

2. GSS MAC 16 – minutes

Committee members present

Jelke Bethlehem	Statistics Netherlands	Frank Nolan	ONS
Martin Brand	ONS	Stephen Penneck	ONS
Robert Crouchley	University of Lancaster	Chris Skinner	Southampton University
Harvey Goldstein	University of Bristol	Andy Sutherland	NHS Information Centre for health and social care
Graham Jenkinson	ONS	Sandy Stewart	Scottish Government
Rachel Leeser	Greater London Authority	Kenneth Wallis	University of Warwick
Jil Matheson	ONS	Martin Weale	NIESR

Presenters

Owen Abbot	ONS	Dick Heasman	ONS
James Brown	Institute of Education, London	Salah Merad	ONS
Sarah Green	ONS	Peter Youens	ONS
Jacqui Jones	ONS		

Others present

Robert Bucknall	ONS	Craig Orchard	ONS (new secretary)
Myrto Miltiadou	ONS	Denise Silva	ONS
Jane Naylor	ONS	Alan Taylor	ONS
Louisa Nolan	ONS (secretary)		

Apologies

David Hand	Imperial College London	Peter Lynn	University of Essex
------------	-------------------------	------------	---------------------

Introduction

Stephen Penneck welcomed new member Andy Sutherland to the committee, and noted that Graham Jenkinson was also present for the first time since his division was incorporated into the Methodology Directorate. Stephen also introduced the new secretary, Craig Orchard.

The minutes of the previous meeting had already been approved.

Comments on progress from GSS MAC 15

A note on progress made in response to GSS MAC comments on NSMAC 14 Paper 3: 'A state space approach to extracting the signal from uncertain data' (Cunningham et al.) was provided by Gary Brown, ONS, as follows.

In response to the committee's helpful suggestions, ONS investigated the potential of revisions to improve estimates of Gross Domestic Product (GDP). Our analyses focussed on three techniques - kernel densities, transition matrices, and component regression.

- Plotting kernel densities of month 1 (M1), month 2 (M2), month 3 (M3) and Blue Book 2 (BB2) estimates showed a gradual positive drift for later maturities, ie indicating positive bias.
- Matrices of the probability of M1, M2 and M3 estimates being in a different quintile of the distribution at BB2 showed that low estimates were more likely to be revised upwards than high estimates downwards, again indicating bias. When the sample was split, a performance measure (based on Shorrocks, 1978) showed the tendency to revise lower quintiles upwards increasing over time.
- Correlation and regression analysis (using Cochrane-Orcutt estimation) of the relationship between component revisions and aggregate revisions provided some useful information about the drivers of the total revision. Although an early stage, this approach may enable ONS to focus resources on reducing revisions in the most influential components of GDP.

However, our overall conclusion was that due to their volatility, and especially in times of economic instability, past revisions do not provide a robust prediction of future revisions, so do not have the potential to adjust GDP. We even question whether GDP needs to be adjusted. We tested mean revisions between M1 and BB2, M2 and BB2, and M3 and BB2 over three time periods: 1993 to 2005, 1993 to 1999, and 1999 to 2005. None of these tests showed evidence of revisions being significantly different from zero.

References

Cochrane, D. and Orcutt, G. H. (1949) "Application of least squares regression to relationships containing autocorrelated error terms" *Journal of the American Statistical Association* **44** 32-61.
Shorrocks, A. F. (1978) "The Measurement of Mobility", *Econometrica* **46** 1013-1024.

Martin Weale noted that this response stated that revisions were likely to go up, not down, but that there was no stability in the revisions to components of GDP, although he thought there was a hint that there was some stability in the original paper. Martin Brand said that this was work that would be returned to, and Stephen Penneck added that the paper showed that the apparent stability was not robust.

Ken Wallis had more general ideas in mind, especially for revisions which are due to known causes, e.g. methodological changes. He asked whether there were any papers in existence which go through the explanations for past revisions. It would be useful if advance warning of revisions, and their expected size, could be given.

Stephen said that National Accounts does publish changes to methodology, but more could be said about the impact. Jil Matheson pointed out that in practice, this would require more resources. Stephen wondered whether more could be done to give advance warning, for example, about the stability of the upward revisions to GDP. He noted that these were stable despite having no known cause, and despite the fact that the same was not true of the individual components of GDP.

Martin said that he would pass the comments on stability and advance warnings of methodological changes on to National Accounts.

Actions

2a	Martin Brand to pass comments on revisions paper back to National Accounts
-----------	--

Comments and news from GSS / ONS

In addition to the items of news presented in the booklet for the 16th meeting of the GSS MAC, Stephen Penneck took the opportunity to advertise the 14th GSS Methodology Conference, at The Church House Conference Centre, London, on the 30th of June, 2009.

Martin Brand added the following news summary.

