KEY-COST DRIVERS SELECTION IN LOCAL PUBLIC BUS TRANSPORT SERVICES THROUGH MACHINE LEARNING

ALESSANDRO AVENALI[1], GIUSEPPE CATALANO[1], TIZIANA D'ALFONSO[1],
GIORGIO MATTEUCCI[1] & ANDREA MANNO[2]
[1]Department of Computer, Control and Management Engineering, Sapienza Università di Roma, Italy
[2]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

ABSTRACT

This paper is aimed at developing a workable model for the identification of key-cost drivers in the Italian Local Public Bus Transport (LPBT) sector. Disaggregated information about costs, technical characteristics and environmental characteristics have been collected by means of questionnaires sent to LPBT companies producing more than 500 million bus revenue kilometres in Italy in 2011. A supervised regression model is built by training a regularized Artificial Neural Network in order to determine the quantitative and qualitative characteristics that contribute to explaining the variability of the driving personnel and the unit cost of the fleet (which usually covers more than 50% of the total economic cost) and the remaining portion of the unit cost. The proposed models could be an effective and simple tool for local authorities to validate reserve prices in tender procedures.
Keywords: standard costs, local public transport, fiscal federalism, cost drivers, machine learning.

1 INTRODUCTION

In Italy, the Local Public Transport (LPT) industry reform bill (1997) touched upon two dimensions: the allocation of public funds to Regions – and, in turns, to Local Authorities (LAs) – as well as the definition of an upper bound to public compensations to LPT firms. Although the reform bill stated that all subsidized LPT services should have been tendered off by January 2004, later legislative interventions left discretion to local governments about whether tendering out concessions or adopting in-house provision. Some competitive tendering took place after 1998 [1]. However, at present, the amount of LPT services that are tendered off is negligible with respect to in house provision that still prevails, especially in large cities [2]. In recent years, a number of policy interventions reaffirmed the political aim to improve market-oriented mechanisms in the allotment of LPT services; moreover, several parliamentary acts established that the maximum economic compensation to LPT firms should be based on standard costs (Law number 135/2012). The standard cost is defined as the cost of a LPT service provided by a reasonably efficient operator given a pre-specified level of service quality. However, at present, the Italian policy makers have not yet promulgated an appropriate methodology for the calculation of unit standard costs.

Our paper contributes to the literature since we develop two Top-Down cost models, based on recent advancements of machine learning approaches [3], in order to identify key cost drivers with respect to different cost items. The proposed models could be an effective and simple tool for LAs to validate reserve prices in tender procedures. Detailed information about costs has been gathered in order to fairly compute the total economic costs of the LPBT services; the collected economic and transport data relates to LPBT companies producing more than 500 million of bus revenue kilometers (BRK) in Italy in 2011 (approximately 30% of the overall amount of BRK offered in Italy for bus LPT services).

This paper relates to a vast empirical research on costs in LPT services (see [4] for a critical review). A first bunch of papers deals with the LPT companies' cost structure, with a focus on the estimation of variable and total costs [5]–[7]. Some papers focus on the most the appropriate output measure, i.e., supply-side indicators – such as vehicle-kilometer [8], seat-

kilometer [9] or total-seat-kilometer [10] – or demand-side measures, such as passenger-trips or passenger-kilometer [11]. More recently, a number of papers included hedonic characteristics among the explanatory variables – such as service frequency, average commercial speed, service size and the average fleet age are the most commonly employed [7], [12], [13], or size [6], [7].

Many of the above-mentioned studies focus on the causes of inefficiencies and the cost structure of firms in order to identify the proper configuration of the operating network. However, they do not focus on the definition and measurement of the standard cost. To the best of our knowledge, Avenali et al. [14], [15] build a Top-Down model for the determination of unit standard costs for the provision of LPT bus services. They find that commercial speed is the most important cost driver, while economies of scale are low and limited to small sized service bunches. Results also highlight a positive correlation between investments in the bus fleet and the cost incurred in service provision. Finally, Avenali et al. [16] develops a hybrid cost model, which follows a combined Bottom-Up and Top-Down approach and allows tuning the required level of efficiency according to policy makers desiderata.

Our paper makes steps further the above-mentioned papers. On the one hand, we identify key cost drivers with respect to different cost items. In particular, we isolate key features that affects the unit cost for the driving personnel, the unit cost of the fleet and the remaining portion of the unit cost. Thus, the applied model should be able to predict cost predictors associated with driver and buses, which usually cover more than 50% of the total economic cost. On the other hand, in order to identify the set of those features that most likely can be candidates to represent effective cost drivers, we empower economic considerations on the production process with the outcome of a *feature selection* algorithm [17], the process of selecting a subset of relevant features (variables, predictors) for use in model construction. Feature selection techniques are used for simplification of models to make them easier to interpret by researchers/users; reach shorter training times, avoid the curse of dimensionality.

