simple approach', they said "Whilst it may reasonably be countered that the cost of simplicity is a certain rough justice, we would expect that cost to be lower at the national level than locally since differences between countries are much smaller than differences between localities within countries." (para. 91)

Our analysis in Approaches 2 and 3 suggests, in the context of that particular dataset for England, that the cost of simplicity is relatively small overall (though there might be significant effects for a small number of English areas).

[54] This model predicts 88% of observations within ±5%, and 98% observations within ±10%. The largest errors are (in absolute terms): 11, 11, 8, 7,and 7.

[55] In other words, our analysis includes all services in one formula, whereas the current funding allocation approach in England for local government, health and education services uses a large number of formulae, each one specific to a particular service activity.

In addition to the analysis undertaken, we also briefly examined an alternative approach where different models are estimated for each of the disaggregated three services (local government, health, and education). Thus, instead of estimating one model for the sum of the three services on a number of explanatory variables and parameters, $y = \boldsymbol{X\beta}$, we estimated one model for each of three services $y^i = \boldsymbol{X^i\beta^i}$, where *i* denotes each of the three services (therefore $y = y^1 + y^2 + y^3$).

We used different specifications with variables selected on a stepwise procedure. Thus, for each of the disaggregated equations we used different variables, and different coefficients were obtained using standard OLS regressions (details are shown in A3.10).

In order to compare the predictions of the individual equations with the predictions of the model for aggregated expenditure *y*, we constructed an aggregate prediction, $\hat{y}^{123}$, from the individual models as the sum of the predictions of the disaggregated models $\hat{y}^i$ (so that $\hat{y}^{123} = \hat{y}^1 + \hat{y}^2 + \hat{y}^3$). In this way, the aggregated coefficient of determination, or AR^2, is calculated from the aggregate prediction of disaggregated models as $AR^2 = 1 - RSS/TSS$, where RSS is the residual sum of squares (or sum of squared errors, $\Sigma (y - \hat{y}^{123})^2$), and TSS is the total sum of squares (or total sample variance, $\Sigma (y - \bar{y})^2$).

Comparisons of R^2 and AR^2 show that there is no significant difference in the goodness of fit of the two methods: R^2 is 95.56% for the aggregated model and AR^2 is 95.53% for the one constructed from the three models[56] (full details are shown in A3.10). It should be noted however that for each of the models estimated here, we started with the same set of potential explanatory variables and then applied the stepwise procedure. For the three disaggregate models it may well be possible to improve on their performance by using alternative, more relevant, variables (e.g. more specifically aimed at explaining variations in health expenditure for the health model). This means that is may be possible to improve on the AR^2 value quoted above, though as the AR^2 value is already relatively high the scope for improvement against this measure must be limited.

[56] One may wonder about the relationship between R^2 and AR^2. By construction, AR^2 is the sum of the predictions of models with different specifications. Thus, it is possible that the predictions obtained from the disaggregated model ($\hat{y}^{123}$) are closer or further away to the observed values than the predictions of the aggregated model ($\hat{y}$). As a consequence the AR^2 could be greater or lower than the R^2 form the aggregated model.

A3.5 Appendix: Description of variables

SCBF

The House of Lords Select Committee on the Barnett Formula has suggested a number of indicators of needs that are used in the analysis contained in this report. All of the indicators relate to the *total number of people* in each category, contained with the geographic boundary of each Local Authority. The variables are:

- The number of people aged under 5 (pop01_u5). This data was taken from the Census 2001.

- The number of people aged 5-16 (pop01_5_19). From the data publicly available on the Census 2001, it was not possible to obtain a breakdown of population consistent with this. For the purpose of this analysis the age category was expanded slightly to fit with the available data. The variable in the analysis encompasses those aged 5-19.

- The number of people aged 65+ (pop01_65). This data was taken from the Census 2001.

- Standardised mortality rate (smortr). This data was provided by the Offices of the Welsh Assembly Government and has been standardised for age profile of each region.

- Work limiting disability among people of working age (wallti01). Publicly available data on people of working age with a limiting long-term illness was used as a proxy for this suggested indicator. This data was taken from the Census 2001.

- Child poverty - The number of children in households with income in less than 60% of UK median (chl1). Publicly available data on this indicator was not readily available at the disaggregated level needed for this analysis. The number of children of IS / IB JSA recipients was used as a proxy for this indicator and was taken from the database for Approach 1 and 2.

- Per capita household income – inverse (ghdi07pc_inv). As a further measure of poverty, per capita gross household disposable income is included in the analysis. The inverse is used to preserve the idea that higher values are linked to higher needs. This data was obtained from the ONS at NUTS 3 level and mapped to the geographic areas contained in our analysis using a mapping provided by the Offices of the Welsh Assembly Government.

- Unemployed (waunemp08). The number of working age people classified as unemployed was obtained from the Annual Population Survey (2008)

Four additional indicators are suggested by the SCBF:

- The number of people aged 75+ (pop01_75). This data was taken from the Census 2001.

- People with limiting long term illness (llti01). This variable covers all people with a limiting long-term illness and was taken from the Census 2001.

- People not in employment (wanemp08). This variable includes all unemployed and economically inactive people of working age in the population. The data was taken from the Annual Population Survey (2008).

