Designing an Efficient Stated Ranking Experiment for *Ex-Ante* Urban Freight Policy Evaluation in a Three AgentType Context: Retailers, Own-Account and Carriers

Edoardo Marcucci[1] Valerio Gatta[2] Amanda Stathopoulos[3] Eva Valeri[4]

[1]Università degli studi di Roma 3, Centro di Ricerca Interdipartimentale di Economia delle Istituzioni (CREI) email emarcucci@uniroma3.it
[2] Università degli studi " La Sapienza"
[3] Università di Trieste, Dipartimento di Scienze Economiche e Statistiche (DISES)
[4]Università di Trieste, Dipartimento di Scienze Economiche e Statistiche (DISES)

ABSTRACT: City logistic policy making requires an understanding of specific issues seldom investigated in current research. Policy implementation produces unsatisfactory results when does not consider behavioural and contextual aspects. Acquiring relevant data is crucial to test hypotheses and forecast agents' reactions. Development and application of appropriate survey instruments to test policy ex-ante acceptability is still lacking despite methodological advances in modelling interactive behaviour. This paper expands and innovates the methodological literature by describing a stated ranking experiment to study freight agent interactive behaviour and discusses in detail the efficient experimental design implemented to incorporate agent specific interaction in city logistic literature.

Keywords: urban freight distribution, group decision making, agent-specific interaction, stated preference, stated ranking experiments.

1 Introduction

Cities are net importers of goods. They are characterised by relevant economies of density and proximity, produce new ideas, innovations and generate economic growth that irradiates to adjacent areas. Cities are characterised both by concentrated research and service production as well as structural goods acquisition that provoke, along with passenger transportation, various negative externalities among which, the most prominent are: congestion, visual intrusion, noise, environmental and acoustic pollution. The impact of freight movement is particularly high in densely populated areas where economic activities are concentrated and generate a consistent, strong and, usually, rigid demand for transportation. Decision makers adopt policies with the intent of optimising the movement of both passengers and freight to foster a sustainable development via the decoupling of economic growth from transport demand. Urban freight policies need to be analysed and evaluated considering a host of aspects and potential impacts including, among others: policy characteristics, strength of the linkages with the problems they should solve, external and secondary effects, distributional impacts among different stakeholders, level of analysis of the phenomenon, data needs and estimated behavioural reactions. In order to take into account all these relevant aspects, it is important to devise modelling frameworks capable of forecasting policy impacts.

This paper illustrates the potential of using a Stated Ranking Experiment (SRE) to acquire data to estimate stakeholders' preferences for potential policy measures within an Urban Freight Transport (UFT) context. We propose an innovative methodology to investigate retailers' and carriers' sensitivity to changes in policy packages simultaneously reputed possible, by the local authorities (transport regulators), and acceptable, by the main stakeholders (retailers, own-account[1] and carriers).

The paper describes the definition, development and administration phases of a SRE in a real-life context where effective policy interventions (e.g. access charging, time windows, loading/unloading bays (l/u), etc.) are envisaged and evaluated for implementation. The experiment proposed enables the identification and measurement of both overall and agent specific *ex-ante* acceptability of the policy mix to be implemented. UFT policies are deeply intertwined with and influenced by interaction effects among the actors involved. The approach proposed identifies not only effective and efficient measures but also, among these, the subset of potentially acceptable ones, if not by all operators, by the majority of the agent types involved.

The innovative features of the paper relate to the contemporaneous consideration and evaluation of both demand and supply agents instead of, as it is usually done, just studying the two groups as separate phenomena. Under this respect our approach proves complementary to the widely used Freight Quality Partnership[2] (FQP) that apply a descriptive and qualitative approach.

The paper is structured as follows. Section 2 reviews the literature on both agent interaction analysis in the freight sector as well as that of stated preference and experimental design in general. The description of the study context is reported in section 3 while section 4 describes the deployment of the survey. Section 5 concludes and describes future research objectives.

2 Literature review

2.1 Modeling multiple agents in freight: an overview

Aggregate models are typically used when modelling freight where little attention is paid to the critical role that individual actors play in the decision making process. This approach is characterised by an inherent limitation especially for policy interventions aimed at changing the reference scenario by, altering agents' relative convenience of past actions. This section reports and discusses recent findings drawn from a behavioural approach to freight modelling, in general, and to UFT, in particular. Hensher and Figliozzi (2007) convincingly argue that standard approaches do not account for the complexity of freight movements at different geographical scales thus missing potentially relevant explanations of the current scenario. What is more, new delivery methods (e.g. JIT) and customer driven freight services (e.g. electronic commerce) have rendered UFT more complex, paving the way to highly specialised third-party logistic providers. Behavioural models, a sub-set of disaggregate models (e.g. inventory models and logistic optimisation), explicitly consider stakeholders' utility maximization efforts. In this case one has to unequivocally identify key decision makers to develop a modelling framework adopting an agent-based micro-simulation approach capable of describing and forecasting the behaviour of the specific actors involved (Liedtke and Schepperle, 2004). Various authors (Gray, 1982;

[1] By own-account we intend a specific group of retailers that transport with their owned vehicles their goods (the transport activity is ancillary).

[2] FQP has been launched by the UK Department for Transport (DfT) to improve involvement of the main stakeholders in urban freight decision-making (DfT, 2007).

