Project

Evidencing Adaptive Sustainability
(EASY)

Evidencing Population Change in the London Thames Gateway

By

Centre for Geo-Information Studies, UEL
Terra Cognita

Evidencing Population Change in the London Thames Gateway

A.J. Brimicombe[1], Y. Li[1] and C.Li[2]

[1]Centre for Geo-Information Studies, University of East London
{a.j.brimicombe y.li} @uel.ac.uk

[2]Terra Cognita
Chao.Li@terra-cognita.com

This working paper provides an insight into the population models for Lower Super Output Area (LSOA) level estimates developed for 10 Boroughs within the London Thames Gateway based on Council Tax, Child Benefit and Schools Census data. The multi-stage multiple regression models are initially constructed using 2001 data and tested against official statistics. The estimates are then moved forward with successive annual data sets to provide an understanding of year-on-year population change. This approach is not intended to displace official statistics but to provide another view through a different route; they can be set alongside each other for evidential decision-support in social infrastructure planning.

1. INTRODUCTION

Official statistics are used extensively for decision-making in both public and private sectors, and can therefore have considerable impact on the lives of citizens[1]. At the core of official statistics is the count of population and monitoring of population change. “Population estimates are central to every national system of official statistics. They are used in statistical formulae that allocate vast sums of public money to the devolved administrations, to local government, the health service and public services. It is therefore a matter of social responsibility to ensure that population statistics are calculated accurately”[2].

For communities to be sustainable, that is desirable places to live in and thrive, they should have a range of services that are accessible and appropriate to residents’ needs[3,4]. However, sustainability is not a static attribute of place. Where the needs of residents rapidly change due to (im)migration, social and economic mobility and transience, there needs to be robust mechanisms for compiling and updating the evidence base on which policy and planning changes must necessarily be founded. Whilst centrally compiled official statistics affect allocations, more granular, locally derived statistics not only inform debates over allocations, but importantly inform local decisions on the nature and distribution of services (always based on finite resources) to equitably meet the needs of residents. Fundamental then are up-to-the-moment population estimates for small area geographies.

Since July 2008, Local Development Frameworks (LDF) coming forward for approval have had to contain a detailed social infrastructure plan – phasing infrastructure growth to match housing and hence population growth or vice versa. The starting point for any such plans is the size, demographic and distribution of current population, the trend in population since 2001 and the projection of population across the planning horizon. Notwithstanding the arguments around the accuracy of the 2001 population figures, the 2001 Census has probably become out-of-date in respect of current population faster than any previous census. But there is also a high level of scepticism in the rolling mid-year estimates of population and any projections based upon them. This then is the backdrop against which new models for evidencing population change have been developed, as presented in this paper. In the next section we summarise current approaches to evidencing population change and develop the argument in support of these new models to provide up-to-the-moment estimates that complement official statistics. We then present the modelling results for 10 Boroughs in the London Thames Gateway. In drawing our conclusions we examine the implications of this new approach and further improvements under development.