- Adult poverty – The number of Adults in households where the household income in les than 60% of UK median. As with child poverty, publicly available data on this indicator was not readily available at the disaggregated level needed for this analysis. The number of adults on benefits was used as a proxy for this indicator and was taken from the Department of Work and Pensions (2008).

ICFFW

A description of the variables used to assess the approach suggested in the Independent Commission on Funding and Finance in Wales' First Report is presented here. These indicators are typically expressed in proportions of the overall population and have been classified in accordance with the categories of indicators given in the report.

Demographics

- Dependency ratio (dep_ratio08). This variable measures the number of children and adults above the retirement age as a proportion of the population. The variable was constructed by subtracting the working age population from the total population and expressing the result as a proportion of the total population. The data was taken from the Annual Population Survey (2008).

- Ethnicity (nwhite08_pct). As a measure of ethnicity, the number of non-white people have been expressed as a proportion of the total population. The data was taken from the Annual Population Survey (2008). It was noted in the ICFFW's report that this variable is likely to be less important for Wales than other regions of England.

Cost

- Sparsity (spar01). The sparsity variable measures the number of people (as a proportion of the population) that live in settlements with more than 10,000 inhabitants. This data was provided by the Offices of the Welsh Assembly Government, source Census 2001. The ICFFW's report notes that this variable is likely to be important for Wales due to the percentage of the population living outside large centres and relative lack of large centres.

- Labour Costs & Other Costs. Due to the absence of publicly available data on these suggested indicators, at the level of disaggregation required, these indicators have not been included in the analysis.

Deprivation

- Gross Value Added, per capita (gva06_inv). As noted in the ICFFW's First Report, GVA can be considered a 'course proxy for deprivation'. As with GHDI, the inverse is used to preserve the idea that higher values are linked to higher needs. This data was obtained from the ONS at NUTS 3 level and mapped to the geographic areas contained in our analysis using a mapping provided by the Offices of the Welsh Assembly Government.

- Skills & Labour Market Performance

  - Adults with no qualifications/unskilled (wanq08_pct). The number of adults of working age in the population with no qualifications (as measured by NVQ) is expressed as a percentage of the total working age population. This data was obtained from the Annual Population Survey (2008).

  - Employment rate (waemp08_pct). The number of adults of the working age in the population in employment is expressed as a percentage of the total working age population. This data was obtained from the Annual Population Survey (2008).

- Health

  - Population with limiting long term illness (llti01). This is measured by the proportion of the population with a limiting long term illness. The data is taken from the Census 2001.

  - Working age population with limiting long term illness (wallti01). This is measured by the proportion of the working age population with a limiting long term illness. The data is taken from the Census 2001.

- Benefit Claimants (wa_adlb2_pct). The number of adults of working age in the population reported to be recipients of benefits, expressed as a proportion of the total working age population. This data is taken from the Department of Work and Pensions (2008).

- Index of Multiple Deprivation. Publicly available data on such an index was found to be unavailable at a complete and matching level of disaggregation needed for our analysis. This indicator has not been included in the analysis.

A3.6 Appendix: Regression models

Figure 18: Summary of variables' names

* SCBF - with additional vars
    * Dep var: Expected Actual Expenditure per capita
    * pop01_u5 Population Under 5
    * pop01_5_19 Population 5 – 15
    * pop01_65 Population 65+
    * pop01_65_74 Population 65 – 74
    * pop01_75 Population 75+
    * ghdi07pc_inv Gross Household Disposable Income, per capita (inverse)
    * waunemp08 Unemployed
    * chl1 Children of parents on benefit
    * smortr Mortality Rate (standardised)
    * wa_llti01 Population of working age with Limiting Long-Term Illness
    * waben Working Age on Benefit
    * wanemp08 People not in Employment
    * llti_tot01 Population with Limiting Long-Term Illness

* ICFFW -
    * Dep var: Total Expected Actual Expenditure per capita
    * dep_ratio07 Dependency ratio
    * nwhite08_pct % Non-Whites in the Population
    * spars01 % Population living in settlements >10,000 inhabitants
    * gva06pc_inv Inverse of GVA, per capita
    * wanq08_pct % People of working age with No Qualifications
    * waemp08_pct % people of working age in Employment
    * llti01_pct % Population with Limiting Long-Term Illness
    * wallti01_pct % Population of working age with Limiting Long-Term Illness
    * wa_adlb2_pct % Working age Population receiving Benefit(s)

* SCBF variables, in % or pc to conform dep var.
    * chl1_pct Percentage of children in the population whose parents are on benefit (IS/IB JSA)

Figure 19: Model 1a

```
. * SCBF - w/o additional vars
.                reg  expm  pop01_u5  pop01_5_19  pop01_65  smortr  wa_llti01  chl1
ghdi07pc_inv waunemp08 if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  8,   109) = 6026.20
       Model |  50299520.8     8   6287440.1           Prob > F      =  0.0000
    Residual |  113725.273   109  1043.35113           R-squared     =  0.9977
-------------+------------------------------           Adj R-squared =  0.9976
       Total |  50413246.1   117  430882.445           Root MSE      =  32.301

------------------------------------------------------------------------------
        expm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pop01_u5 |    .013794   .0030889     4.47   0.000     .0076719     .019916
  pop01_5_19 |  -.0003571     .00145    -0.25   0.806    -.0032309    .0025167
    pop01_65 |   .0027767   .0003964     7.01   0.000     .0019911    .0035622
      smortr |  -36.77723   45.31633    -0.81   0.419    -126.5927    53.03826
   wa_llti01 |   .0105636   .0009792    10.79   0.000     .0086229    .0125043
        chl1 |   .0118235   .0010252    11.53   0.000     .0097916    .0138555
ghdi07pc_inv |   -1656565   406855.4    -4.07   0.000     -2462939   -850190.9
   waunemp08 |   .0074218   .0014821     5.01   0.000     .0044843    .0103592
       _cons |   160.6897   39.34749     4.08   0.000     82.70421    238.6751
------------------------------------------------------------------------------
```

