

Optimisation of transit fares: a multimodal approach based on system and external costs

B. Montella¹, L. D'Acerno¹ & M. Gallo²

¹*Department of Transportation Engineering,
'Federico II' University of Naples, Italy*

²*Department of Engineering, University of Sannio (Benevento), Italy*

Abstract

In this paper a model for the optimisation of transit fares is proposed and tested on a trial network. This model considers a multimodal transportation system under the assumption of elastic demand for simulating the impacts of fare policies on modal split. The model takes into account all system and social costs: (transit and road) user costs, firm costs and external costs. First tests on a trial network show the non-negligible impacts of transit fare design on the overall transportation system conditions and on social costs.

Keywords: transit fares, multimodal optimisation, external costs, elastic demand.

1 Introduction

Fares for public transit services are generally set under the profitability threshold: ticket revenues cover only a (minor) part of operational costs, while the other part is subsidised by public funds. This faring policy is adopted for social and environmental reasons.

Chapleau [1], Hodge [2], Obeng [3] and Parry and Bento [4] studied the social aspects of the problem. More generally, the effects of transit pricing policies were studied by Ballou and Mohan [5], Ferrari [6], Huang [7], Karkaftis and McCarthy [8], Zhou *et al* [9], Wichiensin *et al* [10] and Proost and Van Dender [11].

In this paper an optimization model for establishing transit fares is proposed; this model, in accordance with problem features, satisfies the following requirements:



- it assumes the transportation demand as elastic, for simulating modal split changes;
- it simulates road and transit transportation systems jointly, to take into account the effects of modal split changes on congestion;
- it considers different user classes with different socio-economic attributes;
- it assumes an objective function that considers all costs: operational costs, user costs of both road and transit systems, and external costs.

Therefore, this problem can be seen as a multimodal network design problem (Montella *et al* [12]), in which transit fares assume the role of decisional variables.

A general formulation of pricing problem was provided by Cascetta [13], and the multimodal nature of the analysed problem was highlighted by Huang [7]. Finally Osula [14] showed how changes in transit fares can also modify trip generation.

2 General solution approach

In more complex cases, establishing public transit fares requires five phases (see fig. 1), that can differ according to the area under study.

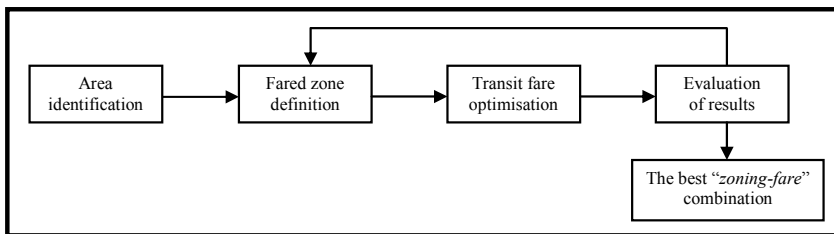


Figure 1: The general methodology of transit fare optimisation.

In the area identification phase the area covered by transit services under study is identified; it can be a town, or part of it, a province or a region.

The fare zone definition phase (see section 3) consists in subdividing the territorial area into zones, if it is so large that the fares have to be differentiated by trip length. The zones should take account of administrative divisions, land uses, pre-existence of fare regulations and user perception of fared zones. The transit fare optimisation phase searches for the optimal configuration of transit fares and is the focus of this paper. Evaluation of results can suggest modifications to area zoning; by comparing different results, the best zoning-fare combination is chosen.

3 Fared zone definition

If the area is so large that it is not fair to fix the same fare for any trip, the area has to be subdivided into fared zones; it is common in metropolitan areas, provinces and regions.

In the literature, several methods have been proposed; they are based on distances (Cervero [15]; Daskin *et al* [16]), time intervals (Cervero [17]) or average travel times (Phillips and Sanders [18]). However, three main methods of fare zoning can be identified (fig. 2): concentric zones, circular rings and sector zones, and alveolar zones.

