

# SEGMENTING FARE EVADER GROUPS BY FACTOR AND CLUSTER ANALYSIS

SARA SALIS, BENEDETTO BARABINO & BRUNO USELI  
Department of Studies and Researches, CTM SpA, Cagliari, Italy

## ABSTRACT

In proof-of-payment transit systems, fare evasion represents a crucial topic and undermines financial viability for public transport companies (PTCs). Two studies segmented all transit passengers by using qualitative research and explorative two-step cluster analysis as well as web-survey data. However, as far as the authors know, no study segmented exclusively the fare evader passengers, with the aim to know deeply characteristics of distinct groups, and to use data gathered from an intercept survey. Moreover, the combined use of a factorial and a cluster analysis was not adopted to solve this issue.

By using 850 on-board personal interviews, gathered from an Italian PTC, a Multiple Correspondence Analysis and a Hierarchical Cluster Analysis have been performed to group the fare evader passengers. Three clusters of evaders' passengers are derived: 1) the chronic, 2) the calculated-risk and 3) the accidental fare evader. The chronic cluster includes passengers who always evade the fare, and represents a medium numerous groups of fare evaders. The calculated-risk cluster includes passengers who usually evade the fare and represents the most numerous group of evaders. The accidental cluster includes passengers who rarely evades the fare and represents the less numerous groups. These outcomes are very relevant and useful for both research and practice. Indeed, this paper contributes to the empirical understanding of the determinants of being a fare evader for a selected group. Moreover, it helps PTCs understand how evaders' groups differ each other, adopt suitable strategies for each group and evaluate their effects.

*Keywords: fare evasion, fare evader characteristics, proof-of-payment, multiple correspondence analysis, cluster analysis, fare evaders groups.*

## 1 INTRODUCTION

Fare evasion undermines financial viability and aggravates subsidization needs for Public Transport Companies (PTCs) [1]. Moreover, it reduces perceived security levels because the actions employed to tackle it can trigger violence from fare evaders. In *proof-of-payment* systems, fare evasion is an old problem that emerged in 1960 in Europe and some years after in USA due to the abandonment of conductors [2]. However, fare evasion is a somewhat new research area in the transit industry as proved by emerging and fascinating studies. Largely, fare evasion can be classified according to the broad research on the unethical behaviour of people e.g., [3] as it somewhat represents a kind of specification. The unethical behaviour of people affects different subjects: economy, crime and psychology, which represent three main topics where studies on fare evasion were focused on. Moreover, economic and criminological studies are usually approached from the PTC's viewpoint, whereas psychological ones are from the passengers viewpoint.

On the one hand, to date, it is no surprise that a plethora of studies on fare evasion have focused attention on how to protect PTCs from monetary losses (Economic studies). This was achieved by evaluating changes to: ticketing systems infrastructure and management e.g., [4], [5]; policy and enforcement issues e.g., [6], [7], [8] and operational issues e.g., [9]. Besides, fare evasion can be associated with criminal behaviour of people. Therefore, several studies examined the effectiveness of situation-specific measures in facing fare evasion by evaluating security issues e.g., [10], enforcement e.g., [11] and deterrence strategies e.g., [11], [13], in the context of the situational crime prevention largely.



On the other hand, until recently, a handful of studies examined fare evasion from a customer's viewpoint (psychology) in the attempt to isolate determinants to draw the portrait of the fare evader. Different approaches are considered using qualitative ([14]) or quantitative research methods using descriptive [15], [16], or inferential statistics [17]–[19].

All previous studies provided evaluable contributions to the field. However, to the best of our knowledge, some gaps in the previous researches persist. First, economic and criminological studies contributed to analyse fare evasion from the PTCs' viewpoint as they focused on how to protect PTCs from economic losses and to counteract it, but neglected both who the possible fare evaders are among passengers and the group of evaders who have similar facets. Second, no study has examined how fare evasion is viewed by group evader passengers only and whether each group presents proper characteristics. Indeed, even if the works of [14], [15], can be considered pioneering as they derived some groups of fare evaders and their own behaviour against fare evasion, they drew conclusions based on a qualitative research or a cluster analysis on the whole of transit users, without focusing only on evaders passengers. In addition, they applied their analysis on web-survey data. Furthermore, though [16], reports groups of fare evaders, this classification is made from the inspector's viewpoint.

Therefore, this study was conducted in response to the lack of research into attitudes regarding fare evasion, as it is crucial to shed new light on fare-evader groups and to identify the different characteristics for each group in order to answer the following research question: What are the main characteristics, which define specific groups of fare evaders *in proof-of-payment* transit systems?