- Good work has been done on the Retail Sales Inquiry (RSI), including testing the chainlinking, and improving the deflation methodology. This comes into effect on the 21st of May.
- The study of outliers and register discontinuities has been successfully completed for the Average Weekly Earnings (AWE) survey.
- Significant progress has been made on work created by the change to Standard Industrial Classification 2007 (SIC (2007)), notably on the creation of the new Monthly Business Survey (MBS) which will replace the four monthly business surveys existing under SIC (2003).
- Construction statistics have now been transferred from BERR to ONS.
- The Eden project, on selective editing, is looking very promising.
- The improved questionnaire and editing rules for the Business Register and Employment Survey (BRES) are being finalised.

Comments from the committee

The committee discussed issues arising from use of the new Publication Hub. It was noted that:

- if a statistic is a National Statistic, it will be published on the Hub, regardless of which department produces the statistic;
- although the Hub contains an ordered list of releases, there was concern that data are difficult to find if the exact name is not known, and were in fact more difficult to find than under the previous system;
- this kind of feedback is very much welcomed, and Stephen Penneck asked that any written feedback be sent to him.

Chris Skinner was interested in the role of committees looking at assessment. He wondered whether the Standards Committee would be taking over any of the role of the GSS MAC. Stephen Penneck responded that the GSS MAC is intended as a forum for the discussion of difficult technical issues, which may be brought by any GSS member. In contrast, the Standards Committee is part of the system of governance. Jil Matheson, who is Chair of the GSS Statistical Policy and Standards Committee (GSS SPSC), added that she wants to strengthen the role of the GSS MAC. There are likely to be some issues which are relevant in more than one situation across the GSS, and she sees the GSS MAC as being important in addressing those issues.

Sandy Stewart asked how the Methodology Consultancy Service (MCS) is accessed. Stephen Penneck replied that Simon Compton is the contact to speak to, and that the service is marketed through the GSS. He added that the MCS will deal with a range of projects or varying sizes, and that currently, many of these projects are providing training for

departments. Other projects include quick quality assessments, intended to identify priorities for action before assessment. He noted that most departments don't usually have specialist methodologists.

Actions

2b	committee members to send feedback about the new Publication Hub to Stephen Penneck
-----------	---

2.1 Cost-benefit analysis of proposed new data requirements

Authors	Craig B. Orchard	ONS
	Sarah Green	ONS
	Bronwen Coyle	ONS
	Jacqui Jones	ONS
Presenter	Sarah Green	ONS
Discussant	Martin Weale	NIESR

The paper presents work in progress developing a cost-benefit modelling tool. This is a conceptually difficult task. The tool aims to allow *comparable* (rather than *accurate*) measurements, and is a move from a one-size-fits-all approach to a more tailored model, constrained to a single use. The output is not a single number, but a list of essential information for informed decision-making.

Discussion

The discussant, Martin Weale, made the following points in response to the presentation.

Firstly, this is very important work. Martin expressed concerns that ONS can be too slow to scrap unnecessary data collection. He hopes that this work is the start of a new framework for cost-benefit analysis, rather than the current system which is often driven by vocal users. He noted that most government departments find it necessary to make trade-offs, and ONS is no different as judgements have to be made.

Martin then commented that users always want more, but often don't pay the costs. This is in opposition to respondents, who always want to give less – this can even be apparent in two parts of the same business. Therefore, this is a sensitive issue, but that is not the same as being difficult.

Martin said that business users need to understand the trade-offs, e.g. accuracy versus timeliness. Where there are a number of factors to be considered, the weighting of these factors is crucial. He strongly believes that it is wrong to give all factors equal weighting. He asked whether there are any plans to find out from businesses what their priorities are.

It is difficult to properly quantify risk. Martin stated that cutting corners is likely to lead to occasional disaster rather than uniform gradual erosion (a 'black swan' event). It is important to estimate the likelihood of unusual disaster, but this requires careful wording on a questionnaire, as businesses are not usually used to thinking about rare events. There is a strong need to be honest about significant risks.

Some benefits are unpredictable and unquantifiable, for example, benefits to academics. Rare large benefits are difficult to model, and some benefits, for example benefits to health, may be quantifiable, but require a convention on valuation.

Martin then asked whether data producers were really able to identify all possible benefits.

When making a judgement using multiple criteria, Martin believes that the ranking of those criteria will depend on the circumstances, for example, accessibility may be less important in some circumstances than accuracy. The ranking of criteria should be reflected in a monetary value. Borda ranking may not be appropriate for this work.

Finally, Martin reiterated that he thought this work was very important, and that he very much hoped it was the start of much more work on the subject.

Comment was then invited from the other committee members.

Andy Sutherland asked whether cost-benefits were attributed to the data collection or the policy. He also pointed out that it is important to look at uncertainties and compare these with the scale of the likely benefits.

Sarah Green, in response to a question from Jelke Bethlehem, said that currently, businesses' perception of burden was measured, rather than the actual burden, and the time to complete a questionnaire was often inflated. Currently, the only robust measurement of burden is through the regular reviews which are required under the Prime Minister's Instructions.

