

Developments in measuring the burden placed on businesses responding to statistical surveys

Ria Sanderson¹, Denise Williams, Megan Pope (Office for National Statistics, UK)

Abstract

The UK Code of Practice for Official Statistics (UK Statistics Authority, 2009) specifies the need to report annually on the burden placed on respondents to surveys of businesses and households. Whereas information on the time taken for a household to respond to a questionnaire can be measured at the point of collection, it is more challenging to measure the time and hence the cost to businesses of responding to surveys used to compile official statistics. A traditional approach to such measurement of surveys conducted using paper questionnaires is to send a short review questionnaire to a sub-sample of businesses. This review questionnaire gathers information both on the time taken to respond to the main survey but also who in the business provides this information; this can then be used to estimate the financial costs to the business.

Such reviews ceased at ONS in 2012 and information on respondent burden was collected through a self-assessment tool used by survey managers to assess the quality of statistical outputs. This used information from previous reviews and knowledge of survey changes to estimate respondent burden. However, it provide difficult to collect high quality information on respondent burden without the data from these review surveys. Therefore, motivated by the aim to improve the measurement of the financial costs to businesses of responding to surveys, a shortened review process has been piloted. In an effort to balance the burden placed on respondents by this process and to make the process as efficient as possible, we have tested the use of statistical modelling to estimate respondent burden for surveys with similar characteristics. If successful, such an approach would reduce the number of review surveys that need to take place whilst still maintaining accurate measurements of respondent burden.

In this paper, we report on the pilot exercise carried out, including the methodology, results and conclusions of this work. We also consider the implications for the future measurement of respondent burden placed on businesses.

1. Introduction

One of the eight principles of the UK Code of Practice for Official Statistics (UK Statistics Authority, 2009) is “proportionate burden”, which places an obligation on producers of statistics to limit and assess the burden placed on respondents to statistical surveys. Under this principle, statistics producers are expected to “report annually the estimated costs (for example, on businesses, service providers, or the public) of responding to statistical surveys”. ONS reports figures for surveys to businesses and Local Authorities through the Online List of Government Statistical Surveys². The OLGSS collates information from across government on statistical surveys and includes information on their frequency, mode of collection and respondent burden. The OLGSS will be expanded in 2016 to also include information on surveys to households and individuals.

The agreed methodology for calculating the cost of complying with government surveys sets out the approach to measuring burden for surveys of households and for surveys of businesses. This

¹ Email: ria.sanderson@ons.gov.uk

² http://www.neighbourhood.statistics.gov.uk/HTMLDocs/OLGSS/OLGSS_interactive.html

methodology applies to all producers of official statistics in the UK and was established by the Government Statistical Service (GSS) Respondent Burden Task Force (2010). In the case of surveys of households or individuals, the required measurement is an estimate of the total time taken in responding to the survey. For surveys of businesses or Local Authorities, the required measurement is the financial cost to the business of responding to the survey. This is further broken down to consider the different components where a cost is experienced by a business responding to a survey and requires knowledge of:

- the time taken to respond to the survey
- the number of respondents to the survey
- any external costs incurred by the business in completing the survey, for example accountancy services used
- the time spent validating survey responses by re-contacting respondents
- the number of respondents contacted for validation

The total respondent burden, in terms of a financial cost, is calculated under this methodology as

$$\begin{aligned}
 \textit{burden} = & \\
 & \left[\{n_{\text{resp,main_surv}} \times \textit{med}(t_{\text{main_surv}})\} + \{n_{\text{val,main_surv}} \times \textit{med}(t_{\text{val}})\} \right] \times \textit{hourly_rate} \\
 & + \\
 & [\textit{prop}_{\text{external costs}} \times n_{\text{resp,main_surv}}] \times \textit{med}(\textit{external cost})
 \end{aligned}$$

where $n_{\text{resp,main_surv}}$ is the number of respondents to the main survey, $\textit{med}(t_{\text{main_surv}})$ is the median time³ taken to respond to the main survey, $n_{\text{val,main_surv}}$ is the number of respondents to the main survey who are re-contacted for the purposes of validation, $\textit{med}(t_{\text{val}})$ is the median time taken for this validation, $\textit{hourly_rate}$ is the appropriate hourly rate for the occupation of the respondent, $\textit{prop}_{\text{external costs}}$ is the proportion of businesses incurring additional costs (such as accountancy fees), and $\textit{med}(\textit{external cost})$ is the median external cost experienced by those businesses who incur additional costs.

