WORKING PAPER

ITLS-WP-15-15

Collecting longitudinal data from freight operators: survey design and implementation

By
Richard B. Ellison, Stephen P. Greaves and David A. Hensher

August 2015

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in

Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-15-15
TITLE:	**Collecting longitudinal data from freight operators: survey design and implementation**
ABSTRACT:	Freight transport research has generally been limited by a lack of data of the breadth and quality available for passenger transport, particularly in terms of behavioural data. Although there are a number of reasons for this lack of data, the challenges of collecting freight data including the expense, participant burden and confidentiality issues are amongst the most significant. Although some improvements in technology and survey design have allowed for the collection of more behavioural freight data more improvements would be beneficial. This paper discusses the survey design and implementation of a survey intended to collect longitudinal behavioural data on the responses of freight transport firms to the introduction of environmental policies. The design of the survey is centred around a hypothetical scenario where respondents are asked how they would complete a given freight task within common constraints including time windows and delivery requirements. One of the key components of the survey design is a dynamic component that is intended to simulate the changing business environment to which firms are required to adapt. This paper also looks at the participant burden involved in completing the survey and compares how this differs depending on how respondents completed the survey.
KEY WORDS:	***Freight transport, longitudinal data, simulation, dynamic responses.***
AUTHORS:	**Ellison, Greaves and Hensher**
CONTACT:	INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (C13) The Australian Key Centre in Transport and Logistics Management The University of Sydney NSW 2006 Australia Telephone: +612 9114 1824 E-mail: business.itlsinfo@sydney.edu.au Internet: http://sydney.edu.au/business/itls
DATE:	August 2015

1 Introduction

The limitations of most existing sources of freight data for use in behavioural models has been repeatedly acknowledged to be one of the primary reasons for the relatively slow advance in freight modelling compared to passenger transport. However, despite improvements in technology enabling the collection of more disaggregate data on freight movements, data on behavioural decisions of freight operators remains a challenge due to the expense, participant burden, and confidentiality concerns. Some advances in overcoming these challenges have been made by Shinghal and Fowkes (2002) and Hensher et al. (2007b) but these methods do not explicitly account for how decisions change (or evolve) over time, something that is crucial in evaluating policies. For this reason, a method that facilitates the collection of data that is suitable for use in behavioural models for freight and can be used to collect 'longitudinal' (defined as repeated measures over time) data is needed. This paper details the design and implementation of a survey that has been developed to collect longitudinal data on freight operators' decisions for a study on the behavioural responses of freight operators to emission mitigation policies.

2 Background and motivation

The primary analytical technique employed for investigating behavioural responses of freight operators to emission mitigation policies was latent curve models (a form of structural equation models). Modelling of behavioural decisions using latent curve models requires a dataset in which data on decisions are collected at repeated intervals along with data on other likely covariates such as the type and size of firm.

The design of this survey was driven by the requirement to collect longitudinal data from freight firms in situations in which it would be impractical (if not impossible) to collect detailed data on decision making. Many of the existing methods for collecting data on freight operators' decisions rely on stated choice (SC) experiments. Although most widely used in freight for mode choice models, stated choice experiments have also been used to investigate freight operators' preferences for travel times, travel time variability and waiting time and route choice (e.g., Puckett et al., 2007; de Jong et al., 2004). Variations of standard SC experiments, such as adaptive SC experiments, have also been used to study freight operators' decisions (Fowkes and Shinghal, 2002). One such example is The Leeds Adaptive Stated Preference (LASP) software that allows the alternatives shown to respondents in a stated choice experiment to change depending on their choices in previous choice-sets. It has been used in a number of studies on freight including for mode choice and journey times (Bolis and Maggi, 2003; Fowkes, 2007; Shinghal and Fowkes, 2002). Similarly, Hensher et al. developed the interactive agency choice experiment (IACE) that allows for interdependent decisions with multiple agents (Brewer and Hensher, 2000; Hensher et al., 2007a). However, for assessing policy-induced changes to freight operations SC experiments do not provide sufficient detail on the adaptation strategies adopted by freight operators that may include responses not covered by the attributes in the experiment and may occur over different lengths of time. Similar limitations apply to data collected using trip diaries which generally collect data only for a single day for each firm and as a result are unable to capture changes over time (Stefan et al., 2005). These limitations are understandable given the expense and difficulties associated with collecting disaggregate freight data, but they remain a key limitation of existing methods.

To investigate the adaptation strategies adopted by freight operators in response to government policies, it is clear that a more practical and less burdensome method to collect time-series data on freight operations was needed. This method needs to allow for the collection of data combining the benefits of stated preference (SP) data and revealed preference (RP) data as well as reducing the work required by respondents.

3 Requirements and design considerations

The primary requirement for the survey design was to collect longitudinal data from freight firms on the decisions required to undertake the freight task both prior to and following the introduction of a government policy imposing additional constraints or costs on their operations. These decisions involve the interaction of a number of complex decisions occurring at various points in time and with different time horizons. As such, the survey instrument needed to be able to collect data on likely decisions of freight operators in a way that is understandable to respondents and straightforward to complete. However, this needs to be done while allowing for as large a variation in allowable decisions as possible without masking time-varying and complementary decisions.

The longitudinal data which the survey had to collect includes both the (multiple) decisions being made at each point in time but also the values of the different aspects of the decision being used to evaluate the alternatives.