Stephen Penneck agreed with Martin that the components in the model should not in general carry equal weight. He also said that the definition of a quality attribute should be clearer.

Robert Crouchley asked if it was possible to do a sensitivity analysis. Stephen suggested that it might be possible.

Sarah Green answered Rachel Leeser, saying that a change in quality could be negative as well as positive.

Rachel also asked how it was known who the users were. Stephen Penneck replied that it was more difficult these days with website accessibility. Martin Weale suggested a survey of users on websites, and Stephen replied that a registration system for users might be possible in the future. At present, users are identified through user groups and networks, but many web users are unknown.

Harvey Goldstein echoed Martin Weale's comment that it would be useful to see a copy of the Wallis equation. He also said that it was not always necessary to convert costs and benefits to cash, there were other options. Stephen Penneck acknowledged the importance of this point. He asked how far the analysis should be turned into a mathematical interpretation, with an answer in pounds. The requirement is for something scaleable.

Sandy Stewart pointed out that different users wanted different things (for example, the Scottish Government pay to top-up certain surveys). The weights on the components of the models must reflect those needs. Further, trade-offs can only be made if a sensible risk measurement is done, and that might change over time.

Robert Crouchley said that it was a fascinating paper. He asked why it had to be a cost-benefit analysis. Why not use a multi-criteria cost effectiveness analysis? This might not need any weighting of components, as judgements could be made by committee. He also pointed out the danger of cost-benefits analysis becoming a 'formalised fiction'. Benefits are often over-stated, and cost under-stated.

Stephen Penneck ended the discussion with a summary of the main points. He said that this problem had to be addressed as it formed part of the new Code of Practice. There was a consensus amongst the committee about investigating non-uniform weights and performing a sensitivity analysis. He acknowledged the difficulty in collecting views from users, and suggested further exploration of other techniques.

Actions for authors

2.1a	to investigate non-uniform weighting of the model components, and to conduct a sensitivity analysis
2.1b	to further explore other techniques

The Wallis equation

The committee asked about the valuation of statistical benefits in the Wallis model. This addendum uses a cut down version of the full method to explain it. It must be emphasized that the Wallis model uses a range of assumptions to generate its estimates.

Wallis first calculates a baseline value for ONS statistical outputs which will be used in producing a value for statistical benefits. To do this Wallis groups the detailed ONS statistical outputs into 15 sets of key outputs. The baseline value of these key outputs is taken as the cost of those outputs in a given year. This is calculated by aggregating value along the Statistical Value Chain (SVC).

To value statistical benefits Wallis next considers which of the key outputs an improvement project will affect. The method then identifies which, of the defined quality criteria (risk and the six European dimensions of quality), the improvement project impacts on and by how much. This impact is recorded as a subjective yet simple score of high, medium or low.

Associated with each quality criteria and impact score is a benefit-to-cost ratio which is used in the final valuation of statistical benefit. Wallis based the ratios on those developed for forecasting the benefits of the Statistical Modernisation Programme. The example of these benefit-to-cost ratios, as given by Wallis, is shown in Table 1.

Table 1: Benefit-to-cost ratios

		Quality Criteria						
		Relevance	Accuracy	Timeliness	Accessibility	Coherence	Comparability	Risk
Impact Score	High	0.05	0.05	0.025	0.025	0.01	0.01	0.05
	Medium	0.025	0.025	0.01	0.01	0.0025	0.0025	0.025
	Low	0.005	0.005	0.0025	0.0025	0.0005	0.0005	0.005

Table 2 shows the method described above and how Wallis uses the components to calculate the total statistical benefit from the improvement project.

Table 2: Example of methodology

Which key output does the project affect?	Total costs of key output excluding overheads?	Which quality criteria are affected?	Impact Score	Which is the benefit-to-cost ratio?	Statistical benefits
LMS	£8,530k	Timeliness	High	0.025	$£8,530 \times 0.025$ = £213k
LMS	£8,530k	Coherence	Medium	0.0025	$£8,530 \times 0.0025$ = £21k
Total Statistical Benefit from Improvement Project					£234k

References

Wallis (2006): A methodology for valuing statistical benefits, *The Office for National Statistics*

2.2 Tackling Bias in the Dual-System Estimator for the Census

Authors	Owen Abbot James Brown	ONS Institute of Education, University of London
Presenter	Owen Abbott	
Discussant	Jelke Bethlehem	Statistics Netherlands

This paper discussed the methodology for assessing and estimating the level of coverage by the Census of the UK population. Particular emphasis was placed on discussing current plans for adjusting for biases in this methodology. These biases are due to failures of assumptions in the methodology. These were discussed in detail, as were the planned approaches for dealing with these biases. The committee was asked to provide any comments on this issue at the meeting, particularly to provide their views on the proposals in the paper.