The contribution of this paper is to answer this question by descriptive (explorative) methods based on field-collected data through an intercept on-board survey of a large sample of people travelling across all bus routes of an Italian PTC network. More in detail, this paper studies the segment of passengers who evaded the fare by classifying evader passengers into different groups with similar facets. Moreover, it helps identify the socio-demographic, travel behaviour and personal knowledge of fare evasion characteristics that most affect the different groups of fare evaders. In doing so, a two step-methodological approach is used. Firstly, a factor analysis enables to obtain a synthetic description of a large set of initial variables. Secondly, a cluster analysis arranges evader passengers into different groups by identifying the common factors that influence their evading behaviour.

Answering the research question, it is expected to enlarge the results from previous research both in theory and in practice, providing insights for both researchers and practitioners interested. The outcomes of this paper will provide researchers with a better refinement of the main characteristics, which identify specific evader groups, thanks to the application of a robust two-step methodological framework. Moreover, due to the common problem of unpaid fines [6], [7], [11], [18], this paper could help in the identification of specific groups who will never pay the fine from the ones who might pay. Furthermore, according to [18], [19], knowing the characteristics which are most affect specific groups can help in understanding whether they are statistically significant in affecting the costs of the PTC as well. These outcomes will be of interest to those practitioners of PTCs who will wish to better understand their non-paying passenger base. This is a relevant topic as a better understanding of how some evader groups differ from each other implies that practitioners can set different strategies to deal with different segments of fare evaders. For example, if a large part of evaders is sensitive to the risk of being caught, the PTC can improve its inspection strategy by providing organized plans to check the large majority of passengers. Furthermore, segmenting the population of fare evaders could originate interests for public authorities and subsidizing agencies in order to revise the norms related to the fare and the

amount of the fine: different fares and/or fines can be re-thought for specific groups of offenders.

The remainder of this paper is organized as follows. In Section 2, materials are presented and methods applied to a bus network managed by a medium-sized PTC in order to profile fare-evader passengers for each cluster. In Section 3, results are presented, discussed and compared with previous works. Finally, in Section 4, conclusions and research perspectives are provided.

## 2 METHODS AND DATA

To assist in achieving the research question, a number of issues was considered: what data are required, how and where will they be collected and the data analyses to be performed.

### 2.1 Research context and survey method

This research was conducted by interviewing bus passengers of Cagliari, a coastal Italian city with 0.4M inhabitants. As shown, the main survey's purpose is to evaluate characteristics that identify different groups of evader passengers. Being a fare evader may depend on many variables, including socio-demographic characteristics, features relating to the trip and to the public transport system chosen, as well as personal knowledge of fare evasion. Hence, we built a three-section questionnaire and gathered data on various aspects according to [18]. The first section contains socio-demographic attributes, the second deals with travel behaviour attributes, the last section contains attributes related to the personal knowledge of fare evasion in general e.g., the inspection frequency in a predefined time window, the knowledge of the fine amount, previous ticket violations. Finally, a question was performed on the possibility to buy the ticket besides i.e., without inspection to evaluate the honesty behaviour. It is worth nothing that according to [14], [18], fare evasion is technically an illegal behaviour. Therefore, barriers to answers are the possible reluctance to admit to undertaking this illegal behaviour without personal gains and/or to lie in response to direct questions. Besides, recent studies assume that people can experience a psychological disutility that holds them back from misreporting due to intrinsic lying costs, honesty, conditional cooperation, etc. e.g., [20], [21].

According to [18], a headway-based route sample plan was built and 16-trained interviewers had administered the questionnaires during two weeks in March 2015, and two more weeks in July 2015, from 0700 AM to 0700 PM. To enhance data quality, accuracy and reliability, a paper and pencil interview (PAPI) was adopted instead of a self-administered interview. As a result, a crucial phase of the survey preparation stage was the interviewers' training activities, both in the asking of a question and in being very persuasive toward respondents through appeals to intentions, attitudes, values and actions sponsored by them. At first, riders were randomly approached, and then the questions were administered. To obtain more accurate information, an intercept on-board survey is made, different from [19], where an intercept at bus stops survey is used, and [15], [17], where a Web-Survey is completed. Several reasons justified the choice to make an on-board survey. First, one removes the shortcomings derived from choosing checkpoints. Second, observations collected at checkpoints are not representative of all possible fare evaders met along the route. Conversely, along the entire route, one can intercept more passengers. Third, the web-survey presents some drawbacks as well. Indeed, web-survey methods are powerful tools to gather data at low cost but the panel of the analysis could be biased as it is confined to passengers who have a web-connection. Besides, several passengers evade the fare as they do not have sufficient money to buy a ticket and/or a pass e.g., migrants, low-income people, as also



shown in a recent analysis in Chile where the 42% rate of fare evasion is found in areas with the lowest socio-economic level [4]. Fourth, the on-board methodology has been preferred over mail and telephone owing to the higher response rates, the ability to collect data from those experiencing the service and so on [22].