The target sample of managers in freight firms with responsibility for decision-making in operations meant the survey design had to take into account the time-pressures many of the managers face as well as the likelihood of interruptions while they were completing the survey. This led to two important considerations for the survey design. First, the time required to complete the whole survey had to be at most 20 minutes including the time to read any instructions. This limited both the number of questions that could be asked and the manner in which the longitudinal data was collected to a method that allowed for quick (but reasonable) responses. Second, it had to accommodate respondents starting the survey and then continuing later in the day (or another day) while ensuring respondents were able to refer back to their previous decisions as well as any instructions they were previously shown. Since it was thought most respondents would complete the survey while they had some time during the work day it was expected that some respondents would be interrupted and would (reasonably) stop the survey until they had more time to complete it. For this reason it was crucial that respondents were able to refresh their memory about the scenario being presented and how they had answered the previous questions. Further, keeping the survey quite short would hopefully reduce the number of times respondents would be interrupted thereby reducing the drop-out rate.

A further requirement for the survey design was that it limited the amount of confidential information requested from respondents regarding their operations. This was because some firms consider their operations to provide them with a competitive advantage and so are reluctant to divulge specific details. This is particularly true for very large firms of which the Australian firms participating in the study are local subsidiaries. Limiting confidential data requested also had the potential benefit of reducing the reluctance of the person (or persons) actually completing the survey on behalf of the firm to participate due to a fear of providing information the firm considers to be confidential.

4 Survey design

The survey design had a broadly similar structure to that of many conventional stated preference surveys in that background questions were first asked, followed by a practice screen (or 'game'), and the main survey and completed with a small number of supplementary questions.

The initial background questions were designed to be straightforward and easy for a manager to complete and requiring only approximate answers to account for firms with changing number of vehicles or drivers based on demand. The questions were based on those asked by Hensher et al. (2013) with some modifications to account for the different focus of the study. Respondents were first asked for information on their position in their firm as well as the length of their experience in the industry and at their firm. The information requested about the firm included the number of drivers who worked or were regularly contracted to the firm, details about the vehicles used and the type of deliveries generally made by the firm. In particular, respondents were asked to enter the approximate number of vehicles of each class and emissions standard used for all operations and those used for urban deliveries (the focus of the study). Respondents were asked to select one of four categories of deliveries comprising of large bulk and pallet deliveries or small boxes and parcels for either perishable or non-perishable goods or general deliveries.

The main part of the survey was where there were substantial differences between this survey and conventional SC surveys. This part of the survey was a simulation game in which respondents were presented with a hypothetical scenario in which a firm is required to complete a specific delivery task with deliveries specified for several customers to be made during specific time windows on an 'average' weekday. The scenario that was presented was designed as a representative subset of customers that a firm would ordinarily deliver to so as to reduce the cognitive burden of both understanding the scenario and making the decision. This was intended to allow respondents to focus on the decisions and their implications while not being overwhelmed by information.

Following the establishment of an initial 'base case' of operations, the respondent was presented with a government policy targeted at improving the environmental outcomes of freight operators that is set to go into effect in one year. Once an initial decision had been made by the respondent, they were presented with updated forecasts for costs and emissions and given the opportunity to make adjustments to their initial decision. This was repeated for a total of five (simulated) six-month time periods with one being the current (or base) period, two being between the announcement of the policy and its implementation and two time periods following the introduction of the policy. Although the survey was initially intended to provide respondents with complete freedom in how they responded to policies, this proved to be too cognitively burdensome during testing and as a result, the presentation of six alternatives with varying levels for different attributes, similar to many SC surveys, was ultimately used. The simulation component of the survey with the six alternatives is shown in Figure 1.

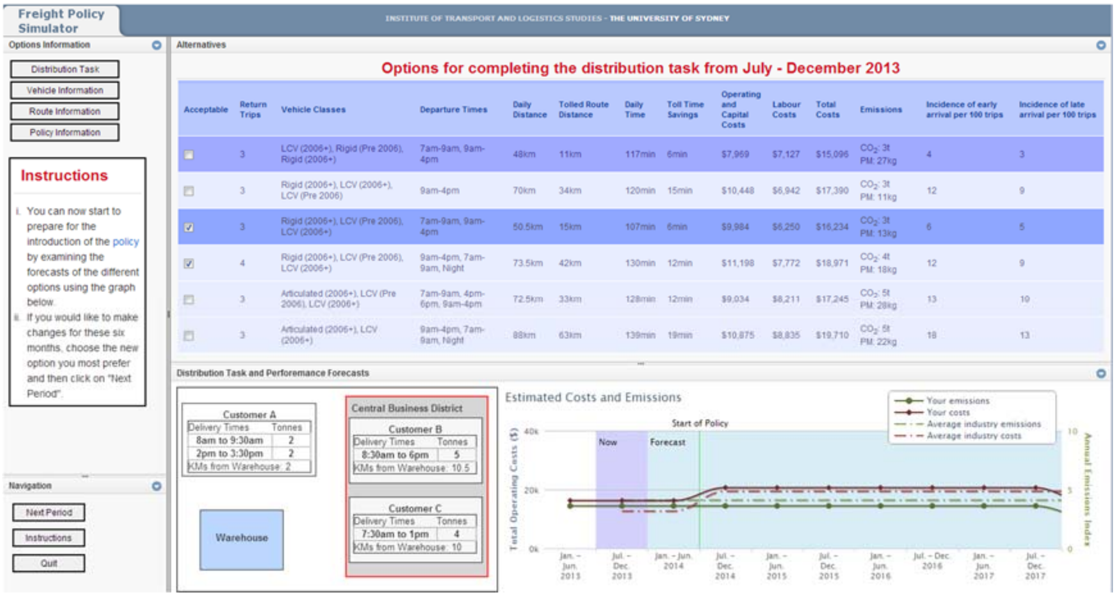

Figure 1: Example of simulation game screen with alternatives during second time period

