Trajectory Modelling of Longitudinal Non-response in Business Surveys

Methodology Development Unit, Statistics New Zealand

Kate Smaill, February 2011 kate.smaill@stats.govt.nz

Abstract

Survey quality is an important goal for National Statistics Offices (NSOs) and non-response negatively affects this quality. Accordingly, it is important to understand the determinants and patterns of non-response, with the aim of providing recommendations as to how to reduce or mitigate non-response. This study looks to better categorise sampled firms based on their previous response behaviour, using trajectory modelling. Trajectory modelling is a tool for distinguishing patterns of behaviour over time, with dual trajectory modelling allowing us to analyse how outcomes in one time period are related to distinct outcomes in time two (Jones and Nagin, 2007). Using dual trajectory modelling to examine longitudinal response patterns is a new application of this model.

Statistics New Zealand's Retail Trade Survey is used as a test case and the model is fitted across the response history 2003 to 2007.We were able to identify three response groups, the good-responders, the group whose likelihood of response improves over time and the non-respondents. These three groups then form the basis for four groups we identified in time two. This then informs some tentative recommendations as how to best target sample members, based on previous response behaviour.

Introduction

Survey quality is an important goal for National Statistics Offices (NSOs) and non-response negatively affects this quality. While Statistics New Zealand frequently meets their response rate targets, non-response still occurs. Accordingly, non-response is something we need to understand to reduce non-response and to understand its impact on survey quality. Non-response is of concern as it reduces the effective survey sample size, thereby resulting in a loss of statistical power (Rogelberg, 2006). Non-response can also affect the perceived credibility of the data (Rogelberg et al, 2000), and therefore the perceived credibility of the collection agency. The most significant problem, though, arising from non-response is bias, which occurs when respondents systematically differ from non-respondents. It is for these reasons that survey researchers should, and are, concerned about non-response.

This paper summarises the results for the trajectory-modelling component of a study into the determinants of non-response in Statistics New Zealand business surveys. Given the panel design of these business surveys, firms are frequently in a sample for a long period. Also, given the size of the New Zealand economy, a firm may be in several surveys at the same time. Therefore, the decision to respond to a current survey is not independent of response decisions to other surveys. This highlights the need to examine response patterns over time.

To examine firms' response histories the following research questions were identified.

- Does early response behaviour predict future response behaviour?

- Can we identify chronic non-respondents and how quickly can we identify them?
- Can we make any recommendations for reducing and dealing with non-response in business surveys?

Dual trajectory modelling techniques allow to answer these questions, as we are able to examine two distinct, but related outcomes which are occurring are two different points in time (Jones and Nagin, 2007). In this application we are able to examine the relationship between response behaviour in time one and time two. If problematic new firms in the sample can be identified early, we can better prioritise our non-response interventions and follow-up. This allows us to maximise response rates, given the limited resources available for non-response follow-up.

Literature Review

Despite the importance of understanding non-response and despite the amount of literature regarding non-response in individual and household surveys, the amount of literature regarding methodology, non-response and data quality in business surveys is limited (Edwards and Cantor, 1991; Lynn and Sala, 2004).

