CREI Working Paper no. 3/2013

URBAN FREIGHT TRANSPORT POLICIES: JOINT ACCOUNTING OF NON-LINEAR ATTRIBUTE EFFECTS AND DISCRETE MIXTURE HETEROGENEITY

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 author(s), not to CREI nor to any institutions of affiliation.
Urban freight transport policies: joint accounting of non-linear attribute effects and discrete mixture heterogeneity

Gatta V., Marcucci E.

ABSTRACT
This paper jointly investigates non-linear attribute effects and discrete mixture heterogeneity. The research context relates to urban freight transport policy evaluation. The paper adopts an agent-specific perspective. The often unforeseen and undesired results deriving from urban freight transport policy implementation have induced many researchers to call for an in-depth analysis of specific agents’ preferences. However, the structural lack of appropriate data has hindered investigations at such a detailed level. This paper contributes to bridging this data and knowledge gap by: constructing an original data set; testing for non-linear effects in attribute level variations; investigating the presence of inter and intra-agent heterogeneity; jointly exploring non-linear attribute effects and intra-agent heterogeneity. The results obtained underline the relevance of intra-agent preference heterogeneity and non-linear effects of attribute variations. More in detail, the paper detects two classes of agents with substantially different preferences with respect to the possible policy interventions. Non-linear sensitivity suggests policy makers should carefully consider the effects induced by the specific status quo level for policy relevant attribute variations. Intra-agent discrete heterogeneity implies different willingness to pay measures for given policy changes. The presence of both non-linear effects and intra-agent heterogeneity suggests policy makers to explicitly consider the status quo level that is to be changed while, at the same time, contemplate differentiated reactions deriving from the implementation of a urban freight policy change. In conclusion, the paper underlines the need for rigorous ex-ante policy analysis if the correct policy outcomes are to be estimated with an adequate level of accuracy.

Keywords: Urban freight transport, discrete mixture heterogeneity, non-linear attribute effects, policy evaluation

Acknowledgments
The Authors would like to thank: 1) Volvo Research Foundation for providing financial support for the “Innovative solutions to freight distribution in the complex large urban area of Rome” research project; 2) Italian Ministry of Education, University and Research for PRIN funding of the “Methods and models for estimating the effectiveness of strategies for urban distribution of goods” research project; 3) Amanda Stathopoulos and Eva Valeri for their active participation in both point 1) and 2) and their suggestions and comments on earlier versions of the present paper.
1. Introduction
Urban freight transport (UFT) is progressively acquiring a prominent position in local policy makers’ agendas throughout the world. Temporal and spatial overlapping of UFT related operations with peaking passenger demand is common and its impact on cities widely acknowledged. This is testified by the growing research efforts and funds, both public and private, dedicated to investigating this issue. More in detail, the EU has financed, among others, the following projects: BESTUFS I, 2000; BESTUFS II, 2004; CUPID, 2000; EFRUD, 2009; EUTP II, 2000; MOST, 2000; OSSA, 2000; PROGRESS, 2000; REVEAL, 1999; STRAIGHTSOL, 2012; SULOGRA, 2000. Private institutions such as, for example, Volvo Research Foundation have supported research by establishing interdisciplinary Centers of Excellence on this specific subject (i.e. CoE-SUFS, 2013 and METROFREIGHT, 2013) and also granting funds for smaller projects. Although UFT policies might have important positive effects on city dwellers’ welfare by stimulating smart urbanization, however they may also produce negative impacts. The need for public intervention is widely acknowledged in order to favor the re-balancing of social costs and benefits.
Various actors collaborate and compete within the UFT system. Each is characterized by unique preferences and priorities. Understanding the main UFT characteristics and identifying policy optimization measures is thus a complex task.
An appropriate investigation of the heterogeneity among and within agent-types is fundamental to gain a clear insight of UFT functioning (Arunotayanun, 2009). Agents involved in UFT can be classified in two main categories: demand (i.e. retailers and own-account operators) and supply actors (i.e. transport providers). The different, and often contrasting, interests of the stakeholders need to be duly considered as well as the potential heterogeneity in preferences among and within them.
In this research field, data needs are often superior to their availability (Samimi et al., 2009). In the EU there is no systematic urban freight activity survey. Ruesch and Glücker (2001), after studying 43 medium sized cities in Europe, discovered that 58% of them were not collecting data on UFT. This, sadly, is also truer for Italy (Filippi and Campagna, 2008) and the analysis of specific agents’ preferences for UFT policies has been structurally under-researched. This is due to the lack of appropriate data, notwithstanding the early suggestions to study specific actors in UFT operations (Ogden, 1992). The purpose of the paper is to bridge this long-standing data and knowledge gap by jointly investigating non-linear attribute effects and discrete mixture heterogeneity in the case of UFT policy evaluation at an intra-agent level.
The main contribution to this articulated field of research (Allen et al., 2013; Arvidsson, 2012; Ballantyne et al., 2013; Dablanc, 2007; Filippi et al., 2010; Marchau et al., 2008; Russo and Comi, 2011; Yamada et al., 2011) is to provide an in-depth analysis of transport providers’ preferences, representing a key actor in UFT\(^1\). This is made possible: 1) by the acquisition of an original, accurate and detailed data set concerning specific transport providers’ preferences with respect to possible policy interventions in the limited traffic zone (LTZ) in Rome’s city center; 2) via the extension of previous results obtained thanks to a sophisticated selection process of the policy attributes considered, articulated attribute specification for the agents considered, efficient design development and appropriate questionnaire administration (see Stathopoulos et al., 2011; 2012). Furthermore, the paper contributes to the literature by: 1) testing non-linear effects of attribute level variations; 2) investigating intra-agent preference heterogeneity; 3) accounting for joint non-linear attribute effects and intra-agent preference heterogeneity; 4) measuring willingness to pay (WTP) distortions attributable to the naïve treatment of all previous points.

