CREI Working Paper no. 6/2013

EX-POST IMPLICATIONS OF EX-ANTE DATA ACQUISITION STRATEGIES IN MULTI-AGENT-TYPE URBAN FREIGHT POLICY EVALUATION

by

Valerio Gatta

University of Roma Tre and CREI

Edoardo Marcucci

University of Roma Tre and CREI

available online at http://host.uniroma3.it/centri/crei/pubblicazioni.html

ISSN 1971-6907

Outputs from CREI research in progress, as well contributions from external scholars and draft reports based on CREI seminars and workshops, are published under this series. Unless otherwise indicated, the views expressed are attributable only to the
Ex-post implications of ex-ante data acquisition strategies in multi-agent-type urban freight policy evaluation

Gatta V.a1, Marcucci E.a

a DISP/CREI, University of Roma Tre, Rome, Italy.

Abstract

Data quality plays a relevant role in policy evaluation. Data quality is also strictly linked to questionnaire development and administration strategies. Willingness to pay estimates might be influenced by the data acquisition method adopted if an ex-post sophisticated modelling cannot compensate for the low quality data acquired. We test this hypothesis by comparing agent-generic vs. agent-specific data acquisition. This methodologically relevant question is investigated with respect to urban freight transport policy evaluation. In this field, in fact, heterogeneity among stakeholders’ (i.e. retailers and transport providers) preferences is allegedly important since retailers, demanding freight transportation services, and transport providers, offering them, are characterised by, structurally and functionally motivated, differences in preferences.

Results indicate that agent-specific data acquisition should be adopted. The analyst ought to take the longer data acquisition route when a strong and well rooted a priori knowledge suggests different agent-types’ preferences are at work. This reflections are particularly valuable given ex-post agent-specific modelling cannot compensate for ex-ante agent-generic data acquisition procedure. Once perpetrated, the original sin cannot be redeemed.
Keywords: Urban freight transport, policy evaluation, experimental design, data acquisition, willingness to pay, agent-specific.
1. Introduction

Populations have heterogeneous preferences. This directly translates in daunting chores for analysts when modelling individual behaviour. Recent methodological advancements in modelling single individual choices have created new research opportunities, underlining both the need for a deeper analysis of choice and its highly informative potential (Frischknecht et al., 2014). However, it is important to recognize that, notwithstanding the ex-post sophistication of the estimation methodologies employed, end-results are highly dependent on the quality of available data. These considerations pertain to policy evaluation issues too. This paper systematically investigates the impact two different data acquisition approaches have on parameter estimates. More in detail, we analyse and compare, in terms of willingness to pay (WTP) measures, the results obtained from an agent-generic vs. agent-specific approach with equally sophisticated ex-post data modelling.

The research question is investigated and discussed within the realm of urban freight transport (UFT) policy implementation. This, in fact, represents, as well documented in the literature, a research field where heterogeneity among stakeholders’ preferences plays an important role. Notwithstanding this paper has also a methodological research bearing, one has to consider that biased estimates might also influence policy implementation. This is especially true considering the growing relevance of freight deliveries in urban areas and the impact they generate on environment, health and economic growth. In fact, poor decisions concerning UFT policies might provoke substantial drawbacks when different stakeholders, characterised by conflicting objectives, are involved.

In particular, the case study used to test this hypothesis refers to a stated ranking exercise (SRE) administered to a sample of 254 interviewees in the limited traffic zone (LTZ) in Rome’s city centre (Stathopoulos et al., 2011; Marcucci et al., 2012). We estimate simple multinomial logit (MNL) models rather than more flexible ones so to emphasise the role data acquisition methods play.

We illustrate the importance and implications of adopting an agent-specific efficient design strategy to elicit agents’ preferences for alternative UFT policies. In particular, the paper shows that the biases in WTP estimates are substantial when inappropriate agent-generic
data acquisition approach is adopted. Furthermore, *ex-post* agent-specific model estimation cannot compensate for an agent-generic data acquisition procedure\(^2\). Once perpetrated, the original sin cannot be redeemed.

The paper is structured as follows. Section 2 proposes a detailed and updated literature review testifying the growing number of papers published on UFT related issues and of their citations. Section 3 describes the two alternative data acquisition processes used. Section 4 reports the econometric results and discusses their policy implications. Section 5 concludes.