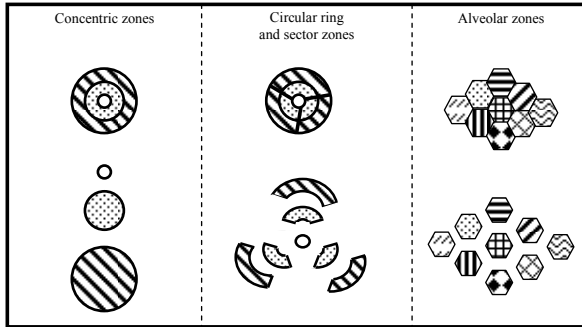


Figure 2: Fared zone types.

Concentric zones are mainly adopted where there is an important centre that attracts and/or generates most trips (e.g. a capital town). In this case, also fares for tangential trips are provided. If tangential trips are considerable, circular rings and sector zones are more used: in this case fares depend on the number of crossed zones. Finally, when there is not a main centre, the area can be subdivided into alveolar zones: fares depend on the number of crossed zones as well.

Although, theoretical speaking, each fare can be independent of the others, in general each fare can be defined according to the minimum fare, that is the fare of a trip inside a single zone, and of number of crossed zones. This relation can be expressed as:

$$T_b = t_b^0 \mathbf{1}^T + \mathbf{Diag}(t_b^*) A \mathbf{Diag}(N) \tag{1}$$

where T_b is the transit fare matrix, of dimensions $(n_{Tickets} \times n_{MaxCrossedZones})$, whose generic element t_i^n is the fare of the i -th ticket type that allows up to n zones to be crossed; t_b^0 is the vector of basic transit fare, of dimensions $(n_{Tickets} \times 1)$, whose elements t_i^0 are the first row elements of matrix T_b and represent fares that allow travel only within the same zone (*intra-zonal trips*); $\mathbf{1}$ is a vector, of dimensions $(n_{Tickets} \times 1)$, whose elements are all equal to 1; t_b^* is the vector of basic variation fares, of dimensions $(n_{Tickets} \times 1)$, whose elements represent the unit variation fare without considering any corrective coefficient; A is the corrective coefficient matrix, of dimensions $(n_{Tickets} \times n_{MaxCrossedZones})$, whose generic element a_i^n is a corrective coefficient related to the i -th ticket type that allows up to n zones to be crossed; N is a vector, of dimensions $(n_{MaxCrossedZones} \times 1)$, whose generic element n_j is equal to $(j - 1)$.

Eqn (1) is equivalent to the following:

$$T_b = \text{Diag}(t_b^0) \Gamma \text{Diag}(N+1) \quad (2)$$

where Γ is a matrix, of dimensions $(n_{\text{Tickets}} \times n_{\text{MaxCrossedZones}})$, whose generic element γ_i^n is a coefficient related to the i -th ticket type that allows up to n zones to be crossed. Indeed, since both previous equations are linear, assuming $\text{Diag}(t_b^*) = \text{Diag}(t_b^0) \text{Diag}(A^*)$ (where A^* is a vector whose generic element a_i^* is equal to the ratio between t_i^* and t_i^0), it is possible to transform eqn (1) into eqn (2) via the following relation:

$$\Gamma = \bar{\mathbf{1}} [\text{Diag}(N+1)]^{-1} + \text{Diag}(A^*) A \text{Diag}(N) [\text{Diag}(N+1)]^{-1} \quad (3)$$

where $\bar{\mathbf{1}}$ is a matrix, of dimensions $(n_{\text{Tickets}} \times n_{\text{MaxCrossedZones}})$, whose elements are all equal to 1. Thus, in this paper we propose to adopt eqn (2) for defining zoned fare criteria using the matrix Γ as the only parameter.