2. EVIDENCING POPULATION CHANGE

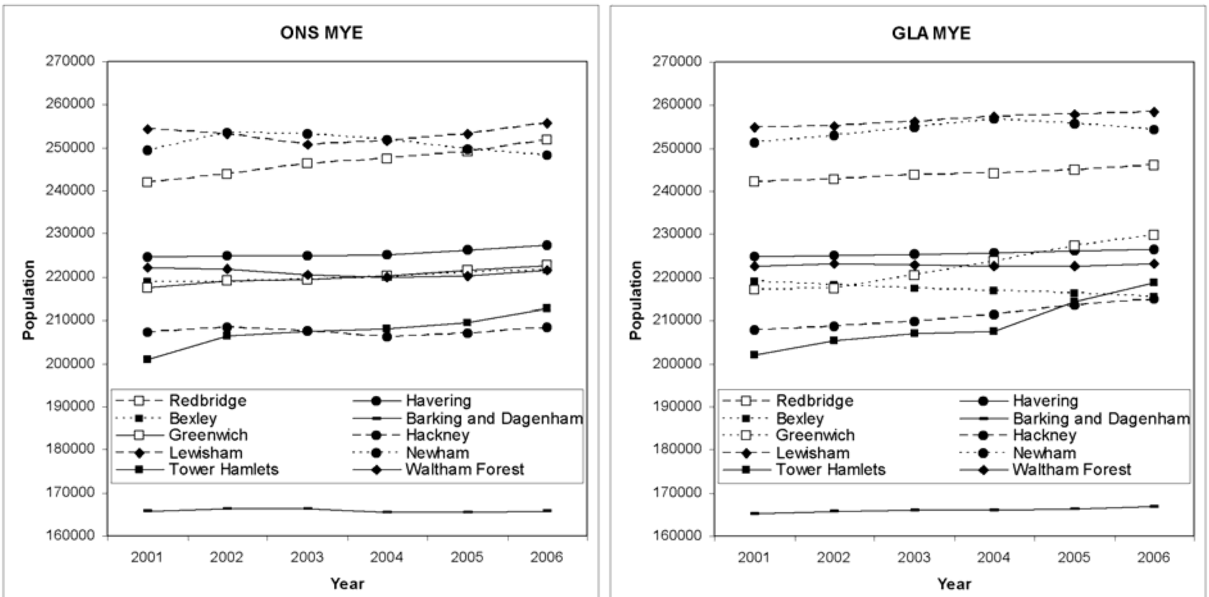
The evidence for population change can come from three broad sources: census, sample survey and official administrative records. Censuses are hugely expensive and logistically complex but provide much more information than just population counts, extending to living conditions, health and travel. Sample surveys are central to social statistics with well-developed theory and associated methods. The cost of such surveys, however, is increasing rapidly whilst response rates are generally declining[6] as the public becomes over-surveyed through telemarketing and other campaigns as well a growing wariness by the public to divulge personal data. Official administrative records are those which are collected under statute such as the registration of births and deaths. These record micro-level events for individuals but have the potential to be aggregated to track changes in population. However, the UK does not maintain sufficient statutory data to fully enumerate population changes on an annual basis.

2.1 Official Population Statistics

Currently in the UK there is a dual approach for ONS population statistics: the decennial census and mid-year estimates. Whilst from a logical perspective the mid-year estimates provide annual estimates with which to track demographic change between censuses, the relationship between the two is not straightforward. The UK census aims to enumerate population on the basis of place of residence rather than where they actually are on the night of the census. Whilst the census results underpin the mid-year estimates, the latter are designed to take more account of 'hard to count' populations that the census might otherwise have missed. Mid-year estimates are intended to reflect *ordinarily resident population* in an area. Thus the LA level mid-year estimates take the 1981 Census as their base and are, broadly speaking, calculated for each subsequent year using the cohort component method in which each age cohort within the population of an area is progressively aged, with births and deaths based on age-specific rates derived from the civil registrations. International migration is estimated from the International Passenger Survey (IPS) and internal migration from the NHS Central Register which records changes in GP registrations.

In 2005 the ONS started introducing mid-year estimates for small area geographies going back to 2001, first for Wards and then for LSOA. Small area estimates are calculated for each LA and are constrained by the totals already calculated for the LA so that they always sum to the LA estimates. The method used to calculate LSOA populations relies on evidencing the ratio of change from the previous year to the current one being estimated using administrative data. Thus, if from GP patient registers there is a ratio of 1.1 in 15-19 year old boys in 2006 compared to 2005 for an LSOA, then the 2005 mid-year estimate of that group for that LSOA is augmented by 10% to give the 2006 estimate.

In London, the Greater London Authority (GLA) also produces LA and Ward mid-year estimates based on projections which are updated periodically. The GLA estimates are broadly based on the ONS estimates in that the population for London is capped to the ONS estimate and that, in the latest round, change components are derived from ONS 2001-2006 mid-year estimates. However, the GLA also includes data from the 2004 London Housing Capacity Survey and from data compiled by the Department for Communities and Local Government (DCLG) on housing projections. In essence then, the GLA population estimates are a redistribution of ONS estimates based on housing projections. The differences that arise for 2001-2007 comparing ONS and GLA estimates for 10 Boroughs in the London Thames Gateway are illustrated in Figure 1.



(a) (b)

Fig. 1. Comparison of population estimates for 2001-2007 for 10 Boroughs in the London Thames Gateway; (a) ONS mid-year estimates; (b) GLA mid-year estimates.