### 2. Literature review

This section reports a detailed, systematic and updated literature review of the scientific papers published within the ISI web of knowledge database\(^3\) on UFT for the 1864-2013 period. Since there is no consistent definition of UFT and, more in general of city logistics, it is important to employ different research keywords to locate the relevant scientific literature (Wolpert and Reuter, 2012). In order to minimize, as much as possible, research arbitrariness a pre-determined plan was systematically followed. The 3 keywords used for literature research priming were: “urban freight transport”, “city logistics” and “urban freight”\(^4\), producing a total of 100 matches. An in-depth analysis of the keywords used in these articles yielded an additional set of 10 items subsequently employed in a wider-ranging database investigation. In particular, the additional keywords employed in the second search effort were: “urban freight delivery”, “urban freight measures”, “urban freight planning”, “urban freight policy”, “urban freight problems”, “urban freight sustainability”, “urban goods delivery”, “urban goods distribution”, “urban goods movement” and “urban supply chain”, generating 9 additional matches.

---

\(^2\) One could suppose to compensate the lower quality data, acquired via an agent-generic approach, by using more sophisticated models explicitly accounting for preference heterogeneity (see *e.g.* Marcucci and Gatta, 2012) such as, for instance, latent class. However, it is important to note that local policy makers usually differentiate UFT policies on the base of clearly determined, stable and easily verifiable characteristics such as being a member of a specific agent-type (e.g. retailer, transport provider, *etc.*).

\(^3\) We concentrated on ISI web of knowledge alone since it is generally recognized as the most exclusive research database given the quality standards journals need to respect in order to be indexed.

\(^4\) The research was performed inserting the keywords between inverted commas so to assure, from a Boolean perspective, the presence of all the words in either the title of the paper or in the keywords used.
The yearly\textsuperscript{5} distribution of the articles is reported in Figure 1. The steady increase in the number of relevant publications is evident.

**Figure 1 – UFT related articles, yearly distribution**

Even steeper is the growth in the number of citations these articles have produced (Figure 2).

**Figure 2 – Citations of UFT related articles, yearly distribution**

\textsuperscript{5}Since 2013 is not over yet the results referring to this year were not included since they would not be comparable to those obtained for the other periods investigated. A total of 13 articles have been published between January and October 2013.
Given the rising number of publications, in order to focus on the most recent research endeavours, the paper concentrates on the 55 works published in the last 4 years only.

Table 1 illustrates the geographical distribution of the articles using the nationality of the corresponding author as an indicator. The results show that the highest number of papers were published in USA (18%) followed by UK, Spain and Italy (11%).

Table 1 – UFT related articles, geographical distribution (2010-2013)

<table>
<thead>
<tr>
<th>Nation</th>
<th>N° of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>10</td>
</tr>
<tr>
<td>Italy</td>
<td>6</td>
</tr>
<tr>
<td>Spain</td>
<td>6</td>
</tr>
<tr>
<td>United kingdom</td>
<td>6</td>
</tr>
<tr>
<td>Sweden</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>3</td>
</tr>
<tr>
<td>France</td>
<td>3</td>
</tr>
<tr>
<td>Greece</td>
<td>3</td>
</tr>
<tr>
<td>Mexico</td>
<td>2</td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
</tr>
<tr>
<td>Austria</td>
<td>1</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
</tr>
<tr>
<td>Colombia</td>
<td>1</td>
</tr>
<tr>
<td>Croatia</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2 reports the list of journals where the above mentioned articles were published and the respective number of occurrences. Journal of Transport Geography (7%) and Transportation Research Part A (7%) hosted the highest number of UFT related articles. Additionally, we underline the relatively high number of review articles in our list (e.g. Cherret et al., 2012, Lindholm and Behrends, 2012; Nuzzolo et al., 2013; Woodburn, 2012). A possible explanation, even without a comparative benchmark and/or a counterfactual example of this phenomenon, is the need of clarification, classification and, as Anand et al. (2012) suggest, clear ontological definition of what UFT research means.