Figure 20: Model 1b

```
. * SCBF - with additional vars
.            reg expm pop01_u5 pop01_5_19 pop01_65_74 pop01_75 smortr wa_llti01
llti_tot01 chl1 waben08 ghdi07pc_inv waunemp08 wanemp08 if sam
> ple==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F( 12,   105) = 4124.91
       Model |  50306532.9    12  4192211.08           Prob > F      =  0.0000
    Residual |  106713.176   105  1016.31596           R-squared     =  0.9979
-------------+------------------------------           Adj R-squared =  0.9976
       Total |  50413246.1   117  430882.445           Root MSE      =   31.88

------------------------------------------------------------------------------
        expm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    pop01_u5 |   .0159483    .003243     4.92   0.000     .0095181    .0223785
  pop01_5_19 |   -.002122   .0016027    -1.32   0.188    -.0052998    .0010558
 pop01_65_74 |   .0086935   .0023654     3.68   0.000     .0040033    .0133837
    pop01_75 |  -.0009025   .0021535    -0.42   0.676    -.0051725    .0033674
      smortr |  -51.43211   46.53764    -1.11   0.272    -143.7076    40.84343
   wa_llti01 |   .0109455   .0061152     1.79   0.076    -.0011799    .0230709
  llti_tot01 |  -.0008382   .0037119    -0.23   0.822    -.0081983    .0065218
        chl1 |   .0131977   .0021368     6.18   0.000     .0089609    .0174345
     waben08 |  -.0002657   .0022484    -0.12   0.906    -.0047238    .0041924
ghdi07pc_inv |   -1450310   465569.5    -3.12   0.002     -2373448   -527171.3
   waunemp08 |   .0075325   .0017649     4.27   0.000     .0040329     .011032
    wanemp08 |  -.0001214   .0005613    -0.22   0.829    -.0012343    .0009915
       _cons |   160.1943   41.54594     3.86   0.000      77.8164    242.5722
------------------------------------------------------------------------------
```

Figure 21: Model 2

```
. * ICFFW -
.          reg exp_ph dep_ratio07 nwhite08_pct spars01 gva06pci_inv wanq08_pct
waemp08_pct llti01_pct wallti01_pct wa_adlb2_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  9,   108) =  165.97
       Model |  21724717.7     9  2413857.52           Prob > F      =  0.0000
    Residual |  1570757.54   108  14544.0513           R-squared     =  0.9326
-------------+------------------------------           Adj R-squared =  0.9270
       Total |  23295475.3   117  199106.626           Root MSE      =   120.6

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 dep_ratio07 |  -13.14503   5.753458    -2.28   0.024    -24.54938   -1.740673
nwhite08_pct |   13.05561   1.528513     8.54   0.000     10.02583    16.08539
     spars01 |   3.108552   .9506943     3.27   0.001     1.224111    4.992993
gva06pci_inv |  -.0003628   .0058274    -0.06   0.950    -.0119137     .011188
  wanq08_pct |  -3.401511   4.624209    -0.74   0.464     -12.5675    5.764474
 waemp08_pct |   7.037158   4.910536     1.43   0.155    -2.696376    16.77069
  llti01_pct |  -24.73368   20.69469    -1.20   0.235    -65.75414    16.28679
wallti01_pct |   51.66653   25.61502     2.02   0.046       .89311    102.4399
wa_adlb2_pct |    82.6689   10.27383     8.05   0.000     62.30438    103.0334
       _cons |   1164.373   454.2205     2.56   0.012     264.0289    2064.716
------------------------------------------------------------------------------
```