Although this method was outwardly similar to a SC experiment, it had a number of key differences. First, the survey had a dynamic aspect where choices were made for several consecutive time periods with information presented to respondents not only for the current time period but also with simulated forecasts of subsequent time periods. This adds an additional element to dynamic stated choice experiments where although the choice-sets are considered to be for different time periods, each is presented separately (Iida et al., 1992). Furthermore, additional information on the policy was presented to respondents only after a decision had been made for the 'base case' with the alternatives and associated attributes adjusted accordingly. The second major difference was that both forecasts and alternatives for subsequent time periods were dependent on the respondent's previous decisions, providing the 'adaptive' element of the survey. As discussed above, the variation between the 'base case' alternatives were intended to be as large as possible allowing a respondent maximum flexibility in how they satisfied customer demands. However, for subsequent time periods some alternatives from the 'base case' may no longer be feasible or relevant depending on which alternative was selected by respondents. For instance, if an alternative had been selected where only newer vehicles (at least Euro III) were used, alternatives involving the use of older vehicles could be removed since respondents were unlikely to switch to older vehicles as a response to the policies being tested. In contrast to the approach used in adaptive SC experiments where subsequent alternatives depend on the range of the attributes of respondents' previous choices (Fowkes and Shinghal, 2002), this approach used the relationship between possible values of the underlying decision attributes and adds the additional element of time-dependent changes to alternatives. Third, the attributes for each alternative were a combination of vehicle and route attributes and summary attributes (including costs and emissions) calculated directly from the lower level attributes which represented the choices of interest. A further element was added by providing respondents with an "industry benchmark" for costs and emissions that they could use to evaluate the performance of their decisions. In this survey the industry benchmarks were simulated benchmarks based on the means of a subset of the available alternatives (including those not shown to a particular respondent).

The final component of the survey was a series of questions relating to their perceptions of the different attributes and their opinions on the survey in general. The first two questions related to their opinion about the attributes of the alternatives shown to respondents. They were first asked to rank five of the attributes in order of their importance to their decisions and were then asked which attributes they attempted to minimise during the simulation. Respondents were also given the opportunity to write in an attribute that was not listed (and not presented to respondents) if desired. They were then asked three open-ended questions to gain a better understanding of how they completed the simulation and how well the simulation would translate into their actual decisions were a similar policy to be introduced in reality. The first of these questions asked if they employed any strategies in completing the simulation and if so, what these were (e.g., "attempted to minimise cost when using LCVs"). The second asked if a similar policy were to be introduced if they would pass any additional costs onto their customers. Lastly, they were asked to describe their ideal alternative if they were to face a similar scenario in reality. Respondents were then given the opportunity to make any further comments about the survey in general or anything else related to the study.

4.1 Scenarios and Policies

Given the complexity of the task and the need to collect reliable data from a hypothetical scenario, it was crucial that the scenario presented to respondents was relevant to their own experience so their decisions within the game mirror the equivalent decisions that would be made within their own organisations. To facilitate this, the background questions asked of respondents were used to select one of two scenarios developed for the surveys. The first of these scenarios was tailored to companies that deliver pallets or make other large deliveries. This scenario used only three customers located in different places with two located in the generic central business district (CBD) used in the survey, each with different time windows and requiring reasonably large deliveries. The second scenario was tailored to couriers and companies making other small deliveries. It was similar to the first scenario except a larger number of customers were used with clusters of customers located in different areas (with one cluster in the CBD). Also in contrast to the first scenario, the deliveries required in this scenario were a certain number of small packages. Delivery time windows were also used in this scenario with different time windows for each cluster of customers. The context of the scenarios was also dependent on whether the respondent indicated the commodities they deliver are primarily perishable or non-perishable. This was used to change the description of the scenario to make it more relevant to the respondent's business but did not change the scenario itself.

The policies that were investigated using this survey were a low emission zone (LEZ) and a congestion charge. Both of these policies were modelled on the comparable policies already in place in London to allow for comparisons between the effects captured by the survey and those that took place in London when these policies were introduced.

The LEZ used in the survey was a cordon charge around the CBD imposed on any vehicles entering the zone at any time of day that did not meet the minimum emissions standard of Euro III (or Euro 3 for light commercial vehicles). The charges imposed are shown in Table 1. Vehicles that met the minimum standard were not charged. In contrast to the LEZ, the congestion charge of $20 applied to all vehicles (including cars) entering the CBD between 7:00 and 18:00 on weekdays. This means a vehicle entering the zone at 6:55 would not pay the charge even if they remained in the zone within the charging period. The congestion charge was intended to test how firms would react if they would need to pay to enter but would in return gain faster and more consistent travel times.