Therefore, to measure respondent burden costs placed on businesses, a number of variables need to be collected. The variables related to validation can be collected from internal systems, however information is required from businesses on the time taken to complete the questionnaire and the level of any external costs where these are incurred.

2. History of the measurement of respondent burden for business surveys at ONS

A traditional approach to measuring the costs incurred by businesses when responding to official statistical surveys is to send a short review questionnaire to a sub-sample of businesses. This was the approach taken at ONS until 2012 under the programme of Triennial and Quinquennial reviews⁴. These were quality reviews that included the collection of information on respondent burden.

In 2012, changes were made to how quality reviews were conducted and the Triennial and Quinquennial reviews were replaced with the Quality, Methods and Harmonisation Tool. This is a

³ The choice of the median was made by the GSS Respondent Burden Task Force (2010). The rationale behind this choice was to limit the impact of outliers on estimates of respondent burden.

⁴ <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/triennial-reviews/index.html>

self-assessment tool that is completed by the managers of statistical outputs and also collected information on respondent burden. This used information from previous reviews and knowledge of survey changes to estimate overall respondent burden for business surveys. However, it proved difficult to collect high quality information on respondent burden without the data from the Triennial and Quinquennial review surveys.

The use of the Quality, Methods and Harmonisation Tool ceased in 2014 following feedback from statistical output managers which indicated that it did not meet their needs. This left a gap in the availability of up-to-date information on respondent burden for business surveys, which in time, is anticipated will be filled by the use of electronic methods for data collection. However, it is important to ensure that accurate up-to-date measures of respondent burden costs can be made until such a time that information is available electronically. The importance of this information is not only in terms of meeting the requirements of the Code of Practice; accurate measurements of respondent burden are invaluable in determining any reductions in burden made as a result of an increase in the use of administrative data or from a change in collection mode as well as in monitoring and managing the level of burden placed on respondents.

3. Methods for measuring respondent burden placed on businesses

To address the need for updated information, a project was established to investigate how the respondent burden could be measured for business surveys in an efficient and effective way. Ideas raised at an early stage identified possible approaches to measuring respondent burden. The pros and cons of these approaches are described in table 1.

Table 1. Possible approaches to measuring respondent burden for business surveys.

Method	Pros	Cons
1. Collect time information on survey questionnaire	<ul style="list-style-type: none"> No additional burden placed on respondents 	<ul style="list-style-type: none"> Only collects time, not external costs Has to be included as a voluntary question
2. Conduct a short review survey	<ul style="list-style-type: none"> Allows all required information to be collected 	<ul style="list-style-type: none"> Cost – both for the questionnaires and to carry out the reviews Could be perceived as additional burden to the respondent
3. Model burden from one survey to another	<ul style="list-style-type: none"> Information does not need to be collected for all surveys 	<ul style="list-style-type: none"> Need to identify similar surveys Relies on variables that are correlated with completion time Still requires the collection of information for some surveys

A pilot was established to investigate options (2) and (3). Option (1) was rejected at this stage as not all the required information can be collected in this way without evaluating and making changes to a large number of questionnaires. The aim of this pilot was twofold:

- To establish whether a shortened review process can be used to measure respondent burden
- To establish whether respondent burden can be modelled for similar surveys

These options were considered in some detail and resulted in a pilot of a new, much shortened, process for reviewing respondent burden; the results of which were used to test a statistical modelling approach to estimating respondent burden. An overview of the pilot and the results of the statistical modelling are presented below.

4. The Pilot

It was decided to adopt a similar approach to the Triennial and Quinquennial review process during the pilot. This meant that a separate, short, voluntary questionnaire was sent to a representative sub-sample of the main survey. The questionnaire collected information on the time taken to complete the main survey (including the time taken to compile the required information), the position in the business of the respondent (used to determine an appropriate hourly rate) and the level of any external costs incurred by the business. The process was streamlined significantly in comparison to the old review process to ensure that it was as efficient as possible. Methodological expertise was sought to design samples for the review that were representative of the main survey.

The process was managed centrally and the relevant survey managers were brought onboard at an early stage. The reviews were scheduled to be sent out approximately two days after the main survey. No changes were made to the response chasing strategy for the main survey; however, where a respondent had not returned the main survey questionnaire or the review questionnaire, when reminded of the need to complete the main survey they were also reminded of the opportunity to complete the review questionnaire. The pilot was carried out between October 2014 and December 2014.