The contributions the paper provides are clarified by a wider framework describing both the research strategy used and the relevance of the issues dealt with. The results are obtained adopting the following strategy: 1) set the benchmark model by assuming linear attribute effects and no preference heterogeneity neither among agents nor within them; 2) test for inter-agent heterogeneity; 3) investigate intra-agent non-linear attribute effects; 4) search for intra-agent preference heterogeneity; 5) explore both intra-agent non-linear attribute effects and intra-agent preference heterogeneity. The relevance of the issues discussed are testified by the substantial differences in WTP measures implicitly deriving from alternative assumptions concerning policy effects (i.e. homogeneity/linearity vs. heterogeneity/non-linearity).

The conclusions reached underline the need for greater accuracy and sophistication in defining and implementing UFT policies.

The paper is structured as follows: section 2 reports a literature review that illustrates previous relevant research and contextualizes the innovative contributions of this paper. The methodology is discussed in section 3 while section 4 provides a succinct data description. The results are illustrated in section 5 together with policy implications. Section 6 concludes and discusses future research endeavors.

\(^1\) The success or failure of UFT policy changes is mainly linked to the reaction these actors have to the specific characteristics of the policies implemented (see, for instance, Danielis and Marcucci, 2007; Holguin-Veras et al., 2006; Marcucci et al., 2007; Marcucci and Danielis, 2008; Paglione and Gatta 2007).
2. Literature review

UFT literature analysis reveals a substantial heterogeneity in the approaches adopted (Comi et al., 2012; Gentile and Vigo, 2007). The main and most evident distinction among the various approaches relates to the public or private perspective considered. In fact, the public approach has mainly focused on the definition and implementation of policies aimed at reducing the negative external effects UFT imposes on specific city dwellers' categories. The private approach, instead, mainly focused on enhancing the efficiency of business operations (Corò and Marcucci 2001; Marcucci and D’Agostino, 2003). Recent modeling advances attempt to develop hybrid approaches. They try to include supply chain elements and considerations in public decision-making modeling. The intent is providing a balanced account of both public and private objectives. Preferences towards policy interventions are elicited in a transparent way and explicitly considered (Roorda et al., 2010) assuming stakeholders’ decision making would benefit from appropriate evaluative systems based on high quality modeling.

Not only is there heterogeneity in UFT modeling approaches but also in the classification adopted in review papers (e.g. Boerkamps and Binsbergen, 1999; de Jong et al., 2004; de Jong et al., 2012; Groothedde, 2005; Regan and Garrido, 2001; Taniguchi et al., 2003; Yang et al., 2010). The criteria usually adopted for categorizing previous works prevalently rely on the modeling approaches used. The main classifications are based on aggregate versus disaggregate modeling, simulation of commodity versus vehicle movements, systemic and operational models (Melo, forthcoming; Nuzzolo et al., 2013; Paglione, 2006).

The approach adopted in this paper relates to a well-established tradition in behavioral and disaggregate freight modeling. Winston (1983; 1985) presents an overview of different freight models and describes the early evolution from aggregate to disaggregate ones. The first group of models were used to forecast the behavior of an entire transport system while the second was used to predict the behavior of individual agents within a given transport system. Disaggregate models quickly developed thanks to their theoretical and empirical advantages. In fact, they rest on individual behavioral analysis and usually provide richer model specifications capable of capturing important decision-maker’s characteristics and warrant a better understanding of policy effects.

Disaggregate models explicitly consider stakeholders’ utility maximization efforts. In the case of freight transport, in particular, one has to identify key decision makers to develop an agent-based micro-simulation modeling framework capable of describing and forecasting the
behavior of the specific actors involved (Liedtke and Schepperle, 2004). Several authors (de Jong and Ben-Akiva, 2007; Gray, 1982; Hensher and Figliozzi, 2007; Wisetjindawat et al., 2006; Yang et al., 2009) indicate UFT as one of the most appropriate fields to develop agent-based micro models.

This wealth of benefits comes at the cost of relevant data needs and high accuracy in defining the relevant attributes to be considered. Previous works testify as to how important and demanding the data acquisition process can be (Stathopoulos et al., 2011; 2012).

This paper adopts a disaggregate behavioral approach at an agent-type level while, for the time being, abstracting from interaction effects among agents collaborating/competing along the supply chain. Modeling interaction effects is postponed to a subsequent paper specifically targeting this issue.

3. Methodology

This section describes the methodology employed. Agents’ preferences are elicited using discrete ranking experiments. The theoretical basis is represented by the micro-economic theory of choice and random utility maximization theory (Louviere et al., 2000). Utility is modeled as a random variable reflecting the assumption that the decision-maker disposes of perfect discriminative capability, while the analyst has incomplete information (Ben-Akiva, Lerman, 1985). Utility \( U \) is composed of a deterministic \( V \) and a stochastic term \( \epsilon \). The former is assumed to be a linear (in the parameters) function of attributes.

The utility that individual \( i \) associates with alternative \( j \) is given by:

\[
U_j = V_j + \epsilon_j = X_j \beta' + \epsilon_j
\]

where \( X_j \) is the vector of attributes and \( \beta' \) the vector of estimated parameters.

Probabilistic utility maximization implies that the choice probability is equal to the probability that the difference between each random term is smaller than the difference between each deterministic part of utility, that is,

\[
P(j) = P[(\epsilon_h - \epsilon_j) < (V_j - V_h)], \quad \forall h \neq j.
\]

Different assumptions about the distribution of the stochastic term lie at the basis of different discrete choice model specifications.