Southworth, 2003; Wisetjindawat et al., 2005; de Jong and Ben-Akiva, 2007; Hensher and Figliozzi, 2007; Samimi et al., 2009; Yang et al., 2009; Roorda et al., 2010) consider UFT one of the most appropriate fields of application for developing agent-based micro models. In fact, freight movements are determined by the underlying motivations of the relative convenience of each stakeholder in making a given choice. Structural behavioural analysis represents a substantial improvement with respect to standard modelling techniques. The benefits of explicitly contemplating behavioural aspects in modelling freight, become evident when considering: 1) network and micro-simulation modelling, 2) land use/transport networks with feedback effects, 3) the relevance of physical characteristics of logistics networks. Freight modelling tends not to consider behavioural implications of the policies introduced. Such considerations are fundamental to understand the motivations of freight stakeholders when: 1) reacting to different policy mixes, 2) dealing with specific constraints (e.g time windows), 3) accounting for incentives (e.g. price rebates for new vehicles), and, in general, 4) interacting with others. Interactions between existing and prospective constraints determined by new policies or reactions to specific strategies or constraints may change when the *Status Quo* (SQ) is altered. For example, policy changes influencing fuel prices, land use patterns and pricing strategies modify the constraints and alter the relative convenience of each option. Puckett and Greaves (2009) argue that it is important to consider jointly both the instruments available to policy makers and the set of attributes influencing freight behaviour to understand the potential impacts that any policy might produce in terms of market outcomes which is exactly what policy makers should/would like to know *ex-ante* before implementing a policy. It is important to identify incentives/disincentives with a relevant impact and quantify their impact on the reference scenario before applying them in a real-life context. To do so one has to pinpoint the type of decision makers involved, discover under which constraints they operate, understand how they interact, and figure out on which set of freight service attributes they finally negotiate and interact.
Some new and interesting approaches have been developed recently to tackle the issues raised in this section. The most prominent promoters of Inter Active Choice Experiments (IACE) are Brewer and Hensher (2000), applying the IACE framework to telecommuting, and Puckett and Hensher (2006, 2008) applying it to freight. Usually, both financial and sample size considerations render this approach difficult to implement for real-life applications. Only a limited number of buyers of road freight transport services or transport providers are usually willing to partake in this type of study and hence it is difficult to guarantee a sufficient participation to obtain statistically reliable parameter estimates[3]. The approach we present tries to contribute to this relevant and daunting question.

2.2 Experimental design: an overview

Stated Choice (SC) experiments have a long standing in applied research. One can trace the first contributions to the works of Louviere and Woodworth (1983) and Louviere and Hensher (1983). A choice experiment aims at acquiring data to generate reliable and useful estimates of parameters of interest. Depending on the research question, one may adopt a different response format among: choice, ranking or rating.

[3] Hensher and Puckett (2008) have provided a solution to this issue by developing Minimum Information Group Inference (MIGI), a less data demanding methodology even if equally capable of producing relevant results. Their research indicates the critical areas where specific efforts are needed to gain a better understanding of UFT related decision making.

The response format impacts on data analysis (Johnson and Desvousges, 1997; Ortúzar and Garrido, 1994; Crask and Fox, 1987; Louviere, 1988, 1992; Aaker and Day, 1990) and on the reliability of the responses obtained.
Estimation of statistically significant parameters, especially when, due either to financial constraints or lack of participation, small samples are used -- as is typical in empirical research--, may be aided (impaired) by a good (poor) experimental design. The choice of an experimental design is relevant for the research conclusions reached. An experimental design is, *de facto*, a matrix of values containing the levels of the attributes that will constitute the SC survey. The analyst needs to optimize the allocation of attribute levels to a design matrix given his research goals. Historically, the most common strategy has been to ensure that attribute levels are uncorrelated or orthogonal (Louviere et al., 2000). However, more recently, efficient designs, an alternative and innovative approach, have been developed by numerous researchers (Huber and Zwerina, 1996; Kanninen, 2002; Kessels et al., 2006; Sándor and Wedel, 2001, 2002, 2005, Ferrini and Scarpa, 2007). An efficient design considers orthogonality in design attributes not to be desirable since discrete choice models (e.g. MNL, MNP) using SC data, are not linear and do not require zero correlation between the attributes of the design.
Huber and Zwerina (1996) were the first to link SC statistical properties to the econometric models. The authors demonstrate that the asymptotic standard errors of the parameter estimates can be reduced by relaxing the SC orthogonality condition. When constructing an efficient design it is easier to define, evaluate and consider a single value instead of assessing the whole Asymptotic Variance Covariance (AVC) matrix and various analysts have proposed different efficiency measures (e.g. d-efficiency, a-efficiency) to express the desirability of the design obtained.

3 The study context: the roman freight limited traffic zone

The formal institution of a Limited Traffic Zone (LTZ) in Rome's historical centre dates back to the late eighties when a 5 km^2 area was restricted to non-resident vehicles. The bans on traffic apply to passenger and freight vehicles alike. The current 4 km^2 area in the historical centre, focus of this study, is characterised by a specific legislation where only Euro 1 and later vehicles are allowed to enter the LTZ with free access awarded only to residents while other agents (e.g. retailers and freight carriers) pay an access fee. The scheme operates during daytime hours (passenger cars: 06.30–18.00 Monday to Friday and 14.00–18.00 on Saturday). Passenger and freight LTZ largely overlaps where the latter is aimed at goods vehicles and operates between 10.00–14.00 and 16.00–20.00. The yearly permit costs 565 € per number plate. The system is based on cameras and optical character recognition software. Specific time windows apply for access and parking of freight vehicles. Nonetheless, a wide range of exemptions applies to freight operators. A synthetic summary of the regulatory regime in place is reported in the Appendix.
Indeed, the regulation is essentially designed to foster the use of third account operations while discouraging lengthy parking of own-account vehicles given the shortage of on-street parking. Time windows are not systematically enforced. The scheme, due to the many exceptions, cannot be considered to be specifically aimed at reducing congestion nor can it be classified as a pure environmental low emission zone (LEZ) since vehicle emissions standards are not currently part of the scheme.