Figure 22: Model 3

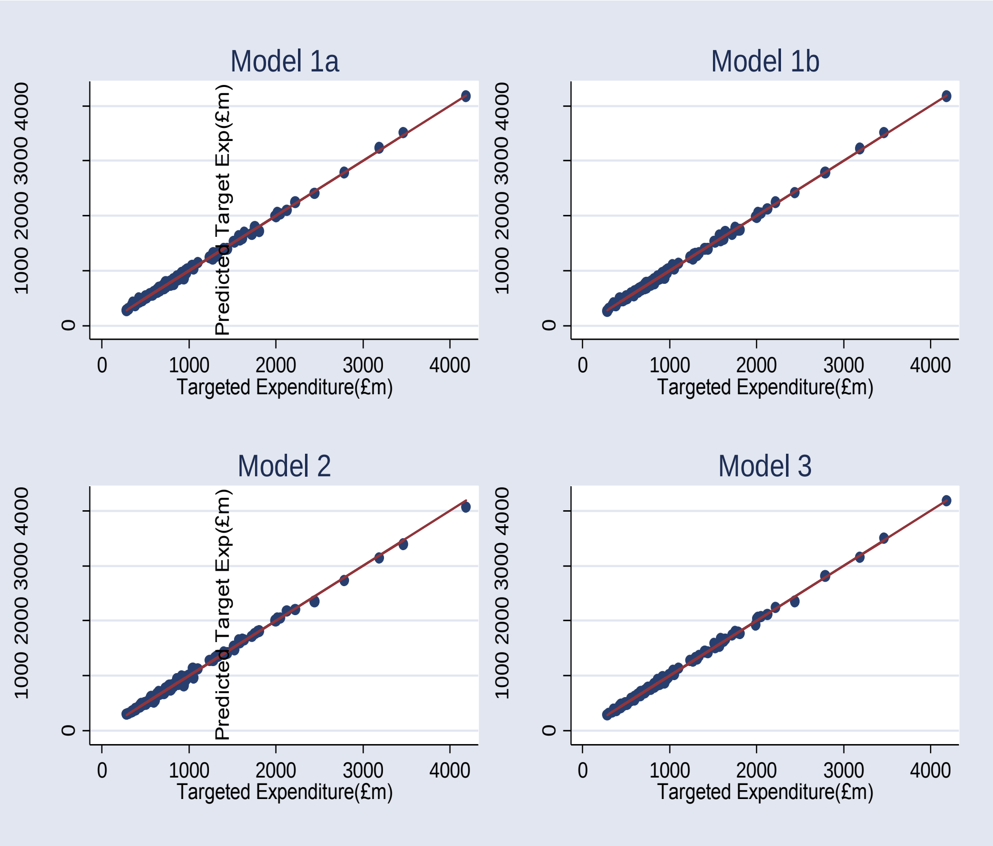
```
. * Model 3
.                  reg  exp_ph  chl1_pct  wallti01_pct  nwhite08_pct  dep_ratio07
ghdi07pc_inv waunemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  6,   111) =  397.93
       Model |  22260572.9     6  3710095.49           Prob > F      =  0.0000
    Residual |  1034902.34   111  9323.44452           R-squared     =  0.9556
-------------+------------------------------           Adj R-squared =  0.9532
       Total |  23295475.3   117  199106.626           Root MSE      =  96.558

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   125.3814   9.734342    12.88   0.000     106.0921    144.6706
wallti01_pct |   64.15157   6.945936     9.24   0.000     50.38773    77.91541
nwhite08_pct |   5.972721   1.193176     5.01   0.000     3.608364    8.337078
 dep_ratio07 |   13.36762   3.788155     3.53   0.001     5.861139     20.8741
ghdi07pc_inv |   -3569756    1195818    -2.99   0.003     -5939351    -1200162
waunemp08_~t |   15.93702   9.504552     1.68   0.096     -2.89688    34.77093
       _cons |   961.9727   130.7841     7.36   0.000     702.8153     1221.13
------------------------------------------------------------------------------
```

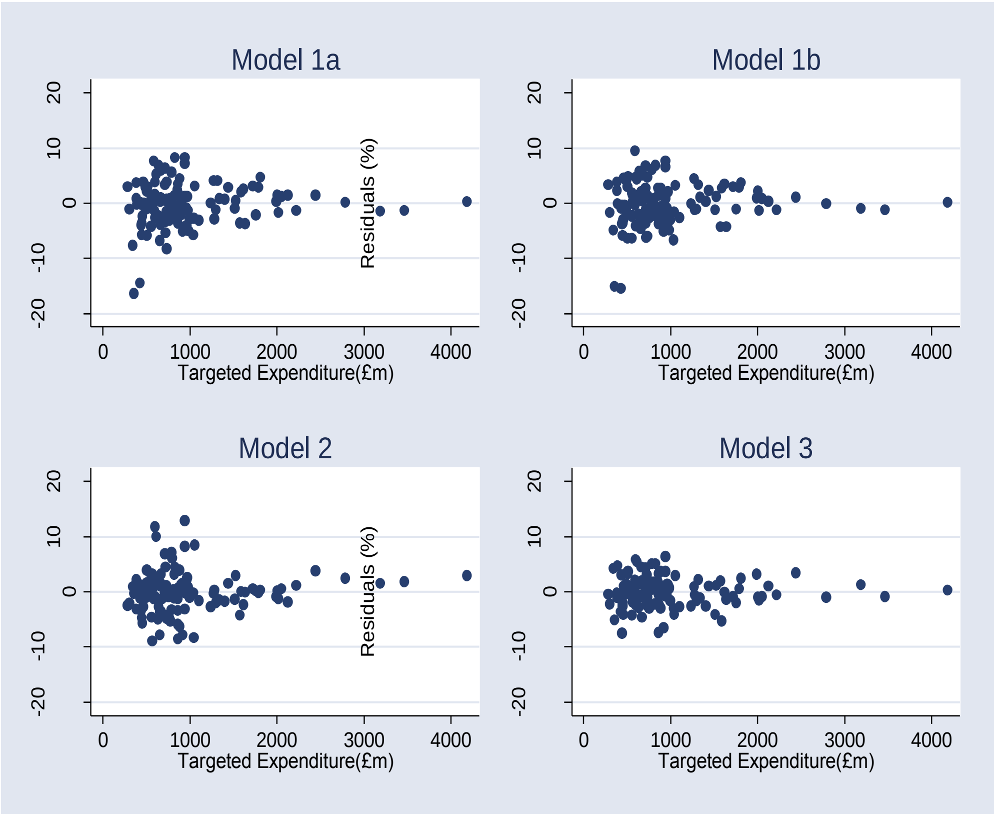
A3.7 Appendix: Analysis of residuals

Figure 23: Actual and predicted values



Source: LE Wales.