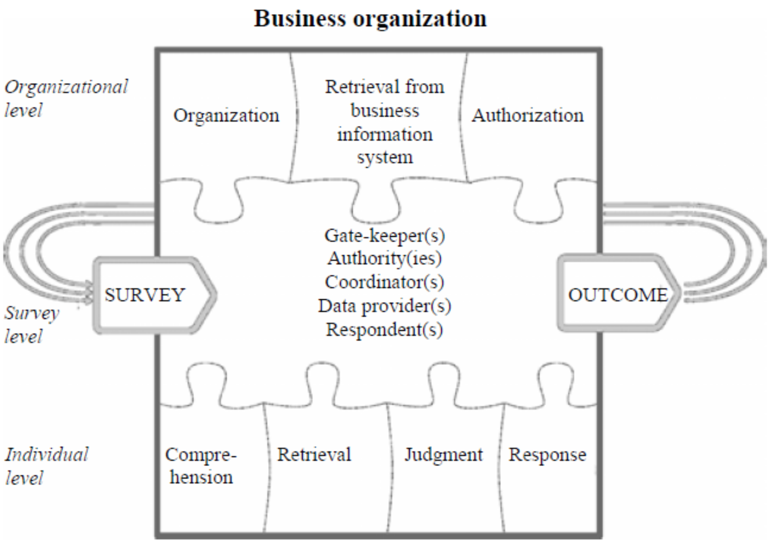
Two internal Statistics New Zealand studies have examined business survey response using the same dataset as our analysis. Fu (2007) found that larger firms are more likely to respond, even before prompting; and businesses that had not expanding or contracted in the last twelve months were found to be more likely to respond overall. Davis & Pihama (2007) then developed logistic regression models to test some of the bivariate relationships that had been examined in Fu (2007). Davis & Pihama (2007), when analysing non-response in the Statistics New Zealand Annual Enterprise Survey (AES), found that larger firms, key firms, multi-establishments, and firms that have a stable Rolling Mean Employment (RME), from the previous year were more likely to respond. Yet, the strongest predictor of response to the current survey was whether a firm had responded to the previous survey.

A number of authors have also developed models regarding business survey response, in order to inform and examine the determinants of response. For example, Tomaskovic-Devey, Leiter and Thompson (1994) developed a theoretical model of business non-response in the context of organisational behaviour and structure. They consider the individual survey respondent's authority to respond to the survey, the firm's capacity to respond through the systems they have in place and the motivation to respond of the firm and the individual. This model then informed which organisational factors were hypothesised to affect the response event. Overall, Tomaskovic-Devey et al (1994:453) concluded that "organizational characteristics do predict the probability of survey response", with small firms that frequently interact with the external environment that they operate within being the most likely to respond.

The limitation of this model, and other response models, is that it just considers one survey period. Additionally, while there are response models that do consider the temporal nature of surveys; these only consider time in terms of the cognitive stages in the survey response process from receiving a survey to participating in the survey (Edwards and Cantor, 1991; Tourangeau, 1984; Willimack and Nichols, 2001).

There is little literature that considers the temporal nature of business surveys across multiple survey periods, despite the repetitive nature of business surveys, but one recent response model by Bavdaz (2010: 83) "allows for the possibility of conceptualizing the response process over several implementations of a survey or over several surveys". This model of non-response referred to as the Multi-dimensional Integral Business Survey Response Model (MIBSR), and is shown below in Figure 1.

Figure 1 Multi- dimensional Integral Business Survey Response Model



Source: Bavdaz (2010)

This model, which draws on other response models (Tourangeau, 1984; Tomaskovic-Devey et al, 1994; Willimack et al, 2002), is the first model to consider the temporal nature of the response process, indicated by the arrows in Figure 1.

Beyond the literature regarding conceptual models of non-response, and determinants of non-response, there is also literature about types of non-response. Rogelberg and colleagues have extensively studied unit non-response and have developed categories of non-response (Rolgelberg et al, 2003). Having examined different reasons for non-response Rogelberg et al (2003) propose that there are two types of non-respondents: passive non-respondents and active non-respondents. While active non-respondents make a conscious choice, passive non-response is due to irregular or random circumstances, such as illness. Rogelberg et al (2003) found the majority of non-respondents are passive non-respondents and only about 15% of the total population were active non-respondents.

This study looks to contribute to the limited literature surrounding business survey non-response, particularly longitudinal patterns of response. Furthermore, it provides a New Zealand context to this work which is perhaps more applicable to other small economies than the work done in the United States, Canada and the United Kingdom.

Data and Methodology

The main data source for this paper is the Statistics New Zealand Respondent Management System (RMS). The RMS is constructed by the Statistics New Zealand Integrated Data Collection team. This dataset incorporates data from all fielded business surveys from 2002 onwards. This data is matched for each enterprise to data from the Statistics New Zealand Business Frame each year. The Business Frame is a comprehensive list of New Zealand businesses and includes limited information about each firm. This dataset therefore contains information regarding both respondents and non-respondents to Statistics New Zealand's business surveys.