4 Development of the survey instrument

Receivers, carriers and forwarders are, traditionally, considered as essential stakeholders in urban freight logistic system analysis (Ogden, 1992). The current SRE concentrates on representing three main supply chain agents: carriers, retailers and own-account operators. The first two are well identified in the literature while previous stakeholder consultations suggest considering own-account operators as well (Stathopoulos et al. 2011).

First of all one has to define, select, develop and customize the attributes to include in the SRE. Here we start by illustrating how we moved from the stakeholder consultation to attribute definition while highlighting and motivating which specific attributes were included in the final questionnaire design. Indeed, the level of joint policy ex-ante acceptability was the main criterion for attribute inclusion. Subsequently, we report how each attribute was defined, structured in levels and ranges and progressively differentiated by agent type to account for real-world agent type specific constraints and preferences. The attribute selection drew on results deriving from stakeholder surveys. The following sections overview the attributes included, describe their characterization and motivates our choices in the following steps.

4.1 Attributes included in the SRE

Each alternative in the SRE is described by a set of attributes that can take several levels to describe ranges of variation when the alternatives are presented to the respondents.

The attributes used in the experiment were derived from three main sources, namely; a) literature survey; b) previous quantitative studies on city freight distribution in Rome; c) focus group meetings with relevant expert stakeholders.

We performed an extensive review of the current *city logistics* literature with an agent-based perspective to identify a set of potentially conflicting policy components when viewed from each of the different agent type perspectives. For instance, night-time deliveries were considered efficiency enhancing by carriers but reputed a mere increase in costs by retailers.

Reviewing previous quantitative studies on city logistics in Rome (STA, 2001; Filippi and Campagna, 2008) and considering the expert stakeholder surveys helped selecting the attributes for the SRE. An important phase of the expert surveys[4] focused on defining the policies considered most appropriate to mitigate the identified UFT problems (Stathopoulos et al. 2011).

Subsequently the results were evaluated according to various criteria to ensure an appropriate attribute selection. The criteria applied were: saliency, shared support and plausibility with respect to changes of the current scenario.

Volvo REPORT (2010) provides a detailed overview of the link between the stakeholder survey results and the attributes used in the SRE. To incorporate the degree of shared support, as a pre-condition for attribute inclusion, it is necessary to look at agent-specific support for policies. Agents were, on the whole, reluctant to propose the use of time windows considered a delicate instrument. Indeed, city access

[4] The results presented here are part of a greater study, VOLVO REPORT - Volvo Research and Educational Foundation (2010) (project SP-2007-50 - Innovative solutions to freight distribution in the complex large urban area of Rome) where specific attention was paid to attribute definition.

time and delivery time restrictions appear to be a core issue behind disagreement among different agent types.
The rationale for the attribute-selection is that a high level of shared support should facilitate the introduction and persistence of a policy (Stathopoulos et al. 2011). Notably there is a strong and mutual support for the eco-vehicle incentive, information provision and number of l/u bays. On the other hand, policies sustained by a single agent type alone (e.g tradable permits, time windows) run the risk of not gaining the necessary support needed for a successful implementation in a real-life context.
Out of the twelve policies, based on the criteria of relevance and acceptability, six attributes were selected to undergo pilot testing with real operators, namely: 1) number of l/u bays; 2) probability to find l/u bays free; 3) time windows; 4) exemption from time windows; 5) entrance fees; 6) exemptions from entrance fees.
Each of these six attributes has been on the political agenda for a long period and all were perceived as realistic measures to include in future policy mixes.
In what follows we discuss in detail the definition and refinement of each attribute.

Loading/unloading bays

L/u bays availability and management was one of the most discussed issues in the focus groups. The main challenge revolved around the definition of the attribute. Critical aspects concerned some attribute dimensions that interviewees considered relevant during the stakeholder meetings. For instance, both the number of bays and their availability were deemed important. Earlier studies in Rome (STA, 1999), testify that both these features are indeed important for operators and, therefore, it was decided to report both these characteristics in the SRE. Although the construction of additional l/u bays has been on the political agenda for decades, the proposals have never been realized. This means that the number of l/u bays in the LTZ is fixed at the restrictive number of 400.

Probability to find l/u bays available

Related to the number of l/u bays, the probability of finding them available was also included. Evidence from stakeholder meetings and a pilot study both indicated that some agents were not so much interested in the number of bays but rather in the probability of finding them available for l/u operations. Policies proposed for implementation foresee the increase of controls to guarantee a correct occupation and high rotation of vehicles using the bays. The focus on the probability suggests operators concentrate on policy outcomes rather than on the instruments used especially since there are alternative methods to contrast illegal or improper use. The desired outcome was to increase the probability of finding the bays free and this also emerged as the most appropriate attribute definition. We identified the probability of finding the bays available, realizing an empirical survey in a section of Rome’s LTZ, to be about 10%. The attribute was formulated as a probability percentage to avoid the

issue of an unequal distribution of bays and freight activity among different areas potentially generating disparity among agents' perceptions[5].
Distinguishing between the number of bays and the probability of finding the bays free allows several modelling options in the estimation phase. Indeed, one can test the fit of models with the two attributes kept separate or interacted, if that corresponds to the prevalent way respondents consider and evaluate the attributes.

Time windows

The importance of time window regulations is unanimously recognised. However, this policy purports a series of important difficulties in its characterization pertaining both to its definition and representation. In fact, the attribute could both be described in terms of: 1) number of open vs. closed hours, or 2) distribution of specific opening or closing hours during the day. The possible definitions and configurations provide the respondent different information on which to compare alternative policy options. The design and refinement of this attribute was carried out in several stages, described in the following: 1) identify the most desired hours for freight delivery; 2) define different variations on the *SQ* which could easily be interpreted by the researcher (e.g. number of hours and their distribution over the day); 3) represent these scenarios to respondents; 4) test the comprehension of the scenarios and their desirability in a pilot study; 5) re-define the attribute in view of pilot study results. Five different scenarios, varying both the number of hours and their distribution according to desirability, were the first set of representation devised.