The paper is organised as follows: Section 2 describes the methodology; Section 3 presents the dataset, while Section 4 contains results associated to cost models; Section 5 briefly provides applications of our work and Section 6 contains concluding remarks.

2 METHODOLOGY

In this section, we provide an overview on supervised learning and neural networks. Then, we illustrate two alternative models for the definition of the unit cost of LPBT services defined in the basis of a suitable neural network procedure.

2.1 An overview on supervised learning and neural networks

Machine Learning concerns with the study and development of computer programs that learn from experience [17]. Suppose to have a process that, given n input $x_1, \ldots, x_n$, produces an output y according to an unknown functional relation $y = f(x_1, \ldots, x_n)$. ML techniques exploit a set of available data to reconstruct a predictive model $m(\cdot)$ that approximates the functional relation $f(\cdot)$ between input and output.

Many widely used ML systems are categorized as *supervised methods*. In supervised learning data are given as set of features-labels instances, p, called training set (TS).

$$TS \coloneqq \{(x^p, y^p), x^p \in \mathbb{R}^n, y^p \in \mathbb{R}, p = 1, \ldots, P\}, \tag{1}$$

representing the input and the output of available samplings of the process. Every instance p is a vector represented by n features. The predictive model $m(\cdot)$ is obtained as a result of a

procedure called training phase during which the TS is used to adjust the parameters of a specific ML architecture, so as to minimize the gap between the output produced by the ML system and the actual labels. Once properly trained, the ML system is able to produce a good approximation of the actual output in correspondence of any possible input of the process.

Generally, according to the type of ML architecture, the training phase is formulated as a specific optimization problem. Among all existing architectures, some of the most popular are the Support Vector Machines (SVMs) [18], [19] and the Artificial Neural Networks (ANNs) [20]. While SVMs are mainly used for classification tasks, ANNs are used for both classification and regression.

The ANNs are characterized by a learning mechanism inspired by the biological neurons. Without entering into technical details, the ANNs are generally structured as networks in which the nodes are processing units called neurons that are organized into hidden ordered layers, while the edges are weighted oriented connections between neurons belonging to the different layers. The orientation of the connections is from the input layer to the output layer. Hence an ANN works as input-output systems that, receiving a vector of input signals $\tilde{x} \in \mathbb{R}^n$, i.e., the features, produces an output $\tilde{y} \in \mathbb{R}$. The output $\tilde{y}$ is the result of the propagation of the n signals from the input layer to the output layer. The number of propagations depend on the number of features and the neurons. For instance, assume $\tilde{x} \in \mathbb{R}^2$, $\tilde{x} := (\tilde{x}_1, \tilde{x}_2)$ i.e., each instance is characterized by 3 features, and 3 neurons in the first hidden layer.

The two signals, $\tilde{x}_1$ and $\tilde{x}_2$, generate 6 propagations (each signal is propagated towards 3 neurons). During this forward propagation, each neuron processes the weighted sum of all signals coming from its ingoing connections by means of a nonlinear *activation function*, producing an output signal that is transmitted to the neurons of the successive layer, $\tilde{y} = \tilde{y}(w, x^p)$ where $w \in \mathbb{R}^m$ are the weight associated to the total number of propagations m within the network.

In Leshno et al. [21], it is proved that ANNs can approximate any continuous function with arbitrary precision, if the activation function is continuous and nonpolynomial. This makes ANNs very attractive for regression tasks and, for this reason, we decided to use ANNs for our application. For a given TS, the training of an ANN consists in adjusting the weights of the network so as to minimize the overall discrepancy between all output signals produced by the network in correspondence of each x^p and its corresponding y^p, where $(x^p, y^p) \in TS$. Formally, the training concerns the minimization of the function:

$$E(w) = \sum_{p=1}^{P} (\tilde{y}(w, x^p) - y^p)^2, \tag{2}$$

where $E(w)$ is the overall square error and $\tilde{y}(w, x^p)$ is the output produced by the network for a given weight vector $w \in \mathbb{R}^m$ and a given features vector $x^p \in \mathbb{R}^n$. Function (1) is nonconvex with many local minima, at regions and steep sided valley, so its minimization is a very challenging task. In the last decades, many methods have been proposed in literature for the minimization of function (1) involved in the training of ANNs. Much of them relies upon a technique called *backpropagation* [22] in which the signals produced by the networks are propagated backward in order to compute the gradient vector of (1) with respect to w. These methods suffer very slow convergence, attraction to bad local minima, high cpu-time needed to perform the backpropagation procedure. In 2006 a technique called *Extreme Learning Machine* [23] was introduced for the training of ANNs, based on the random selection of the weights associated to the input and internal layers, combined with the optimization of the weights of the output layer. While achieving good training performance in a very short time, this technique may fail when high levels of training accuracy are requested.