This dataset is the only dataset, to our knowledge, which provides the opportunity for a NSO to examine response patterns longitudinally within, and across surveys, in the context of firm characteristics. This paper focuses on applying trajectory models to better understand non-response in the Retail Trade Survey (RTS). The monthly RTS postal sample consists of about 3,500 businesses. It covers firms in the Retail trade, Accommodation, Cafe and Restaurants and Personal Services industries.

The variable of interest in this study is whether a firm responded to a specific postal business survey. We investigated bivariate relationships between the variable of interest and a number of organisational and survey factors. One of the survey factors was whether the firm has responded to the survey in the previous period. We then used to logistic regression to test which variables where significant when controlling for the other variables. The logistic model was performed using the SAS® Proc Survey Logistic, as we were drawing from sample surveys. For brevity, these logistic models are not detailed here.

Then we began modelling response patterns over time to see if we could categorise enterprises based on their previous response history. To examine response patterns we required a sample we could follow from their first time in a survey. For this we used a subsample of the Retail Trade Survey which was redesigned in 2003 and introduced just over 2,500 firms into the sample. These firms provided us with a sizeable sample that we could follow from their first time in the RTS, in August 2003.

To examine response patterns over time we undertook two approaches. First, we used trajectory modelling. Trajectory models are a form of mixture models that involve analysing unobserved heterogeneity in a population over time to identify distinct subpopulations with differing patterns of, in this case, response and non-response. We undertook trajectory modelling using a SAS® add-on Proc Traj[1] (Jones, 2010). The trajectory model we developed was a dual model, which models the first six months of response, with a second model being fitting over the following thirty-three months. The dual modelling approach allows us to see the relationship between initial response in the first six months and whether this is a strong predictor of future patterns of response. This approach assumes the two time periods are not

[1] SAS® Proc Traj is available for download from http://www.andrew.cmu.edu/user/bjones

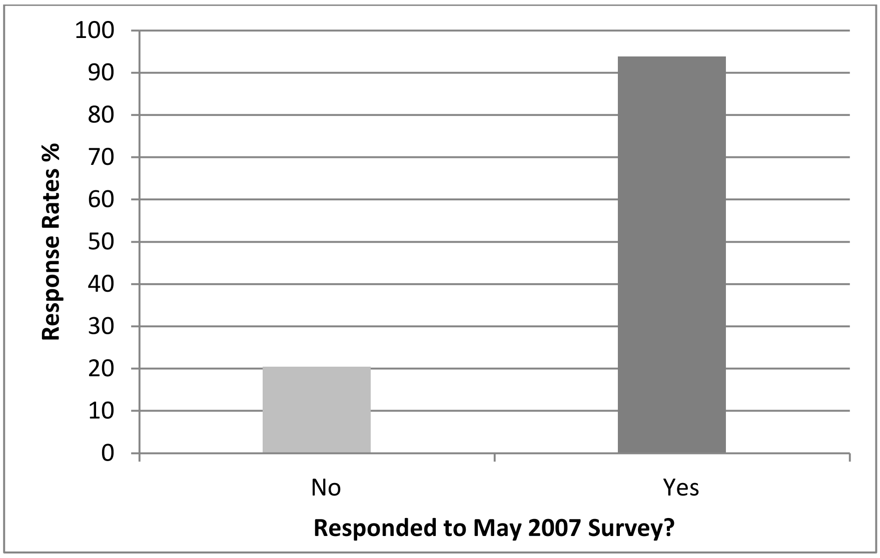
independent, and it outputs information on the conditional and joint probabilities of group membership. For a more detailed explanation of dual-trajectory modelling see Jones and Nagin (2007).