Entrance fees

A price attribute is usually included when creating a choice or ranking experiment to calculate implicit prices. The importance of the entrance fee was confirmed during the discussions with stakeholders. Carriers were particularly sensitive since they are directly influenced. This attribute was very much considered especially due to the large increases in recent years. In fact, the annual fee shifted from 35€ euro to 565€ for each number plate. The attribute was represented as upward and downward variations from the SQ, calculated as percentage increases/decreases.

4.2 Agent specific SRE

Respondent-type differentiations of the SRE were necessary after the piloting with operators and also to get the most out of the efficient design strategy adopted.
The main agent type diversification is the inclusion of the time window attribute only for own-account operators due to an anchoring affect around the SQ condition. Indeed, only own-account operators are *de facto* facing time window restrictions since carriers, operating as third account, can access the LTZ at all times. The SRE choice set consisted of three policy options always including the SQ alternative. Agents were asked to rank policy bundles according to their preferences and were solicited to indicate whether a policy was considered unacceptable thus not part of their ordering.

[5] Defining an attribute in probabilistic terms may provoke an excessive cognitive burden for respondents, but was a necessary condition to ensure a general interpretation of the perceptions of this complex issue.

Respondents were also asked, for each choice task, to rank the exact same alternatives according to their best guess of their "typical" commercial partner's preferences. This requires respondents to state, to the best of their knowledge, the ranking of their freight partners and whether any of the alternatives would be considered unacceptable by their partners. Table 1 reports an example of a SRE task.

Table 1 - Example of a ranking task

	Policy 1	**Policy 2**	**Status Quo**
Loading/Unloading bays	400	800	400
Probability to find L/U bays free	20%	10%	10%
Entrance fee	1000 €	200 €	600 €
Own policy ranking			
Which ranking of the policies, *in your view*, would your partner provide?			

After selecting the attributes to include in the SRE, the next important step is to determine the appropriate levels and ranges for each attribute.
The levels that characterize the attributes should ideally be both plausible and policy relevant, although a choice experiment may also test currently unavailable, but possible, alternatives (e.g. a new mobility control policy). In defining the levels it is important to consider their number, how they are spaced among them and their overall range of variation. The attributes, levels, distribution and range are illustrated in Table 2.

Table 2 - Attribute levels and ranges used in the SRE

Attribute	**Number of levels**	**Level and range of attribute (Status Quo underscored)**
Loading/unloading bays:	3	400, 800, 1200
Probability to find l/u bays:	3	10%, 20%, 30%
Time windows:	3	OPEN from 18:00 to 08:00 e from 14:00 to 16:00; OPEN from 20:00 to 10:00 e from 14:00 to 16:00; OPEN from 04:00 to 20:00
Fees:	5	200€, 400€, 600€, 800€, 1000€

The first issue is to determine the number of levels to include and analyse the implications for subsequent estimation. For instance a two-level attribute only allows for the estimation of linear effects. Yet, the indirect utility function of an attribute may exhibit non-linear effects and for this reason it is often more informative to include more than two levels to describe an attribute, when appropriate, and to allow for the estimation of non-linearities in the utility.
A second issue is how to distribute the levels. The literature recommends that levels be evenly spaced to aid interpretation of the coefficients. What is more, if levels are

also symmetrical with respect to the *SQ*, this allows for the control of asymmetrical effects related to gains and losses.
The ranges of the levels are of particular importance. Indeed, a sufficiently wide range of levels should be used to avoid respondents ignoring the attribute due to a lack of variation. The level range is particularly important for the price attribute which is used to calculate implicit prices of other attributes using willingness to pay (WTP) estimates. Moreover, the payment vehicle should be chosen to match the empirical setting.
As may be observed in Table 2 all attributes are characterized by at least three levels. This allows for controls for non-linear effects as explained above. Such effects are of great importance in particular when considering reactions to policies since there might be large effects on well-being derived from specific levels.
Joint stakeholder meetings were an important source of information concerning the attribute distribution and range. On this occasion the six selected attributes were presented and agents asked to provide indications of ranges. Typical questions posed were: "*What is the minimum increase in the number of l/u bays you would consider necessary?*" for each attribute. Based on the ranges provided by the stakeholders a maximum increase for each attribute was defined for the two l/u bays and the fees. For the time windows, instead, stakeholders were asked to suggest two alternative scenarios to the current one: the first representing a minimum increase desirable for operators of freight distribution and the second defining a maximum sustainable reduction concerning the number of hours of access to the LTZ. Moreover, a meeting with local policy makers, responsible for promoting and planning changes to the LTZ regulations, was organized. In these meetings both the feasibility of fee increases and the likely construction of l/u bays were discussed. Based on opinions expressed during stakeholders meeting, these attributes were further redefined to achieve realism and properly mirror plausible policy changes.
Drawing on these results the minimum and maximum points of the attribute ranges were defined. For the l/u bay attributes the minimum coincides with the current situation. Instead the range is extended to reflect the stakeholder opinions and the three levels are then equally distributed implying that the policy scenarios only proposed an increase with reference to the *SQ* levels. The time window attribute was reduced from five to three levels due to its complexity. Great effort was dedicated to define one improved and one deteriorated level for the time window attribute. Due to the qualitative nature of the attribute it was not possible to ensure that the levels were evenly spaced. Lastly, the entrance fee attribute was defined to vary in both directions with respect to the *SQ* level of approximately 600€. Since past policy changes have been quite abrupt, the attribute proposed for the SRE had a wide range of variation going from 200€ to 1.000€. The quantitative nature made it a simple task to ensure that the levels were both symmetrical and evenly spaced over the five levels.