There is an understandable lag in generating official statistics since they are usually released simultaneously for all areas of the country and the base data must be collated, quality checked and then processed. Thus at any time the ONS LA estimates are 12 to 24 months in arrears and LSOA estimates 24 to 36 months in arrears.

The main criticism levelled at the ONS mid-year estimates is the inability of the current methodology to adequately take account of turnover through (im)migration. “The amount of population turnover, both nationally and locally has made it increasingly difficult for the current methods of counting population to estimate the numbers of people in an area and on what basis they are there”[2]. The 2001 census showed that in London 13.7% of those enumerated had changed house in the previous 12 months with the national figure for churn at 11.6% [5]. Higher transience within city populations is going to make them harder to count let alone monitor in a timely manner.

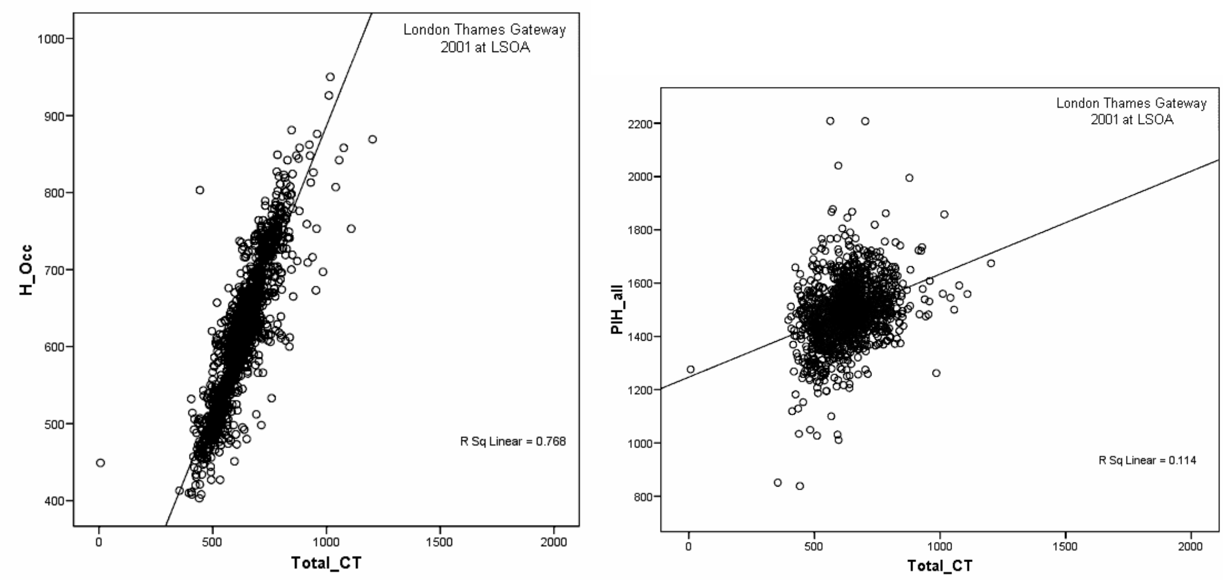
2.2 Administrative Records as Latent Variables

Judson[6], in a recent critique of social statistics in general, states that “A ‘data problem’ of the 21st century can be stated in simple terms: the official statistical systems .... deliver data not sufficiently *fast*, not sufficiently *local*, not sufficiently *granular* and not sufficiently *integrated*” [original emphasis]. He proposes an administrative records (AR) method of generating statistics which can also be an inexpensive, more up-to-the-moment solution given the rising costs of surveys and falling response rates. The AR method relies on data integration, that is, “using multiple data sets *in concert* to construct statistical estimands for answering questions about those estimands”[6]. These data sets have been collected for other purposes and were not intended as measures even estimates of the target population. The AR method particularly relies on the identification of suitable latent variables, that is, variables which may not directly measure the target population(s) but can reflect them in some understandable and consistent way. Use of small area geographies as a database key is a particularly effective way of integrating disparate data sets. The AR method offers opportunities but is not without problems. One is that the multiple data sets may not be harmonised in terms of dates with the potential for subtle inconsistencies. Another is in assessing the level of uncertainty in any estimands because as well as each data set having its own level of uncertainty there may be amplification or even dampening of some uncertainties as a result of data integration. This makes it harder to determine confidence intervals.