Discussion

The discussant, Jelke Bethlehem noted that he comes from a country that has not had a real Census since 1971 as it has a population register.

Jelke gave a presentation that highlighted the following questions and issues for discussion.

- Jelke requested further information on levels of non-response in the Census and the Census Coverage Survey (CCS) as none were given in the paper.
- He highlighted the issue of memory effects, due to the distance in time between the CCS and the Census, biasing the results of the CCS.
- He raised concerns about how the assumptions of the methodology could be tested in practice.
- He thought that the methodology for estimating the number of movers looked sound, but might be unnecessarily complicated. This was mainly because of the inclusion of a number of different levels of aggregation in the equation for the mover estimation. The rationale behind this was not completely clear. Jelke also noted that the methodology of matching movers across geography posed a very large computational problem. One possible approach for overcoming this would be make estimates based on smaller sub-samples.
- Jelke noted that there is an issue of respondents claiming that they had not moved since Census night in order to avoid additional questions.
- He was also aware that rates of moving were likely to vary by area. He asked if it would it be possible to model these area differences or include them using some form of stratification. He noted it would be useful to think about using any structure in moving behaviour to assist estimation.
- An analysis of movers by age groups may also be a good idea, but this should take account of birthdays that occur in the period between the Census and the CCS.
- Jelke asked whether it was possible to identify new arrivals to the country in the six week period between Census and CCS.
- He requested further clarification on matching weights. It was not clear to him how they were defined and it may be worth investigating whether these could be improved.
- The statement that the false positive rate was not relevant required justification.
- He noted that the quality of the measured variables defining the weights was important. These could be subject to varying levels of measurement error.
- The independence of Census and CCS was very important, and it was not clear to him how to test this assumption.
- He noted that it was important to recognise that simulation studies were only informative if the model reflected the true situation. The validity of the simulations therefore required further justification.
- In terms of using alternative data sources, it would be important to assess the quality of these data to guarantee that their inclusion would result in improvements to the estimation process.

- The assumption made in the paper that the sex ratios were constant across areas was likely to be incorrect. For example, he would expect to find a difference between these ratios calculated for urban and rural areas.
- Overall, he felt the work proposed in the paper would require a large amount of resource for potentially only small improvements. He therefore recommended assigning priorities and carrying out some cost-benefit analysis.

Comments were then invited from other committee members.

Harvey Goldstein asked whether memory effects in the CCS, due to the time lag between the Census and the CCS, could be investigated by separating the CCS fieldwork into time blocks. If these blocks were carried out, for example, three, five and seven weeks after Census night, this would allow memory effects to be estimated. Alternatively, could the natural variation in the timing of the fieldwork be systematically exploited?

Harvey Goldstein also asked how feasible it would be to publish estimates of the interval for bias, and propagate these in the Census outputs. Owen Abbott responded that this did happen in 2001, and the plan for 2011 was to ensure this had higher visibility in the Census outputs.

Martin Weale noted that the quality of the address register was crucial. He asked whether this was now of a good enough quality to obviate similar issues to those that occurred in 2001, particularly those that arose due to multiple households per address in places like Westminster and Manchester.

The chair noted this point and tabled it for further discussion with Graham Jenkinson, Divisional Director of ONS Statistical Frame Work Division, which is responsible for creating the address register.

Rachael Leaser noted that an address frame was not the same as a household frame and that these should not be considered equivalent. She gave the example of a household that dissolves between Census and CCS. The parts of this dissolved household are unlikely to have the same probability of being counted in both the Census and the CCS.

She asked whether research had been carried out to determine the effect of postal and internet data collection, compared with the CCS interviewer-based collection.

She also requested the Government Office Region adjustments make a distinction between inner and outer London. Owen responded that this would, in general, be the case.

In response to Jelke's questions Owen noted that non-response, in Jelke's terminology, was equivalent to what he termed under-coverage.

With regard to the timing between the Census and the CCS, Owen noted that the Census team wanted to extend the time they had to carry out follow-up field work in order to maximise response to the Census. He had carried out some simulations which indicated that moving the time between the Census and CCS to four, five or six weeks had little effect on the quality of Census estimates. This work was based on a range of reasonable assumptions regarding Census and CCS response rates under different scenarios. There would be an increase in Census response due to lengthening the field period and a decrease in CCS response rate as the "halo" effect of proximity to the Census decayed over time. At about a 10 week separation there was a marked reduction in the efficiency of the estimates. A decision has therefore been taken to make the time lag between the Census and the CCS six weeks. Therefore, it will therefore not be possible to phase the CCS field work over any large timescale. It might be possible to include phasing of the CCS in a smaller-scale experiment as part of the Census and CCS field procedures. Owen thought that utilising the natural variation in CCS response times was an idea worthy of further investigation. He noted that this was an issue he was concerned about, but as yet there was no hard evidence of the likely impact these memory effects would have on the quality of the CCS data collection.