In this work, we adopt a training approach recently proposed in [3]. This training algorithm, thanks to an intense decomposition of the optimization phase, coupled with a proper regularization strategy [17], is able to produce in reasonable time very accurate predictive models that is the best configuration of the learning machine in order to obtain the best predictive model in reasonable time. In particular, we make use of:

- a network with a single internal layer made up of 20 neurons and an output layer with a single neuron;
- a sigmoidal activation function of the form $g(z) = 1/(1 + e^{-z})$ where $z \in \mathbb{R}$ is the weighted sum of all signals coming from its ingoing connections;
- a TS made up of 54 training data (see Section 3), i.e., $P = 54$. We remark that ANNs are frequently used for the training of such small dataset, mainly related to healthcare problems often characterized by scarcity of data [24].

2.2 The cost models

In order to identify the main drivers of the cost of the local public bus transport services we apply the following three steps procedure

2.2.1 Step 1
We first derive the unit cost of LPBT services. The total cost of a service in a given period obviously depends on the quantity of the service that is produced in the period. Therefore, we focus on identifying the features of the service and of the production process, which explain crucial cost differences in terms of unit cost. The unit cost of local public bus transport services in the contracts designed by Italian Local Authorities is usually defined in term of total cost per BRK. Thus, in the analysis we apply this definition of unit cost. At this end, for any considered instance we compute the ratio between the total cost of the observed service and the number of BRK. Table 1 provides the categories of expenses included in the cost basis. We remark that, following the literature [7], [13], the number of vehicles owned by each firm has been used as a physical proxy of the invested capital, while the net value of those vehicles has been considered as a proxy of the net invested capital.

2.2.2 Step 2
For any considered instance, we gathered data related to several elementary attributes, i.e., features, which can impact on the unit cost of LPBT services. Some attributes represent characteristics of the provided service while the other ones represent features of the production process (the productivity of some inputs). In order to identify the set of those features that most likely can be candidates to represent effective cost drivers, we empower economic considerations on the production process with the outcome of a *feature selection* algorithm [17]. This revealed to be very helpful due to the small dimension of the training set. Specifically, 4 features, i.e., input signals, are selected and used for the training. Following notation in Section 2.1 we have $x^p \in \mathbb{R}^4, p = 1, \dots, P, P = 54$. Namely, features are:

- *commercial speed* (*km/h*). This variable is defined as the ratio between the total number of bus kilometers supplied yearly and the total number of net driving hours provided yearly (i.e., hours necessary to drive BRK from the beginning to the end of the line). This is a qualitative (hedonic) characteristic of a service, which can be barely controlled by the LPT firm;
- *average fleet age* (years). This variable represents the ratio between the sum of the ages of every used bus and the number of buses;

- *degree of renewal of the feet* (€/km). This variable is defined as the ratio between a monetary value and bus kilometers. The monetary value is the sum of all depreciations of owned vehicles (gross of non-repayable public contributions and adjusted by assuming a 15 years depreciation life) and rents/leasing for non-owned vehicles. This variable identifies a qualitative characteristic, which can be controlled by the LPT firm [14], [15];
- *average driver's salary* (€). This variable denotes the yearly average salary per driver.

2.2.3 Step 3

We use the 4 features selected at the end of STEP 2 to develop two alternative models which are specified according to the methodology described in Section 2.1.

Model 1 divides the unit cost of LPBT services in:
- the unit cost for the driving personnel, which is the ratio between the total cost for the driving personnel and BRK;
- the remaining portion of the unit cost, which is the ratio between the total cost (net of the cost for the driving personnel) and BRK.

Model 2 divides the unit cost of LPBT services in:
- the unit cost of the fleet, which is the ratio between the total cost to use/manage the fleet and BRK. In particular, the total cost to use/manage the fleet include: depreciation of owned vehicles, cost of maintenance, the cost of rented/leased vehicles, the (equivalent) yearly rent for vehicles given to LPT firms free of charge and the cost of capital (see Table 1). Actually, the unit cost of the fleet does include the unit cost of depots and related maintenance. However, this portion is negligible in our sample [14], [15];
- the remaining portion of the unit cost, which is the ratio between the total cost (net of the cost of the fleet) and BRK.

Table 1: The total cost of LPBT services.