4 Transit fare optimisation model

The proposed transit fare optimisation model can be formulated as follows:

$$T_b^{\wedge} = \arg \min_{T_b \in S_{T_b}} Z(T_b, f_c^*, f_b^*) \quad (4)$$

subject to:

$$(f_c^*, f_b^*) = \mathcal{A}(T_b, f_c^*, f_b^*) \quad (5)$$

with:

$$f_c^* \in S_{f_c} ; f_b^* \in S_{f_b} \quad (6)$$

where T_b^{\wedge} is the optimal value for T_b ; S_{T_b} is the feasibility set of matrices T_b ; $Z(\cdot)$ is the objective function; f_c^* is the equilibrium link flow matrix referred to the road system, of dimensions $(n_{\text{RoadLinks}} \times n_{\text{UserCategories}})$; f_b^* is the equilibrium link flow matrix referred to the transit system, of dimensions $(n_{\text{TransitLinks}} \times n_{\text{UserCategories}})$; $\mathcal{A}(\cdot)$ is the multimodal assignment function; S_{f_c} is the feasibility set for f_c^* ; S_{f_b} is the feasibility set for f_b^* .

Constraint (5) represents the multimodal equilibrium assignment; it constrains road and transit flows to be in multimodal equilibrium for each configuration of transit fares, T_b . At equilibrium, the road and transit flows generate road and transit generalised costs that produces a modal split and path choices such that the same flows are reproduced. For solving the multimodal equilibrium assignment problem, in this paper we adopt the fixed-point model and the solution algorithms proposed by D'Acerno *et al* [19].

Constraint (6) indicates that flows have to belong to feasibility sets that express consistency of flows.

The proposed model shows a bi-level formulation: the upper level is the optimisation model (4) and the lower level is the assignment problem (5) subject to constraint (6).

If matrix Γ is fixed, through eqn (2), the optimisation model (4) can be simplified as:

$$t_b^0 \wedge = \arg \min_{t_b^0 \in S_{t_b^0}} Z(\mathbf{Diag}(t_b^0) \Gamma \mathbf{Diag}(N+1), f_c^*, f_b^*) = \arg \min_{t_b^0 \in S_{t_b^0}} Z(t_b^0, f_c^*, f_b^*) \quad (7)$$

subject to:

$$(f_c^*, f_b^*) = A (\mathbf{Diag}(t_b^0) \Gamma \mathbf{Diag}(N+1), f_c^*, f_b^*) = A (t_b^0, f_c^*, f_b^*) \quad (8)$$

with:

$$f_c^* \in S_{f_c} ; f_b^* \in S_{f_b} \quad (9)$$

where $t_b^0 \wedge$ and $S_{t_b^0}$ are respectively the optimal value and the feasibility set of t_b^0 .

This second model has less complexity since the vector t_b^0 has only $n_{Tickets}$ elements, while matrix T_b has $n_{Tickets} \times n_{MaxCrossedZones}$ elements to be optimised.

In this paper we test two objective functions; the first is the objective function adopted in several transit network design problems that considers only system costs and user costs, adapted to our multimodal problem:

$$Z_1(T_b, f_c^*, f_b^*) = OTC(T_b, f_c^*, f_b^*) + UGC(T_b, f_c^*, f_b^*) \quad (10)$$

where $OTC(\cdot)$ is the Operational Transit Cost and $UGC(\cdot)$ is the User Generalised Cost of all transportation systems.

The second also considers the external costs, $EC(\cdot)$, produced by road traffic:

$$Z_2(T_b, f_c^*, f_b^*) = OTC(T_b, f_c^*, f_b^*) + UGC(T_b, f_c^*, f_b^*) + EC(T_b, f_c^*, f_b^*). \quad (11)$$

In both functions the transit revenues and the ticket costs for users are not present since they annul each other.

5 First results and discussion

The proposed optimisation model is tested on the simple network of fig. 3, where the problem may be solved in a closed form; this network has one link, a , shared by road and transit systems, and there is a single bus line l . Moreover, it is assumed that there is a single fared zone, a single ticket type, a single user category and hence a single fare, t_b^0 .