4.1 Selecting the surveys

It was agreed to carry out the pilot on three monthly surveys. Monthly surveys were chosen as they presented regular opportunities for sending out questionnaires. The surveys were chosen on the basis of being relatively similar both in terms of theme and similarity in the number of questions; this was an important decision from the point of view of the statistical modelling. The surveys chosen were:

- Monthly Business Survey (Retail Sales Index)
- Monthly Wages and Salaries Survey
- Vacancies Survey

These are all surveys on the economy that have a small number of questions. The Monthly Business Survey (Retail Sales Index) collects information on monthly retail turnover; the Monthly Wages and Salaries Survey collects information on salaries and is used to estimate Average Weekly Earnings; the Vacancies Survey collects information on the number of vacancies in the economy. On the basis of past data, these three surveys were expected to show similar completion times. This is indicated in figure 1, which shows the median completion times based on past data for a selection of monthly and quarterly business surveys. They also all use the same sampling frame, the Inter-Departmental Business Register (IDBR) which means that the same potential explanatory variables are available for all three surveys.

Figure 1. Median completion times for a selection of monthly and quarterly surveys. The three surveys selected for the pilot are shown as solid circles.



4.2 Sample sizes and response rates

The sample sizes and response rates for the pilot are shown in table 2. Note the differing response rates between the three review surveys. The reason for this is not clear but it may be due to the dates when the surveys were dispatched; both RSI and Vacancies were dispatched in December whereas MWSS was dispatched in November.

Table 2. Sample sizes and response rates for the review surveys.

	Monthly Business Survey (Retail Sales Index)	Monthly Wages and Salaries Survey	Vacancies Survey
Main survey sample size	4,959	9,295	6,030
Review survey sample size	501	773	600
Review survey response rate	41%	71%	47%

5. Statistical modelling of respondent burden

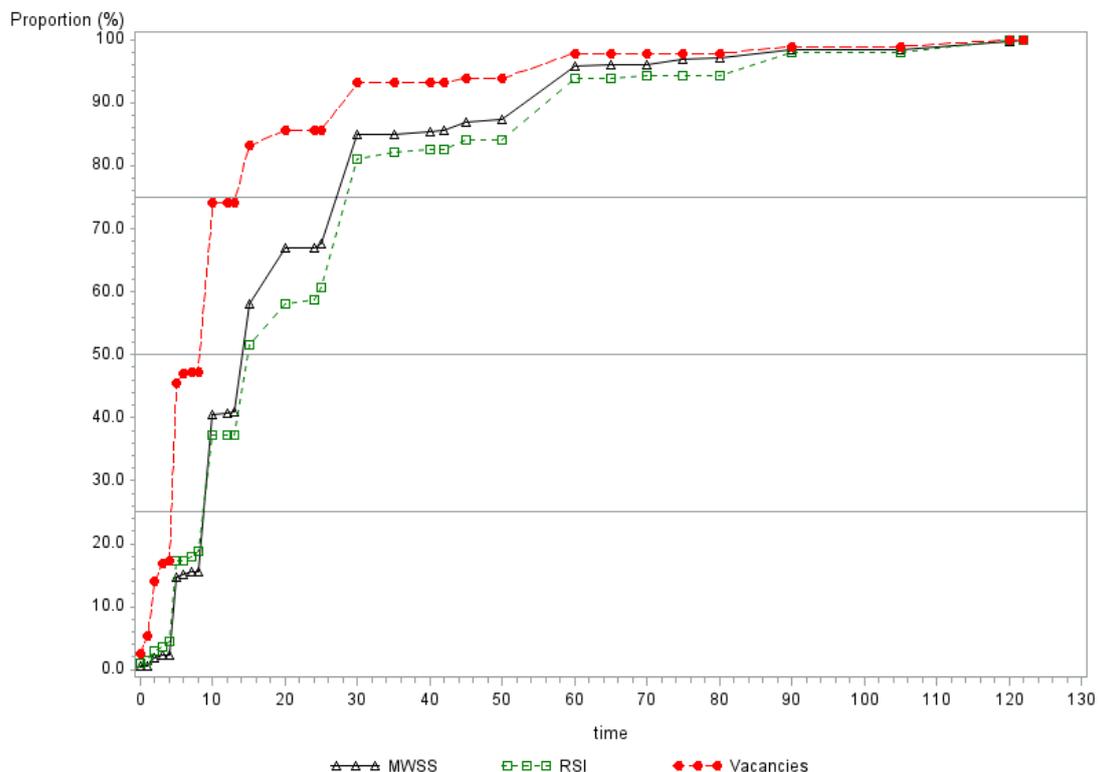
One of the ideas put forwards at an early stage was to consider whether information on respondent burden from one business survey could be used to estimate the burden faced by respondents to a different business survey. The motivation for this approach was to limit the costs of carrying out

reviews of all surveys, but also to limit the burden placed on respondents by the additional review activity.