The second approach was to develop a 'risk-group' measure that approximates the trajectory model results, while being easier to operationalise. This measure simply categorises respondents into three respondent categories based on their response behaviour in the first five survey periods. The first group consists of firms who have responded to none of the first five surveys, while the second group consists of firms who have responded to some of the first five, and finally the third group is firms who have responded to all of the first five surveys.

Results

Our logistic modelling of business survey response found that of the variables we considered only previous response (whether the business had responded to the survey in the previous survey) was a significant and consistent predictor of response. For example, in Figure 2, over 90% of firms that responded to the RTS in May 2007 responded in June, compared to 20% who didn't respond in May.

Figure 2 Retail Trade Survey 2007 June Response Rates by 2007 May Response Status



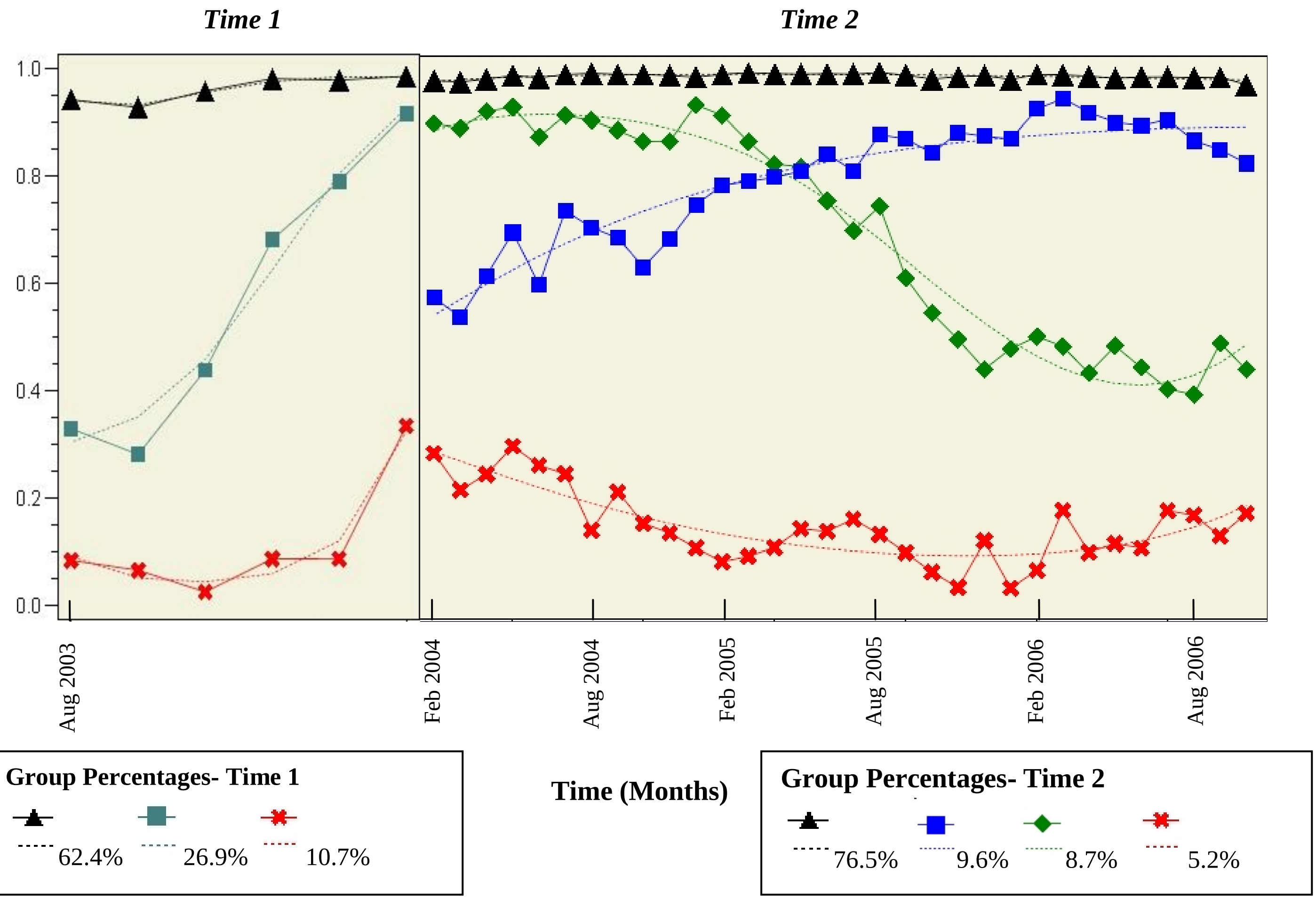
For brevity these results are not further discussed here. Yet, given the significance of previous response in the logistic models, it is important to understand response patterns over time and to understand how these response patterns can be used to predict future behaviour and if we can establish who is likely to fall into which response behaviour group, and how quickly we can identify this. Here we firstly outline the results from the trajectory modelling, and then we show how well our risk-group measure matches and predicts the trajectory modelling groups.