At the end of survey, 2,171 usable questionnaires were collected, including people who have only tickets and no passes, as pass-holders are usually considered as not-evaders in this study. As the goal is to analyse evader passengers only, we disregarded from the dataset ticket-holders who revealed they had never evaded the fare. The final dataset contains 850 questionnaires and represents only the evader passengers at the different levels.

## 2.2 Summary of variables

Table 1 describes all the variables used in the analysis in depth. The variables are distinguished according to the previous sections:

It is worth noting that three variables are categorical (age, educational qualification, perceived inspection frequency and fare evasion frequency) and the remaining are dichotomous. Table 1 also provides descriptive statistics on the sample for each variable considered. All descriptive statistics are self-explanatory.

## 2.3 Data analysis

Data analysis followed a two-step approach in which the Multiple Correspondence Analysis (MCA) is performed before the Hierarchical Cluster Analysis (HCA).

The first step reports the result of MCA in order to reduce the variables used to describe the evader passengers, and leads to a minor number of latent variables (or factors) uncorrelated to each other. As an output of MCA, the second step provides results on some groups of evader passengers according to similarity criteria by the HCA. The characteristics observed within each group are as homogeneous as possible. MCA and HCA were performed by the free software, Tanagra [23].

### 2.3.1 Multiple correspondence analysis

The MCA is a factor analysis approach, which is appropriated when the observations in the dataset are described by a set of categorical variables (and binomial of course). The aim is to represent the dataset in a reduced space dimension, which enables to highlight the links between the observations and the variables.

The method builds a set of latent variables (or factors) which are linear combinations of the original variables and allows the expression of some concepts not directly observable, but which are the results of the combined measurement of the initial set of variables. In this way, the original variables are replaced by a reduced set of latent variables able to account for a large percentage of the total variance of the data. The number of latent variables is much lower than the number of original variables because original variables could be more or less correlated. Conversely, MCA is carried out to minimize multicollinearity effects between the factorial scores obtained as they are mutually orthogonal. The analysis is done on a matrix obtained from the Burt's one [24]. The  $p$  eigenvalues are extracted from it and the first  $q \leq p$  eigenvalues are retained as "latent variables" (factors) which reproduce, in decreasing order, the highest variability - inertia - among the observations in the Burt's matrix.

Table 1: Variables list.

Variables	Abbreviation	Description	%
<b>Socio-demographic characteristics</b>			
Gender	Gen F	Female	51.96%
	Gen M	Male	48.04%
Age	Above 50	Above 50 years old	4.39%
	26_50	Between 26 and 50 years old	28.87%
	Under 26	Under 26 years old	66.74%
Educational qualification	Upper_sc	Upper school graduated	43.65%
	Middle_sc	Middle school graduated	52.77%
	Middle_sc_n	Middle school not graduated	3.58%
Employment	Work_y	Worker	21.59%
	Work_n	Not worker	78.41%
	Stud_n	Not student	44.92%
	Stud_y	Student	55.08%
	Unempl_n	Employed	82.10%
	Unempl_y	Unemployed	17.90%
Car availability	Car_y	Has a car	17.90%
	Car_n	Does not have a car	82.10%
Reason to use bus	Other_use_bus	Use of the bus for other reasons (not related to the lack of trip alternatives)	23.09%
	No_alter_use_bus	Use of the bus because there are no alternatives	76.91%
<b>Travel behaviour characteristics</b>			
Trip purpose	Syst_trips_y	Systematic trips for work or study	42.03%
	Syst_trips_n	Not systematic trips (for other purposes)	57.97%
	Leis_trips_n	Not leisure trips (shopping, sport, amusement, etc.)	67.44%
	Leis_trips_y	Leisure trips	32.56%
Time of the day	Rush_hour_n	Not rush hour trips	84.41%
	Rush_hour_y	Rush hours trips (7.00-9.00 and 13.00-14.00)	15.59%
In-vehicle time	In_vehicle_time_more15	Travel time more than 15 minutes	68.71%
	In_vehicle_time_less15	Travel time less than 15 minutes	31.29%
Other transit systems use	Other_transit_y	Use of other transit systems	36.14%
	Other_transit_n	No use of other transit systems	63.86%
Bus use frequency	Freq_traveler_y	The user travels more than 3 days a week	85.10%
	Freq_traveler_n	The user travels less than 3 days a week	14.90%
	Daily_trips_under2	Number of daily trips under 2	60.85%
	Daily_trips_over2	Number of daily trips over 2	39.15%
Quality rating	Satisf_y	Satisfied user (grade on the overall service more than sufficient)	94.00%
	Satisf_n	Not satisfied user (vote on the overall service more than sufficient)	6.00%
<b>Personal knowledge of fare evasion</b>			
Perceived inspection frequency	Insp_freq_more5	The user has seen the inspectors boarding the vehicle to perform control activities more than 5 times in 4 months.	14.43%
	Insp_freq_1_5	The user has seen the inspectors boarding the vehicle to perform control activities a number of times between 1 and 5 in 4 months.	56.47%
	Insp_freq_null	The user has never seen the inspectors boarding the vehicle to perform control activities in 4 months.	29.10%
Amount of the fine	Know_fine_y	The user knows the amount of the fine	89.15%
	Know_fine_n	The user does not know the amount of the fine	10.85%
Fines in the past	Fine_past_y	The user has already been fined in the past	59.70%
	Fine_past_n	The user has not been fined in the past	40.30%

Table 1: Continued.