Figure 24: Plot of residuals



Source: LE Wales.

A3.8 Appendix: Descriptive analysis

Figure 25: Analysis of correlations (different variables)

```
             |   exp_ph gh~c_inv wallti~t dep_r~07 nw~8_pct p~65_pct po~9_pct waemp0~t waunem~t chl1_pct
-------------+------------------------------------------------------------------------------------------
      exp_ph |   1.0000
ghdi07pc_inv |   0.3575   1.0000
wallti01_pct |   0.7116   0.7166   1.0000
 dep_ratio07 |  -0.2856   0.3922   0.2087   1.0000
nwhite08_pct |   0.4515  -0.3656  -0.2030  -0.6238   1.0000
pop01_65_pct |  -0.3613   0.3304   0.1988   0.8087  -0.7184   1.0000
pop01_5_19~t |   0.3405   0.5764   0.4138   0.4384  -0.0382  -0.0323   1.0000
 waemp08_pct |  -0.8282  -0.2914  -0.5834   0.3902  -0.5026   0.3517  -0.1598   1.0000
waunemp08_~t |   0.7086   0.3928   0.5536  -0.1630   0.3269  -0.2171   0.2751  -0.7687   1.0000
    chl1_pct |   0.9354   0.2545   0.5495  -0.4902   0.5578  -0.5223   0.2685  -0.8627   0.6541   1.0000
```

Figure 26: Univariate analysis (different variables) (1)

```
.         reg exp_ph ghdi07pc_inv if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =   17.00
       Model |  2976944.8     1   2976944.8            Prob > F      =  0.0001
    Residual |  20318530.5   116  175159.745           R-squared     =  0.1278
-------------+------------------------------           Adj R-squared =  0.1203
       Total |  23295475.3   117  199106.626           Root MSE      =  418.52

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ghdi07pc_inv |   1.33e+07    3216586     4.12   0.000      6889754    1.96e+07
       _cons |   1853.784   233.4923     7.94   0.000     1391.323    2316.245
------------------------------------------------------------------------------

.         reg exp_ph wallti01_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  119.01
       Model |  11796944.2     1  11796944.2           Prob > F      =  0.0000
    Residual |  11498531.1   116  99125.2681           R-squared     =  0.5064
-------------+------------------------------           Adj R-squared =  0.5021
       Total |  23295475.3   117  199106.626           Root MSE      =  314.84

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
wallti01_pct |     110.76   10.15291    10.91   0.000     90.65092    130.8691
       _cons |   1391.789   132.5829    10.50   0.000     1129.192    1654.386
------------------------------------------------------------------------------
```

Figure 27: Univariate analysis (different variables) (2)

```
.         reg exp_ph nwhite08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =   29.70
       Model |  4748862.97     1  4748862.97           Prob > F      =  0.0000
    Residual |  18546612.3   116  159884.589           R-squared     =  0.2039
-------------+------------------------------           Adj R-squared =  0.1970
       Total |  23295475.3   117  199106.626           Root MSE      =  399.86

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
nwhite08_pct |     14.074   2.582416     5.45   0.000     8.959204     19.1888
       _cons |   2615.558   50.39951    51.90   0.000     2515.735     2715.38
------------------------------------------------------------------------------

.         reg exp_ph dep_ratio07 if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =   10.31
       Model |  1900654.56     1  1900654.56           Prob > F      =  0.0017
    Residual |  21394820.7   116  184438.109           R-squared     =  0.0816
-------------+------------------------------           Adj R-squared =  0.0737
       Total |  23295475.3   117  199106.626           Root MSE      =  429.46

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 dep_ratio07 |  -34.13432   10.63322    -3.21   0.002    -55.19476   -13.07388
       _cons |   4070.102   396.6364    10.26   0.000     3284.514     4855.69
------------------------------------------------------------------------------

.         reg exp_ph pop01_65_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =   17.42
       Model |  3041083.79     1  3041083.79           Prob > F      =  0.0001
    Residual |  20254391.5   116  174606.823           R-squared     =  0.1305
-------------+------------------------------           Adj R-squared =  0.1230
       Total |  23295475.3   117  199106.626           Root MSE      =  417.86

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pop01_65_pct |  -58.37717   13.98812    -4.17   0.000    -86.08242   -30.67193
       _cons |   3698.435   217.9398    16.97   0.000     3266.778    4130.092
------------------------------------------------------------------------------

.         reg exp_ph pop01_5_19_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =   15.22
       Model |  2701537.13     1  2701537.13           Prob > F      =  0.0002
    Residual |  20593938.1   116  177533.949           R-squared     =  0.1160
-------------+------------------------------           Adj R-squared =  0.1083
       Total |  23295475.3   117  199106.626           Root MSE      =  421.35

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pop01_5_19~t |   109.8101   28.14994     3.90   0.000     54.05558    165.5646
       _cons |   1240.121   402.5645     3.08   0.003     442.7914    2037.451
------------------------------------------------------------------------------
```