2.3 Building Alternative Population Models

Our objective in developing new population models was not to compete with official statistics but to provide an alternative view, derived in a different way, that might corroborate, or otherwise, population estimates for small area geographies (LSOA) as well as providing more timely (up-to-the-moment) estimates. It was clear that within the time and resources available an AR approach using existing data sets would need to be adopted. Therefore, rather than attempting an accounting approach of enumerating population from AR, we statistically modelled populations from latent variables. This is an important distinction. Nevertheless, an AR approach was going to be speculative given the existing level of uncertainty in population estimates and the wide range of AR that can be accessed. Modelling population change can be an uncertain endeavour as is not clear at the outset when a method will perform well and there are dangers of model misspecification[7]. Thus trying to keep the models simple is a justifiable aim particularly as in the history of population modelling "more complex models have been no more accurate than simpler models"[8].

We started with the fundamental assumption that resident populations live in dwellings and that the dwelling stock is the key starting point. The official measure of the national dwelling stock from the Land Registry is compiled from Council Tax data. Council Tax lists are kept up-to-date by LAs because they are an important means of collecting revenue. However, whilst dwelling stock can be a reasonable predictor of households, they are a poor predictor of population. This is illustrated in Figure 2. In Figure 2(a) 2001 data on total count of Council Tax dwellings is plotted against households occupying dwellings by LSOA in the Thames Gateway (n=1448) with R^2=0.768; a good fit. In Figure 2(b) is a similar plot of Council Tax dwellings against population in households with R^2=0.114; a very poor fit. Thus, if households are modelled from counts of dwellings, can population be modelled from households? Our study identified Child Benefit data and Pupil Level Annual School Census (PLASC) data as being two data sets that are kept up-to-date and were available by small area geography. Uptake of Child Benefit is considered to be almost universal[6] though there could be an important lag in registering a family's change of address. PLASC data are collected twice a year as part of official returns to the Department for Children, Schools and Families (DCSF) and are available aggregated by pupil home address rather than school attended. Importantly, all three data sets just discussed are available annually going back to 2001 and thus models can be built that take otherwise unknowns from the 2001 Census and then brought forward year-on-year to identify trends.



(a) (b)

Fig. 2. Modelling Council Tax against Census variables by LSOA in the London Thames Gateway, 2001; (a) Council Tax dwellings against households in dwellings; (b) Council Tax dwellings against population in households (Total_CT is total count of Council Tax dwellings; H_Occ is households occupying a dwelling; PIH_all is total population in households).

The structure of the model is given in Figure 3. It is a multi-stage, multiple regression model. A separate model is built for each Borough so they can better reflect local rateable values (on which Council Tax bands are based and the demographics of residents. The models are constructed using 2001 LSOA-level data (Council Tax, Child Benefit, PLASC) with unknowns on numbers of households and population coming from the 2001 Census in the following steps:

- *Step one*. A regression model using counts of Council Tax dwellings in Bands A to H and the percentage of all dwellings in multiple occupancy is created to derive the number of households occupying dwellings. The multiple occupancy variable comes from the 2001 Census and is the only variable that has not been possible to subsequently update. The ONS definition of dwellings in multiple occupancy does not conform with statutory requirements for registration of properties in multiple occupation nor with the Housing Capacity Survey.
- *Step two*. From Child Benefit records it is possible to identify the number of families with dependent children. Assuming each of these families is a household, then by subtraction the number of households derived in step one can be split into households without dependent children and those with.
- *Step three*. A regression model using number of households without dependent children is used to derive the population in households without dependent children.
- *Step four*. Five variables – the number of families with one, two, and three or more children, together with the number of children attending primary school and secondary school – are used in a regression model to derive the population in households with dependent children.
- *Step five*. The results from steps three and four are summed to give total estimated population.