The multinomial logit (MNL) expression that assumes an i.i.d. Gumbel distribution for \( \epsilon \), is:

\[
P(j) = \left[\prod_{h \neq j} \exp(-e^{-(\epsilon_j - V_j - V_h)})\right]e^{-\epsilon_j} \exp(-e^{-\epsilon_j}) = \frac{e^{X_j \beta'}}{\sum_{h=1}^{H} e^{X_h \beta'}}.
\]

MNL is characterized by important advantages (e.g. closed form, ease of interpretation, etc.) as well as by relevant drawbacks linked to the assumption of preference homogeneity across respondents (McFadden, 1974). The estimated parameters represent the marginal utility of
each attribute variation (even if confounded with the scale) assuming all agents have an equal
taste for a given attribute (Marcucci, 2005).

Given the well-known restrictions characterizing MNL that limit its ability to detect random
variations of agents’ preferences, it is advisable to test also more flexible and sophisticated
ways to treat preference heterogeneity. The typical way of allowing for variations in behavior
across decision makers is to incorporate taste heterogeneity via the systematic component of
utility (Marcucci and Gatta, 2012). In this case, one can rely on the assumption of either
continuous or discrete mixture structure.

In more detail, the Mixed Multinomial Logit (MMNL) model assumes a continuous mixing
distribution of parameters across individuals that represents an important advantage in
terms of flexibility. Besides the disadvantages related to the estimation process2, the use of
such models implies a risk of producing misleading results when making an inappropriate
choice of distribution (Hess et al., 2005).

On the contrary, the Latent Class (LC) model investigates heterogeneity assuming a discrete
mixing distribution of preference parameters where a small number of mass points (c) are
interpreted as different groups/segments of agents (Boxall and Adamowicz, 2002; Kamakura
and Russell, 1989). It makes use of two sub-models, one for class allocation, and one for
within class choice. The former models the probability of an individual being assigned to a
specific class as a function of respondents’ attributes and, possibly, of the alternatives in the
choice set. The within class model is then used to compute the class-specific choice
probabilities for the different alternatives, conditional on the tastes within that class.

In this case, the choice probability is the expected value, over classes, of the choice probability
within each class as analytically reported below:

\[
P_j(j) = P(j | c) P(c) = \frac{\exp(X_j \beta^*_c)}{\sum_{n=1}^{df} \exp(X_{nj} \beta^*_n)} \sum_{c=1}^{C} \left( \frac{\exp(q_c \phi^*_c)}{\sum_{c=1}^{C} \exp(q_c \phi^*_c)} \right)
\]

where \(q_i\) denotes a set of observable characteristics that enter the model for class
membership (e.g. socio-economic variables) and \(\phi\) their relative parameters. When no such
covariates enter into the model, the only element used for class allocation is the constant term.

When opting for a discrete mixture hypothesis for detecting preference heterogeneity a
crucial issue relates to the number of classes to be estimated (Bujosa et al., 2010). The

---

2 MMNL have the disadvantage that their choice probabilities take on the form of integrals that do not possess a
closed form solution, thus requiring simulations. However, recent developments in simulation methods coupled
with low-cost computational power, now available, allow the estimation of open-form discrete choice models
with relative ease.
conventional specification tests cannot be used to determine the optimal number of classes. In particular, the likelihood ratio or Wald tests cannot be employed. In fact, within this context, these tests do not satisfy the regularity conditions for a limiting chi-square distribution under the null hypothesis adopted. The optimal number of classes to be used should thus be determined on the basis of some information criteria statistics such as those developed by Hurvich and Tsai (1989). The entropy index has also been commonly employed when choosing the number of segments to be estimated (Thacher et al., 2005). Nonetheless the relevance of these considerations other issues to be considered relate to the: 1) significance of the estimated parameters, 2) plausibility of model results (e.g. signs and magnitudes of the parameters) and 3) a priori information concerning existent groups.

To sum up, LC is somewhat less flexible than MMNL since it approximates the underlying potentially continuous distribution with a discrete one. However, it does not require the analyst to make specific assumptions about the distributions of parameters across individuals. Both approaches are characterized by their own pros/cons and none can be considered unambiguously preferable. In the present study, we opted for a LC specification since it allows for differences in sensitivities across population groups producing a more effective result in terms of policy intervention.

As it is for non-linear effects one has to make sure of including, at least, three levels for each attribute when developing the experimental design. In fact, in the case of two levels it is not possible to detect non-linear effects of attribute levels variations since a single parameter can be estimated. When three or more levels have been used two are the approaches that can be employed to test for non-linear effects. The most commonly used approach is to dummy code the attributes and produce n-1 parameter estimates (n = number of levels) where the impact of the base level cannot be explicitly estimated since its effect is confounded with the grand mean. On the other hand, effects coding the attribute levels allows for the estimation of all levels' effects at the cost of constraining the sum of all parameters' values to zero. In other words, the impact of the base level is equal to the negative of the sum of the non-base estimated parameters (Hensher et al., 2005).

4. Survey instrument and data description
The data used in this paper were acquired in Rome’s LTZ between March and December 2009 (VRF, 2009). The LTZ in the city center of Rome, first implemented in the late ’80s, encompasses a 5km² area originally banned to non-resident vehicles only. Nowadays, Euro1 and more fuel-efficient vehicles only can enter. Residents are granted free access while others
(e.g. retailers and transport providers) pay an entrance fee. Enforcement relies on cameras and optical character recognition software. The system operates diurnally with a yearly entrance fee of 565€ per number plate.

An extensive list of prohibitions generically applies while a wide ranging of ad hoc exemptions pertain to third party freight operators. The regulation, after detailed examination and careful interpretation of the exemptions conceded, appears as prevalently aimed at dissuading own-account operators’ entrance.

The stated ranking experiment (SRE), that constitutes the elicitation mechanism adopted, was progressively developed starting from attribute definition. A focus group with relevant stakeholders was conducted to define the most relevant UFT problems, determine appropriate policy intervention and forecast likely stakeholder reactions (Stathopoulos et al., 2011). The efficient experimental design3 adopted was developed using Ngene 1.1 software (Rose and Bliemer, 2012). A SRE rather than a stated choice exercise was used for two reasons. On the one hand, this response format was considered the most appropriate given the aim of the research was unveiling agents’ preferences concerning UFT policies. In this peculiar study context, in fact, one cannot de facto “choose” any policy while it seems reasonable to ask for a preference ranking of various options. On the other, rank-exploded data were considered useful to contrast the impact the likely small number of observations would produce.