Variables	Abbreviation	Description	%
Honesty	Honesty_y	The user buys the ticket also if no checks are performed	45.15%
	Honesty_n	The user does not buy the ticket if no checks are performed	54.85%
Fare evasion frequency	Fare_evasion_always	The user evades the fare every time he travels (100% of trips)	7.04%
	Fare_evasion_often	The user often evades the fare (more than 50% of trips)	14.90%
	Fare_evasion_mildly	The user very little evades the fare (between 10% and 50% of trips)	12.82%
	Fare_evasion_rarely	The user evades the fare rarely (less than 10% of trips)	65.24%
It is right to check tickets	Right_check_y	The user thinks that it is right to check the tickets	90.99%
	Right_check_n	The user thinks that it is not right to check the tickets	9.01%
Bother for checking activities	Bothered_check_y	The user is bothered by the checking activities	21.94%
	Bothered_check_n	The user is not bothered by the checking activities	78.06%

Maximizing the inertia means searching the straight line “closer” to the points that represent the data, and measuring the distance between the vectors of the space in which the data are represented by a Chi-square metric. The inertia can be a criterion for the choice of the size of solution. Thus, a threshold for the inertia can be established (e.g., 80%) and the first  $q \leq p$  eigenvalues are selected when the cumulative percentage of inertia explained reaches or exceeds this threshold. The total number of eigenvalues  $p$  depends on the dataset dimension. More formally:

$$p = n - k, \quad (1)$$

where  $k$  represents the number of variables and  $n$  the total number of possible values that the categorical variables may assume. As a result, the increasing of the analysis size results in the increasing the number of eigenvalues, thus reducing the percentage of data variability explained by them. In the literature, some formulas have been proposed to correct the measurement of inertia explained by each eigenvalue, to make the criterion of the threshold more applicable. In this paper, the correction formula of inertia rate proposed by Benzécri [25], is considered.

### 2.3.2 Hierarchical cluster analysis

Based on the output of MCA, a HCA has been performed to find homogeneous groups of evader passengers. In this paper, a hierarchical agglomerative clustering algorithm is adopted, in which the observations in the same cluster are similar and those in distinct groups are different, according to a similarity (or dissimilarity) measure, which is often a measure of distance. Strategies for hierarchical clustering of agglomerative type use a “bottom up” approach; starting from one observation per cluster, successive agglomerations are built until the whole dataset belong to one cluster. In the large majority of hierarchical clustering methods, this is achieved by the use of an appropriate measure of distance between pairs of observations and the one of linkage criterion which specifies the distance among groups as a function of the pairwise distances between observations. Since numerical variables are considered (i.e., factorial scores), the Euclidean distance is used, whereas the Ward’s linkage

criterion is adopted in building the hierarchy [26]. The goal of Ward's linkage method is to merge observations into a cluster in order to minimize variance within that cluster. This means that an observation is selected to belong to a cluster if its inclusion in that cluster produces the least increase in the error sum of squares. The distance of the Ward's method, i.e.,  $D_{AB}$  is calculated as follows:

$$D_{AB} = \frac{\|\bar{X}_A - \bar{X}_B\|^2}{\frac{1}{N_A} + \frac{1}{N_B}}, \quad (2)$$

where  $A, B$  are the Set of clusters;  $N_A$  and  $N_B$  represent the number of observations in clusters  $A$  and  $B$ , respectively;  $\bar{X}_A, \bar{X}_B$  are the Mean Vectors representing clusters  $A$  and  $B$ ;  $\|\bar{X}_A - \bar{X}_B\|^2$  is the squared Euclidean distance between vectors  $\bar{X}_A$  and  $\bar{X}_B$ .

### 3 RESULTS AND DISCUSSION

In this section, the outcomes of the two analyses previously described are presented.

#### 3.1 Multiple Correspondence Analysis

From the application of MCA over twenty-three variables, two factors explain 84.27% of the total data variance according to the Benzecri's index. The results are reported in Table 2, which describes the influence of the most important variables in the determination of the two factors. Moreover, the coordinates on the related factor of the significant variables are also reported. The percentage of inertia shows factors that mostly explain the variability of the original data.