Cost elements	
Operations and maintenance expense	Cost of the driving personnel (includes health and social insurance, and retirement funds)
	Cost of the depots personnel and the movement personnel (includes health and social insurance, and retirement funds)
	Cost of fuel
	Cost of vehicles inspections, insurances, tolls
	Depreciation of owned vehicles
	Cost of rented/leased vehicles and (equivalent) yearly rent for vehicles given to LPT firms free of charge
	Cost of depots (depreciation, rents and leasing charges)
	Cost of maintenance (cost of maintenance outsourced to third parties, the cost of spare parts, labor costs for in-house maintenance, the cost of equipment, machinery and other fixed assets used for in-house maintenance net of extraordinary maintenance capitalized within the fiscal year).
Overheads	Overall management, economic planning and control costs, membership fees, business consulting and information systems costs, other labor cost for personnel employed in overhead activities, cost of overhead activities outsourced to third parties
Cost of Capital	The cost of a company's funds (debt and equity) including tax effects

In particular, we remark that our models are defined by following a "representative" paradigm (instead of a "benchmark" paradigm), in the sense that all the observed instances are used to train and test the neural network (while in a benchmark paradigm only the services produced according to the best practices are used while the others are discarded). In such a way, our models can be more safely exploited by local authorities in order to agilely determine reserve prices in auction-based competitive tenders for local public bus transport services. In fact, they enlarge potential competition by allowing both best operators and averagely efficient ones to take part in the auction while blocking the access to the other (less efficient) firms. We also note that to furtherly raise the competition, LAs could allow for multiple winning operators/coalitions by simultaneously allocating single services through combinatorial auctions; however, such an approach is hardly applicable in the practice as identifying the winning bids requires a high computational effort [25].

2.2.4 Step 4

We use the 4 features selected at the end of STEP 2 to develop two alternative models which are specified according to the methodology described in Section 2.1. Each trained model was used to perform sensitivity analysis on 4 fictitious LPBT services (see Table 2). In order to fix ideas, the reader might assume that Services 1 and 2 (3 and 4) are in the urban segment (intercity segment). A urban service is defined as a service provided within high or low-density municipalities, typically, with high frequency and close stops. As opposite, an intercity service connects two or more municipalities either within a region or in two neighboring regions (typically, they are scheduled services and stops are sufficiently far removed). Moreover, a correlation exists between high commercial speed and intercity services, as well as low commercial speed and urban services.

The degree of renewal of the bus fleet has been set equal to the standard value. For instance, the average price of a representative vehicle used for urban services is € 257,000. The average productivity of any vehicle is 44,000 km per year and the depreciation life is 15 years. In such a scenario, the degree of renewal of the feet is 257,000/ (15×44,000) =0.39 €/km in the case of urban services in metropolitan cities.

3 DATA

In order to fairly compute the total economic cost of LPBT services, disaggregated information about costs (e.g., labor, energy, materials and services, capital) and about technical and environmental characteristics (e.g., average fleet age, average commercial speed, service size) have been collected in 2011 by means of questionnaires sent to managers of 45 Italian private (20) and public-owned (25) companies, producing more than 500 million of bus revenue kilometres. The questionnaire has been later adopted by the national Observatory on Local Public Transport Policies.

Table 2: LPBT services used to perform sensitivity analysis.

LPBT services	Commercial speed (km/h)	Average driver's salary (€)	Average fleet age (years)	Degree of renewal of the fleet (€/km)
Service 1	16	44,000.00	7.5	0.39
Service 2	19	44,000.00	7.5	0.39
Service 3	32	44,000.00	7.5	0.34
Service 4	35	44,000.00	7.5	0.34

The resulting Dataset consists of a cross section of 54 instances, in the following referred to as service bundles [14], [15]. A service bundle is the set of one or more service contracts for which the firm is able to measure (only) jointly its direct and indirect costs. Indeed, Italian local authorities are allowed to design the geographical boundaries of the service areas to be assigned to firms through a single or multiple service contracts. In our Dataset, 39 firms in our database operate 1 bundle only; 4 firms operate 2 service bundles (accounting for 8 instances); 1 firm operates 3 bundles (accounting for 3 instances); and 1 firms operates 4 bundles (accounting for 4 instances). As for the type of service, companies provide only urban transport service in 14 bundles; companies provide only an intercity transport service in 27 bundles and in the remaining ones, companies provide both services. 27 service bundles are localized in the Northern Regions, while 15 and 12 service bundles are localized in Central and Southern Italy, respectively. Finally, as for size 13 service bundles are large (more than 10 million of bus-kilometres), 15 medium-sized (4 to 10 million) and 26 small (less than 4 million).