The policy alternatives used in the SRE are characterized by a set of policy elements taking several levels. Attributes selection was based on: 1) literature survey; 2) previous UFT studies performed in Rome; 3) focus groups with experts. A literature review permitted the identification of a set of eligible attributes that embodied potentially conflicting policy instruments4.

Attribute selection was further refined on the base of previous UFT research results carried out in Rome (Filippi and Campagna, 2008; STA, 2001) and focus groups with experts/stakeholders5. The attributes finally selected were considered as central by at least two stakeholders categories (Stathopoulos et al., 2011) and were also validated via a pre-test with real operators. Table 1 reports attributes, levels, and ranges. Attributes are characterized

---

3 For an in-depth discussion of efficient experimental design please refer to Rose et al., 2008.
4 Night-time deliveries were excluded since they are considered efficiency enhancing by carriers but are opposed by retailers that see them as a mere increase in costs. For an in-depth analysis of this issue and relevant policy proposals please refer to Holguin-Veras, 2008.
5 Expert surveys focused on the definition of the most appropriate policies to mitigate the identified UFT problems (Stathopoulos et al., 2011).
by, at least, three levels. All policy attributes used are contemplated both as promising levers of intervention by decision-makers and perceived as acceptable by operators.

**INSERT TABLE 1 ABOUT HERE**

An SRE is used to elicit the effects of currently not implemented policy options. The alternatives presented to respondents include two policy options plus the status quo (SQ) situation. Table 2 reports an SRE sample task.

**INSERT TABLE 2 ABOUT HERE**

In total, 252 interviews were completed and 229 used in modeling after discarding pilot interviews. We sampled 66 transport providers representative of 8 main macro-freight sectors (see Figure 1), namely: 1) food (e.g. fresh, canned, drinks, tobacco, bars, hotels and restaurants, etc.); 2) personal and house hygiene (e.g. detergents, pharmaceuticals, cosmetics, perfumes, watches, barbers, etc.); 3) stationery (e.g. paper, newspapers, toys, books, CDs, etc.); 4) house appliances (e.g. dish washers, computers, telephones, metal products, etc.); 5) services (e.g. laundry, flowers, live animals, accessories and animal food, etc.); 6) clothing (e.g. clothes, leather, etc.); 7) construction (e.g. cement, scaffold, chemical products, etc.); 8) cargo (general cargo).6

**INSERT FIGURE 1 ABOUT HERE**

5. Econometric results and policy implications

This section reports the results of the models estimated for transport providers, reviews the WTP measures and discusses their policy implications. Following the research strategy illustrated in section 1, we briefly summarize the main findings of previous works so to contextualize those described in this paper and illustrate their relevance.

More in detail, we first find that an agent-specific approach is relevant when studying UFT operators’ preferences for policy interventions intended to modify the extant regulatory framework. In fact, comparing the results derived from the agent-generic effects model (MNL assuming no agent-specific heterogeneity and linear effects) with agent-specific ones we find

---

6 This specific 8 macro-area sub-division was adopted to obtain comparable results with previous research conducted in the LTZ in Rome (Filippi and Campagna, 2008).
that no one-size-fits-all policy can be implemented. In fact, the adoption of the agent-generic model would cause, on one hand, serious discontent to own-account operators and, on the other hand, less public revenues signaling substantial inter-agent heterogeneity (Gatta and Marcucci, 2013).

Taking the analysis further we investigated non-linear attribute effects and preference heterogeneity at an intra-agent level. Marcucci and Gatta (2013b) show, with respect to retailers located within Rome’s LTZ, that not only is it appropriate to adopt an agent-based perspective but it also necessary to contemplate potential non-linear effects of the policy attributes considered. The findings suggest that policies might produce very dissimilar effects depending on the given SQ attribute level the policy is trying to modify. The comparison between linear and non-linear effects models shows that potentially relevant biases could pervade the results if non-linearities were ignored.

Furthermore, Marcucci and Gatta (2013a) discovered, with respect to own-account operators in Rome’s LTZ, the presence of relevant intra-agent heterogeneity. Using a simple MNL specification for estimating WTP measures would induce, from a public policy perspective, the adoption of inefficient policy measures.

The present paper focuses on transport providers. We preliminary tested the potentially non-linear effects of the variations of the attributes considered. Table 3 reports a description of the variables used in all models presented.

**INSERT TABLE 3 ABOUT HERE**

The first model estimated (M1) adopts an MNL specification of the utility function; all attributes are linear and normalized. The results for M1, reported in Table 4, are in line with expectations both for variables’ signs and for their relative weight. In fact, EF, with a negative sign, has the highest explanatory power while both LUB and PLUBF, with approximately half EF weight, have a positive sign and an almost equal impact on utility. It is important to underline the aversion towards the SQ situation testified by the positive sign of the two alternative specific constants (ASC1 and ASC2) for the unlabeled hypothetical alternatives.

**INSERT TABLE 4 ABOUT HERE**

---

7 In the case of linear effects, attributes are normalized by dividing each level by the value of its own minimum to account for unit-of-measurement effects on coefficients estimates and facilitate a direct comparison among them.
The second model (M2), reported in Table 5, investigates potential non-linear effects by effects-coding the explanatory variables.