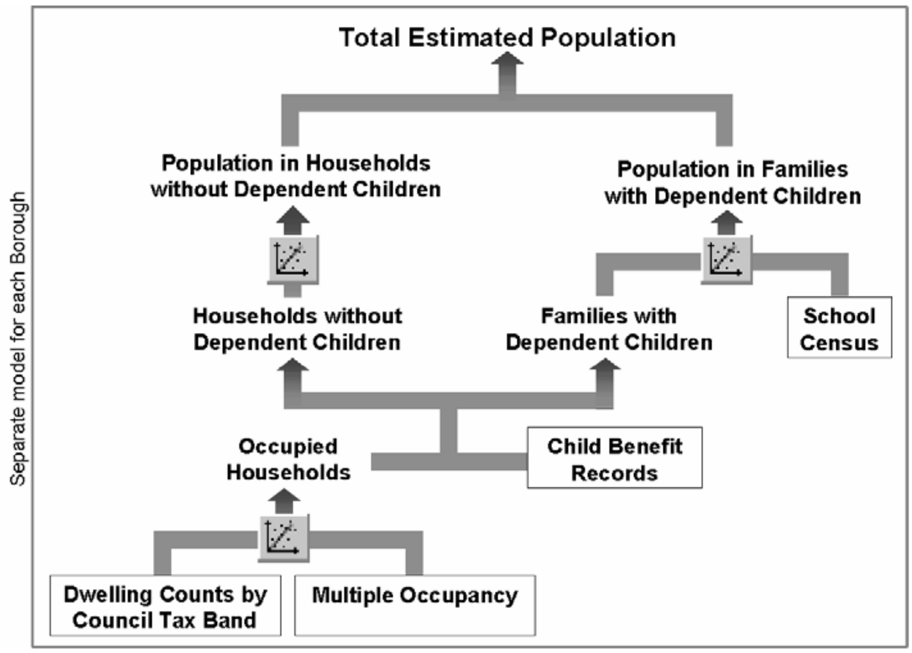


Fig. 3. Overall structure of the population models.

The regression models are constructed using stepwise linear regression whereby independent variables are sequentially added to the model on the basis of their significance until any further change in the fit (R^2) is not significant. This means that the model for each Borough need not contain all the independent variables where some are found to be not significant, and therefore each Borough's model is effectively tailored to its local situation of dwelling stock and demographics. Once the model for a Borough is constructed based on 2001 data, population change can be evidenced by using data for the independent variables from 2002 onwards.

3. RESULTS FOR LONDON THAMES GATEWAY

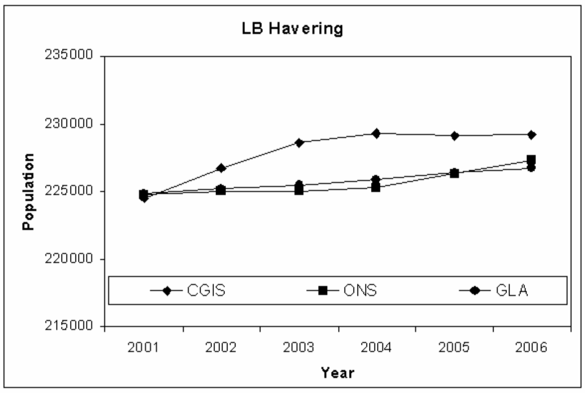
LSOA-level population models for the ten Boroughs comprising the London Thames Gateway (Figure 4) were constructed and tested against ONS and GLA figures as means of evidencing year-on-year population change.

Fig. 4. Location of the ten Boroughs comprising the London Thames Gateway.

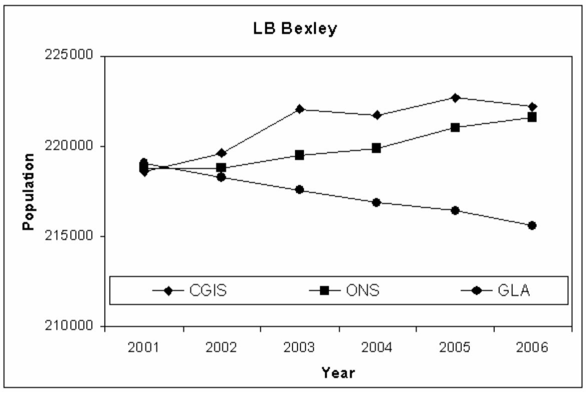
Table 1 compares the ONS 2001 mid-year estimate of population with that from the new models just described (denoted CGIS). On the whole the new models very slightly undercount the ONS estimates. The largest percentage difference is for Greenwich where the model estimate is 0.6% lower than the ONS estimate, whilst the overall variance for the London Thames Gateway is -0.234%. *t*-tests comparing the LSOA figures in each Borough found no significant difference between the new models and the ONS mid-year estimates. However, with subsequent years' data, significant differences do start to emerge as the estimates from the new models start to deviate from the respective ONS mid-year estimates. Thus by 2005, there are statistically significant differences in Barking & Dagenham, Hackney, Havering and Waltham Forest.