Collected data allows reconstructing cost items in *Model 1* and *Model 2*, as well as the set of the 4 features used for the training for each of the 54 instances

4 RESULTS

In this section, we show results related to the sensitivity analysis performed with respect to *Model 1* and *Model 2* on the 4 fictitious LPBT services.

4.1 Sensitivity analysis on the the unit cost for the driving personnel and other unit costs

With respect to *Model 1*, the machine learning method highlights that the unit cost of the driving activities raises with the average driver salary while reduces with commercial speed. As shown in Fig. 1(a), where four services with different commercial speeds are considered, the unit cost of the driving activities linearly rises (approximately) with the yearly salary per driver.

As expected, the higher the yearly salary per driver, the higher the labor cost for the driving personnel, and thus the higher the unit cost of the driving activities. Fig. 1(b), where the four considered services have the same average driver salary (and thus curves collapse), highlights that the unit cost of the driving activities decreases with the commercial speed at decreasing marginal rates. In fact, when the commercial speed is low every driver produces a few BRK. Thus, service provision requires a higher number of drivers (the width of the daily schedule of any driver has a regulated upper bound which cannot be extended even if the commercial speed is low, i.e., 6 hours per day).
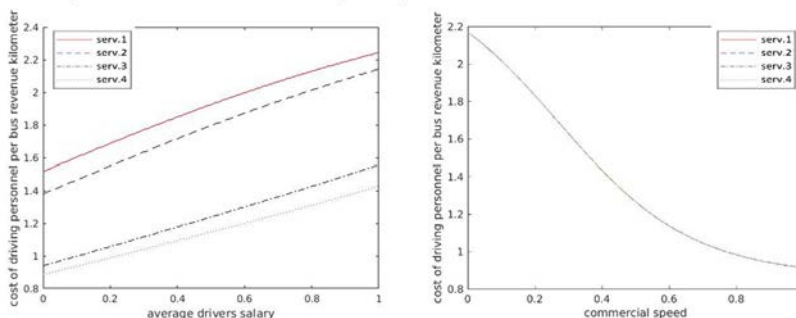

Figure 1: (a) The impact of average driver's salary on the unit cost of the driving personnel; (b) The impact of commercial speed on the unit cost of the driving personnel.

As commercial speed rises, any driver wastes less time in congested rides and thus he produces more kilometers. In such a way, the optimal number of drivers necessary to provide the service and thus and the overall labor cost for driving personnel decreases decrease (ceteris paribus), as well as the unit cost of the driving activities. Obviously, since the upper bound on the width of the driver daily time sheet is fixed, the possibility of minimizing the number of drivers by optimizing the daily time sheets of the drivers shrinks as the commercial speed rises. Essentially, when commercial speeds are high, any driver produces yearly a large number of BRK and thus it is very hard to reduce the number of necessary drivers by raising the drivers' productivity through a rearrangement of their daily time sheets (even if the commercial speed further increases). Moreover, machine learning approach highlights that the unit cost of the non-driving activities raises with the degree of renewal of the fleet, as well as with the yearly average driver salary, while reduces with commercial speed (see Fig. 2(a)–(c)).

Costs related to the non-driving activities contains the labor cost of other categories of personnel (e.g., personnel for maintenance activities, technical and administrative staff) and the possible cost of outsourced activities (e.g., outsourced maintenance activities, outsourced technical and administrative activities). Such costs are related to the driving labor cost, as the higher the number of drivers, the higher the overall value of these other costs. Thus, the neural network learns and signals that the yearly average driver salary and the commercial speed have roles in explaining the differences of the non-driving unit cost that are similar to the ones played under the driving unit cost. In addition, the non-driving activities contain also the costs for the depreciation and maintenance of the fleet. This implies a raise of the non-driving cost as the degree of renewal of the fleet increases (e.g., due to a higher quality of the owned/rented buses).