Table 1: Low Emission Zone cost structure

	Simulation LEZ		London LEZ	
Vehicle Class	**Pre Euro III**	**Euro III+**	**Non-compliant**	**Compliant**
Light Commercial Vehicle	$100	$0	£100 ($165)	£0
Rigid	$200	$0	£200 ($330)	£0
Articulated	$200	$0	£200 ($330)	£0

4.2 Generation and selection of alternatives

Although this survey was similar in some ways to conventional SC surveys, it differed most significantly in the generation and selection of alternatives. Each alternative was defined as a unique combination of tours, vehicles, toll roads and departure times which together formed a set of attributes which included both the underlying decisions but also the aggregate figures used to provide performance measures to respondents.

The selection of alternatives presented to participants followed a two-stage process. The first stage involved the generation of many possible (feasible and satisfactory) alternatives using a set of algorithms. Together these algorithms can be considered to be solving an extended version of the well-known vehicle routing with time windows problem. The second stage used data collected from respondents in the background questions and their responses to each of the prior time period alternatives to select alternatives to show.

4.2.1 Generating alternatives

The algorithms used to generate the alternatives first generated a data structure representing a set of tours. It was populated with the available values for the decision attributes varying the values so that a sufficiently large variation in combinations of decisions was used. Once the data structure had been populated with the required values, it was used within a second algorithm that tested to see if all customer demands could be met within the required constraints. This was done by taking the selected vehicle and departure time window for each tour within the alternative, and the relevant travel times (based on the time of day and the use of a toll road) to check if each stop on the tour could be served. The algorithm then made additional adjustments to departure time (if within the allowed departure time) and to waiting time (up to a maximum of 30 minutes) if some customers on the tour could not be served due to the time windows. Partial deliveries were also used if required due to vehicle capacity constraints. For each tour, the most efficient (specific) departure time and waiting times were used to calculate the tour-level attributes before the alternative-level attributes were calculated using the sum of the tour-level attributes.

To reduce the number of possible alternatives to consider, each of the decision attributes were reduced to a set of possible values. The possible values used are summarised in Table 2. It should be noted that the range for the number of tours was chosen based on the set of scenarios used and that for some applications this range may not be appropriate. It should also be noted that of the decision attributes, it is the number of tours/routes (and the range of possible values) that had the largest effect on the necessary computing time needed to generate the alternatives.

Table 2: Range of values for decision attributes

Decision Attribute	Possible Values
Origin	Warehouse, Customers
Destination	Warehouse, Customers
Toll	Yes / No
Vehicle class	LCV, Rigid, Articulated
Vehicle emissions standard	Pre Euro III, Euro III or later
Time of day	Before 7:00, 7:00-9:00, 9:00-16:00, 16:00-18:00, After 18:00
Number of stops	1 – Number of customers
Entries into policy zone	0 or more
Number of tours	2 – 8

Even with the (relatively) constrained range of values used for the decision attributes, the number of possible alternatives (without considering feasibility) was extremely large and could be potentially problematic even when only a small number of customers were used. For example, a set of alternatives where there are only three customers and the number of tours is allowed to vary from two to eight, while ignoring all other decision attributes results in a number of unique combinations of tours of 490,298. Since each tour could use 30 possible combinations of vehicles and departure times, the total number of tours to be considered was over 110 million or more than 15 million unique alternatives. Rather than attempt to check every possible unique alternative, an approach using distributions of each of the decision attributes in vectors was used to create the data structure for the potential alternatives.

The data structure was generated in R, an open-source statistical package (R Development Core Team, 2012) using a set of functions to generate random samples for different distributions for each of the decision attributes. This was undertaken separately for the alternatives used in each of the scenarios. A Weibull distribution (Weibull, 1951) with a shape parameter of 2 and a scale parameter of 2.2 was first used to generate a vector of random values for the number of routes required for each alternative to ensure there were a sufficient number of feasible alternatives with a small number of routes. These values were then increased by two and then rounded to the nearest integer to generate the number of routes for each alternative to be tested. This resulted in a distribution with most alternatives having three or four routes with five routes being the next most common. The Weibull distribution was chosen because the distribution resulting from a shape parameter of two has a positive slope where the value of x is zero (Weibull, 1951) meaning that the distribution can easily be transformed to the desired minimum value without affecting the shape of the distribution.

The resultant vector can be defined as $\boldsymbol{R} = (R_1 \;\; R_2 \;\; \cdots \;\; R_n)$ where R is the number of routes and n is the sample size of potential alternatives. Separate vectors of length $\max(\boldsymbol{R}) \cdot n$ were then generated for vehicle classes, emissions standards, number of stops and departure times by taking a random sample of the available values. These vectors were then converted into matrices where the rows corresponded to an alternative and the columns corresponded to each route in the alternative. Given that not all alternatives have the same number of routes, any cells for non-existing routes were ignored in the final data structure.

The generation of the variables determining the number of stops in each route and which stops are made on each route differ slightly depending on which scenario the alternative relates to. For the scenario related to pallets and other large deliveries, the length of the routes ranges from one stop to three (i.e., the number of customers). A random sample of this range was taken and used to populate a vector for the number of stops in each route with a length of $\max(\boldsymbol{R}) \cdot n$. The specific routing for each of the routes was then generated by selecting a random permutation of the set of customers repeated for every route in every alternative with the number of customers used in each route limited by the (now) defined route lengths.