Owen noted that although a household frame is required for the Census, it is only possible to create an address list. The primary issue is how the Census procedure will be able to overcome this discrepancy. One method that will be used is, where a potential difference between the household and address frames is identified, the form will be hand-delivered.

Graham Jenkinson commented on the development and quality of the address register. The register should be able to pick up addresses in houses of multiple occupation. In 2001, the register suffered from inaccuracies and was not up-to-date. For 2011, the address list is being created by merging the Royal Mail Postcode Address File (PAF) and the National Land and Property Gazetteer. There are several stages to the matching and checking process, and these have

been piloted in a number of areas, including a number of challenging areas. For example, ONS staff have checked addresses in areas where there is found to be a mismatch between the two lists. Lists will be sent to Local Authorities for a final quality check.

Rachael Leeser asked whether the final cut from these sources had been taken. Graham replied this would not be done until much closer to the Census. At the moment, ONS was sending a list of anomalies to the original list suppliers for resolution. The final list will continue to be updated until as close to Census day as is practical for field procedures. This will certainly occur up until the end of 2010. There will be an initial Census form print run in the middle of 2010 and a later run to include later list updates so that changes, such as new housing, are include in the Census post-out. Graham noted that ONS thinks it has a model that meets the Census requirements. This is a very challenging process but he is confident the final register will be significantly better than in the last Census and will meet the Census objectives.

Owen noted that the Census in 2001 did not use an address register and was not a postal survey. The address list was rather an aid for enumerators. However, in 2011 most of the delivery of forms will be based on the address register. Stephen Penneck noted that this made the address register a critical part of the Census.

Stephen thanked Owen and James, as well as Jelke for his comments. He suggested that a further discussion would be valuable.

Actions for authors

2.2a	to have further discussion regarding the useful comments made by the panel
2.2b	to look at modelling estimates of moves on small sub-samples
2.2c	to model differences between areas' mover rates, or to account for them via stratification
2.2d	to investigate whether matching weights can be improved
2.2e	to investigate memory effects using the natural variation in response times of the CCS

2.3 When to benchmark short term surveys to annual

Authors	Martin Brand	ONS
Presenter	Martin Brand	
Discussant	Kenneth F. Wallis	University of Warwick

This paper examines the issue of whether or not to benchmark sub-annual (usually monthly) surveys to annual. It concludes by suggesting a possible policy for ONS surveys.

Questions arising from the paper are listed below.

- Is the list of criteria for deciding when to benchmark correct?
- Is the list of pre-requisites for benchmarking correct?
- Does GSS MAC concur that the ONS benchmarking policy should not be global i.e. it should be elective, judging each case on its merits?

Discussion

Kenneth Wallis said that the answer to Martin's questions in his mind is 'yes, but....'

- The issue is how to weight the different criteria and pre-requisites that are given, in order to form a judgement about the merits of one case over another. What is the framework which might be used to answer the questions on criteria from a statistician's perspective?
- The problem has been around in statistics for a long time (see list of references at the end of this document). Early work tended to assume that the annual data was the more reliable, and the practical problem was how to interpolate annual series using monthly observations. The monthly data was assumed to be generally less reliable or inferior to the annual series.
- The latest work is built on combination forecasts. Here, data are always improved by combining series even if one is of better quality than the other (in terms of mean square error). Therefore, quality improvement is always possible.
- The metric needed to weight criteria and combine series arises from consideration of the uncertainties of the series.
- The immediate question arising from Martin's slides is: what is the standard error of those numbers? Are the differences:
 - a. important (should we be bothered about them?)
 - b. is there a combined estimate that we might prefer? (i.e. treat them both as uncertain).
- This is the framework that Ken finds the most appealing. He suggests that uncertainties should always be reported, and this leads to answers to subsidiary questions.
- Within a simplified theoretical framework we can then find a way to calculate estimates, which gives us a sensible way of thinking about any judgements to be made, e.g. on coverage, subjective assessments, timing etc..
- Investigating what effect any judgements have on the standard errors rather than the numbers is the sensible way forward. To answer the policy questions in the paper, ONS should use this framework for treating the uncertainty of the data.
- This approach will cause revisions to preliminary data, and monthly data will be revised when annual benchmarking is done.
- This raises the question of reporting the uncertainty in terms of revisions and giving some estimates on the likely size of future revisions. This is a familiar problem in National Accounts.
- Users have to be convinced. There is a protocol on revisions in the office that recommends that information should be given on past revisions.

Comments were then invited from other committee members.

Chris Skinner pointed out that there many more options than the conventional benchmarking ones. For example, the annual data can be used to calibrate the monthly series to adjust for non-sampling errors. He said that it is more honest to report inferior and different variables in sub-annual series as the variables they are than it is to adjust them to something was not measured in the first place. We need to make a judgement to decide how this is done, but this judgement should be informed by analysis of the available options.