In particular, Fig. 2(a) shows that an urban service raises with the degree of renewal of the fleet at a decreasing marginal rate (see Services 1 and 2). Conversely, an intercity service raises with the degree of renewal of the fleet at an increasing marginal rate (see Services 3 and 4). Usually, an urban service is characterized by an incidence of the labor cost on the total cost larger than under an intercity service. Thus, the same increase in the degree of renewal of the fleet has a lower impact on the unit cost of an urban service than on the unit cost of an intercity service (ceteris paribus). Since a large amount of the labor cost characterizes the non-driving costs (e.g., the cost of the personnel for maintenance activities, technical and administrative staff), we can expect that the same increase in the degree of renewal of the fleet has a lower impact on the non-driving unit cost of an urban service than on the non-driving unit cost of an intercity service.
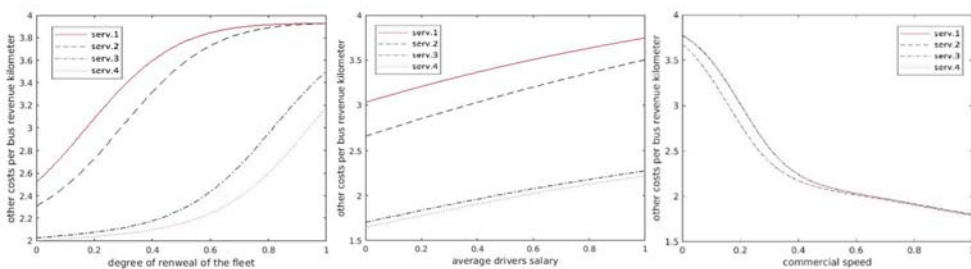

Figure 2:   (a) The impact of the degree of the degree renewal of the bus fleet on other unit costs; (b) The impact of the average driver's salary on other unit costs; (c) The impact of commercial speed on other unit costs.
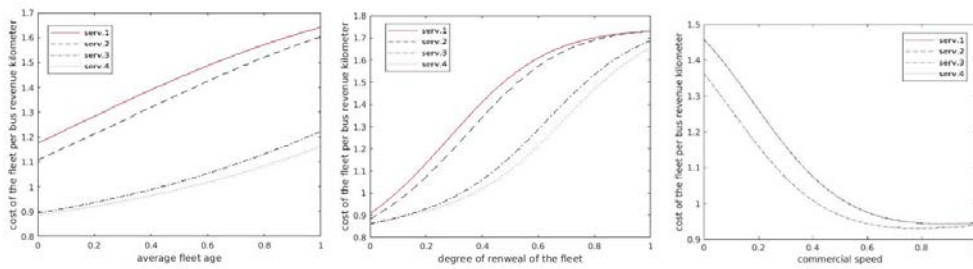
Figure 3:   (a) The impact of the average fleet age on the unit cost of the fleet;
             (b) The impact of the degree of renewal of the fleet on the unit cost of the fleet;
             (c) The impact of the commercial speed on the unit cost of the fleet.

4.2 Sensitivity analysis on the the unit cost of the fleet and other unit costs

In *Model 2* we focus on the unit cost of the fleet. As regards this model, the machine learning method highlights that the fleet unit cost raises with the average fleet age and with the degree of renewal of the fleet (see Fig. 3(a)–(c)).

Since the average age of the rolling stock in our sample, as well as for the majority of the Italian operators, is high above the European average age, the neural network mainly reflects how the maintenance costs raise when an aged fleet gets even older. We assume that, by considering a different dataset with more variance regarding the average fleet age, the impact of the fleet age on the unit cost would be U-shaped. In this sense, the functional form between the average fleet age and the fleet unit cost fits well the Italian context but could not fit well other scenarios where the variance of the average fleet age is large. In addition, the neural network shows that: the fleet unit cost first decreases at a marginal decreasing rate as the commercial speed increases; then, the fleet unit cost stops decreasing and begins to raise for very high levels of the commercial speed (see Fig. 3(c)).

Indeed, when commercial speed gets higher, buses can increase their productivity in terms of yearly BRK provided; in such a case, with a lower number of buses, it is possible to offer the same amount of service and thus the fleet unit cost reduces. However, it is hardly impossible to raise bus productivity when commercial speed is above certain feasible limits. Moreover, when commercial speed is very low or very high some mechanical components of the buses are stressed and thus more maintenance activities are needed. Therefore, when commercial speed gets too high the fleet unit cost stops decreasing and raises. As regard the non-fleet unit cost, the machine learning approach finds that it raises with the average driver salary while decreases with commercial speed (see Fig. 4(a), (b)). The main part of the non-fleet costs is represented by the labor cost (and, in particular, the driving labor cost). As discussed above, the unit driving labor cost rises with the yearly salary per driver and decreases with the commercial speed.

5 APPLICATION

In this section, we perform cost simulation on the 4 fictitious LPBT services (see Table 2) in order to show how our models could be applied to support the design of competitive tenders for the allotment of LPBT services in terms of the definition of reserve prices.