Figure 28: Univariate analysis (different variables) (3)

```
.         reg exp_ph waemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  253.36
       Model |  15979358.4     1  15979358.4           Prob > F      =  0.0000
    Residual |  7316116.86   116   63069.973           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6832
       Total |  23295475.3   117  199106.626           Root MSE      =  251.14

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 waemp08_pct |  -66.68845   4.189694   -15.92   0.000    -74.98667   -58.39023
       _cons |   7679.516   307.2264    25.00   0.000     7071.015    8288.017
------------------------------------------------------------------------------

.         reg exp_ph waunemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  117.00
       Model |  11697815.7     1  11697815.7           Prob > F      =  0.0000
    Residual |  11597659.6   116  99979.8242           R-squared     =  0.5021
-------------+------------------------------           Adj R-squared =  0.4979
       Total |  23295475.3   117  199106.626           Root MSE      =   316.2

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
waunemp08_~t |   229.2657   21.19546    10.82   0.000     187.2854     271.246
       _cons |    1663.38   109.3201    15.22   0.000     1446.857    1879.902
------------------------------------------------------------------------------

.         reg exp_ph chl1_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  811.68
       Model |  20382531.6     1  20382531.6           Prob > F      =  0.0000
    Residual |  2912943.66   116  25111.5832           R-squared     =  0.8750
-------------+------------------------------           Adj R-squared =  0.8739
       Total |  23295475.3   117  199106.626           Root MSE      =  158.47

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   180.6696   6.341516    28.49   0.000     168.1094    193.2298
       _cons |   1903.572   34.78317    54.73   0.000     1834.679    1972.464
------------------------------------------------------------------------------

.         reg lexp_ph lchl1_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  868.63
       Model |  2.50896859     1  2.50896859           Prob > F      =  0.0000
    Residual |  .335055786   116  .002888412           R-squared     =  0.8822
-------------+------------------------------           Adj R-squared =  0.8812
       Total |  2.84402438   117  .024307901           Root MSE      =  .05374
```

A3.9 Appendix: Alternative Model 3 specifications

Figure 29: Alternative specifications to Model 3 (1)

```
.         reg exp_ph chl1_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  1,   116) =  811.68
       Model |  20382531.6     1  20382531.6           Prob > F      =  0.0000
    Residual |  2912943.66   116  25111.5832           R-squared     =  0.8750
-------------+------------------------------           Adj R-squared =  0.8739
       Total |  23295475.3   117  199106.626           Root MSE      =  158.47

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   180.6696   6.341516    28.49   0.000     168.1094    193.2298
       _cons |   1903.572   34.78317    54.73   0.000     1834.679    1972.464
------------------------------------------------------------------------------

.         reg exp_ph chl1_pct wallti01_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  2,   115) =  774.77
       Model |  21686024.7     2  10843012.4           Prob > F      =  0.0000
    Residual |  1609450.52   115  13995.2219           R-squared     =  0.9309
-------------+------------------------------           Adj R-squared =  0.9297
       Total |  23295475.3   117  199106.626           Root MSE      =   118.3

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   150.6213   5.666283    26.58   0.000     139.3975    161.8451
wallti01_pct |   44.06614   4.566049     9.65   0.000     35.02168    53.11061
       _cons |   1491.667   49.95937    29.86   0.000     1392.707    1590.627
------------------------------------------------------------------------------

.         reg exp_ph chl1_pct wallti01_pct nwhite08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  3,   114) =  695.54
       Model |  22088686.6     3  7362895.52           Prob > F      =  0.0000
    Residual |  1206788.71   114  10585.8659           R-squared     =  0.9482
-------------+------------------------------           Adj R-squared =  0.9468
       Total |  23295475.3   117  199106.626           Root MSE      =  102.89

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |     107.38   8.569838    12.53   0.000     90.40317    124.3567
wallti01_pct |   70.59096   5.853751    12.06   0.000     58.99472    82.18719
nwhite08_pct |   7.278356    1.18012     6.17   0.000     4.940547    9.616166
       _cons |   1271.951   56.18758    22.64   0.000     1160.644    1383.258
------------------------------------------------------------------------------
```