Factor 1 - "No alternatives to bus use" - identifies evader passengers that do not have a car and use the bus as the only transportation mode owing to lack of alternatives. Indeed, the analysis of the negative signs of the coordinates of *Car\_y* and *Other\_use\_bus* shows that those who have the availability of the car and use the bus for various reasons are negatively associated to Factor 1. The coordinates of Factor 1 are also negative for *Work\_y* and *Upper\_sc* variables, while they are positive for *Under\_26* and for *Fare\_evasion\_often*. This means that individuals who belong to this factor do not work, are not upper school graduated, are less than 26 years and often evade the fare.

Factor 2, "Employment", reflects the employment condition of users. In particular, it includes passengers that are Not engaged in Education, Employment or Training (NEET). With positive coordinates are also the variables *26\_50* and *Fare\_evasion\_always*. Thus, Factor 2 discriminates the interviewed people between 26 and 50 years old and those that always evade the fare.

#### 3.2 Cluster analysis

Using a hierarchical classification method with the factorial scores of the factors obtained by MCA, the cluster analysis identifies three groups of individuals. Table 3 shows the detailed results of HAC. In this table, for each group, the most significant variables are reported and the percentage of members are computed for the whole dataset and within each group.

The detection of the differences among clusters is highlighted with the *t-test*, which is a criterion used for the characterization of a group of observations according to different variables/categories. For a 5% significance level, the difference among the values of a descriptive statistic indicator computed on the whole sample and one computed on a sub-sample related to the group is significant if the absolute value is larger than 2. Therefore, the *t-test* is used as a criterion for the ranking of the attributes in order to identify those that play an essential role in the interpretation of the groups. What follows is the analysis of groups.

Table 2: Variables of both MCA factors with the higher contributions.

	Variable contribution (%)	Coordinate on the related Factor	% Inertia explained	% Accumulated inertia
<b>Factor 1: No alternatives to bus use</b>			<b>68.51</b>	<b>68.51</b>
Car availability	12.435			
<i>Car_y</i>		-1.417		
Reason to use bus	12.039			
<i>Other use bus</i>		-1.193		
Employment (Worker)	9.913			
<i>Work_y</i>		-1.132		
Age	9.808			
<i>Under 26</i>		+0.416		
Educational qualification	8.240			
<i>Upper sc</i>		-0.610		
Fare evasion frequency	7.161			
<i>Fare evasion often</i>		+0.787		
<b>Factor 2: Employment</b>			<b>15.76</b>	<b>84.27</b>
Employment (Student)	25.420			
<i>Student n</i>		+0.832		
Employment (Unemployed)	17.346			
<i>Unemployed y</i>		+1.325		
Age	10.280			
<i>26_50</i>		+0,671		
Fare evasion frequency	7.391			
<i>Fare evasion always</i>		+1.160		

Table 3: Description of clusters in terms of percentages for the most significant variables.

Cluster 1			Cluster 2			Cluster 3		
Variable	% Group [t-test]	% Overall	Variable	% Group [t-test]	% Overall	Variable	% Group [t-test]	% Overall
<i>Car_y</i>	83.0 [19.21]	18.0	<i>Stud_y</i>	87.6 [25.18]	55.1	<i>Unempl_y</i>	90.8 [19.11]	18.0
<i>Other use bus</i>	71.4 [18.09]	23.1	<i>Under_26</i>	70.0 [17.68]	66.7	<i>Stud_n</i>	58.9 [18.12]	44.9
<i>Work_y</i>	68.9 [16.38]	21.5	<i>Work_n</i>	61.8 [14.67]	78.5	<i>Syst_trips_n</i>	38.5 [7.96]	57.8
<i>26_50</i>	51.2 [12.21]	28.9	<i>Unempl_n</i>	59.0 [12.91]	82.0	<i>Fare evasion always</i>	64.4 [6.45]	6.9
<i>Upper_sc</i>	43.3 [12.14]	44.0	<i>No_alter_use bus</i>	59.9 [12.09]	76.9	<i>Fine_past_y</i>	35.9 [6.26]	60.0
<i>Stud_n</i>	40.3 [10.51]	44.9	<i>Car_n</i>	58.2 [12.02]	82.0	<i>26_50</i>	42.3 [5.91]	28.9
<i>Syst_trips_y</i>	40.1 [9.83]	42.2	<i>Leis_trips_y</i>	69.5 [8.36]	32.0	<i>Middle_sc_n</i>	70.4 [4.98]	3.2
<i>Leis_trips_n</i>	32.2 [8.80]	68.0	<i>Middle_sc</i>	57.5 [5.47]	52.8	<i>Car_n</i>	31.4 [4.74]	82.0
<i>Fare evasion rarely</i>	32.6 [8.68]	65.3	<i>Freq_traveler_y</i>	52.1 [4.83]	84.9	<i>Middle_sc</i>	34.5 [4.48]	52.8
<i>Honesty_y</i>	36.6 [8.19]	45.1	<i>Know_fine_y</i>	51.1 [4.17]	89.4	<i>Honesty_n</i>	31.8 [4.18]	54.9
<i>Know_fine_n</i>	57.8 [8.14]	10.6	<i>Fare evasion often</i>	61.9 [3.24]	14.8	<i>Bothered_check_y</i>	39.8 [4.05]	21.9
<i>Bothered_check_n</i>	27.9 [5.79]	78.1	<i>Honesty_n</i>	53.5 [3.18]	54.9	<i>No_alter_use bus</i>	31.0 [3.60]	76.9