Table 3 shows the outcome of our machine learning approach applied to *Model 1* and *Model 2*. In particular, according to previous literature, the unit cost of urban services (e.g., Service 1 and 2) is higher than the unit cost of intercity services (e.g., Service 3 and 4) due
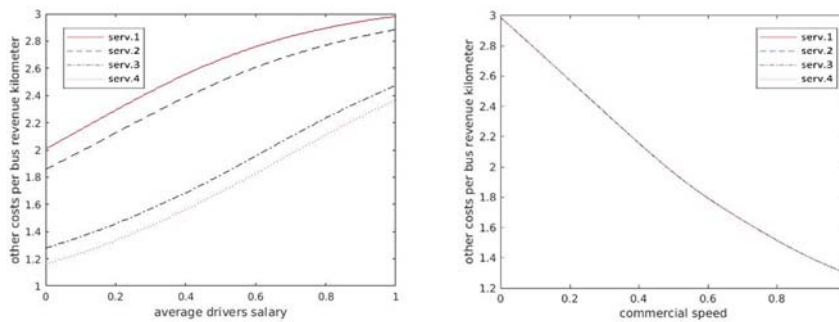
Figure 4:   (a) The impact of the average driver's salary on other unit costs;
             (b) The impact of the commercial speed on other unit costs.

Table 3: Application.

LPBT services	Service 1	Service 2	Service 3	Service 4
Model 1				
Cost of the driving personnel per BRK [€/km]	2.046	1.924	1.345	1.239
Other costs per BRK [€/km]	3.554	3.258	2.116	2.062
Total cost per BRK [€/km]	5.600	5.182	3.460	3.301
Model 2				
Cost of the fleet per BRK [€/km]	1.383	1.313	0.985	0.961
Other costs per BRK [€/km]	2.814	2.672	2.057	1.929
Total cost per BRK [€/km]	4.197	3.984	3.042	2.890
Cost difference				
[€]	1.403	1.198	0.418	0.410
[%]	33.42	30.07	13.74	14.20

to a lower commercial speed and a higher degree of renewal of the bus fleet. Moreover, we observe that, in the considered cases, *Model 1* provides a higher estimate of the total unit cost than *Model 2.* A possible interpretation of this results is that *Model 1* aims at isolating the cost of the driving personnel which, indeed, represents the most important cost component [14], [15]. As opposite, in *Model 2* such component is estimated together with other cost elements. Thus, the cost of the driving personnel could be underestimated according to *Model 2* or overestimated according to *Model 1*. Finally, we note that the difference in the model estimates seems affected by commercial speed in the sense that it is higher for urban services.

The proposed models could be an effective and simple tool for Local Authorities to validate reserve prices in tender procedures; for instance, the two estimates can represent suitable upper and lower bounds.

6 CONCLUDING REMARKS

In this paper, we develop a model for the determination of key drivers of unit costs for the Italian local bus transport sector, in order to provide an innovative and approach to define the maximum economic compensation in competitive tendering procedures or for in-house provision of LPBT services. To this purpose, our research develops two machine learning-based cost models which are able to predict the unit cost of an Italian LPBT service by taking into account both the main characteristics of the provided service and of the production

process. In particular, a neural network learns from the gathered sample how to combine the effects of the different features in order to generate an effective estimate of the unit cost of the service.

We propose two different models, where the unit cost for the driving personnel and the unit cost of the fleet, respectively, are separately estimated. The two models highlight which of the different features related to the provision of LPBT services are the most relevant in determining the overall unit cost of the service. In particular, confirming the existing literature, we found that commercial speed is without doubt the most important cost driver. We also found that an important role in driving the unit cost is played by the average age of the rolling stock and the degree of renewal of the fleet. This should induce policy makers towards the promotion of mechanisms that spurs firms in improving investments to lower the average age of buses (that in Italy is high above the average of other EU member States). As expected, we finally found that the yearly average salary per driver is another relevant cost driver. This feature can be significantly different across Italian Regions, due to several reasons (e.g., the average prosperity of a Region, the bargaining powers of operators and workers inside a Region). Therefore, a model based on the yearly average salary per driver is able to take into account the local cost differences in Italy regarding the labor cost of the driving personnel.

We remark that the developed cost models could also be an effective and simple tool for Local Authorities to validate reserve prices in tender procedures, in the sense that the two provided estimates can represent suitable upper and lower bounds on the unit cost of the LPT bus services.