The sampling frame (the IDBR) was used to provide supplementary information for the statistical modelling. Exploratory analysis was conducted to examine whether the time taken to complete the questionnaire correlated with any of the available auxiliary variables. This showed that there is little to no correlation between the time taken to complete the questionnaire and employment size or turnover.

The returned data showed that it is common for respondents to report the time taken to complete the questionnaire to the nearest five or ten minutes as the returned data were clustered around these points. The distributions of completion times across the three surveys differs as indicated in figure 2, which shows the cumulative proportion of businesses completing the survey as time increases. This figure also highlights the steps in the recorded data. This hides some of the true variation in the data. These non-sampling errors could be a result of the delay between receiving the main questionnaire and the review questionnaire, meaning that the respondent could not recall the true completion time. They may also occur if a different person completes the review questionnaire in comparison to the main survey questionnaire or it may result from rounding on the part of the respondent.

Figure 2. Cumulative proportion of businesses as a function of completion time. The horizontal lines are at proportions of 25%, 50%, and 75%. Only times shorter than 150 minutes are shown.



Despite the weak correlation with the available auxiliary variables, a linear regression model was fitted to the data to try to predict completion time. The aim was to fit a model to one survey and then use this to predict completion times for another survey. Models were fitted using employment and the number of questions in the survey as explanatory variables. The number of questions was chosen as a possible explanatory variable under the assumption that the time to complete a single

question of similar complexity may be fairly stable across surveys. As expected given the poor correlations, the models did not perform well. The rounding of the data also masked some of the true variation, which made fitting a successful model more challenging. The analysis was extended to combine data from two surveys to fit the model. This led to an improvement in the R^2 value, but this was still poor (between 0.16 and 0.19). As the pilot concentrated on three surveys that were chosen due to their similarities, it is highly unlikely that this approach would therefore be applicable to other surveys. The analysis showed therefore that trying to model respondent burden from one survey to another, in terms of the completion time, was unviable.

6. Next steps

The pilot showed that the streamlined process for measuring respondent burden worked effectively. The streamlined process will be used over the coming months to provide updated measurements of respondent burden for a number of business surveys. The information collected will help to inform future savings in respondent burden that are expected through the use of electronic methods of data collection and through the increasing use of administrative data.

7. Conclusions

There is an obligation under the UK Code of Practice for Official Statistics (UK Statistics Authority, 2009) to report on the burden placed on respondents to government surveys. Whereas this information is relatively easily collected for surveys of households and individuals, where these surveys are typically administered by an interviewer using a computer, it is more difficult to collect the relevant information from businesses who receive paper questionnaires. The GSS methodology for calculating respondent burden for business surveys requires information on the completion times of the survey, external costs incurred by the business and information on the time taking validating the responses through re-contacting businesses where required. The information on validation can be collected internally, but the remaining information still needs to be collected directly from businesses.

In time, this information could be collected through electronic means as surveys are moved from paper questionnaires to electronic methods of data collection. However, due to changes and improvements made to quality review methods at ONS, there is currently a need to update the information collected on the respondent burden placed on businesses. A pilot of a process of sub-sampling respondents and sending them a short questionnaire asking for information on completion times and external costs has been piloted and worked well. The data collected were used to try to model respondent burden from one survey to another by formulating a statistical model of completion time using auxiliary variables from the sampling frame. However, due to weak correlations between completion time and possible explanatory variables and the fact that the way the data were reported masked some of their true variation, the models had very poor predictive power. This approach has therefore been discarded. Over the coming months, the streamlined process of sub-sampling respondents will be used, in combination with existing data on completion times collected via some survey questionnaires, to update estimates of respondent burden.

References

GSS Respondent Burden Task Force, 2010, "Guidance on Calculating Compliance Costs", available from: <http://www.ons.gov.uk/ons/guide-method/best-practice/gss-best-practice/gss->

[respondent-burden-guidance/gss-respondent-burden-task-force-guidance/guidance-on-calculating-compliance-costs.doc](#)

UK Statistics Authority, 2009, Code of Practice for Official Statistics, available from:

[http://www.statisticsauthority.gov.uk/assessment/code-of-practice/code-of-practice-for-official-statistics.pdf](#)