**INSERT TABLE 5 ABOUT HERE**

M2 passed a log-likelihood ratio test (LLRT) with respect to M1 (restricted versus unrestricted) thus showing a substantially better fit to the data and the presence of non-linear effects when moving from one attribute level to another. Equally scaled variations in attribute levels have, thus, heterogeneous effects on the ranking probability of a given policy change. M2, the best fitting parsimonious model, uses effects coding for LUB and EF while keeping PLUBF linear. In fact, an MNL model, not reported for reasons of brevity, where all the variables were effects coded, shows that PLUBF2 has a non significant coefficient and close to zero (0.026) implying a linear effect of different levels of probability of finding a bay free. A graphical analysis suggests the presence of non-linear effects for both LUB and EF (see Figure 2).

**INSERT FIGURE 2 ABOUT HERE**

Subsequent tests performed by re-estimating the same model adopting a selective linear normalization of the variables confirms the intuitions induced by direct visual inspection. The signs of the attributes are in line with expectations and the coefficients of the attribute levels considered are statistically significant. EF, given the range considered, is the most significant attribute. The previously detected SQ aversion is confirmed while the other two hypothetical policies considered have no distinct impact *per se*. Effects coding the variables facilitates interpretation. In fact, the constant term can only reflect the utility associated with the base case alternative thus avoiding any possible misinterpretation. Furthermore, the paper investigates the presence of heterogeneity in preferences via a discrete mixture model that relaxes the restrictive independence of irrelevant alternatives (IIA) assumption characterizing the stochastic part of utility in a MNL specification. To facilitate model comparison, using M1 and M2 as a reference, two LC models (M3 and M4,

---

8 We also estimated an Error Component model that produced no interesting results. No random effects linked to the alternatives were detected. The result can intuitively be attributed to the non-labeled nature of the ranking exercise performed.

9 Heterogeneity in preferences is important in itself. One should, however, recall that the hypotheses and procedures used to search for it have also an impact on the final results. See Marcucci and Gatta (2012) on this point.
respectively) are estimated coding the variables in the same way as in the MNL specification. The statistical analysis of our data, performed using NLOGIT 4.0 (2007), suggests for both models the presence of two separate latent classes of transport providers each characterized by a clearly differentiated behavioral profile. If more than two classes are included in the model the interpretability of results drastically diminishes, the parameters cannot easily be associated to specific behavioral characteristics and the additional groups include only a small portion of the total respondents with scarcely significant parameters.

**INSERT TABLE 6 ABOUT HERE**

Looking at M3 and M4 results similar conclusions can be drawn (see Table 6 and 7). EF has, in general, a marked impact. Class 1 (C1) comprises more price-sensitive transport providers while class 2 (C2) includes agents that are more interested in bay-based policies implemented either via construction or by increasing the probability of finding them free. Moreover, the estimated LC probabilities are almost equal in both models. As for the MNL specification, the SQ is perceived negatively.

**INSERT TABLE 7 ABOUT HERE**

It is now important to characterize the agents belonging to the two classes. We first used socio-economic variables to estimate conditional class probabilities. None of the socio-economic variables had a statistically significant impact in class allocation\(^\text{10}\). We, then, opted for a CART analysis\(^\text{11}\). The higher of the two available conditional class probabilities was used to allocate agents to classes. Initially we have 56% of the observations falling in C1 and 44% in C2. We feel quite comfortable with this pragmatic criterion since, in our case, the highest conditional class probability, with only 2 exceptions, was greater than 80%. The socio-economic variables used for segmentation are: 1) number of customers in the LTZ (*Customers*), 2) number of daily deliveries (*Deliveries*), 3) number of employees (*Employees*), 4) number of vehicles used for deliveries within the LTZ (*Vehicles-LTZ*), 5) freight sector

---

\(^{10}\) In our opinion, the inability of the socio-economic variables in explaining class probability is mostly due to the limited number of observations available rather than to the intrinsic capability of the variables in explaining class membership.

\(^{11}\) CART is a classification method (Breiman et al., 1984). It generally uses historical data with pre-assigned classes to construct so-called decision trees, which are subsequently used to classify new data. The CART algorithm searches among all variables those capable of producing the best split of the data into two parts with maximum internal homogeneity. The process is iterated for each of the resulting data fragments. The classification tree is built in accordance to a splitting rule.
(Sector), 6) total number of vehicles (Vehicles-TOT), 7) ratio between the number of customers served within the LTZ and the number of vehicles used for deliveries within the LTZ (Customers/Deliveries), 8) Vehicle emission standard (Standard).

The CART analysis reported in Figure 3, stops the tree-building process when one of the following cases applies: 1) there is only one observation in each of the child nodes; 2) all observations within each child node have the same distribution of the predictor variable; or 3) the improvement (i.e. decrease of impurity) is smaller than a 0.05 threshold\textsuperscript{12}.

\textbf{INSERT FIGURE 3 ABOUT HERE}

The variables that proved most useful in segmenting the sample were Customers and Deliveries. More in detail all the transport providers with more than 145 customers (100\%) belong to C2 while those with less than 145 customers belong in 65\% of the case to C1 and 35\% to C2. For transport providers with less than 145 customers Deliveries is the variable providing the best subdivision. In fact, in presence of a high number of deliveries (i.e. more than 4.5 deliveries per day) the original percentage of transport providers belonging to C1 and C2 is reversed with 56\% falling into C2 and 44\% in C1. On the other hand, with less than 4.5 deliveries per day the original percentage of observations split between C1 and C2 further polarizes with 79\% belonging to C1 and 21\% to C2.

In conclusion, one can state that transport providers belonging to C2 either serve more than 145 customers in the LTZ or, if they serve a smaller number of customers, they perform more than 4.5 deliveries per day. Transport providers belonging to C1 either have less than 145 customers or perform less than 4.5 deliveries per day. This information greatly aids LC results interpretation from a policy-oriented perspective. In fact, C1 represents the price sensitive class whereas C2 members are interested in bays-based policies. It is reasonable to assume that a transport provider frequently travelling to the LTZ and serving more customers should be more interested in LUB and PLUBF whereas one seldom serving the LTZ and with a limited number of customers should care more about the EF level since, in this last case, he/she cannot easily amortize the cost incurred for entering the LTZ.