Figure 30: Alternative specifications to Model 3 (2)

```
.          reg  exp_ph  chl1_pct  wallti01_pct  nwhite08_pct  dep_ratio07  if
sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  4,   113) =  554.44
       Model |  22166060.1     4  5541515.03           Prob > F      =  0.0000
    Residual |  1129415.16   113  9994.82441           R-squared     =  0.9515
-------------+------------------------------           Adj R-squared =  0.9498
       Total |  23295475.3   117  199106.626           Root MSE      =  99.974

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   121.9765   9.841939    12.39   0.000     102.4778    141.4751
wallti01_pct |   61.29298   6.597028     9.29   0.000     48.22308    74.36289
nwhite08_pct |   7.290896   1.146711     6.36   0.000     5.019054    9.562737
 dep_ratio07 |    10.4751   3.764867     2.78   0.006     3.016223    17.93398
       _cons |   928.7929   134.8786     6.89   0.000     661.5741    1196.012
------------------------------------------------------------------------------

.            reg  exp_ph  chl1_pct  wallti01_pct  nwhite08_pct  dep_ratio07
ghdi07pc_inv if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  5,   112) =  469.36
       Model |  22234359.3     5  4446871.85           Prob > F      =  0.0000
    Residual |  1061115.99   112  9474.24995           R-squared     =  0.9544
-------------+------------------------------           Adj R-squared =  0.9524
       Total |  23295475.3   117  199106.626           Root MSE      =  97.336

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   127.0093   9.763824    13.01   0.000     107.6635    146.3551
wallti01_pct |   67.08734   6.775794     9.90   0.000     53.66197     80.5127
nwhite08_pct |   6.566576   1.148579     5.72   0.000     4.290813    8.842339
 dep_ratio07 |   13.34955   3.818653     3.50   0.001     5.783379    20.91572
ghdi07pc_inv |   -3172369    1181540    -2.68   0.008     -5513439   -831298.9
       _cons |   959.9913   131.8322     7.28   0.000     698.7827      1221.2
------------------------------------------------------------------------------

.            reg  exp_ph  chl1_pct  wallti01_pct  nwhite08_pct  dep_ratio07
ghdi07pc_inv waunemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  6,   111) =  397.93
       Model |  22260572.9     6  3710095.49           Prob > F      =  0.0000
    Residual |  1034902.34   111  9323.44452           R-squared     =  0.9556
-------------+------------------------------           Adj R-squared =  0.9532
       Total |  23295475.3   117  199106.626           Root MSE      =  96.558

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   125.3814   9.734342    12.88   0.000     106.0921    144.6706
wallti01_pct |   64.15157   6.945936     9.24   0.000     50.38773    77.91541
nwhite08_pct |   5.972721   1.193176     5.01   0.000     3.608364    8.337078
 dep_ratio07 |   13.36762   3.788155     3.53   0.001     5.861139     20.8741
ghdi07pc_inv |   -3569756    1195818    -2.99   0.003     -5939351    -1200162
waunemp08_~t |   15.93702   9.504552     1.68   0.096     -2.89688    34.77093
       _cons |   961.9727   130.7841     7.36   0.000     702.8153     1221.13
------------------------------------------------------------------------------
```

A3.10 Appendix: Analysis of the effects of aggregation

Figure 31: Original model

```
. * Regs
.         * Original model
.                 reg  exp_ph  chl1_pct  wallti01_pct  nwhite08_pct  dep_ratio07
ghdi07pc_inv waunemp08_pct if sample==1

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  6,   111) =  397.93
       Model |  22260572.9     6  3710095.49           Prob > F      =  0.0000
    Residual |  1034902.34   111  9323.44452           R-squared     =  0.9556
-------------+------------------------------           Adj R-squared =  0.9532
       Total |  23295475.3   117  199106.626           Root MSE      =  96.558

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   125.3814   9.734342    12.88   0.000     106.0921    144.6706
wallti01_pct |   64.15157   6.945936     9.24   0.000     50.38773    77.91541
nwhite08_pct |   5.972721   1.193176     5.01   0.000     3.608364    8.337078
 dep_ratio07 |   13.36762   3.788155     3.53   0.001     5.861139     20.8741
ghdi07pc_inv |   -3569756    1195818    -2.99   0.003     -5939351    -1200162
waunemp08_~t |   15.93702   9.504552     1.68   0.096     -2.89688    34.77093
       _cons |   961.9727   130.7841     7.36   0.000     702.8153     1221.13
------------------------------------------------------------------------------
```

Figure 32: Separate models (1)

```
.         * Separate Models (SW)
.            sw reg exp_ph ghdi07pc_inv wallti01_pct dep_ratio07 nwhite08_pct
waunemp08_pct chl1_pct if sample==1, pr(.1)
                      begin with full model
p < 0.1000            for all terms in model

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  6,   111) =  397.93
       Model |  22260572.9     6  3710095.49           Prob > F      =  0.0000
    Residual |  1034902.34   111  9323.44452           R-squared     =  0.9556
-------------+------------------------------           Adj R-squared =  0.9532
       Total |  23295475.3   117  199106.626           Root MSE      =  96.558

------------------------------------------------------------------------------
      exp_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ghdi07pc_inv |   -3569756    1195818    -2.99   0.003     -5939351    -1200162
wallti01_pct |   64.15157   6.945936     9.24   0.000     50.38773    77.91541
 dep_ratio07 |   13.36762   3.788155     3.53   0.001     5.861139     20.8741
nwhite08_pct |   5.972721   1.193176     5.01   0.000     3.608364    8.337078
waunemp08_~t |   15.93702   9.504552     1.68   0.096     -2.89688    34.77093
    chl1_pct |   125.3814   9.734342    12.88   0.000     106.0921    144.6706
       _cons |   961.9727   130.7841     7.36   0.000     702.8153     1221.13
------------------------------------------------------------------------------

.         predict exphat
(option xb assumed; fitted values)

.            sw reg pct_ph ghdi07pc_inv wallti01_pct dep_ratio07 nwhite08_pct
waunemp08_pct chl1_pct if sample==1, pr(.1)
                      begin with full model
p = 0.6824 >= 0.1000  removing nwhite08_pct
p = 0.3621 >= 0.1000  removing waunemp08_pct

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  4,   113) =  489.86
       Model |  4138447.03     4  1034611.76           Prob > F      =  0.0000
    Residual |  238663.447   113   2112.0659           R-squared     =  0.9455
-------------+------------------------------           Adj R-squared =  0.9435
       Total |  4377110.47   117  37411.2006           Root MSE      =  45.957

------------------------------------------------------------------------------
      pct_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ghdi07pc_inv |   -2658311     542260    -4.90   0.000     -3732625    -1583996
wallti01_pct |   47.40921   2.757482    17.19   0.000     41.94614    52.87228
 dep_ratio07 |   4.196378   1.799493     2.33   0.021     .6312567    7.761499
    chl1_pct |     42.725   3.276698    13.04   0.000     36.23327    49.21673
       _cons |   870.1121   59.88604    14.53   0.000     751.4671    988.7572
------------------------------------------------------------------------------

.         predict pcthat if sample==1
(option xb assumed; fitted values)
(19 missing values generated)
```