Specific thematic concentrations emerge from a first qualitative article clustering based on title, keywords and abstract inspection. In particular, the most numerous group of articles relate to vehicle routing and efficiency maximization of freight movements within city boundaries (e.g. Ehmke et al., 2012; Hemmelmayer et al., 2012; Motraghi et al., 2012; Nguyen et al., 2012; Perboli et al., 2013; Pillac et al., 2012). Another set concentrates on UFT environmental impacts and its regulation (e.g. Arvidsson, 2013; Figliozzi, 2011; Lee et al., 2012; Sathaye et al., 2010a, 2010b). Data acquisition issues have also catalysed consistent research efforts (e.g. Allen et al., 2012; McCabe et al., 2013; Roorda, 2011), while minor attention has been paid to the effects of UFT disruptions (e.g. Friesz et al., 2011; Mamasi et al., 2013) and multi-agent modelling (e.g. Ballantyne et al., 2013; Teo et al., 2012).
In general, independently from the specific time span considered and database investigated, UFT is reputed by a consistent number of authors (de Jong and Ben-Akiva, 2007; Gray, 1982; Hensher and Figliozzi, 2007; Roorda et al., 2010; Samimi et al., 2009; Southworth, 2003; Wisetjindawat et al., 2005; Yang et al., 2009) one of the most promising field for agent-based modelling. Freight movements should be explained via the underlying motivations deriving from the relative convenience each stakeholder has when choosing. These considerations are important for understanding freight agents’ motivations when reacting to different policy
mixes, dealing with specific constraints (e.g. time windows) and accounting for varying alternative incentives (e.g. price rebates for new vehicles). It is essential to jointly consider both the tools available to policy makers and the elements influencing freight operators if one aims at understanding the potential impacts policies might produce (Puckett and Greaves, 2009). One needs to identify incentives/disincentives, considered acceptable by agents involved, in order to quantify the impact they might produce on the status quo (SQ). Behavioural models explicitly consider stakeholders’ utility maximization efforts. This approach rests upon an unambiguous identification of the decision maker. In fact, only in this case agent-based models will be able to describe and forecast specific stakeholders’ behaviour (Liedtke and Schepperle, 2004). Notwithstanding the recognition of a deep rooted need for developing a behaviourally consistent approach to UFT modelling, it is interesting to note that only a limited number of articles have explicitly modelled agent-specific behaviour (e.g. Bolis and Maggi, 2003; Hensher and Puckett, 2005; Hensher et al., 2007) and even shorter is the list of those adopting an agent-specific perspective for both data acquisition and modelling (i.e. Gatta and Marcucci, 2013; Marcucci and Gatta, 2013, 2014; Marcucci et al., 2013; Stathopoulos et al., 2012). Furthermore, to the best of our knowledge, the present paper alone systematically compares two different data acquisition methods and describes their respective impact on WTP measures. These reflections assume a particular relevance considering, as Hensher and Figliozzi (2007) persuasively contend, that standard approaches, not accounting for the intricacy of freight movements, are structurally unfit to predict the most likely reactions to perturbations of the SQ situation.

3. Data acquisition

The data acquisition process described in this section refers to a SRE administered in Rome's LTZ between September 2009 and February 2010 aimed at eliciting agents’ preferences with respect to possible UFT policy variations. In the late 80s, access to a 5-km² area in the city centre was restricted to non-resident vehicles. Nowadays a specific legislation limits entrance to passenger and freight vehicles alike. The constraints essentially refer to Euro-vehicle standards. A yearly access fee of 565€
is levied for each vehicle; time windows apply, albeit not systematically enforced, to freight vehicles while numerous exemptions are contemplated (Stathopoulos et al. 2011). Administering questionnaires on UFT policy is not an easy task. The main difficulties arise from: 1) privacy concerns; 2) limited participation; 3) high interviewing costs. Attribute definition, selection and customization are discussed in previous papers where a detailed description of the Bayesian efficient design developed is also reported (Marcucci et al., 2012, 2013; Stathopoulos et al., 2011). Each SRE alternative is characterised by attributes taking at least three levels to describe variation ranges presented to respondents. The attributes included are: 1) number of loading and unloading bays (LUB); 2) probability of finding a loading and unloading bay free (PLUBF); 3) entrance fee (EF). LUB and PLUBF have 3 levels while EF has 5. Notwithstanding the agent-specific level investigated, attributes were defined on the base of agent-types’ shared relevance. This choice is motivated by the recognition that UFT policy interventions are usually uniformly applied to all agents involved. Sufficiently broad level ranges were used in order to avoid possible behavioural misinterpretations. In fact, seemingly non-trading behaviour among interviewees could, de facto, be the result of insufficient variation in level ranges producing an undetectable impact on utility. Joint stakeholder meetings were an important source of information concerning attribute ranges. SREs reported three policy options always including the SQ alternative. This paper tests the impact two alternative data acquisition techniques have on WTP measures. More in detail, we compare an agent-specific design with an agent-generic one. Retailers, transport providers and own-account agents constitute the level of disaggregation adopted in investigating UFT policies (Marcucci et al., 2013). The first two agent-types are well identified in the literature and we concentrate on them alone since a more articulated utility specification (i.e. including time windows) was used for own-account operators thus impeding the comparison among all three agent-types.