In contrast, for the courier alternatives a two-stage process was used which first selected the number of customer clusters the route will stop at with the second stage selecting how many customers to deliver to during each stop at the cluster. The approach used for selecting the customers in the pallet scenario alternatives was also used to select those used in the courier scenario alternatives with the addition of an interim selection of the clusters before selecting the specific customers to be delivered to within each cluster. Since in this scenario all customers were in a cluster, each cluster can be expressed as $\boldsymbol{C}_u = \left(C_{u_1}\ C_{u_2}\ \cdots\ C_{u_3}\right)$ where u is the cluster and l is the number of customers in the cluster. Each route was then defined using the combination of the clusters and the customers.

The last step in generating the data structure for the alternatives was the choice to use a toll road between each stop. For the scenarios used in this study it was assumed that a toll route would be available between any two locations (warehouse or customers) at least eight kilometres apart. Using this criterion limited the number of origin-destination pairs for which a toll could be paid but still allowed the majority of the distance to be done using a tolled route. For situations in which a toll route was available a random binary choice was made that determined if the toll route would be used.

Given the definitions above the alternatives can be defined using the equations in Equation Set 1 where n is the total number of alternatives, i is the index of the alternative, and j is the index of the route. Vector $\boldsymbol{R}$ is the vector with the number of routes in each alternative. Each alternative A_i is made up of a set of vectors each related to a different decision. $\boldsymbol{D}_i$ is a vector of the departure times for each route in alternative i, $\boldsymbol{V}_i$ are the vehicle classes and $\boldsymbol{E}_i$ are the emissions standards. $\boldsymbol{G}_i$ is a vector of vectors $\boldsymbol{Q}_{ij}$, a vector with the stops for route j in alternative i. Similarly, $\boldsymbol{P}_i$ is a vector of vectors $\boldsymbol{H}_{ij}$, a vector with each element representing the use of a toll road for each trip in route j in alternative i. Vector $\boldsymbol{M}_i$ contains the number of stops in each route in alternative i.

Equation Set 1: Equations for alternatives

$$
\begin{aligned}
A &= (A_1\ A_2\ \dots\ A_n)\\
A_i &= \{D_i, V_i, E_i, G_i, P_i\}\\
D_i &= \left(D_{i1}\ D_{i2}\ \dots\ D_{i[R_i]}\right)\\
\boldsymbol{V}_i &= \left(V_{i1}\ V_{i2}\ \dots\ V_{i[R_i]}\right)\\
\boldsymbol{E}_i &= \left(E_{i1}\ E_{i2}\ \dots\ E_{i[R_i]}\right)\\
\boldsymbol{G}_i &= \left(\boldsymbol{Q}_{i1}\ \boldsymbol{Q}_{i2}\ \dots\ \boldsymbol{Q}_{i[R_i]}\right)\\
\boldsymbol{P}_i &= \left(\boldsymbol{H}_{i1}\ \boldsymbol{H}_{i2}\ \dots\ \boldsymbol{H}_{i[R_i]}\right)\\
\boldsymbol{R} &= (R_1\ R_2\ \dots\ R_n)\\
\boldsymbol{Q}_{ij} &= \left(Q_{ij1}\ Q_{ij2}\ \dots\ Q_{ij[M_{ij}]}\right)\\
\boldsymbol{H}_{ij} &= \left(H_{ij1}\ H_{ij2}\ \dots\ H_{ij[M_{ij}+1]}\right)\\
\boldsymbol{M}_i &= \left(M_{i1}\ M_{i2}\ \dots\ M_{i[R_i]}\right)
\end{aligned}
$$

4.2.2 Calculating final attributes and selecting alternatives

Once the data structure had been generated, it was converted into a PHP: Hypertext Pre-processor (PHP) array and run through a routing and scheduling algorithm to check which of the generated alternatives were feasible for completing the scenario. The algorithm also calculated the values of the attributes including costs, travel times and reliability. Feasible alternatives were returned to R and then inserted into a database to be used by the online survey. The routing and scheduling algorithm iterates through each of the routes for a given alternative checking if a delivery can be made to each of the stops on the route given constraints on vehicle capacity, delivery requirements and time windows for the vehicle and departure time to be used for the route. This was done by calculating the earliest and latest times the vehicle will be at each point on the route. For each stop spare capacity on the vehicle was compared to delivery requirements and the minimum and maximum times were compared to the delivery time windows.

Once the base attribute values have been calculated and the alternatives checked for feasibility the algorithm was re-run for each of the possible policies. This results in three sets of attribute values for each alternative, one for the base (or current) situation, one for the costs after the implementation of a congestion charge, and one after the implementation of the LEZ. These fixed attribute values formed the basis of the method used to select which alternatives are shown to respondents for each of the time periods.

The alternative generation algorithms resulted in thousands of feasible alternatives being found. Since it is not reasonable to show all of these alternatives and have respondents choose one, a method of choosing a selection of six alternatives from which respondents can choose was required. There were several considerations for choosing the six alternatives. These include that there was sufficient variation in the alternatives to provide respondents with some flexibility in responses, that the alternatives were reasonably comparable in terms of the major attributes of total costs or time, and that the alternatives were those which have the potential to be chosen in reality. In addition, for the base situation the alternatives needed to take into account the vehicle fleet of the respondent's firm and for subsequent time periods, the alternatives needed to take into account their previous decisions. One final consideration was that the respondent must have been able to decide to make no changes to their previous decisions meaning that the alternative they chose during the previous alternative needed to still be available to be chosen with updated attributes after the introduction of the policy. In addition, because the vehicles available for purchase vary over time with newer vehicles becoming a larger part of the metropolitan and national vehicle fleet, the alternatives must be selected such that newer vehicles comprise a larger proportion of those used in the potential alternatives as the survey progresses through the time periods.