### 3.2.1 Cluster 1: Accidental and/or unintentional fare evader passengers

Cluster 1 includes 199 users that represent 23.4% of the sample. From a socio-demographic perspective, the members of this cluster are car owners that use the bus for reasons not related to the lack of travel options (e.g., difficulty in parking, etc.), that work, are between 26 and 50 years old and are upper school graduated. From the viewpoint of travel behaviour characteristics, these passengers use the bus for systematic trips (e.g., work and/or university). According to the personal knowledge of fare evasion, the cluster comprises of passengers who rarely evade as in [15], would buy a ticket even if controls were not performed, do not know the amount of the fine and are not bothered by the ticket inspections. Therefore, this cluster mainly consists of people who use the bus by choice while having the car availability, and who rarely evade the fare, probably due to accidental causes and not for personal utility. Probably, they could have evaded the fare in the past year due to some contingent reasons (e.g., crowded buses that impede the ticket validation, ticket validator machine being out of service). Compared with the other group, this cluster is the less numerous in the group and represents about one in four fare evaders. It is worth noting that the work of [14], distinguishes two segments of evader passengers that can be included in our cluster. In these segments, an evasion occurrence is detected, rare and occasional respectively. They included the accidental evaders who have strong views against fare evasion and evaded the fare by accident only, and the unintentional evaders who meant to pay but sometimes find themselves fare evading because of difficulty or barriers to obtaining valid tickets.

### 3.2.2 Cluster 2: Calculating fare evader passengers

Cluster 2 includes 413 users that represent 48.6% of the sample. Focusing on the socio-demographic perspective, the members of this cluster are students, young (under 26 years old), do not work, do not have a car and they move by bus for lack of trip options. This result differs from that of [14]. In their study, deliberate evaders are significantly more likely to be employed full-time. The passengers who frequently travel by bus and mostly move for leisure trips belong to this cluster. According to the personal knowledge of fare evasion, the cluster comprises passengers who know the amount of the fine, often evade and who would not buy the ticket if controls are not performed. As a result, this cluster of passengers largely evades the fare as there are no ticket barriers and they believe it is unlikely they will be caught. Moreover, this cluster is less likely to hold strong honesty beliefs. Indeed, calculating passengers were also much more likely to believe it was acceptable to bend the rules to save money. These results are in accordance with [14], [15]. Therefore, this cluster mainly consists of people who have no trip alternative to the use of public transport and who often evade the fare, probably for their own economic benefit, well aware of the amount of the fine that they must expect if checked by the inspectors. This represents the most representative groups in the analyses, i.e., about one in two fare evaders is calculated risk-related.

### 3.2.3 Cluster 3: Chronic fare evader passengers

Cluster 3 includes 238 users that represent 28.0% of the sample. The members of this cluster are unemployed, do not study, are between 26 and 50 years old, do not have a car, are at most middle school graduated and move with the bus because they do not have other trip options. These passengers travel by bus for unsystematic trips. Referring to the personal knowledge of fare evasion, the cluster comprises passengers who always evade, have already been fined, would not buy a ticket if controls are not performed and are bothered by the check activities.

Therefore, this cluster mainly consists of people who habitually evade the fare, probably irrespective of the level of network inspection and the amount of the fine. Indeed, they usually



do not pay for bus or train tickets, they never pay their fines, and they are not deterred by even tougher sanctions as most of them could already have an extensive rap sheet. They make no effort at escaping from detection as they do not incur losses. They simply can sit and give a (true or false) name and address, and never open their mail. This means that these people are not affected by increasing the fines or inspection rates. For these people, there is no choice element at all: they will always evade as they can't be bothered. This class of passengers is relevant as it represents about 1 in 3 evader passengers. According to [14], this cluster will always fare evade. However, while their study draws conclusions about this segment using first-hand accounts, in this study, this segment was derived by considering information achieved from the personal knowledge of fare evasion.

### 3.3 Strategies for reducing fare evasion

As shown, the three fare evasion segments reflect different characteristics for fare evading. As a result, this fact suggests three strategies to try to decrease fare evasion; one strategy for accidental evaders, one for calculator evaders and another one for chronic ones.