5 Efficient design strategy and motivation

5.1 Overview

In experimental research applications the attributes and levels of a design are defined in advance on the basis of personal judgement and prior findings, and choice sets are generated by a randomized procedure (Louviere, 1988). The current study instead is based on sequential efficient experimental design theory.

The survey context, characterized by difficulty in performing interviews due to privacy concerns and general scarce interest to participate on behalf of the operators taken together with the high cost of *face-to-face* interviews lead us to use an efficient design strategy. Moreover, the lack of prior information concerning the sensitivities of single agent types induced us to apply a gradual approach with progressive refinement of the experimental design across several so-called waves. This means there is an evolution of the design, which is upgraded in several, waves, where each wave incorporates findings from previous interviews. The following sub-sections give a detailed overview of the design criteria used in each stage along with a rationale for decisions made, summarising the main aspects of the efficient design strategy among waves in Table 3.

5.2 Pilot

The novelty of the selected attributes and the lack of any previous study to rely on in the definition of the sign and dimension of the coefficients lead us to test three different design approaches (all developed in Ngene 1.02[6]).

First, a d-efficient design with assumptions on the coefficient signs was tested. Due to the low precision of the priors, characterised by large standard deviations of the coefficients, it proved impossible to make the design converge based on the limited sample size planned for the first wave of interviews. Secondly, an orthogonal design was tested where each attribute is perfectly uncorrelated with every other attribute (Louviere Woodworth, 1983). It proved impossible to generate such a design using only 9 tasks and it was considered inconvenient to work with blocks given the small sample-size foreseen for the pilot. In third place, we tested a fractional factorial design where a subset of the possible level combinations appears in the design. Given that six attributes were present in the initial design, the number of combinations of the design would be equal to $2^5 \times 2^3 \times 2^2 = 1{,}024$. No differentiation of the design according to agent types – own-account retailers, third account retailers and carriers – was carried out due to the lack of prior information regarding differences in utility among these. Drawing on Multinominal Logit choice-models estimated using data from the pilot sample we were subsequently able to delineate the first efficient design with priors.

5.3 First wave efficient design

For the first wave efficient design some important novelties were incorporated. Based on the estimates from the pilot study interviews it was possible to obtain indications of the magnitude and sign of each attribute coefficient leading to differentiations in the SRE design properties. A first layer of differentiation concerns the attributes used to characterize the utility function of the three agents considered. Some of the attributes originally tested were eliminated following the pilot and, for the four selected attributes, agent-specific considerations were introduced. The time windows attribute was, in fact, used only for own-account operators since they represent the only group subject to the policy in the current regulatory framework. Table 3 describes the complete utility functions used in the experimental design, attribute by attribute, for each wave and agent type. This includes assumptions on prior Bayesian

[6] ChoiceMetrics Pty Ltd, http://www.choice-metrics.com/index.html.

mean estimates and standard deviations, distributional assumptions and presence of non-linearities.
Concerning the first wave, due to the lack of statistical difference in the taste parameters between carriers and retailers, identical priors were used for these operators. As can be observed in Table 3 linear effects were hypothesized for each attribute (the annual entrance fee, number of l/u bays and probability of finding bays available). Bayesian priors were used to reflect uncertainty in the estimation of parameters and the degree of uncertainty was reflected in the spread of their respective probability distribution. Parameter distributions were assumed normal $\beta_{attribute} \sim N(\mu_{attribute}, \sigma^2_{attribute})$ to reflect the wide distribution of preferences around the mean.
Instead, the own-account operators were treated in a separate manner, both in terms of their utility function (additionally containing the time-window attribute) and the Bayesian priors imposed for each attribute. Due to the qualitative nature of the time-window attribute, characterized by three descriptive levels, this was effects-coded[7]. Thereby n-1 levels needed to be specified with a relative prior for each of them. In our case, due to the small sample size in the pilot, these were simply fixed to be negative for the lowest attribute level, positive for the intermediate with the most favourable level omitted for identification, e.g. time.effects[(u,-1,0)|(u,0,1)]*time[0,1,2].
In each case, to select the final design, we applied the widely used d-efficiency criterion, based on calculating the determinant of the AVC matrix along with a-efficiency, taking the trace of the AVC matrix, thereby looking only at the variances and not at covariances. The criterion of level balance, where each attribute appears equally often, was respected for the three-level attributes but not for the entrance fee given the use of 9 choice sets. During the design process, we controlled for utility balance to ensure options in each choice set had similar probabilities of being chosen. Finally, an important rationality test was included to control respondent consistency, where one task was duplicated.
Based on the data collected at this stage additional models were estimated, including controls for non-linear effects, to guide the ensuing design choices.

5.4 Second wave efficient design

The second wave should ideally confirm and solidify the coefficient estimates derived from the previous wave in view of the final and most comprehensive one. Guided by the estimation results from wave 1 non-linearities were included in attribute level effects for the entrance fee. Another important innovation was the definition of agent-specific priors. To illustrate this, for the case of the retailer, the non-linear sensitivity for the yearly entrance fee was written as follows: tar.effects[(u,0.8,1.1)|(u,0.5,0.8)|(u,0.01,0.4)|(u,-0.9,-0.5)]*Tar[1,2,3,4,5] where the last level corresponding to a uniform distribution (u,-0.8,-1.2) was omitted. The specification reflects the large drop in utility when reaching the levels above the SQ fee, causing significant non-linearity compared to the positive utility in the lower ranges. As a comparison, the same attribute for the carrier was given a quite different definition, where the lowest levels (e.g. 200€ and 400€) were not statistically different and were given the same prior. Overall, compared to the retailer, the sensitivities are marked by a notably wider range of variation. This observation again underscores the importance

[7] An advantage of effects coding over dummy coding is that it avoids correlation with the baseline estimate.

of performing separate analysis by agent types and to progressively refine the characterization of the utility structure of different operators.