Figure 3 shows the results from our dual trajectory model. Time 1 of the model shows month 1 (August 2003) to month 6 (January 2004); then time 2 shows month 7 (February 2005) to month 38 (October 2006). The vertical axis represents the estimates probability of response. In the first time period we found three distinct response trajectory groups. The largest group

(62.4%) consists of the good respondents, shown at the top of the left side of Figure 3. The second largest group (26.9%) is what we refer to as the ‘learners’. This group’s estimated probability of responding steadily increases over the first six months that they are in the sample. The third group (10.7%) consists of the chronic non-responders. These chronic non-respondents consistently have a low estimated probability of responding in the first six months.

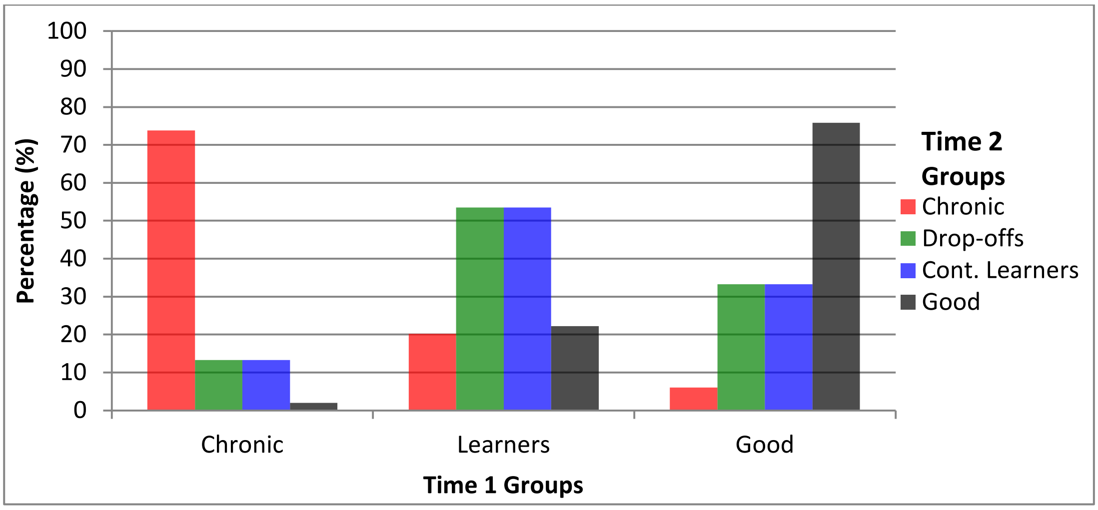
In the second time period, covering month 7 to month 38, four distinct respondent groups were identified. Ideally, we would see a high degree of consistency between time 1 and time 2, indicating that early response behaviour does predict future response behaviour. Consistent with time 1 we see the largest respondent group is the good responders (76.5%). Additionally, we see a consistency between the ‘chronic’ non-respondent group, but this group now only consists of 5.2% of the population. While in time 1 we found one ‘learners’ group, we find two groups who are in between the chronic and good respondents groups. We refer to the group which continues to improve over months 7 to 38 as the ‘continued learner’ group (9.6%). The other group has been coined the ‘drop-off’ group (8.7%) as over time their probability of responding declines.