---

6 Volvo Report (2010) provides, in more detail, an overview of the relation between stakeholder surveys and attributes used.

7 The attribute selection criterion chosen had to ensure all stakeholders were, at least in principle, interested in the policy characteristics presented in the SREs. Policy attributes considered significant by one actor only (e.g. tradable permits for policy makers) were excluded.
The advantages of adopting an efficient design strategy are well documented in the literature (Bliemer and Rose, 2005; Huber and Zwerina, 1996; Sándor and Wedel, 2001). Taking the beneficial impacts of this strategy for granted, we investigate the effects of an agent-specific approach. In other words, we generate multiple designs corresponding to different subsegments of the sampled population (see e.g. Rose and Bliemer, 2006; Sándor and Wedel, 2005). In fact, we underline the strong a priori nurtured concerning the structurally different preferences that supposedly characterise retailers and transport providers. The ex-ante, theoretically based, considerations, suggesting a consistent distinction between these two agent-types’ preferences, are based on their respective structural and functional characteristics. In fact, while retailers demand freight transportation services, transport providers offer them. This is considered sufficient to assume that they are characterised by contrasting interests explaining heterogeneous policy preferences. These different preferences for UFT policy attributes should, in our opinion, also be reflected in the efficient design development process too. The results confirm the correctness, pertinence and empirical relevance of the a priori hypothesis.

The two datasets used, related to an agent-generic and agent-specific approach, were acquired by employing a Bayesian efficient design with Ngene 1.1 software (Rose and Bliemer, 2012). In particular, they incorporate indications on magnitude, sign and distribution for each attribute coefficient derived from MNL model estimates using previously acquired data. Bayesian priors account for uncertainty in parameter estimates. The size of the spread for the a priori assumptions was determined using the estimated confidence intervals from previous rounds of interviews. D-efficiency (i.e. AVC matrix determinant minimization) and a-efficiency (AVC matrix trace minimization) were the main drivers for design development while level and utility balance criteria were also considered. Both datasets include an equal number of interviews for each agent-type. Notwithstanding both samples used are small in size, they are in line with the commonly suggested rules-of-thumb to determine a sufficient number of observations to get statistically significant parameters estimates (Orme, 1998). Furthermore, sampling more respondents will, at the margin, have only a small effect on the standard errors of the parameter estimates (Rose and Bliemer, 2013). These considerations
justify the adoption of a Bayesian efficient design and the shunning of an orthogonal one since robust results can be obtained also with a small number of interviews.

Thanks to the specific data acquisition processes adopted, the paper also shows that using an agent-generic efficient design strategy, producing low quality data, cannot be compensated ex-post via an agent-specific model estimation approach.

4. ECONOMETRIC RESULTS AND POLICY IMPLICATIONS

This section reports the results, deriving from the SRE, which are shown to be influenced by the data acquisition method used and modelling assumptions. One can either follow an agent-generic or an agent-specific approach as described in the previous section.

The choice between the two alternatives depends on the a priori assumptions made concerning the homogeneous/heterogeneous preference hypothesis concerning the agents to be interviewed. If the interviewees are considered having homogenous preferences, a single MNL model is estimated on the base of the responses obtained from both agent-types, otherwise two MNL models, one for each agent-type, are estimated.

We describe and compare the impact on WTP estimates of three different data acquisition/data modelling strategies; namely: agent-generic data acquisition and agent-generic modelling (S1); agent-generic data acquisition and agent-specific modelling (S2); agent-specific data acquisition and agent-specific modelling (S3). The methods described are compared on the base of their relative model fit indicators (i.e. Akaike (AIC), Schwarz-Bayesian (SBIC) and Hannan-Quinn (HQIC) information criteria) and policy effects (i.e. WTP measures).