\textsuperscript{12} The results reported are based on the Gini splitting function. Twoing’s rule was also tested without detecting any major differences.
5.1 WTP measures

WTP measures are used to: 1) compare the effects of different model specifications; 2) measure the implicit biases of possible misspecifications; 3) circumvent scale problems that would, otherwise, fraud the comparison.

We calculated, using the Krinsky and Robb (KR) percentile method\(^1\), the confidence interval for the estimated WTP measures. In particular, we generated 10,000 pseudo-random draws from the unconditional distribution of the estimated parameters and subsequently calculated the simulated estimates for each draw.

Table 8 reports WTP measures based on the four models previously presented.

**INSERT TABLE 8 ABOUT HERE**

WTP measures represent the additional amount of € transport providers are willing to pay to enter the LTZ given the policy changes proposed and their comparisons underline the differences deriving from the consideration of: 1) non-linear effects, 2) heterogeneity, 3) joint accounting of non-linear effects and heterogeneity. Results demonstrate the methodological relevance of the issues raised\(^4\).

Using the Wilcoxon rank sum test\(^5\) one notices that the simulated WTP distributions derived from the KR percentile method are statistically distinct both in the intra-model (across classes) and intra-class (across models) case as well. These results further confirm the necessity of testing methodologically articulated hypothesis.

In the following, we report detailed observations regarding the impact of the different specifications.

---

\(^1\) The KR percentile method (Krinsky and Robb, 1986; 1990), also known as the parametric bootstrap (Mooney and Duval, 1993; Efron and Tibshirani, 1993), adopts a simulative approach to the estimation of confidence intervals for WTP measures. In fact, coefficient estimates are produced by taking a large number of draws from a multivariate normal distribution with means calculated by the estimated coefficients and covariances given by the estimated covariance matrix of the coefficients (Hole, 2007). Subsequently, simulated WTP values, calculated by taking \(r\) draws from the joint distribution of the coefficients, are used to compute the percentiles of the simulated distribution reproducing the desired level of confidence.

\(^4\) WTP measures imply some calculations before drawing policy conclusions due to either normalization or effects-coding the variables. When linear attributes are assumed the ratio between the coefficients has to be multiplied by the value that the minimum level of cost attribute takes (i.e. 200) and divided by the value of the minimum level of the attribute at the numerator (i.e. 400 for LUB and 10 for PLUBF). When effects coding the variables, the WTP measures to move from one level to another is calculated as the difference in the corresponding valuations. One first calculates the marginal effect of the attribute level variations on utility and then compute the ratio. The result obtained is finally multiplied by the difference in € between the two levels of cost attribute (i.e. 200). In our case, since all non-cost attributes included positive variations we used the part-worth utility variation from the SQ level to the next step up (i.e. from 600€ to 800€).

\(^5\) The Wilcoxon rank sum test also known as rank sum test is used to test for a difference between two samples and represents the nonparametric counterpart to the two-sample Z or \(t\) tests (Wilcoxon, 1945). Instead of comparing two population means, the test compares two population medians.
5.1.1 Non-linear effects

The impact of explicitly accounting for non-linear effects can be determined by comparing WTP measures derived from M1 vs. M2 and M3 vs. M4 for C1 and C2 respectively. With reference to the comparison between M1 and M2 one notices that there is not a substantial difference for 20% additional PLUBF (149€ - M1; 143€ - M2). In the case of LUB, instead, the effects of both increments (+400 and +800) are substantially different. In fact for +400 LUB transport providers are willing to pay 95€ in M1 compared to 113€ in M2 while 191€ in M1 and just 142€ in M2 for +800 LUB. The erroneously assumed linear impact of LUB increments provokes a progressively larger over-estimation the further away one moves from the SQ.

Comparing the WTP in M3 and M4 one notices, in the case of C1, an inversion of the order of magnitude when moving from the first to the second increment of LUB as seen in the M1 vs. M2 comparison. This does not apply for C2 where, in both cases, M3 estimates are smaller than M4 and similar considerations apply to PLUBF.

5.1.2 Heterogeneity

A discrete mixture approach is adopted to investigate heterogeneity. A M1 vs. M3 and M2 vs. M4 comparison is performed to test the preference homogeneity hypothesis implicitly assumed in the standard MNL model. The comparisons proposed, accounting for non-linear effects, allow focusing on heterogeneity alone. M1 estimates are, as expected, an approximate average of two substantially different values for each of the two classes in M3. In particular, the differences between the results obtained in M1 and M3 (C1/C2) are +47€/-151€ for additional 400 LUB and +95€/-301€ for additional 800 LUB while +113€/-368€ for additional 20% PLUBF. M1 produces a rough approximation of two distinctively different WTP measures thus potentially provoking a relevant policy bias. Similar considerations ensue when comparing M2 vs. M4. M2 estimates are an average of two even more different classes. MNL-based WTP measures are always biased. The bias is greater when not accounting for heterogeneity with respect to assuming linear effects.

5.1.3 Joint accounting of non-linear effects and heterogeneity

M1 vs. M4 results are compared to evaluate the impact joint accounting for non-linear effects and heterogeneity has on WTP measures.
M1 overestimates with respect to C1 and underestimates for C2. The size of over- and under-estimation is considerable. In particular, for LUB we have +43€, -392€ (additional 400) and +129€, -485€ (additional 800) while for PLUBF +92€ and -601€ (additional 20%).
In conclusion, not accounting for heterogeneity and non-linear effects produces the greatest bias and induces the most relevant policy distortions.