Figure 3 Dual Trajectory Model for a Sample of Retail Trade Survey Firms from (1) August 2003 to January 2004 to (2) February 2004 to September 2006



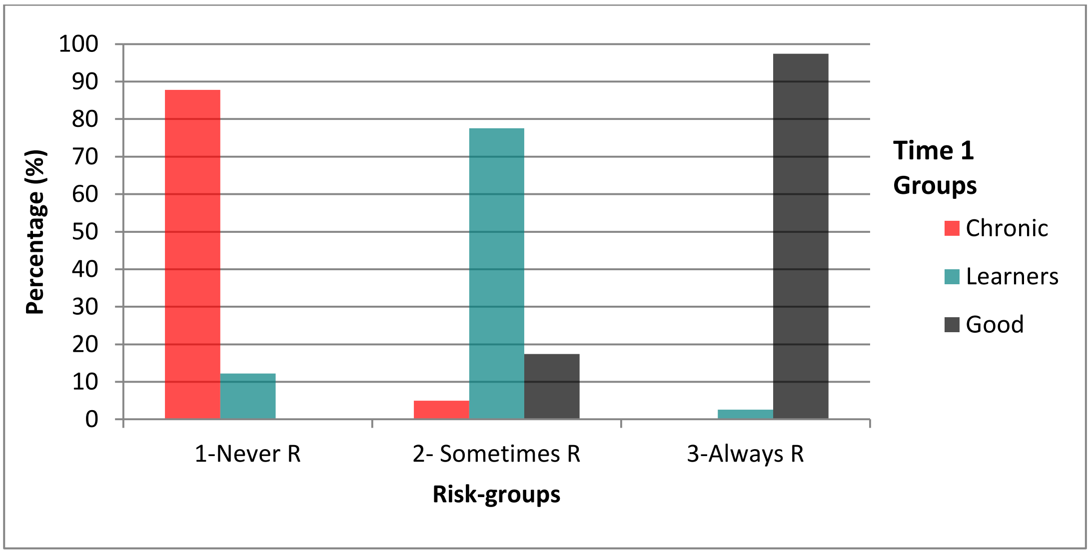
To understand if the groups are consistent between time 1 and time 2, in Figure 3, we need to examine where firms in time 1 transit to in time 2. Figure 4 shows the probability of a firm belonging to a particular time 2 group, given its group membership in time 1. Figure 4 shows that the majority of chronic non-responders and good responders remain chronic and good, respectively, in time 2.

Figure 4 Probability of being in a Trajectory Time 1 Group Conditional on the Trajectory Time 2 Groups



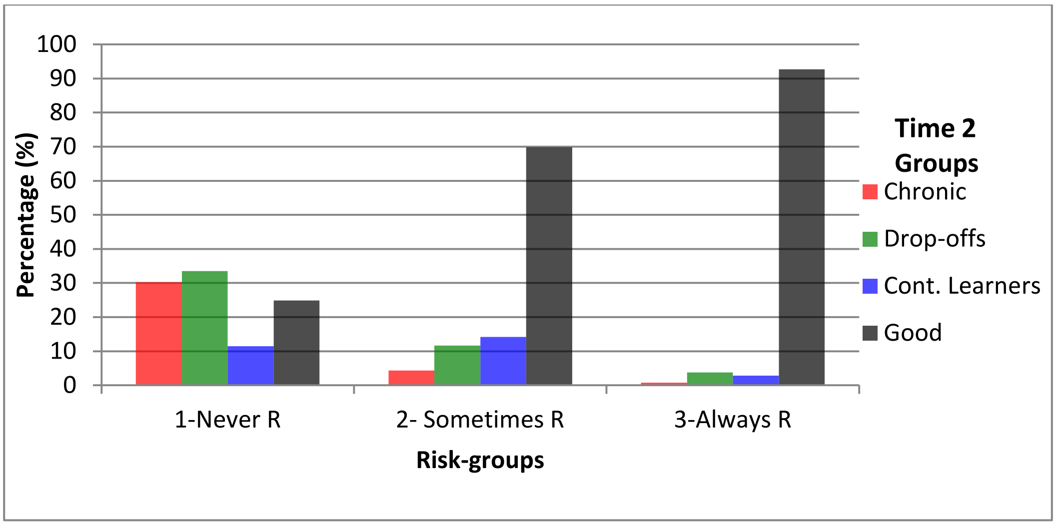
In order to examine how our own 'risk-group' variable matches with the groups found through trajectory modelling we are able to compare the groups that are categorised. Firstly, we compare the risk-groups with the time 1 trajectory groups, shown in Figure 5, to see if the risk-group measure is able to serve as a simple proxy of the time 1 trajectory groups. There is a high degree of consistency between the time 1 trajectory groups and the risk groups.