For own-account, even though the underlying utility specification did not change, there were some important evolutions in the Bayesian priors used. In each of the linear variables, there was a shift from normal to uniform distributions on parameters, to reflect the more consistent findings and avoid any outlying sensitivities. Moreover, while the range of standard deviations remained similar for fare and for the l/u bay probability, there was a marked narrowing of the prior range for the number of bays (e.g. from ??piaz[(n,?0.554,0.5)]?*?Piaz[1,2,3] to ??piaz[u,0.1,0.3]?*?Piaz[1,2,3]).

Table 3 – Efficient design for the 3 waves

Wave	d-error		a-error		Carrier			
	fixed	mean	fixed	mean	**annual entrance fee**	**Probability to find l/u bay free**	**numbero of l/u bays**	
Wave 1	0,2291	0,2721	0,2488	0,2985	tar[(n,-0.50,0.14)] * Tar[1,2,3,4,5]	prob[(n,0.5,0.32)] * PrPiaz[1,2,3]	piaz[(n,0.39,0.28)] * Piaz[1,2,3]	
Wave 2	0,6663	0,6739	1,0537	1,0691	tar.effects[(u,1.2,1.7)|(u,1.2,1.7)|(u,0.4,0.9)|(u,-1.7,-1.2)]* Tar[1,2,3,4,5]	prob[u,0.5,0.8] * PrPiaz[1,2,3]	piaz[u,0.3,0.5] * Piaz[1,2,3]	
Wave 3	0,6219	0,6344	1,5537	1,5895	tar.effects[(u, 1.328, 1.711)|(u,1.144,1.471)|(u,0.130,0.407)|(u,-1.039,-0.708)]* Tar[1,2,3,4,5]	prob.effects[(u,-1.309,-0.914)|(u,0.010,0.315)]* PrPiaz[1,2,3]	piaz.effects[(u,-1.282, -0.886)|(u, 0.255, 0.508)]* Piaz[1,2,3]	
					Retailer			
					entrance fee	**bays prob.**	**bays number**	
Wave 1	0,2291	0,2721	0,2488	0,2985	tar[(n,-0.5,0.14)] * Tar[1,2,3,4,5]	prob[(n,0.5,0.32)] * PrPiaz[1,2,3]	piaz[(n,0.39,0.28)] * Piaz[1,2,3]	
Wave 2	0,5852	0,5891	0,7772	0,7835	tar.effects[(u,0.8,1.1)|(u,0.5,0.8)|(u,0.01,0.4)|(u,-0.9,-0.5)]* Tar[1,2,3,4,5]	prob[u,0.3,0.5] * PrPiaz[1,2,3]	piaz[u,0.1,0.3] * Piaz[1,2,3]	
Wave 3	0,9715	0.987454	0,2040	0,2084	tar.effects[(u,1.144,1.471)|(u,0.130,0.407)|(u,-1.039,-0.708)|(u,-2.520,-1.923)]* Tar[1,2,3,4,5]	prob.effects[(u,-0.603,-0.091)|(u,-0.103,0.229)]* PrPiaz[1,2,3]	piaz.effects[(u,-0.569,-0.173)|(u,0.004,0.257)]* Piaz[1,2,3]	
					Own account			
					entrance fee	**bays prob.**	**bays number**	**time window**
Wave 1	0,2472	0,3465	0,3048	0,4293	tar[(n,-0.32,0.2)] * Tar[1,2,3,4,5]	prob[(n, 0.1852,0.15)] * PrPiaz[1,2,3]	piaz[(n, 0.554,0.5)] * Piaz[1,2,3]	time.effects[(u,-1,0)|(u,0,1)]*finor[0,1,2]
Wave 2	0,2812	0,3405	0,3405	0,3466	tar[u,-0.6,-0.3] * Tar[1,2,3,4,5]	prob[u,0.1,0.4] * PrPiaz[1,2,3]	piaz[u,0.1,0.3] * Piaz[1,2,3]	time.effects[(u,-0.8,-0.3)|(u,-0.1,0.5)]*finor[0,1,2]
Wave 3	0,7016	0,7169	1,0936	1,1211	tar.effects[(u,1.457,1.950)|(u,0.382,0.896)|(u,0.271, 0.740)|(u,-0.836, -0.264)]* Tar[1,2,3,4,5]	prob.effects[(u,-1.313,-0.800)|(u,0.025,0.357)]* PrPiaz[1,2,3]	piaz.effects[(u,-1.154,-0.758)|(u,0.210,0.463)]* Piaz[1,2,3]	time.effects[(u,-0.773, -0.551)|(u,-0.317,-0.081)]*finor[0,1,2]

5.5 Third wave efficient design

The design of the third wave chiefly confirmed the approach previously used. In conclusion, the criteria used to model the design in the previous waves were characterised by agent-specific models with priors based on estimates of ranking data in previous waves including effects-coded attributes where appropriate. Important changes from the previous wave include the specifications of non-linear effects for all attributes concerning retailers and carriers.

Since the last wave of interviews involved, by far, the largest share of interviews an additional feature was introduced to ensure the quality of the data gathered. In previous waves one set of ten identically ordered ranking tasks was administered to all respondents for each agent type. However, for the third wave, to avoid acquiring low data quality due to problems deriving from specific task positioning (e.g incomplete comprehension of early task or fatigue in the later) we developed an algorithm for shuffling the tasks to ensure each task appeared in different positions within the SRE sequence.