To allow for alternatives to be selected in real-time depending on respondents' previous responses a weighted index was developed that ranks the alternatives based on some of the critical attributes and respondents' responses on the vehicle fleet with an adjustment factor for the current time period. The index was calculated for all possible alternatives generated for the scenario the respondent was completing each time they continued to the next time period. The index was the sum of two components, $w_{\text{calculated}}$ comprises the calculated attributes for the alternative and w_{vehicles} is based on the vehicle mix used in the alternative where $w_{\text{calculated}}$ and w_{vehicles} are defined as in the equations below.

$$w_{\text{calculated}} = 1.2_{\rho_i} + 0.1_{\kappa_{\sigma i}} + 0.2_{\kappa_{pi}} \cdot \left(1 + \frac{4t}{5}\right) + 100\alpha_{\varphi i} + 0.5w_{\sigma i}$$

$$w_{\text{vehicles}} = \sum_{v=1}^{3} \left(\sum_{e=1}^{2} \left| \frac{B_{vei}}{R_i} - \frac{B_{ve[t-1]}}{R_{t-1}} \right| \cdot 550 \right)$$

Each of the variables subscripted by *i* represent variables related to the alternative for which the index is being calculated in the current time period while variables with subscripts of t=1 represent variables based on a respondent's previous responses. In the first of these equations, ρ_i is the greenhouse gas emissions in tonnes per period, $\kappa_{\sigma i}$ is the total cost of that alternative in 2013 (Australian) dollars, κ_{pi} is the cost for the policy in dollars when it is introduced, t is the current time period (zero being the first time period), $\alpha_{\varphi i}$ is the proportion of distance using toll roads and $w_{\sigma i}$ is the total time required per day in hours. In the second of these equations, the value of w_{vehicles} was calculated by multiplying 550 by the sum of the absolute values of the difference in the proportion of the vehicle mix made up of each combination of vehicle class (v) and emissions standard (e) in the alternative i compared to the respondents' previously chosen alternative. When t = 0 the proportion of the fleet used for urban deliveries in the respondents' firm was used instead of their previous response. This means that for the 'base' period, the alternatives are selected to have a similar mix of vehicles as used in the firm they manage so as to maximise the relevance of the scenario and alternatives to their experience in the industry.

After the weighted index of all potential alternatives was calculated, the six that are actually shown to respondents were selected by placing the alternatives in ascending order of w then choosing three pairs of alternatives in the top 15 percent of alternatives with the three pairs approximately evenly distributed. For each subsequent time period the proportion of alternatives considered for selection was gradually decreased resulting in 'better' alternatives being shown for later time periods. This was done to ensure respondents were shown a range of alternatives, which would mitigate against the increased cost of the policy when this was introduced. Once the three pairs of alternatives were selected they were split into their two component alternatives and then randomly ordered resulting in the 'best' alternative (according to the index) not always appearing in the same position. With the exception of the alternatives shown in the base period, the previously selected alternative then replaced the new alternative in the same position. For instance, if a respondent chose the alternative shown on the third row during the base time period, the alternative shown on the third row during the second period would remain the same as that shown during the base period with the five other alternatives changing.

5 Respondent reaction and burden

This survey design was used to conduct a field survey from late January to early June, 2013 involving respondents in freight firms from urban areas around Australia. It included both an initial pilot in January and February involving computer-assisted personal interviews (CAPIs) and a main phase from March to June involving a combination of computer-assisted telephone interviews (CATIs) and self-administered online surveys. Respondents were managers and others involved in decision making in freight firms.

Recruitment of respondents proved far more challenging than had been initially anticipated. Of the 1,080 firms who were initially contacted, 450 nominally agreed to participate and had an eligible respondent available. Particularly in larger firms, getting into contact with a manager eligible to complete the survey required several telephone calls and e-mails to several different people until the organisation decided who was best placed to complete the survey. Of those who agreed to participate, 120 respondents started the survey (i.e., clicked on the link to start) with 62 completing all parts of the survey. The relatively large drop-out rate was found to be largely a result of respondents deciding it was not relevant to their business after having read the introduction and instructions and some self-administered survey respondents having difficulty understanding the scenario. Although the relatively small sample was disappointing, the recruitment rate is in line with many other surveys of freight operators (Puckett and Hensher, 2008; Smalkoski and Levinson, 2005). The final sample included firms of a range of sizes from firms with fewer than five drivers to those with more than 100 (see Figure 2). The final sample contained more large firms than the industry as a whole but these large firms also comprise a substantial proportion of the vehicle fleet.