LA_NAME	2001 ONS_MYE	2001 CGIS	% difference
Barking and Dagenham	165,654	165,761	*0.064%*
Bexley	218,757	218,578	*-0.082%*
Greenwich	217,460	216,154	*-0.600%*
Hackney	207,246	206,883	*-0.175%*
Havering	224,717	224,471	*-0.110%*
Lewisham	254,336	252,969	*-0.538%*
Newham	249,411	249,239	*-0.069%*
Redbridge	241,893	241,312	*-0.240%*
Tower Hamlets	201,090	201,088	*-0.001%*
Waltham Forest	222,015	221,777	*-0.107%*
London Thames Gateway	2,202,579	2,197,427	*-0.234%*

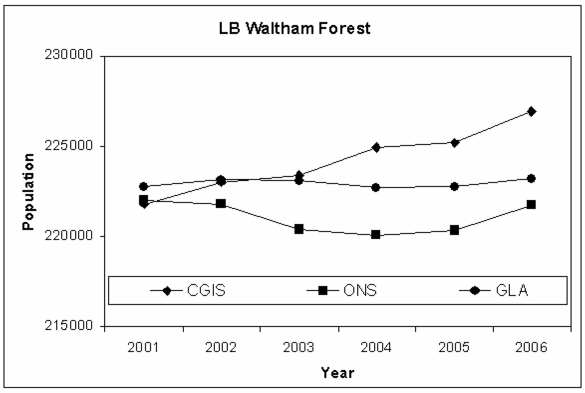
Table 2. Comparison of 2001 mid-year estimates and the new population models (CGIS) for the ten Boroughs of the London Thames Gateway.



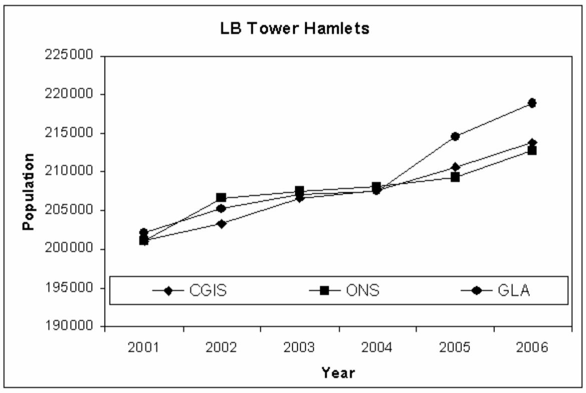
(a)



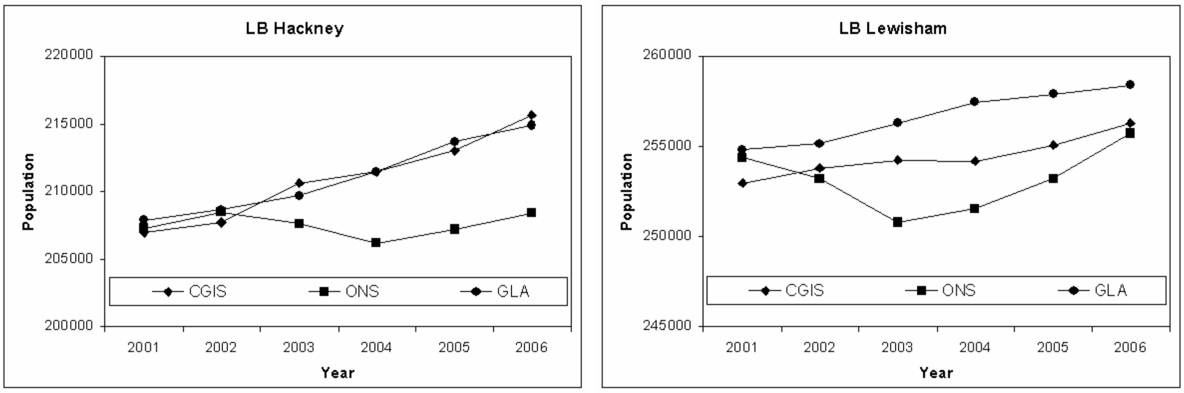
(b)