Figure 5 Composition of the Risk- groups by the Time 1 Trajectory Groups



Then we compare the risk-groups variable with the time 1 trajectory groups, shown in Figure 6. This allows us to see if early response behaviour, as shown by the risk-group measure, is a good indicator of future behaviour, as shown by the time 2 trajectory groups. In Figure 6 we see a degree of consistency between the risk-groups and the time 2 groups. Chronic non-responders are largely from the 'never respond' risk-group, while good responders either, mainly, always responded or sometimes responded to the first five surveys. There is a degree of variation, though, between the drop-offs and continued learners time 2 groups and their risk-group category.

Figure 6 Comparison of the Time 2 Trajectory Groups with the Risk-groups Measure



Discussion

Survey quality is an important goal for National Statistics Offices (NSOs), and non-response negatively affects this quality. Therefore we need to understand what drives non-response to more effectively target non-respondents. Through the examination of response rates and logistic modelling it was found that previous response was the largest predictor of non-response. Given how significant previous response was as a predictor of response, we then focused our research on examining response patterns over time. Through trajectory modelling we found three distinguishable groups in the first few months firms were in the survey. These three groups were the good respondents, the learners and the chronic non-respondents. In the second part of the dual trajectory model we found four distinguishable respondent groups. These were the good respondent group, the continued learners group, the drop-off group and the chronic group.

These groups do match what has been found in the literature, if we draw parallels with the chronic group and active non-respondents. That is firms who actively choose not to participate, and the learners group (later the continued learners and the drop-offs groups) can be thought of in the context of passive non-respondents, that is firms who may have intended to respond but due to other circumstances they did not. Rogelberg et al (2003) found the majority of non-respondents are passive non-respondents and only about 15% of the total population were active non-respondents, this is consistent with our findings, were more people were learners than there were chronic non-respondents, although we only found 10% (and later 5%) of the population to be chronic non-respondents.

In terms of how early response behaviour is a good predictor of future response behaviour we found that there was a high degree of consistency between the time 1 trajectory groups and the time 2 trajectory groups, although a number of firms do shift groups. Furthermore, using our own risk-group measure as a simple proxy for the time 1 groups we found that even by just considering whether a firm responded to all, some, or none of the first five surveys they were in provided a good indication of future response behaviour. Given the consistency between early response behaviour and future response behaviour it therefore enables us to provide a potential method for prioritising non-respondents for follow-up.

Based on these findings we can draw some tentative recommendations about how to treat survey recipients. From the groups in the trajectory model statistical organisations should target the learner group early on to ensure their likelihood of response does increase over time. Similarly, the chronic non-responders should be targeted early before the behaviour becomes engrained. Past a certain point though we may have to accept that 5% of the sample will be a waste of resources. In the second time period, we also identify a 'drop off' group which should be a particular concern, as we must maintain a willing supply of information. As we unable to distinguish between the drop-offs and continued learners we must continue to focus on this group over time, so that if a firm starts to miss a couple survey periods we have to intervene to try and prevent them from falling into the drop-offs group.