It is important to note, that while the d-error and a-error were used as criteria to select designs within a wave, as described above, this reasoning does not apply across waves. Indeed the changes in the utility structure, or/and in the prior values and distributions hampers an appropriate comparison of the goodness of the design in terms of its evolution from earlier waves. The advantage obtained in the use of the approach for the design updating was instead visible from analysing the evolution of the estimations on the data obtained in different waves. In fact, if the utility definition is kept fixed, we observe a marked improvement in the estimates in terms of their precision, i.e. the shrinking size of the standard errors of coefficients.

6 Summary, conclusions and future research

In this paper we illustrates the potential of using a SRE to explore the acceptability of main UFT stakeholders for innovative urban freight policy measures. The paper reports a selected literature review of both agent interaction in freight and experimental design followed by a description of the study context and the roman freight LTZ. This motivates and justifies our approach aimed at modelling preferences of three different agent types and their likely interactions with their "typical" business partners. The section overviewing the development of the survey instrument includes a description of the essential activity of organizing focus group meetings with local policy makers, demand (retailers) and supply (transport providers). These proved fundamental for identifying the main freight distribution problems in Rome's LTZ and provide a clear view of problem perceptions and view of possible solutions. The main output from this consultation phase was the identification of the attributes considered most critical for inclusion in potential policy-mixes to be implemented. Several criteria were employed in selecting the specific attributes used in the SRE. This approach assured two positive outcomes. On the one hand it provided attributes considered relevant by interested stakeholders and, on the other, it identified attributes viewed as significant and important for a balanced group of stakeholders. In fact, policy evaluations ought to address both relevant and collectively important issues/attributes aimed at providing policy makers with indications of potentially effective and acceptable solutions. Subsequently, the paper describes in detail the various phases of the development and refinement of a SRE for the survey in Rome’s LTZ. In fact, a major innovation of the present research is the sub-division of the analysis to consider three different agent types: carriers, retailers and own-account. Most of the recent literature on *city logistics* acknowledges, in principle, the importance of agent-specific measures. The present study has, for the first time, to the best of our knowledge, acquired the necessary data to formulate analytically sound and empirically *ex-ante* verifiable proposals incorporating knowledge of agent-specific behaviour. The main problems and potentially feasible solutions identified in stakeholder surveys were extremely useful in the progressive specification of the various attributes conceived to map the preferences of each agent type. Innovative solutions were also adopted in the questionnaire design strategy pertaining to a novel use of prior preference data to capture the trade-offs of different agent types. More precisely, the design strategy relied on state-of-the-art efficient design theory.

The data acquired will allow for the estimation of agent-specific models that are useful, in particular for policy maker, in analyzing the most promising and *ex-ante* acceptable policy-mixes. The results obtained are not only reliable but also relevant under a policy implementation and evaluation scenario. The research produced is not only innovative under several aspects but also provides socially relevant results.

In brief, the research approach described in this paper allows for the: 1) identification of the most relevant problems for Rome’s LTZ for the most important stakeholders; 2) enumeration of potentially feasible and relevant policies based on stakeholders’ opinions and preferences; 3) the design of a SRE differentiated by using agent-specific attributes and specification.

The data acquired open the door to several promising future research explorations. A central extension concerns the estimation of the potential for shared acceptability of policy interventions by “couples of agents”, namely retailers and freight carriers. Moreover, it would be of interest to detect potential distribution channel effects for each category of goods. Another important extension would be to include and evaluate other potentially relevant attributes in the policy mix scenarios such as time window exemptions, entrance fee exemptions, etc. The reactions to such policies are likely to be strongly differentiated for different agents and have rarely been explored experimentally in past research. A further point that would be important to investigate relates to the reaction to extended “what if” scenarios. This would allow practitioners to predict the degree of acceptance and foresee behavioural adjustments as a response to wider contextual changes, such as fuel-price changes, tax restructurings or changes in related policies such as parking.

Finally, we would like to stress the great benefits provided by the methodology proposed in terms of greater accuracy of the estimates obtainable given a specific budget for interview administration or, alternatively, the reduction of the budget needed to reach a predetermined level of accuracy. This last aspect may be crucial in different empirical research situations.

7 Bibliographical references

Aaker, D., Day, G., (1990), Marketing Research, Wiley and Sons.

Brewer, A. and Hensher, D.A. (2000). “Distributed work and travel behaviour: The dynamics of interactive agency choices between employers and employees”, *Transportation*, Vol. 27, pp. 117-148.

Crask, M., Fox, R., (1987), “An exploration of the interval properties of the three commonly used marketing research scales: a magnitude estimation approach”, *Journal of the Market Research Society*, 29 (3), pp. 317-39.

de Jong, G., M. Ben-Akiva (2007) A micro-simulation model of shipment size and transport chain choice. Transportation Research Part B 41: 950–965.

DfT, (2007) A guide on how to set up and run Freight Quality Partnerships. Department for transport, http://www.freightbestpractice.org.uk/case-studies (accessed on 15/01, 2010).

Ferrini, S., and R., Scarpa (2007) Designs with a-priori information for non-market valuation with choice-experiments: a Monte Carlo study, Journal of Environmental Economics and Management, 53:342-363.

Filippi, F., and Campagna, A. (2008). "Indagine sulla distribuzione delle merci a Roma, nell’ambito dello Studio di settore della mobilità delle merci a Roma." CTL: Centre for studies on Transport and Logistics and ATAC: Agency of mobility of Rome’s municipality.

Gray, R. (1982) Behavioural Approaches to Freight Transport Modal Choice. Transport Reviews, 2(2), pp. 161-184.

Hensher, D. A. and Puckett, S. M. (2008) “Power, concession and agreement in freight distribution chains: subject to distance-based user charges”, International Journal of Logistics Research and Applications, 11: 2, 81-100

Hensher, D., and Figliozzi, M. A. (2007). "Behavioural insights into the modelling of freight transportation and distribution systems." *Transportation Research Part B: Methodological*, 41(9), 921-923.