All model results reported are based on stated choice data. Estimated coefficients refer to normalized attributes so to circumvent unit of measurement issues and facilitating utility impact assessments.

WTP point estimates for an additional unit of the desired attribute are reported together with their confidence intervals calculated via the Delta method.

---

8 In fact, we explode a ranking exercise of (n) alternatives into (n-1) choice data.

9 Attributes are normalised dividing each attribute level by its own minimum value (see Marcucci et al., 2013). As an example, unnormalised LUB=400, 800, 1200 while normalised LUB=1, 2 and 3 (e.g. 1200/400).
Table 3 reports S1 results. All coefficients, with the exception of the two alternative specific constants (ASCs) referring to the hypothetical non-SQ alternatives, are statistically significant and with the expected sign: positive for LUB and PLUBF while negative for EF. In fact, an increase in either the number of loading and unloading bays or in the probability of finding them free has a positive impact on utility. On the contrary, an increase in the annual cost negatively affects utility. As imaginable EF, given the attribute level ranges used in the SRE, plays the lion part in explaining preferences. Moreover, agents are willing to pay 0.22€ annual fee for each additional loading and unloading bay and nearly 11€ for an additional unit of probability of finding loading and unloading bays free.

Table 3 - S1: MNL, agent-generic data acquisition and agent-generic modelling

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>St. Err.</th>
<th>t-stat</th>
<th>WTP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>0.337</td>
<td>0.079</td>
<td>4.24</td>
<td>-0.22 (0.05)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.419</td>
<td>0.083</td>
<td>5.05</td>
<td>-11.13 (2.10)</td>
</tr>
<tr>
<td>EF</td>
<td>-0.753</td>
<td>0.057</td>
<td>-13.18</td>
<td>-</td>
</tr>
<tr>
<td>ASC1</td>
<td>0.096</td>
<td>0.173</td>
<td>0.55</td>
<td>-</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.017</td>
<td>0.181</td>
<td>0.09</td>
<td>-</td>
</tr>
</tbody>
</table>

Log likelihood function: -461.53
N° of observations: 653
Info. Criterion AIC: 1.429
Finite Sample AIC: 1.429
Info. Criterion BIC: 1.463
Info. Criterion HQIC: 1.442

*WTP for an additional unit of the corresponding attribute. In brackets the standard error estimated using Delta method.

S2 adopts an agent-specific model investigation strategy. In particular, it differentiates between the two agent-types considered (Table 4).

---

10 In a choice-modeling framework, WTP for a given attribute is obtained by dividing its marginal coefficient by that of cost. Due to attribute normalization, the ratio between the two coefficients has to be multiplied by the minimum level of cost attribute (i.e. 200) and divided by the minimum level of the attribute at the numerator (i.e. 400 for LUB and 10 for PLUBF).

11 The Delta method relies on the joint and strong normality assumption for both the maximum likelihood estimates and their ratio. In practice, the Delta method is usually recommended when the cost coefficient is highly significant (Gatta et al., 2013).
Similar considerations on significance and interpretation, as expressed for S1, apply to this case. Only small WTP measures differences are detected. However, accounting for an inter-agent heterogeneity one discovers that the two agent-types considered have opposite sensitivities to policy attributes. Retailers are more interested in PLUPF suggesting they would prefer the adoption of a light intervention policy relying more on regulation rather than on infrastructural investments. On the contrary, transport providers are more interested in LUB.

Table 4 - S2: MNL, agent-generic data acquisition and agent-specific modelling

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>St. Err.</th>
<th>t-stat</th>
<th>WTP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retailers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUB</td>
<td>0.264</td>
<td>0.104</td>
<td>2.54</td>
<td>-0.21 (0.08)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.377</td>
<td>0.109</td>
<td>3.46</td>
<td>-11.99 (3.33)</td>
</tr>
<tr>
<td>EF</td>
<td>-0.629</td>
<td>0.070</td>
<td>-8.97</td>
<td>-</td>
</tr>
<tr>
<td>ASC1</td>
<td>-0.053</td>
<td>0.226</td>
<td>-0.23</td>
<td>-</td>
</tr>
<tr>
<td>ASC2</td>
<td>-0.132</td>
<td>0.236</td>
<td>-0.56</td>
<td>-</td>
</tr>
</tbody>
</table>