Also, statistical organisations must continue with their current interventions. While it is tempting to conclude from Figure 2 that good respondents do not require follow-up activity, as this study did not consider the impact of response interventions, we cannot draw this conclusion. There is potential, though, to consider reducing the level of contact with these respondents over time and measuring the impact. Another limitation of this work is that is only considers a sample of the Retail Trade Survey. This work needs to be extended to other sample within the Retail Trade Survey and across other surveys, to see if results are consistent. Of particular note is that we need to understand how trajectory models can be applied to quarterly surveys and produce useful results in a timely manner. A further extension of this work is developing a model that looks at response across surveys, therefore enabling us to distinguish between firm-effects and survey-specific effects.

Conclusion

It is important to understand non-response in business surveys to gain a better appreciation of survey quality. This paper looked to develop a new application for trajectory modelling by examining non-response patterns over time. Through trajectory modelling we have been able to show there is a strong relationship between early and future response behaviour and therefore this provides the opportunity to target sample member based on their early response behaviour. This provides a good foundation for considering future non-respondents interventions. This work has therefore contributed to the existing literature by extending, and applying, the model of business survey non-response by considering the temporal nature of the business survey response decision, it has also outlined practical measures and approaches which could be utilised in the future.

References

Bavdaz, M, (2010). The multidimensional integral business survey response model. Survey Methodology, 36, (1), 81-93.

Davis, W R, & Pihama, N, (2009, July). Survey response as organisational behaviour: an analysis of the annual enterprise survey, 2003- 2007. Paper presented at New Zealand Association of Economists conference, Wellington, New Zealand.

Edwards, W S, & Cantor, D, (1991). Toward a Response Measurement at the Bureau of Labor Statistics. In P B Biemer, R M Groves, L E Lyberg, N A Mathiowetz & S Sudman (Eds), Measurement Errors in Surveys. New York: John Wiley & Sons.

Fu, N, (2007). Determinants of Response in Business Surveys, Statistics New Zealand, Wellington.

Jones, B L (2010). *Traj: Group based modeling of longitudinal data*. Available from www. www.andrew.cmu.edu/user/bjones/index.htm.

Jones, B L & Nagin, D S (2007). Advances in group-based trajectory modeling and a SAS procedure for estimating them. Sociological Methods & Research, 35, 542- 571.

Lynn, P & Sala, E (2004). The contact and response process in business surveys: lessons from a multimode survey of employers in the UK. Economic & Social Research Council Research Methods Programme Working Paper Series, no. 10.

Rogelberg, S G (2006). Understanding nonresponse and facilitating response to organizational surveys. In A I Kraut (ed) Getting action from organizational surveys: new concepts, technologies, and applications (312-325). San Francisco: Jossey-Bass.

Rogelberg, S G, Luong, A, Sederburg, M E, & Cristol, D S (2000). Employee attitude surveys: examining the attitudes of noncompliant employees. Journal of Applied Psychology 85(2), 284: 293.

Rogelberg, S G, Conway, J M, Sederberg, M E, Spitzmiller, C, Aziz, S, Knight, W E (2003). Profiling active and passive nonrespondents to an organizational survey. Journal of Applied Psychology 88,6, 1104-1114.

Tomaskovic-Devey, D, Leiter, J, and Thompson, S, (1994). Organizational survey nonresponse. Administrative Science Quarterly, 39, 3, 439-457.

Tourangeau, R, (1984). Cognitive sciences and survey methods. In T Janine, M Straf, J Tanur, R Tourangeau (Eds) Cognitive aspects of survey methodology: building a bridge between disciplines. Washington, D.C.: National Academy of Science.

Willimack, D, & Nichols, E, (2001). Building an alternative response process model for business surveys. Proceedings of the Annual Meeting of the American Statistical Association.

Willimack, D, Nichols, E, & Sudman, S (2002). Understanding unit and item nonresponse in business surveys. In R M Groves, D A Dillman, J L Eltinge , R J A Little (Eds) Survey non-response. New York: Wiley.