Accidental (unintentional) fare evaders are “sufficiently at ease”. They are well available to pay the fare but are very quick to evade the fare if the ticket system is too hard to understand or is not well managed, e.g., the ticket validation machine is not in order. Moreover, this group includes those passengers who forget to buy or validate their tickets. Therefore, it consists of passengers who do not pay either for problems that are PTC-dependent or passengers-carelessness. According to our models, this represents the less numerous groups. This segment seems to be a simpler cluster to target as it mainly responds to organizational problems that are PTCs-dependent. Moreover, according to [15], traditional strategies such as increased inspection activities or awareness and information campaigns might be unproductive to tackle this group. In this area, the PTC could reduce fare evasion by making it more easy and simple to access the system. For example, if a possible excuse to avoid the payment of the fare is due to the ticket validation machine which is not in order, a PTC can avoid this fact by checking that vehicles have their “devices” in order before leaving the depot. In this way, the occurrence of in-vehicle problems is expected to reduce considerably and thus passengers cannot use this excuse to avoid the payment of the fare. Therefore, if the PTC wants to reduce this fact, it should reduce barriers to easy ticketing.

Calculator evaders are the cleverest in finding solutions to defraud the ticketing system consciously. They are always on guard and ready to benefit from drawbacks of the systems by looking to not be caught by inspectors. For example, they stay next to the ticket validation machine and validate their ticket when the inspectors get on board. According to our models, this evader group is the most numerous. This segment is a medium difficult cluster to target, besides response to actions taken by PTCs such as increased ticket inspections or changes to access systems e.g., use of barrier or turnstiles. Indeed, according to [6], this segment is sensible to the perception of controls. Therefore, increasing the perception that the offender will be caught and fined may influence the behaviour of this group and discourage it from committing the offence. This is also according to the literature on deterrence as potential evaders pay more attention to the certainty of being inspected than to the severity of the punishment if caught e.g., [27].

Chronic fare evaders habitually evade the fare and unlike previous clusters, they completely “ignore” the ticketing system and are not deterred by even tougher sanctions, as most of them could already have an extensive rap sheet. Moreover, as pointed out by [11], in the Netherlands chronic fare evaders usually generates disturbance and violence on-board. In our opinion, this segment is a very difficult cluster to target as it does not respond to neither

traditional strategies or to barriers to easy ticketing. Indeed, this cluster makes no effort at escaping from detection as they do not incur in losses. Therefore, strategies impeding the access to transportation system could be useful to counteract this cluster. Besides, this cluster is always present and it is largely accepted by PTCs e.g., [12].

#### 4 CONCLUSIONS AND DIRECTION OF FUTURE RESEARCH

In proof-of-payment transit systems, fare evasion is a relevant concern that heavily affects PTCs in terms of lost fare revenues, damage to corporate image, social iniquity and personal security. Besides, little research has examined characteristics of the population of fare evaders for a specific segment. In this context, two recent studies provided evidence that at least three clusters of fare evaders emerged in the passengers taken as a whole (evaders and not): accidental and/or unintentional, career and calculator-risk related. These studies evaluated these segments by qualitative and explorative researches using a web survey. Conversely, to the best of our knowledge, no study examined characteristics and attitudes that discriminate specific groups of exclusively fare evader passengers by means of an intercept survey. Therefore, in order to shed additional light on this fact, using 850 on-board personal interviews collected among passengers of an Italian PTC, by MCA and HCA, we derived three clusters of evader passengers: 1) the accidental fare evader passengers, 2) the calculating and 3) the chronic. Accidental evaders use the bus by choice and rarely evade the fare, probably due to fortuitous causes. Calculator-risk evaders use only the bus and often evade the fare, probably for their own economic benefit. Chronic evaders always evade the fare, probably irrespective of the level of network inspection and the amount of the fine. Although this research was small in scale (i.e., the dataset is not very large and includes data on the passengers of an Italian PTC), and further studies are recommended to confirm the results, it was large enough to establish patterns for some fare evaders' groups. Moreover, even if this study draws portraits of the fare evader groups in Cagliari, it contributes to a clearer sketch of the attitudes and characteristics that identify the different types of fare evaders. It also shows that PTC can adopt different strategies for each group of passengers and evaluating the effects. For example, the PTC can consider some actions such as 1) reducing the barrier to avoid the not-payment fare, for accidental/unintentional evaders; 2) increasing the perception that fare evaders will be caught and fined for the calculator-related risk; 3) impeding the access to the public transportation system for chronic fare evaders.

Moreover, this study confirms that passengers that evade the fare more frequently were more likely to be students or unemployed, without alternative to bus as in [18], [19], while it differs from [14], [15], [17], where deliberate evaders were actually more likely to be educated full-time workers. This difference can be explained owing to differences in the survey method. Nevertheless, these results are preliminary steps in our agenda and raise a relevant issue for future research in a psychological issue. An in-depth analysis of some general segments (e.g., the chronic ones) and further refined analysis of specific segments (e.g., analysing differences by gender) may open new insights on the psychological attributes of fare evaders.