Log likelihood function: -283.00
N° of observations: 370
Info. Criterion AIC: 1.557
Finite Sample AIC: 1.557
Info. Criterion BIC: 1.610
Info. Criterion HQIC: 1.578

* WTP for an additional unit of the corresponding attribute. In brackets the standard error estimated using Delta method.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>St. Err.</th>
<th>t-stat</th>
<th>WTP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport providers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUB</td>
<td>0.461</td>
<td>0.126</td>
<td>3.65</td>
<td>-0.23 (0.06)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.501</td>
<td>0.132</td>
<td>3.79</td>
<td>-10.12 (2.51)</td>
</tr>
<tr>
<td>EF</td>
<td>-0.990</td>
<td>0.104</td>
<td>-9.53</td>
<td>-</td>
</tr>
<tr>
<td>ASC1</td>
<td>0.322</td>
<td>0.275</td>
<td>1.17</td>
<td>-</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.283</td>
<td>0.286</td>
<td>0.99</td>
<td>-</td>
</tr>
</tbody>
</table>

Log likelihood function: -169.78
N° of observations: 283
Info. Criterion AIC: 1.235
Finite Sample AIC: 1.236
Info. Criterion BIC: 1.300
Info. Criterion HQIC: 1.261

* WTP for an additional unit of the corresponding attribute. In brackets the standard error estimated using Delta method.
S3 assumes an agent-specific formulation both for data acquisition and modelling. Results, reported in Table 5, are policy relevant and indicate a remarkable goodness-of-fit. S3 is characterised, when compared to S1 and S2, by the lowest information criteria values and entails substantially different WTP measures. In fact, in this case, EF has a much higher impact on utility (this is more evident for retailers) thus implying lower WTP measures for both LUB and PLUBF. Moreover, the two ASCs are positive and statistically significant indicating, ceteris paribus, a negative attitude towards the SQ alternative.

Table 5 - S3: MNL, agent-specific data acquisition and agent-specific modelling

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>St. Err.</th>
<th>t-stat</th>
<th>WTP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>0.281</td>
<td>0.151</td>
<td>1.86</td>
<td>-0.12 (0.05)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.390</td>
<td>0.175</td>
<td>2.23</td>
<td>-6.51 (2.60)</td>
</tr>
<tr>
<td>EF</td>
<td>-1.197</td>
<td>0.117</td>
<td>-10.23</td>
<td>-</td>
</tr>
<tr>
<td>ASC1</td>
<td>0.686</td>
<td>0.350</td>
<td>1.96</td>
<td>-</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.658</td>
<td>0.308</td>
<td>2.14</td>
<td>-</td>
</tr>
</tbody>
</table>

Retailers

Log likelihood function: -251.98
N° of observations: 452
Info. Criterion AIC: 1.137
Finite Sample AIC: 1.137
Info. Criterion BIC: 1.183
Info. Criterion HQIC: 1.155

WTP for an additional unit of the corresponding attribute. In brackets the standard error estimated using Delta method.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>St. Err.</th>
<th>t-stat</th>
<th>WTP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>0.548</td>
<td>0.159</td>
<td>3.46</td>
<td>-0.20 (0.05)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.468</td>
<td>0.183</td>
<td>2.56</td>
<td>-6.83 (2.28)</td>
</tr>
<tr>
<td>EF</td>
<td>-1.372</td>
<td>0.149</td>
<td>-9.23</td>
<td>-</td>
</tr>
<tr>
<td>ASC1</td>
<td>0.255</td>
<td>0.315</td>
<td>0.81</td>
<td>-</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.612</td>
<td>0.325</td>
<td>1.88</td>
<td>-</td>
</tr>
</tbody>
</table>

Transport providers

Log likelihood function: -199.75
N° of observations: 357
Info. Criterion AIC: 1.147
Finite Sample AIC: 1.148  
Info.Criterion BIC: 1.201  
Info. Criterion HQIC: 1.169  

* WTP for an additional unit of the corresponding attribute. In brackets the standard error estimated using Delta method.