(c)



(d)

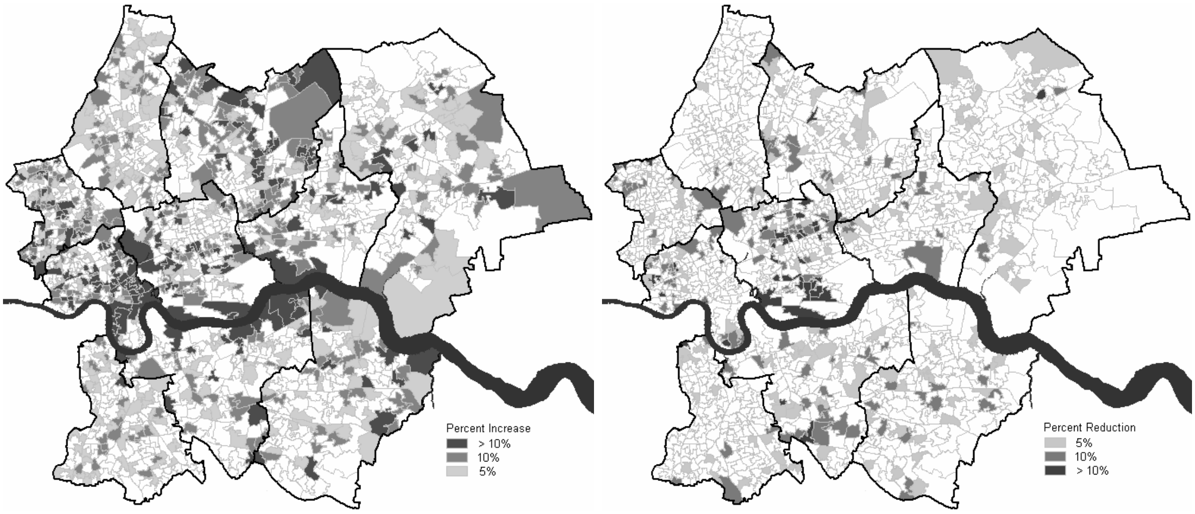


(e) (f)

Fig. 5. Results of year-on-year comparisons of the new models (denoted CGIS) against ONS and GLA figures for six Boroughs.

Year-on-year comparisons with the ONS and GLA estimates produce a number of patterns as illustrated for six of the Boroughs in Figure 5. The aggregate estimates for Havering (Figure 5(a)) show that the ONS and GLA figures for 2001 to 2006 closely conform with each other and that all three estimates are very close indeed in 2001. However the new model (CGIS), whilst agreeing with the other estimates for 2001, evidences initial stronger growth which then levels out from 2004. In Bexley, the GLA estimates show a falling population from 2001 whilst ONS estimates show an opposite trend. The new models would suggest that the population is slightly higher than the ONS estimates, but evidences roughly the same rate of growth. In Waltham Forest, the ONS estimates show a dip in population whilst the GLA figures show a more-or-less flat population. The new model shows a continually increasing population based on the variables used. In Tower Hamlets, population is growing quite fast compared with the other Boroughs shown here, with the GLA figures the highest. The new model seems to corroborate the ONS estimates. In Hackney, the new model again appears to corroborate the ONS estimates whilst the GLA estimates drop between 2001 and 2004 to rise again thereafter. Finally, in Lewisham, the ONS estimates this time show an initial fall in population with a rise thereafter. The GLA estimates continue to grow throughout. The new model for Lewisham starts lower in 2001 than the other two but shows a growth rate similar to the GLA estimates but converging with the ONS estimates in 2006. One Primary Care Trust (PCT) manager attending a project workshop on the new models indicated that every individual unaccounted for in population estimates equated to approximately £1,700 per annum shortfall in budget. Taking the cumulative differences from 2001 onwards between the ONS estimates for Lewisham (on which funding is based) and the new model, the shortfall in budget for health services may have been £13 million rising to £36 million if the GLA estimates are taken as a truer reflection of actual population.