Sandy Steward offered an example from the Scottish Government. They take annual figures from Regional Accounts and monthly and quarterly figures from the Retail Sales Inquiry. The quarterly data are used to reflect quarterly changes and the annual data are used for benchmarking. By shifting the annual figures down or up, discontinuities appear at Q4 and Q1, so the quarterly figures are modelled, and the trend replaced with a cubic spline. This works for their purposes., preserving the annual and quarterly fix. Sandy Steward also suggested that more work on congruence might be needed, i.e. an investigation of why the differences occur between monthly/quarterly and annual figures on the first place.

Stephen Penneck noted that there has been a tendency to try to understand differences between individual returns from businesses, rather than understanding differences in the aggregate data.

Martin Weale added that Sandy's method can be extended to address questions on bias as well. Work on this has been done by Chow and Lin (GRSS series C 1969). The question is whether it is the sums of squares of the changes or the

sums of squares of the deviations that need to be minimised. the discontinuities that Sandy talked about are brought about because levels are being used. Some work has been done to resolve this issue.

In response to a further question by Sandy Stewart, Stephen Penneck noted that if both annual and quarterly figures have revisions, there would be a user perception issue, and users would find it difficult to understand why it happened. Martin Weale pointed out that if Ken Wallis' suggestion was used, the annual and monthly data would always be consistent.

Graham Jenkins raised the issue of what users want. There is scope here to ask users. Do they prefer a simple story that gives them a consistent picture or two different stories depending on where they look? Benchmarking enables us to confront and explore differences between short period and annual estimates. This can lead to a better quality of outputs.

Martin Brand responded:

- in principle, the proposal to treat both series as uncertain and produce combined estimates is something that can be investigated. There are issues with the fact that the annual surveys have more variables than the short term surveys. It will be possible to reconcile and produce combined estimates for the turnover variable but how would annual variables that are not collected sub-annually be adjusted?
- standard errors are available for every survey and they are published. This is going to be used to compare precision. However, bias measurement is trickier, and not currently handled very well.
- how should a consistent time series be created? In ONS, there are standard methods for benchmarking, like the cubic spline method, and the functionality is built in our software.

Martin also added that he is grateful for the useful comments he received.

Actions for authors:

2.3a	to treat both estimates as uncertain
2.3b	to use standard errors to measure precision
2.3c	to report uncertainty
2.3d	to measure and report revisions
2.3e	to explore why differences occur between annual and monthly or quarterly series

References provided by Kenneth F. Wallis:

- Bates, J.M. and Granger, C.W.J. (1969), 'The combination of forecasts', *Operational Research Quarterly*, 20, 451-468
- Chow, G.C. and Lin, A. (1971), 'Best linear unbiased interpolation, distribution, and extrapolation of time series by related series', *Review of Economics and Statistics*, 53, 372-375
- Denton, F.T. (1971), 'Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization', *Journal of the American Statistical Association*, 66, 99-102
- Friedman, M. (1962), 'The interpolation of series by related series', *Journal of the American Statistical Association*, 57, 729-757
- Hillmer, S.C. and Trabelsi, A. (1987), 'Benchmarking of economic time series', *Journal of the American Statistical Association*, 82, 1064-1071
- Whittle, P. (1963), 'Prediction and regulation by linear least-square methods', London: English Universities Press.

2.4 Developing expertise in record linkage within ONS Methodology Directorate

Authors	Dick Heasman	ONS
	Briony Eckstein	ONS
	Peter Youens	ONS
Presenter	Dick Heasman	
Discussant	Harvey Goldstein	University of Bristol

This paper outlines initiatives in developing capacity for record linkage within ONS, illustrating the procedures necessary for data sharing across Government. The specific example of a project sharing data on school pupils with the Department for Children, Schools and Families (DCSF) is discussed. Linking this shared data is of benefit to both National Statistics and research on pupil attainment. The paper is a follow-up to the paper presented at NSMAC 12 by Marie Cruddas '*Combining Data: Developing a Centre in MD to meet the challenges*'. The presenter, Dick Heasman noted that Methodology Directorate at ONS fits within the data sharing trend of the GSS, and should have a role in advising on record linkage.

Discussion

Harvey Goldstein began by noting that he was new to the area of record linkage and learnt a lot from the paper. He went on to make the following points.