From a purely statistical point, one should suggest policy makers to have more faith in S3 results given its higher fit to the data. S3 is thus taken as a benchmark while S1 and S2 are used for measuring the potential bias. A quantification of WTP distortions is reported in Table 6 where policy interventions are simulated on the base of the highest possible level that can be adopted for each attribute in the SRE. Using homogeneous data acquisition and modelling would induce the adoption of inefficient policy measures. In fact, in the case of LUB, policy makers would increase the annual entrance fee by nearly 180€ which is twice as much the WTP retailers are willing to pay (+104%). This would, most likely, cause serious discontent to them while transport providers would mildly oppose the policy change due to a small WTP overestimation (+12%). With respect to PLUBF, noticeable distortions emerge for both retailers (+71%) and transport providers (+63%). Similar considerations apply when looking at the full policy intervention package based on LUB and PLUBF level variations. The error margin is not reduced when considering S2. All WTP measures are overestimated in this case too. Compared to S1, the distortions are, in some cases, even higher (e.g. PLUBF for retailers) while lower in others (e.g. PLUBF for transport providers). Moreover, policy makers would incorrectly conclude that retailers are willing to pay more than transport providers for the full policy package.

**Table 6 – S1, S2 and S3 WTP comparisons**

<table>
<thead>
<tr>
<th>Policy</th>
<th>Agent-type</th>
<th>Benchmark</th>
<th>Distortion</th>
</tr>
</thead>
<tbody>
<tr>
<td>+800 LUB</td>
<td>Retailers</td>
<td>S3</td>
<td>vs. S1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>88€</td>
<td>+104%</td>
</tr>
<tr>
<td>+20% PLUBF</td>
<td>Transport providers</td>
<td>160€</td>
<td>+12%</td>
</tr>
<tr>
<td></td>
<td>Retailers</td>
<td>130€</td>
<td>+71%</td>
</tr>
<tr>
<td></td>
<td>Transport providers</td>
<td>137€</td>
<td>+63%</td>
</tr>
</tbody>
</table>
Results suggest the adoption of an agent-specific perspective in both data acquisition and treatment. In fact, employing an undifferentiated approach would provoke severe policy distortions. Even more important is considering heterogeneity only when modelling data alone. In fact, this would not reduce the risk of erring thus suggesting the adoption of a well-defined data acquisition strategy too.

### 5. DISCUSSION AND CONCLUSION

This paper analyses and compares the impact an agent-specific vs. agent-generic data acquisition process might have on WTP estimates.

The datasets used in this paper are acquired via a SRE performed in Rome’s city centre with respect to potential freight policy changes in the LTZ regulation. The paper investigates both retailers and transport providers’ preferences. It shows the relevant distortions that a simplistic data acquisition process might provoke. In our case, in fact, it is reasonable to assume that supply (i.e. transport providers) and demand (i.e. retailers) agents have structurally different preferences and this should also be duly considered in questionnaire development and administration. Results show that once an *ex-ante* agent-generic data acquisition strategy is adopted and lower data quality acquired, no *ex-post* agent-specific data modelling can compensate for the original choice made. This result is of particular relevance given the biases in WTP measures it implies. These considerations should induce the analyst to take the longer data acquisition route (i.e. agent-specific approach) when a strong and well rooted *a priori* knowledge indicates profound differences in agent-types’ preferences. These reflections are pertinent independently of the experimental design sophistication adopted (Bayesian d-efficient design in our case).

The conclusions reached are, in our opinion, also important for other policy evaluation context such as, for instance, the case of LTZ creation, the imposition of a road pricing scheme,
in other words in all those cases where a general-purpose policy has effects on structurally different agents.

The main weakness of the paper is the small number of interviews that limits the robustness of the conclusions reached, their transferability to other geographical/dimensional contexts and the possibility to perform out-of-sample prediction tests.

Future research will: 1) increase the dataset size used for estimation by interviewing a larger number of both retailers and transport providers; 2) extend the investigation to other cities of different dimension and characterised by dissimilar urban freight transport policies; 3) also include citizens' preference analysis.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of Volvo Research Foundation ("Innovative solutions to freight distribution in the complex large urban area of Rome") and that of the Italian Ministry of Education, University and Research ("Methods and models for estimating the effectiveness of strategies for urban distribution of goods"). We are also grateful to Amanda Stathopoulos for her precious suggestions and comments on an earlier version of the paper.

BIBLIOGRAPHICAL REFERENCES