Figure 33: Separate models (2)

```
.            sw  reg  lea_ph  ghdi07pc_inv  wallti01_pct  dep_ratio07  nwhite08_pct
waunemp08_pct chl1_pct if sample==1, pr(.1)
                      begin with full model
p = 0.6900 >= 0.1000  removing waunemp08_pct
p = 0.2640 >= 0.1000  removing ghdi07pc_inv

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  4,   113) =   73.77
       Model |  863438.668     4  215859.667           Prob > F      =  0.0000
    Residual |  330638.112   113  2926.00099           R-squared     =  0.7231
-------------+------------------------------           Adj R-squared =  0.7133
       Total |  1194076.78   117  10205.7844           Root MSE      =  54.093

------------------------------------------------------------------------------
      lea_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    chl1_pct |   32.45287   5.325131     6.09   0.000     21.90282    43.00291
wallti01_pct |  -6.804313   3.569423    -1.91   0.059    -13.87598     .267358
 dep_ratio07 |   16.23023   2.037039     7.97   0.000     12.19449    20.26597
nwhite08_pct |   4.037023   .6204456     6.51   0.000     2.807808    5.266237
       _cons |  -112.2042   72.97812    -1.54   0.127     -256.787     32.3786
------------------------------------------------------------------------------

.         predict leahat if sample==1
(option xb assumed; fitted values)
(19 missing values generated)

.             sw  reg  lg_ph  ghdi07pc_inv  wallti01_pct  dep_ratio07  nwhite08_pct
waunemp08_pct chl1_pct if sample==1, pr(.1)
                      begin with full model
p < 0.1000            for all terms in model

      Source |       SS       df       MS              Number of obs =     118
-------------+------------------------------           F(  6,   111) =  178.92
       Model |  4527430.51     6  754571.751           Prob > F      =  0.0000
    Residual |  468137.833   111  4217.45795           R-squared     =  0.9063
-------------+------------------------------           Adj R-squared =  0.9012
       Total |  4995568.34   117  42697.1653           Root MSE      =  64.942

------------------------------------------------------------------------------
       lg_ph |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ghdi07pc_inv |   -1694134   804271.1    -2.11   0.037     -3287851   -100417.3
wallti01_pct |    24.0671   4.671625     5.15   0.000     14.80997    33.32424
 dep_ratio07 |  -6.307861   2.547797    -2.48   0.015    -11.35649    -1.25923
nwhite08_pct |   1.461985   .8024935     1.82   0.071    -.1282098    3.052179
waunemp08_~t |   14.59272   6.392472     2.28   0.024     1.925608    27.25983
    chl1_pct |    53.6055   6.547022     8.19   0.000     40.63214    66.57886
       _cons |   221.6583   87.96139     2.52   0.013      47.3569    395.9596
------------------------------------------------------------------------------

.         predict lghat if sample==1
(option xb assumed; fitted values)
(19 missing values generated)

.         g exp2hat=pcthat+leahat+lghat if sample==1
(19 missing values generated)

.         summ exp_ph if sample==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      exp_ph |       118    2803.178    446.2137       1997       4303
```

Figure 34: Analysis goodness of fit (constructed R^2)

```
. * LER
.         * Aggregated
.                 g stot0=(exp_ph-2803.178)^2 if sample==1
(19 missing values generated)

.                 g sres0=(exp_ph-exphat)^2 if sample==1
(19 missing values generated)

.                 egen ssres0=total(sres0)

.                 egen sstot0=total(stot0)

.                 g ler0=1-ssres0/sstot0

.
.         * Disaggregated
.                 g stot=(exp_ph-2803.178)^2 if sample==1
(19 missing values generated)

.                 g sres=(exp_ph-exp2hat)^2 if sample==1
(19 missing values generated)

.                 egen ssres=total(sres)

.                 egen sstot=total(stot)

.                 g ler=1-ssres/sstot

.
. summ ler0 ler*

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        ler0 |       137     .955575           0     .955575    .955575
         ler |       137    .9553357           0    .9553357   .9553357
```

LE Wales

Sophia House,
28 Cathedral Road
Cardiff CF11 9LJ
Tel: 02920 660 250
Fax: 02920 660 251
Email: info@le-wales.co.uk