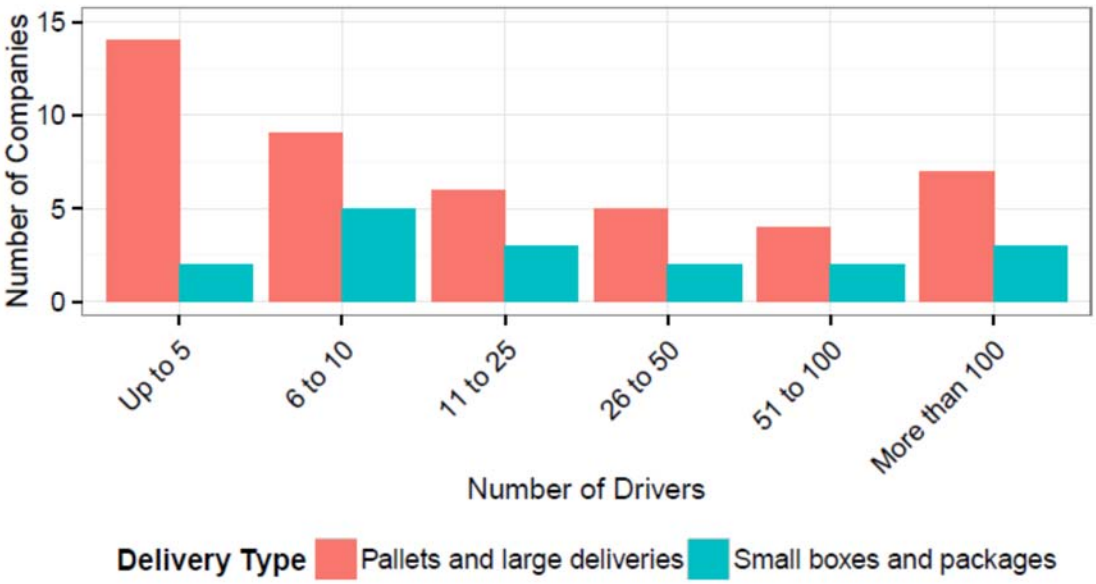

Figure 2: Completed responses by size of firm and delivery type

Given the time pressures many respondents were under, the survey was initially designed to be able to be completed within 15 to 20 minutes, although many respondents took slightly longer. Since respondents may have left the survey open to attend to more urgent matters before later returning to the survey, it is not possible to accurately determine how long respondents spent actively completing the survey. However, an approximate time spent can be calculated using time-stamped activity logs for the survey. Excluding pilot respondents who generally took longer due to the more in-depth discussions during the interviews, this yields a median completion time of 23.68 minutes and a mean of 28.89. The fastest respondent took approximately 10 minutes with the slowest taking 77 minutes (see Figure 3). It should be noted that all the respondents who took 45 minutes to complete the survey appear to have spent more than 10 minutes on at least one of the information screens, suggesting they may not have been actively completing the survey during part of this time. There was only a very small difference between the survey completion methods in the time taken to complete the survey with CATI respondents taking an average of approximately 30 seconds longer to complete than self-administered online survey respondents. However, there is a distinct difference in the time taken depending on the job title of the respondent further highlighting the importance of experience in completing such a complex survey. General Managers and other business-focused respondents took less time to complete the survey while managers less involved in the day-to-day operations took longer (primarily environmental managers). Interestingly, owners took about the same time to complete the survey as operations and transport managers.

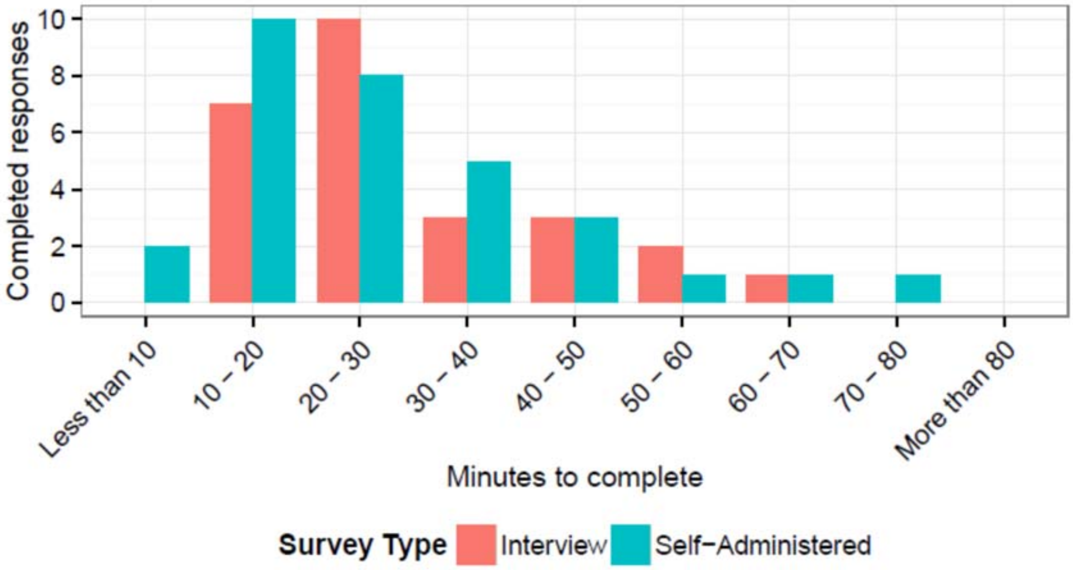

Figure 3: Time taken to complete survey

One of the final questions in the survey gave respondents the opportunity to provide feedback on the survey as well as providing comments to the interviewer directly for CATIs and CAPIs. Feedback for those who completed the survey was generally positive with several respondents saying the survey was easy to complete once they understood the scenario and the different options. One respondent suggested it would be a useful exercise for training new operations managers since it made them consider the implications of different decisions. Some respondents who opted to complete the survey as a self-administered survey indicated it was somewhat confusing at times suggesting that future surveys using this method would benefit from using an interviewer (either as a CATI or CAPI). Other feedback generally focused on the policies presented with several saying the policies, if implemented, would cause problems for the industry. Feedback from those who did not complete the survey focused primarily on the perceived irrelevance of the survey to their business. Interestingly, none cited confidentiality as a reason for not completing the survey, an indication that the use of hypothetical scenarios achieved one of its objectives in limiting the confidential information required.