- The point of record linkage is to determine values of variables not included in a data set. A convenient way of doing this is to match to a different dataset that does contain these variables.
- It is important to know what the record linkage is for and what analysis will be done. It is also important to know what the linked file looks like – is it produced for a particular analysis? Therefore, should linking be done just when required for a particular analysis, developing a methodology for each analysis task independently? Is the linkage a file, or a process, which produces a virtual, updated file?
- In the National Pupil Database (NPD), children outside the maintained sector are not included. Families of children in the maintained sector are likely to have different characteristics from families of children in private schools. However, private school pupils' exam results are included in the NPD, and there is a case for the private sector to be included in the NPD completely.
- There are further questions to be addressed before birth-linked data could be used to improve understanding of migration and to improve census counts. What would be done about unmatched records? Could probabilistic matching be used?
- Probabilistic linking is widely used, and widely mis-used. There must be some error attached to the process, which is not always carried over into the analysis, although this can in principle be done.
- Even though records could be matched there may still be a problem of missing data. This could be dealt with by imputation.
- The Pupil Level Annual School Census (PLASC) / NPD is a file of children, but their membership of families can change over time, so linking with household and area data may be complex. At the moment it is not possible to link children of the same family together. It would be of great benefit to be able to do this.
- Date of birth errors in PLASC/NPD may lead to biases if they are associated with other variables of interest. It is known that these data contain many inaccuracies and biases.
- There is already a lot of experience on pupil attainment using PLASC / NPD. Literature exists to guide future analysis. The PLASC / NPD Users' Group (PLUG) seminar series brings this together.
- The usefulness of PLASC / NPD for enhancing the census seems problematic as only children from maintained schools are included.
- Harvey expects that the ONS Record Linkage team will be very much in demand as record linkage of official data sets is likely to become more important. The problems are far from trivial and there is a lot of work taking place internationally in Australia and Europe.

- Harvey recommended that the team maintain strong contacts with ADMIN, as they have considerable experience e.g. in NPD linking.
- Attendance by members of the ONS record linkage team at international meetings was strongly recommended.

Stephen Penneck commented that this is part of an attempt to provide a general capability for data linking which the ONS and GSS will need over the next few years as part of the move to use more administrative data sources and to look at the complex relationships between data sources.

Comments were then invited from other committee members.

Chris Skinner reinforced Harvey's points about the importance of looking at the quality implications of this work, in particular, analysing the occurrence of false positives and false negatives. For a range of applications coverage issues are important. It would be useful to get an idea of matching errors; investigating possible errors of that kind would be worthwhile, and commented that in 2001 Census/Census Coverage Survey matching these had been estimated using double clerical matching.

Rachel Leeser echoed Harvey's comments about groups missing from the dataset and reported that in Westminster for some wards 93% of children are not on the database. For detailed geographies this is a major problem. Rachel also pointed out that when linking to other datasets, addresses may be different. Therefore, datasets put together for different specific purposes might look different from each other. How should this be dealt with when linking? It is important to keep in mind that the data must be fit for purpose.

Jelke Bethlehem commented that Statistics Netherlands have a population register and unique identity codes, and therefore do not experience record linkage issues. Bill Winkler from United States has a lot of experience in the field.

Robert Crouchley pointed out that ONS was probably the only body able to do record linkage, due to privileged access to the microdata. As soon as the data are released, they are anonymised, which only allows linking to occur before release. Robert also highlighted the harmonisation issue of reconciling different responses when surveys ask similar, but not identical, questions.

Sandy Stewart reiterated Rachel's point about the necessity for doing a specific job for a specific purpose. In addition, he also suggested that a generalised research dataset could be produced, where as much data are matched as possible, missing data are imputed, and the dataset of anonymised records is released for analysis on an ad-hoc basis. This data set would be made available in parallel with the other more specific information.

Stephen Penneck commented that not all research can have a clear purpose. There is value in speculative research which looks at data, identifies what is interesting and then doing something more focussed. He acknowledged Rachel Leeser's comment that it would be a massive undertaking, but expected there to be a demand for it.

Andy Sutherland suggested that it would be very useful to keep a register of records that have accrued from this linkage work.

Stephen Penneck agreed that this was important. It takes a long time to acquire data, due to the time it takes to prepare business cases and persuade departments that a project is worthwhile. It is important to develop clearer statements of benefit to the community as this will strengthen the case for subsequent requests.

Dick Heasman responded.

Regarding Harvey's question about whether separate methodologies should be used for each linked file that needs to be created, Dick pointed out that this is what has happened in ONS until now. There is something to be said for that approach – the purpose should be kept in mind. There is also a place for having a pool of expertise.

Dick was not aware that PLASC/NPD data did not contain membership of families, but pointed out that complete families are included on the Labour Force Survey (LFS).

Dick is hoping that the DCSF matching project will investigate date of birth errors on a selected extract from the date of birth data. Where matches are confidently made on other variables, date of birth errors can be analysed.