REFERENCES
[1]   Boitani, A. & Cambini, C., To bid or not to bid, this is the question: the Italian experience in competitive tendering for local bus services. *European Transport*, **12**(33), pp. 41–53, 2006.
[2]   Boitani, A., Nicolini, M. & Scarpa, C., Do competition and ownership matter? Evidence from local public transport in Europe. *Applied Economics*, **45**(11), pp. 1419–1434, 2013.
[3]   Grippo, L., Manno, A. & Sciandrone, M., Decomposition Techniques for Multilayer Perceptron Training. *IEEE transactions on neural networks and learning systems*, **27**(11), pp. 2146–2159, 2016.
[4]   Daraio, C., Diana, M., Di Costa, F., Leporelli, C., Matteucci, G. & Nastasi, A., Efficiency and effectiveness in the urban public transport sector: a critical review with directions for future research. *European Journal of Operational Research*, **248**(1), pp. 1–20, 2016.
[5]   Berechman, J., Cost structure and production technology in transit: an application to Israeli bus transit sector. *Regional Science and Urban Economics*, **17**(4), pp. 519–534, 1987.
[6]   Fraquelli, G., Piacenza, M. & Abrate, G., Regulating public transit networks: how do urban-intercity diversification and speed-up measures affect firms' cost performance? *Annals of Public and Cooperative Economics*, **75**(2), pp. 193–225, 2004.

[7]   Cambini, C., Piacenza, M. & Vannoni, D., Restructuring public transit systems: evidence on cost properties from medium and large-sized companies. *Review of Industrial Organization*, **31**(3), pp. 183–203, 2007.

[8]   Cambini, C. & Filippini, M., Competitive tendering and optimal size in the regional bus transportation industry: An example from Italy. *Annals of Public and Cooperative Economics*, **74**(1), pp. 163–182, 2003.

[9]   Farsi, M., Fetz, A. & Filippini, M., Economies of scale and scope in local public transportation. *Journal of Transport Economics and Policy*, **41**(3), pp. 345–361, 2007.

[10]   Gagnepain, P. & Ivaldi, M., Incentive regulatory policies: the case of public transit systems in France. *RAND Journal of Economics*, **33**(4), pp. 605–629, 2002.

[11]   Bhattacharyya, A., Kumbhakar, S.C. & Bhattacharyya, A., Ownership structure and cost efficiency: a study of publicly owned passenger-bus transportation companies in India. *Journal of Productivity Analysis*, **6**(1), pp. 47–61, 1995.

[12]   Fraquelli, G., Piacenza M. & Abrate, G., Regulating public transit networks: how do urban-intercity diversification and speed-up measures affect firms' cost performance? *Annals of Public and Cooperative Economics*, **75**(2), pp. 193–225, 2004.

[13]   Piacenza, M., Regulatory contracts and cost efficiency: stochastic frontier evidence from the Italian local public transport. *Journal of Productivity Analysis*, **3**(25), pp. 257–277, 2006.

[14]   Avenali, A., Boitani, A., Catalano, G., D'Alfonso, T. & Matteucci, G., Un modello per la determinazione del costo standard nei servizi di trasporto pubblico locale su autobus in Italia. *Economia e Politica Industriale*, **33**, pp. 181–213, 2014.

[15]   Avenali, A., Boitani, A., Catalano, G., D'Alfonso, T. & Matteucci, G., Assessing standard costs in local public bus transport: Evidence from Italy. *Transport Policy*, **52**, pp. 164–174, 2016.

[16]   Avenali, A., Boitani, A., Catalano, G., D'Alfonso, T. & Matteucci, G., Assessing standard costs in local public bus transport: A hybrid cost model. Transport Policy, Online, https://doi.org/10.1016/j.tranpol.2017.03.011. Accessed on: 7 May 2017.

[17]   Bishop, C.M., Pattern recognition. *Machine Learning*, **128**, pp. 1–58, 2006.

[18]   Boser, B.E., Guyon, I.M. & Vapnik, V.N., A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152, 1997.

[19]   Manno, A., Sagratella, S. & Livi, L., A convergent and fully distributable SVMs training algorithm. *Proceedings of the Neural Networks (IJCNN), 2016 International Joint Conference*, pp. 3076–3080, 2016.

[20]   C.M. Bishop (ed.), *Neural networks for pattern recognition*, Oxford University Press: Oxford, 1995.

[21]   Leshno, M., Lin, V.Y., Pinkus, A. & Schocken, S., Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, **6**(6), pp. 861–867, 1993.

[22]   Rumerhart, D.E., Hinton, G.E. & Williams, R.J., Learning representations by back-propagation errors. *Nature*, **323**, pp. 533–536, 1986.

[23]   Huang, G.B., Zhu, Q.Y. & Siew, C.K., Extreme learning machine: theory and applications. *Neurocomputing*, **70**(1), pp. 489–501, 2006.

[24]   Chen, Y., McCall, T.W., Baichwal, A.R. & Meyer, M.C., The application of an artificial neural network and pharmacokinetic simulations in the design of controlled-release dosage forms. *Journal of controlled release*, **59**(1), pp. 33–41, 1999.

[25]   Avenali, A., Resolution Branch and Bound and an Application: The Maximum Weighted Stable Set Problem. *Operations Research*, **55**(5), pp. 932–948, 2007.