5.2 Policy implications
The findings obtained have relevant policy implications. The non-linear effects detected confirm previous results in this research field (Rotaris et al., 2012) supported by the more general loss aversion hypothesis where declining marginal impacts of changes away from the SQ are postulated along with different weights characterizing gains and losses (Kahneman and Tversky, 1979). In particular, the linearity in the attribute assumption should be rejected and the marginal impacts on utility of both LUB and EF are not constant.
Transport providers show, in absolute terms, greater sensitivity to increases with respect to reductions in EF. Taken together with the non-linear sensitivity to increases in LUB it clearly emerges that transport providers are not willing to pay nearly as much for a total increase of 800 bays as they are for the first 400. In contrast, the sensitivity increases in PLUBF is constant indicating that, within the range of variation considered, each percentage unit of probability of finding the bays free is equally valued.
Heterogeneous valuations characterizing two different classes of transport providers for the same policy variations implemented signal the need of specifically tailoring intervention policies with respect to the objectives pursued.
Furthermore, considering both the non-negligible heterogeneity and the presence of non-linear effects linked to attribute variations the sophisticated policy options possibly available to local policy-makers are numerous and they, operating under a re-election constraint, are warned against assuming homogenous effects of the policies implemented and invited to explicitly consider the amount of variations the policy propose with respect to the SQ. The investigation of preferences reported in this paper testifies the need for a deeper understanding of the multi-faceted world of UFT if well-tailored and effective policies are to be implemented by local public administrators.

6. Caveats, conclusion and future research
To qualify the results derived one has to underline three limitations of this paper: 1) the one-case-only and small dimension of the sample bounds the transferability of results to other
contexts; 2) the analysis focuses only on deliveries performed by manufacturers residing out of the city using third party transport providers whereas we have evidence that, in general (Patier and Routhier, 2009) and especially for Rome (Filippi and Campagna, 2008), both own-account transportation and direct deliveries by manufacturers using their own vehicles (e.g. internalizing the freight distribution function) play an important role; 3) there is no explicit consideration of the specific distribution channel freight uses while recent evidence (Danielis et al., 2010) indicates that this is an important element in explaining different logistic constraints that could also justify heterogeneous WTP measures.

This paper jointly tests for non-linear effects and discrete mixture heterogeneity in the case of transport providers’ preferences for alternative policy options in Rome’s LTZ. The main findings relate to: 1) non-linear effects ascribable to attribute level variations measured via effects-coding of the variables; 2) heterogeneity among transport providers detected by estimating a LC model; 3) WTP comparable measures used for the joint accounting of non-linearity and heterogeneity thus clarifying the implicit biases an erroneous utility specification might imply.

The contributions to the literature relate to the: 1) acquisition of a new and relevant data set; 2) development of an agent-specific approach to the evaluation of WTP measures for UFT policies; 3) determination of specific WTP measures for the various policy options considered under the assumptions tested.

The results obtained testify to the need for a sophisticated treatment of the policy options that might be implemented at a local level. Both heterogeneity and non-linear effects cannot be overlooked if major biases are to be avoided.

The paper underlines the need for a deeper knowledge of the subtleties influencing the end-results a policy intervention might produce in the complex UFT environment. Future research might improve the analysis proposed in different ways. The most important, in our judgment, are to: 1) test the impact alternative level ranges have on non-linear effects; 2) control for heteroschedasticity by selecting a larger sample and collecting all relevant information potentially aiding richer model specifications; 3) account for agent-interaction effects.
Bibliographical references


Table 1 - Attribute levels and ranges used in the SRE

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Number of levels</th>
<th>Level and range of attribute - <em>(Status Quo underscored)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Loading/unloading bays</td>
<td>3</td>
<td>400, 800, 1200</td>
</tr>
<tr>
<td>Probability of free l/u bays</td>
<td>3</td>
<td>10%, 20%, 30%</td>
</tr>
<tr>
<td>Fees</td>
<td>5</td>
<td>200€, 400€, 600€, 800€, 1000€</td>
</tr>
</tbody>
</table>
Table 2 - Example of a ranking task

<table>
<thead>
<tr>
<th></th>
<th>Policy 1</th>
<th>Policy 2</th>
<th>Status Quo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loading/Unloading bays</td>
<td>400</td>
<td>800</td>
<td>400</td>
</tr>
<tr>
<td>Probability to find L/U bays free</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Entrance fee</td>
<td>1000 €</td>
<td>200 €</td>
<td>600 €</td>
</tr>
<tr>
<td>Policy ranking</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3 - Description of the variables used in the various models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Level variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>Number of L/U bays: linear attribute (level/400)</td>
<td>(1-3)</td>
</tr>
<tr>
<td></td>
<td>Probability of finding L/U bays free: linear attribute (level/10)</td>
<td>(1-3)</td>
</tr>
<tr>
<td>PLUBF</td>
<td>Entrance fee: linear attribute (level/200)</td>
<td>(1-5)</td>
</tr>
<tr>
<td>LUB2</td>
<td>Number of L/U bays: effects coded attribute (level 2)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>LUB3</td>
<td>Number of L/U bays: effects coded attribute (level 3)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>PLUBF2</td>
<td>Probability of finding L/U bays free: effects coded attribute (level 2)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>PLUBF3</td>
<td>Probability of finding L/U bays free: effects coded attribute (level 3)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>EF1</td>
<td>Entrance fee: effects coded attribute (level 1)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>EF2</td>
<td>Entrance fee: effects coded attribute (level 2)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>EF4</td>
<td>Entrance fee: effects coded attribute (level 4)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>EF5</td>
<td>Entrance fee: effects coded attribute (level 5)</td>
<td>(-1;0;1)</td>
</tr>
<tr>
<td>ASC1</td>
<td>Alternative specific constant: dummy coded variable (1st alternative showed)</td>
<td>(0;1)</td>
</tr>
<tr>
<td>ASC2</td>
<td>Alternative specific constant: dummy coded variable (2nd alternative showed)</td>
<td>(0;1)</td>
</tr>
</tbody>
</table>
Table 4 – Model estimates for MNL with linear effects, M1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>0.558</td>
<td>9.16</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.435</td>
<td>6.31</td>
</tr>
<tr>
<td>EF</td>
<td>-1.170</td>
<td>-16.85</td>
</tr>
<tr>
<td>ASC1</td>
<td>0.686</td>
<td>3.97</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.709</td>
<td>4.46</td>
</tr>
</tbody>
</table>