#### REFERENCES

- [1] Bonfanti, G. & Wagenknecht, T., Human factors reduce aggression and fare evasion. *Public Transport International* **59**(1), pp. 28–32, 2010.
- [2] Diebel, L.E., The introduction of European fare collection techniques in the United States. *Journal of Advanced Transportation* **15**(1), pp. 55–60, 1981.
- [3] Mazar, N., Amir, O. & Ariely, D., The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, **45**(6), pp. 633–644, 2008.



- [4] Guarda, P., Galilea, P., Paget-Seekins, L. & de Dios Ortúzar, J., What is behind fare evasion in urban bus systems? An econometric approach. *Transportation Research Part A: Policy and Practice*, **84** pp. 55–71, 2016.
- [5] Sasaki, Y., Optimal choices of fare collection systems for public transportations: Barrier versus barrier-free, *Transportation Research Part B* **60** pp. 107–114, 2014.
- [6] Barabino, B., Salis, S. & Useli, B., A modified model to curb fare evasion and enforce compliance: Empirical evidence and implications. *Transportation Research Part A* **58**, pp. 29–39, 2013.
- [7] Barabino, B., Salis, S. & Useli, B., Fare evasion in proof-of-payment transit systems: Deriving the optimum inspection level. *Transportation Research Part B* **70**, pp. 1–17, 2014.
- [8] Boyd, C., Martini, C., Rickard, J. & Russell, A., 1989. Fare evasion and non-compliance: A simple model. *Journal of Transport Economics and Policy* **23**(2), pp. 189–197.
- [9] Yin, Z., Jiang, A.X., Johnson, M., Tambe, M., Kiekintveld, C., Leyton-Brown, K., Sandholm, T. & Sullivan, J., TRUSTS: Scheduling Randomized Patrols for Fare Inspection in Transit Systems Using Game Theory. *AI Magazine* **33**(4) pp. 59–72, 2012.
- [10] Smith, M.J. & Clarke, R.V., Crime and public transport. In: Tonry, M. (ed) *Crime and Justice. A Review of Research*, **27**, pp. 169–233. University of Chicago Press, Chicago, 2000.
- [11] Bijleveld, C., Fare dodging and the strong arm of the law. *Journal of Experimental Criminology* **3**(2), pp. 183–199, 2007.
- [12] Clarke, R.V., Contre, S. & Petrossian, G., Deterrence and fare evasion: results of a natural experiment. *Security Journal* **23**(1), pp. 5–17, 2010.
- [13] Killias, M., Scheidegger, D. & Nordenson, P., The effects of increasing the certainty of punishment: A field experiment on public transportation. *European Journal of Criminology* **6**(5), pp. 387–400, 2009
- [14] Delbosc, A. & Currie, G., Four types of fare evasion: A qualitative study from Melbourne, Australia. *Transportation Research Part F*, **43**, pp. 254–264, 2016.
- [15] Delbosc, A. & Currie, G., Cluster analysis of fare evasion behaviours in Melbourne, Australia. *Transport Policy*, **50**, pp. 29–36, 2016.
- [16] Suquet, J.B., Drawing the line: how inspectors enact deviant behaviours. *Journal of Service Marketing*, **24**(6), pp. 468–475, 2010.
- [17] Currie, G. & Delbosc, A., An empirical model for the psychology of deliberate and unintentional fare evasion. *Transport Policy*, **54**, pp. 21–29, 2017.
- [18] Barabino, B., Salis, S. & Useli, B., What are the determinants in making people free riders in proof-of-payment transit systems? Evidence from Italy. *Transportation Research Part A*, **80**, pp. 184–196, 2015.
- [19] Bucciol A., Landini F. & Piovesan M., Unethical behaviour in the field: Demographic Characteristics and beliefs of the cheater. *Journal of Economic Behaviour & Organization* **93**, pp. 248–257, 2013.
- [20] Abeler, J., Becker, A. & Falk, A., Representative evidence on lying costs. *Journal of Public Economics*, **113**, pp. 96–104, 2014.
- [21] Traxler, C., Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, **26**(1), pp. 89–103, 2010.
- [22] Transit Cooperative Research Program (TCRP), Synthesis 63, On Board and Intercept Transit Survey Techniques – A Synthesis of Transit Practice, Transportation Research Board, Washington, DC, 2005.

- [23] <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>.
- [24] Abdi, H. & Valentin, D., Multiple correspondence analysis. *Encyclopedia of measurement and statistics*, pp. 651–657, 2007.
- [25] Benzécri, J.P., Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers de l'Analyse des Données*, **4**, pp. 377–378, 1979.
- [26] Ward, J. H., Jr., Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, **58**, pp. 236–244, 1963.
- [27] Von Hirsch, A., Bottoms, A.E., Burney, E. & Wikström, P.O.H., *Criminal Deterrence and Sentence Severity: An Analysis of Recent Research*. Hart Publishing, Oxford, UK, 1999.