The operational transit cost, OTC , is calculated as:

$$OTC = c_{oc}^l L_l \varphi_l \tag{12}$$

where c_{oc}^l is the operational cost of line l ; L_l is the length of line l ; φ_l is the service frequency of line l .

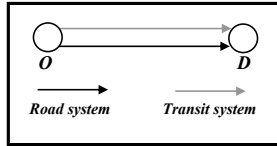


Figure 3: Simple network framework.

Road travel time, $UTC_{c,a}$, on the link is expressed by means of cost function $c_a(f_{c,a})$ and a road monetary cost (road/parking pricing), RMC_a , can be applied on the link.

Transit travel time, $UTC_{b,a}$, is the sum of the on-board time, assumed equal to the road travel time, $c_a(f_{c,a})$, the waiting time, equal to the ratio between a regularity term η_l and the frequency φ_l of the line, and the access-egress time, c_p , that is:

$$UTC_{b,a} = c_a(f_{c,a}) + \eta_l / \varphi_l + c_p. \tag{13}$$

The mode choice model is expressed by means of a Multinomial Logit model, where the systematic utility of road [transit] users is equal to:

$$\begin{aligned} & -\beta_a^{TIME} UTC_{c,a} - \beta_a^{COST} RMC_a / \delta_c - \beta_a^{SE} MSE_c, \\ & [-\beta_a^{TIME} UTC_{b,a} - \beta_a^{COST} t_b^0 - \beta_b^{SE} MSE_b]. \end{aligned} \tag{14}$$

where MSE_c [MSE_b] is the socio-economic variable of modal choice model for road [transit] system and the β terms are the parameters of the model. Therefore the road travel demand is equal to:

$$d_c = d / (1 + \exp(-(\beta_a^{TIME} (\eta_l / \varphi_l + c_p) - \beta_a^{COST} RMC_a / \delta_c + \beta_b^{SE} MSE_b + \beta_c^{SE} MSE_c) / \theta_M)) \exp(-\beta_a^{COST} t_b^0 / \theta_M). \tag{15}$$

and the transit travel demand is equal to:

$$d_b = d - d_c. \tag{16}$$

Assuming that capacity constraints are not present on the transit system, the transit (user) flow, $f_{b,a}$, is equal to the transit demand, d_b ; moreover, the road (vehicle) flow, $f_{c,a}$, is equal to the ratio between the road demand, d_c , and the

occupancy index, δ_c , assumed constant. Therefore, objective function (10) can be written as:

$$Z_1(t_b^0, f_{c,a}, f_{b,a}) = c_{oc}^l L_l \varphi_l + d(\eta_l/\varphi_l + c_p) \beta_a^{TIME}/\beta_a^{COST} + (RMC_a/\delta_c + (\eta_l/\varphi_l + c_p) \beta_a^{TIME}/\beta_a^{COST}) f_{c,a} + d c_a(f_{c,a}) \beta_a^{TIME}/\beta_a^{COST}. \tag{17}$$

Since objective function (17) is continuous with continuous first and second partial derivatives, the road travel time function is continuous with continuous first and second partial derivatives, and the feasibility set of transit fares is a closed interval $[0, t_b^{0,max}]$, solution t_b^0 of problem (7) is one of the points among endpoints of feasibility interval (i.e. 0 and $t_b^{0,max}$) and values t_b^0 (of the above feasibility set) that satisfy conditions:

$$\begin{cases} [\partial Z_1 / \partial t_b^0]_{t_b^0 = \bar{t}_b^0} = 0 \\ [\partial^2 Z_1 / \partial (t_b^0)^2]_{t_b^0 = \bar{t}_b^0} \geq 0 \end{cases} \tag{18}$$