Log likelihood function: -690.6266
Log likelihood (constants only): -926.3309
N° of observations: 1128
Rho2: 0.254
Adj. Rho2: 0.252
### Table 5 – Model estimates for MNL with non-linear effects, M2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB2</td>
<td>0.238</td>
<td>3.07</td>
</tr>
<tr>
<td>LUB3</td>
<td>0.491</td>
<td>6.40</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.613</td>
<td>8.05</td>
</tr>
<tr>
<td>EF1</td>
<td>2.220</td>
<td>15.19</td>
</tr>
<tr>
<td>EF2</td>
<td>1.586</td>
<td>13.62</td>
</tr>
<tr>
<td>EF4</td>
<td>-1.131</td>
<td>-10.40</td>
</tr>
<tr>
<td>EF5</td>
<td>-3.260</td>
<td>-14.56</td>
</tr>
<tr>
<td>ASC1</td>
<td>1.041</td>
<td>4.92</td>
</tr>
<tr>
<td>ASC2</td>
<td>0.972</td>
<td>5.28</td>
</tr>
</tbody>
</table>

Log likelihood function: -661.1001
Log likelihood (constants only): -926.3309
N° of observations: 1128
Rho2: 0.286
Adj. Rho2: 0.281
Table 6 – Model estimates for LC with linear effects, M3.

<table>
<thead>
<tr>
<th>Variable</th>
<th>CLASS 1</th>
<th></th>
<th></th>
<th>CLASS 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-stat</td>
<td>Coeff.</td>
<td>t-stat</td>
<td>Coeff.</td>
<td>t-stat</td>
</tr>
<tr>
<td>LUB</td>
<td>0.545</td>
<td>4.90</td>
<td>0.920</td>
<td>13.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.203</td>
<td>1.54</td>
<td>0.966</td>
<td>10.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF</td>
<td>-2.271</td>
<td>-13.55</td>
<td>-0.747</td>
<td>-10.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC1</td>
<td>0.839</td>
<td>2.94</td>
<td>0.847</td>
<td>3.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC2</td>
<td>0.740</td>
<td>2.99</td>
<td>0.869</td>
<td>3.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Log likelihood function: -571.2232
Log likelihood (constants only): -926.3309
N° of observations: 1128
Rho2: 0.383
Adj. Rho2: 0.377
Estimated latent class probabilities: Class 1 = 0.50; Class 2 = 0.50
### Table 7 – Model estimates for LC with non-linear effects, M4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>CLASS 1</th>
<th></th>
<th>CLASS 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-stat</td>
<td>Coeff.</td>
<td>t-stat</td>
</tr>
<tr>
<td>LUB2</td>
<td>0.236</td>
<td>1.63</td>
<td>0.339</td>
<td>3.64</td>
</tr>
<tr>
<td>LUB3</td>
<td>0.405</td>
<td>3.03</td>
<td>0.982</td>
<td>10.37</td>
</tr>
<tr>
<td>PLUBF</td>
<td>0.480</td>
<td>3.61</td>
<td>1.277</td>
<td>10.23</td>
</tr>
<tr>
<td>EF1</td>
<td>4.105</td>
<td>12.60</td>
<td>1.724</td>
<td>9.43</td>
</tr>
<tr>
<td>EF2</td>
<td>3.003</td>
<td>11.53</td>
<td>1.138</td>
<td>8.92</td>
</tr>
<tr>
<td>EF4</td>
<td>-2.204</td>
<td>-8.56</td>
<td>-0.363</td>
<td>-2.71</td>
</tr>
<tr>
<td>EF5</td>
<td>-6.079</td>
<td>-9.39</td>
<td>-2.818</td>
<td>-12.57</td>
</tr>
<tr>
<td>ASC1</td>
<td>1.534</td>
<td>4.15</td>
<td>0.735</td>
<td>2.24</td>
</tr>
<tr>
<td>ASC2</td>
<td>1.359</td>
<td>4.20</td>
<td>0.770</td>
<td>2.61</td>
</tr>
</tbody>
</table>

Log likelihood function: -525.8435
Log likelihood (constants only): -926.3309
N° of observations: 1128
Rho2: 0.432
Adj. Rho2: 0.423
Estimated latent class probabilities: Class 1 = 0.50; Class 2 = 0.50
### Table 8 – WTP estimation for the different models considered

<table>
<thead>
<tr>
<th>Model</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Policy</strong></td>
<td></td>
<td></td>
<td><strong>Class 1</strong></td>
<td><strong>Class 2</strong></td>
</tr>
<tr>
<td>400 additional LUB</td>
<td>95 (77-114)</td>
<td>113 (83-147)</td>
<td>48 (30-66)</td>
<td>246 (218-282)</td>
</tr>
<tr>
<td></td>
<td>191 (155-227)</td>
<td>142 (112-179)</td>
<td>96 (61-132)</td>
<td>492 (435-566)</td>
</tr>
<tr>
<td>800 additional LUB</td>
<td>149 (108-188)</td>
<td>143 (112-178)</td>
<td>36 (2-79)</td>
<td>517 (454-589)</td>
</tr>
<tr>
<td>20 additional PLUBF</td>
<td></td>
<td></td>
<td><strong>Class 1</strong></td>
<td><strong>Class 2</strong></td>
</tr>
<tr>
<td></td>
<td>52 (26-79)</td>
<td></td>
<td>62 (40-85)</td>
<td>57 (27-86)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>676 (517-942)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>750 (623-969)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1 – Transport provider by main freight sector
Figure 2 – Part-worth utilities for different policy attributes
Figure 3 – CART analysis