6 Conclusions

The key feature of this survey method is the simultaneous preservation and updating of alternatives across different simulated time periods allowing for *status quo* bias in firm (and individual) decision-making (Erixon, 2007) to be integrated into the survey. This ability to take into account how decisions change over time provides a method of collecting data on decisions that are not made independent of time and the *status quo* in a way that other common methods are unable to achieve. Furthermore, the direct linking of the calculated attributes to the underlying decisions means that the decision of (for instance) vehicle class has a direct effect on the costs. At the same time choosing an alternative based on cost limits the available vehicle class options both in the current and future time periods.

The survey method should, in general, be applicable to other scenarios and policies than those used here albeit possibly with some minor modifications. These decisions are not limited to decisions made by freight operators nor those related only to environmental and other government policies, but are found in many decisions in which an external force changes the costs and benefits of previous decisions. However, it must be emphasised that as with any hypothetical study, the description and context of the scenarios are of crucial importance and other applications may need a somewhat different approach to the design of the scenarios and alternatives. In the same manner, serious consideration should be given to simplifying the design and generalising the scenarios so as to reduce the likelihood of respondents thinking the survey irrelevant or too much work to complete. Furthermore, depending on the complexity of the scenarios and decisions being studied, surveys using this method would likely benefit from the use of CAPIs or CATIs to complete the survey rather than relying on self-administered surveys alone.

7 References

Bolis, S. and Maggi, R. (2003), “Logistics strategy and transport service choices: an adaptive stated preference experiment”, *Growth and Change*, Vol. 34 No. 4, pp. 490–504.

Brewer, A.M. and Hensher, D.A. (2000), “Distributed work and travel behaviour: the dynamics of interactive agency choices between employers and employees”, *Transportation*, Vol. 27 No. 1, pp. 117–148.

Erixon, L. (2007), “Even the bad times are good: a behavioural theory of transformation pressure”, *Cambridge Journal of Economics*, Vol. 31 No. 3, pp. 327–348.

Fowkes, T. (2007), “The design and interpretation of freight stated preference experiments seeking to elicit behavioural valuations of journey attributes”, *Transportation Research Part B: Methodological*, Vol. 41 No. 9, pp. 966–980.

Fowkes, T. and Shinghal, N. (2002), *The Leeds adaptive stated preference methodology* (No. 558), Working Paper, available at: http://eprints.whiterose.ac.uk/2018.

Hensher, D.A., Collins, A.T., Rose, J.M. and Smith, N.C. (2013), “Direct and cross elasticities for freight distribution access charges: Empirical evidence by vehicle class, vehicle kilometres and tonne vehicle kilometres”, *Transportation Research Part E: Logistics and Transportation Review*, Elsevier Ltd, Vol. 56, pp. 1–21.

Hensher, D.A., Puckett, S.M. and Rose, J. (2007a), “Agency decision making in freight distribution chains: Establishing a parsimonious empirical framework from alternative behavioural structures”, *Transportation Research Part B: Methodological*, Vol. 41 No. 9, pp. 924–949.

Hensher, D.A., Puckett, S.M. and Rose, J.M. (2007b), “Extending stated choice analysis to recognise agent-specific attribute endogeneity in bilateral group negotiation and choice: a think piece”, *Transportation*, Vol. 34 No. 6, pp. 667–679.

Iida, Y., Akiyama, T. and Uchida, T. (1992), “Experimental analysis of dynamic route choice behavior”, *Transportation Research Part B:*, Vol. 26 No. I, pp. 17–32.

De Jong, G., Gunn, H. and Walker, W. (2004), “National and International Freight Transport Models: An Overview and Ideas for Future Development”, *Transport Reviews*, Vol. 24 No. 1, pp. 103–124.

Puckett, S. and Hensher, D. (2008), “The role of attribute processing strategies in estimating the preferences of road freight stakeholders”, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 44 No. 3, pp. 379–395.

Puckett, S.M., Hensher, D. a., Rose, J.M. and Collins, A. (2007), "Design and development of a stated choice experiment for interdependent agents: accounting for interactions between buyers and sellers of urban freight services", *Transportation*, Vol. 34 No. 4, pp. 429–451.

R Development Core Team. (2012), "R: A Language and Environment for Statistical Computing", Vienna, Austria, available at: http://www.r-project.org/.

Shinghal, N. and Fowkes, T. (2002), "Freight mode choice and adaptive stated preferences", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 38 No. 5, pp. 367–378.

Smalkoski, B. and Levinson, D. (2005), "Value of Time for Commercial Operators", *Journal of the Transportation Research Forum*, Vol. 44 No. 1, pp. 89–102.

Stefan, K.J., McMillan, J.D.P. and Hunt, J.D. (2005), "Urban commercial vehicle movement model for Calgary, Alberta, Canada", *Transportation Research Record: Journal of the Transportation Research Board*, Trans Res Board, Vol. 1921, pp. 1–10.

Weibull, W. (1951), "A statistical distribution function of wide applicability", *Journal of applied mechanics*, Vol. 18 No. 3, pp. 293–297.