Actions for authors:

2.4a	to investigate what could be done with unmatched records when using birth-linked data to understand better migration and improve census counts
2.4b	to investigate appropriate analysis methods following probabilistic linking
2.4c	to further investigate date of birth errors
2.4d	to seek wider research on pupil attainment using PLASC / NPD - see PLUG seminar series and work in Australia and Europe
2.4e	to investigate the quality implications of this work and to analyse matching errors
2.4f	to consider producing a dataset of anonymised records for analysis on an ad-hoc basis to be made available in parallel to more specific information
2.4g	to keep a register of records that have accrued from this linkage work

2.5 Developing an apportionment method for financial variables based on returned and synthetic local unit turnover data

Authors	Salah Merad	ONS
Presenter	Salah Merad	
Discussant	Sandy Stewart	Scottish Government

Salah Merad presented his paper about the use of new site turnover data that will be available from BRES in the Annual Business Inquiry Part 2 (ABI/2) for apportioning enterprise level data to local units in order to produce regional estimates. The proposed apportionment method consists in fitting a model for predicting turnover value for each local unit based on BRES data. A large pilot project has been implemented, using a random sample of 11,000 businesses. However, only about two thirds of the multi-unit sites returned useable data.

Discussion

Sandy Stewart welcomed the pilot exercise and noted the importance and relevance of the project. He stated that a good quality business register was vital for regional analysis, and that BRES will hopefully improve the Inter-Departmental Business Register (IDBR) by keeping it more up-to-date.

The discussant noted that there may be a lot of noise in the data from small units in the pilot exercise, and that economic variables have in general skewed distribution.

He added that it is important to understand the reasons for zero returns. In the case of large firms, which are generally ONS special arrangements, they may be due to accounting procedures or business operations. Some business types may have irregular or seasonal turnover (for example, in construction or agriculture). The discussant suggested that a more detailed analysis by Standard Industrial Classification (SIC) could shed some light on zero return patterns. He also raised concerns about the quality of the addresses recorded in the IDBR, especially if they are to be used as a key variable to define location of business units when producing regional or local statistics.

In relation to the overall modelling procedure, the discussant agreed with the option of treating zero and positive returns separately but recommended that future work for providing more evidence of the model behaviour by SIC and region should be carried out. Sandy also noted that a final report could include more detailed information on model diagnostics, including residual analysis and the analysis of variance to identify how different variables contribute to explain the total variation. He also suggested the development of some sensitivity analysis, investigating the large units and zero returns in more depth. Zero returns may be due to seasonal effects, accounting problems or no trading in a given period.

When asked to consider the use of synthetic BRES turnover for all units to produce annualized local unit turnover, he recommended that both methods described in the paper should be tested and compared, with differences by region and economic activity highlighted.

In relation to relying on the model prediction function, for variables other than turnover, for multiple-site units, Sandy noted that the model and analysis presented are based on single units and proposed the inclusion of other enterprise characteristics/variables in the model to capture this differential effect.

Other GSS MAC members added comments and suggestions as follows.

There was a general agreement that the zero returns must be investigated, and that ONS may be trying to collect local unit data that large enterprises cannot provide regularly for all sites. In addition, they questioned the concept of “local unit turnover”, indicating that is vital to understand the non-response for big enterprises, and recommended that ONS should consider a improvement in the survey questionnaire if necessary.

Expert members also noted that turnover is an “end of the line” measure and does not provide information about the value added in each stage of the process. It was also noted that, in this case, missing data is non-ignorable and analysing only non-missing data could be misleading and introduce bias.

There are alternative ways of measuring regional GDP. Wages and employment figures can be used as additional and relevant information.

Actions for authors:

2.5a	to model turnover behaviour by SIC and region
2.5b	to test and compare both models suggested in the paper
2.5c	to investigate the reasons for non-response in large businesses
2.5d	to improve the questionnaire if investigating the non-response indicates that this would be useful

Summary of actions and suggestions:

Section	Participant	Action
2a	Martin Brand	to pass comments on revisions paper back to National Accounts
2b	committee members	to send feedback about the new Publication Hub to Stephen Penneck
2.1	Craig Orchard Sarah Green Bronwen Coyle Jacqui Jones	to investigate non-uniform weighting of the model components and to conduct a sensitivity analysis to further explore other techniques
2.2	Owen Abbott James Brown	to have further discussion regarding the useful comments made by the panel to look at modelling estimates of moves on small sub-samples to model differences between areas' mover rates, or to account for them via stratification to investigate whether matching weights can be improved to investigate memory effects using the natural variation in response times of the CCS
2.3	Martin Brand	to treat both estimates as uncertain to use standard errors to measure precision to report uncertainty to measure and report revisions to explore why differences occur between annual and monthly or quarterly series
2.4	Salah Merad	to model turnover behaviour by SIC and region to test and compare both models suggested in the paper to investigate the reasons for non-response in large businesses to improve the questionnaire if investigating the non-response indicates that this would be useful
2.5	Dick Heasman Briony Eckstein Peter Youens	to investigate what could be done with unmatched records when using birth-linked data to understand better migration and improve census counts to investigate appropriate analysis methods following probabilistic linking to further investigate date of birth errors to seek wider research on pupil attainment using PLASC / NPD - see PLUG seminar series and work in Australia and Europe to investigate the quality implications of this work and to analyse matching errors to consider producing a dataset of anonymised records for analysis on an ad-hoc basis to be made available in parallel to more specific information to keep a register of records that have accrued from this linkage work