Huber, J. and K. Zwerina (1996) The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research* 33, 307-317.

Johnson, L., Desvouges, W., (1997), Estimating Stated Preferences with Rated-Pair data: Environmental, Health, and Employment Effects of Energy Programs, *Journal of Environmental Economics and Management*, 34, 79-99.

Kanninen, B.J. (2002) Optimal Design for Multinomial Choice Experiments. *Journal of Marketing Research* 39, 214-217.

Kessels, R., Goos, P. and Vandebroek, M. (2006) A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research*, 43(3), 409-419.

Liedtke, G., and H. Schepperle (2004) Segmentation of the Transportation Market with Regard to Activity–based Freight Transport Modelling, International Journal of Logistics: Research and Applications, Vol. 7, No. 3, pp. 199-218.

Louviere, J., (1988), Analyzing Decision Making Metric Conjoint Analysis, *Quantitative Applications in the Social Sciences* (67), Sage University Paper, SAGE Publication Inc.

Louviere, J., (1992), Experimental Choice Analysis: Introduction and Overview, *Journal of Business Research*, 24, pp. 89-95.

Louviere, J.J. and D.A. Hensher (1983) Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. *Journal of Consumer Research* 10, 348-361.

Louviere, J.J., and G. Woodworth (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregated data. *Journal of Marketing Research* 20, 350-367.

Louviere, J.J., D.A. Hensher, and J.D. Swait (2000) *Stated Choice Methods—Analysis and Application*. Cambridge University Press, UK.

Ngene 1.0, Choicemetrics Pty Ltd, http://www.choice-metrics.com.

Ogden, K. W. (1992). *Urban goods movement: a guide to policy and planning,* Ashgate, Aldershot, UK.

Ortuzar, J., Garrido, R. (1994) “A practical assessment of stated preferences methods”, *Transportation*, 21, pp. 289-305.

Puckett, S. M., and Greaves, S. (2009). "Estimating Policy-Induced Changes in Freight Vehicle Emissions in Sydney." Institute of Transport and Logistics Studies Working Paper, The University of Sydney.

Puckett, S.M., Hensher, D.A. (2006) Modelling interdependent behaviour utilising a sequentially-administered stated choice experiment: analysis of urban road freight stakeholders, Paper presented at the 2006 Conference of the International Association of Travel Behaviour Research, Kyoto, Japan.

Puckett, S.M., Hensher, D.A. (2008) The role of attribute processing strategies in estimating the preferences of road freight stakeholders under variable road user charges. Transport. Res. E., 44,379-395.

Roorda, M. J., R. Cavalcante, S. McCabe, and H. Kwan (2010) A conceptual framework for agent-based modelling of logistics services. Transportation Research Part E: Logistics and Transportation Review, Vol. 46(1), pp. 18-31.

Samimi, A., A. Mohammadian, and K. Kawamura (2009) Integrating supply chain management concept in a goods movement microsimulation. IEEE International Conference on Service Operations, Logistics and Informatics, July 22 - 24, 2009, Chicago, IL, USA.

Sándor, Z., and M. Wedel (2001) Designing Conjoint Choice Experiments Using Managers’ Prior Beliefs. *Journal of Marketing Research* 38, 430-444.

Sándor, Z., and M. Wedel (2002) Profile Construction in Experimental Choice Designs for Mixed Logit Models, *Marketing Science* 21(4), 455-475.

Sándor, Z., and M. Wedel (2005) Heterogeneous conjoint choice designs. *Journal of Marketing Research* 42, 210-218.

Southworth, F. (2003) Freight Transportation Planning: Models and Methods. Transportation systems planning: methods and applications, by K. G. Goulias.

STA. (2001). "Quadro conoscitivo del problema della distribuzione delle merci nel Centro Storico di Roma." Rome.

Stathopoulos, A., E. Valeri, E. Marcucci, V. Gatta, A. Nuzzolo, A. Comi (2011) “Urban freight policy innovation for Rome’s LTZ: A stakeholder perspective”, in Macharis, C., and Melo, S. (Eds.), Multiple views on City Distribution: a state of the art, Edward Elgar Publishing, London, UK.

VOLVO REPORT - (2010), *Innovative solutions to freight distribution in the complex large urban area of Rome*, Final report, project SP-2007-50.

Wisetjindawat, W., S. Matsumoto, and K. Sano (2005) Supply Chain Simulation for Modeling the Interactions in Freight Movement. Journal of the Eastern Asia Society for Transportation Studies, 6: 2991 – 3004.

Yang, C. H., J. Y. J. Chow, and A. C. Regan (2009) State-of-the Art of Freight Forecasting Modeling: Lessons Learned and the Road Ahead. Proceedings, 88th Annual Meeting of the Transportation Research Board, Washington DC, 11–15 January 2009.

Appendix

Main regulatory characteristics of Rome freight LTZ

General regulation	
Laden weight < 35 q	Laden weight > 35 q
Transit and parking allowed from 20.00 to 10.00 and 14.00 to 16.00 and prohibited otherwise	Transit and stopovers permitted from 20.00 to 7.00 and prohibited otherwise
Exceptions from time window (around the clock transit and parking)	
Laden weight < 35 q	Laden weight > 35 q
1. Transport of perishable foods, pharmaceuticals, newspapers and precious goods	1. Trucks with justified request detailing time, place and route (for instance house moving)
2. All courier and transport companies operating as third account (if enrolled in the “National registry of auto transporters”)	
3. Trucks involved in cleaning and maintenance services on account of the municipality or ATAC	
Fee reductions 50% reductions offered for electric cars and 25% reduction for CH4, GPL and hybrid motor/fuel	