At the LSOA level, population change can be studied for subtle changes that are taking place in the demographics. The shaded LSOA in Figure 6(a) are those that have experienced more than 2.5% increase in population between 2001 and 2006 according to the new models. These strongly reflect areas of new build along the Thames, such as in Tower Hamlets and Greenwich. Figure 6(b) shows those LSOA where population is falling, sometimes due to decanting in preparation for regeneration (e.g. Canning Town in Newham) but is also due to falling family sizes (fewer dependant children from the Child Benefit data) in other areas.



(a) (b)

Fig. 6. Percentage change in population by LSOA 2001–2006; (a) increases in population; (b) decreases in population.

4. DISCUSSION AND CONCLUSIONS

Population estimates can only be approximations just as models are always simplifications of reality. Nevertheless, it is desirable that estimates consistently reflect the true situation for all places, over a range of official geographies and over time. This is particularly important for population estimates which are fundamental for sustainable communities as they are at the core of both funding formulae and planning processes – meeting the needs of residents now and in the future. We have explored an AR approach at LSOA-level using Council Tax data at its core. To our knowledge this is the first time Council Tax data have been used to derive population estimates. Council Tax data will capture both new build in an area as well as sub-division of existing properties into separate units, something which even the GLA models may fail to reflect since there is no systematic monitoring of sub-division and its effect on population numbers. The models (one for each Borough) are built using 2001 data with unknowns coming from the 2001 Census. Year-on-year estimates can then be tracked using annual updates and in theory can be up-to-the-moment as the data sets used are continually updated. These new estimates perform well against the ONS 2001 mid-year estimates as a benchmark. In some Boroughs, the year-on-year population change conforms closely to that modelled by the ONS, but in other Boroughs there is a marked, statistically significant deviation. Where such deviations occur then there appears to be reasonable grounds for a detailed analysis of how the discrepancy arises.

This new approach to modelling population estimates is actively being applied to further Boroughs and Districts at the request of Primary Care Trusts (PCTs) as they enter new rounds of healthcare planning. The models themselves are going through various refinements such as using GP registrations to proportionally disaggregate the LSOA population totals into gender and age bands. The results from London Thames Gateway have been studied by the ONS, GLA and DCLG as part of their on-going commitment to improve understanding of population change.

ACKNOWLEDGEMENTS

This project was funded by the Department for Innovation, Universities and Science and the Higher Education Funding Council for England through the UrbanBuzz programme. We would like to give due recognition to the inputs from our project partners: Dr Sue Batty and Professor Paul Longley at University College London and Chris Horton at Thames Gateway London Partnerships (TGLP); and to thank all those at TGLP and in the Boroughs who provided assistance. The ONS data used in this paper along with the boundary data in the maps are Crown Copyright. GLA data are GLA Copyright.

REFERENCES

1. HOLT D.T. Official statistics, public policy and public trust. *Journal of the Royal Statistical Society A*, 2008, **171**, 323-346
2. HOUSE OF COMMONS TREASURY COMMITTEE. *Counting the Population*. Report HC 183-1. The Stationary Office, London, 2008.
3. OPDM. *Sustainable Communities: Peoples, Places and Prosperity*. Office of the Duputy Prime Minister, London, 2005.
4. NHS LONDON HEALTHY URBAN DEVELOPMENT UNIT. *The Case for Social Infrastructure Planning*. NHS London, London, 2006.
5. GORDON I., TRAVERS A., WHITEHEAD C. *The Impact of Churn and Migration Population Estimates for Central London*. London School of Economics, London, 2007.
6. JUDSON D.H. Information integration for constructing social statistics: history, theory and ideas towards a research programme. *Journal of the Royal Statistical Society A*, 2008, **170**, 483-501
7. BOOTH H. Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 2006, **22**, 547-581
8. SMITH K.S. Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting*, 1997, **13**, 557-567