If road travel time is estimated by means of a BPR function, that is:

$$c_a(f_{c,a}) = c_a^0 (1 + \alpha_a^{BPR} (f_{c,a})^{\beta_a^{BPR}} / (\delta_c Cap_a)^{\beta_a^{BPR}}) \tag{19}$$

it is possible to state that eqn (18) is satisfied with:

$$\begin{aligned} \bar{t}_b^0 = & -\frac{\theta_M}{\beta_a^{COST}} \ln \left(1 / \beta_a^{BPR} \sqrt[1]{\frac{(\delta_c Cap_a)^{\beta_a^{BPR}}}{c_a^0 \alpha_a^{BPR} \beta_a^{BPR} d^{\beta_a^{BPR}}} \left(\eta_l + c_p - \frac{RMC_a \beta_a^{COST}}{\delta_c \beta_a^{TIME}} \right) - 1} \right) +, \\ & -\frac{\beta_a^{TIME}}{\beta_a^{COST}} \left(\frac{\eta_l}{\varphi_l} + c_p \right) + \frac{RMC_a}{\delta_c} - \frac{\beta_b^{SE} MSE_b - \beta_c^{SE} MSE_c}{\beta_a^{COST}}. \end{aligned} \tag{20}$$

only if:

$$RMC_a \in \begin{cases} \left[\frac{\delta_c \beta_a^{TIME}}{\beta_a^{COST}} \left(\eta_l + c_p - \frac{c_a^0 \alpha_a^{BPR} \beta_a^{BPR} (d^0)^{\beta_a^{BPR}}}{(\delta_c Cap_a)^{\beta_a^{BPR}}} \right); \frac{\delta_c \beta_a^{TIME}}{\beta_a^{COST}} \left(\eta_l + c_p \right) \right] \\ \text{if } \frac{\eta_l}{\varphi_l} + c_p \geq \frac{c_a^0 \alpha_a^{BPR} \beta_a^{BPR} (d^0)^{\beta_a^{BPR}}}{(\delta_c Cap_a)^{\beta_a^{BPR}}} \\ \left[0; \frac{\delta_c \beta_a^{TIME}}{\beta_a^{COST}} \left(\eta_l + c_p \right) \right] \\ \text{if } \frac{\eta_l}{\varphi_l} + c_p < \frac{c_a^0 \alpha_a^{BPR} \beta_a^{BPR} (d^0)^{\beta_a^{BPR}}}{(\delta_c Cap_a)^{\beta_a^{BPR}}} \end{cases} \tag{21}$$



where c_a^0 is the free-flow road travel time on link a , that is equal to the ratio between the length of link a and the free-flow average speed on the same link; α_a^{BPR} and β_a^{BPR} are parameters of the BPR function; Cap_a is the capacity of link a . Therefore if condition (21) is not satisfied, the solution of problem (7) is one of the endpoints of interval $[0, t_b^{0,max}]$.

Adopting the parameters in table 1, the solution of problem (4) may be analysed. In particular, fig. 4 shows objective function (17) by different TWT values, with:

$$TWT = UTC_{b,a} - UTC_{c,a} = \eta_l / \varphi_l + c_p \tag{22}$$

where the TWT variation is obtained by φ_l variation.

Table 1: Parameter values.

$d = 3,000$ user/h	$\beta_a^{TIME} = 1.00$
$Cap_a = 2,000$ veh./h	$\beta_c^{COST} = 0.0833$
$\delta_c = 1.30$ user/veh.	$\beta_b^{SE} \cdot MSE_b = \beta_c^{SE} \cdot MSE_c$
$\alpha_a^{BPR} = 0.15$	$\theta_M = 0.04$
$\beta_a^{BPR} = 4.00$	$\varphi_l = 4.0$ bus/h
$c_a^0 = 4.80$ min	$\eta_l = 0.50$
$RMC_a = 1.50$ €	$c_p = 2.50$ min
$L_l = 4.00$ km	$c'_{oc} = 4.00$ €/km

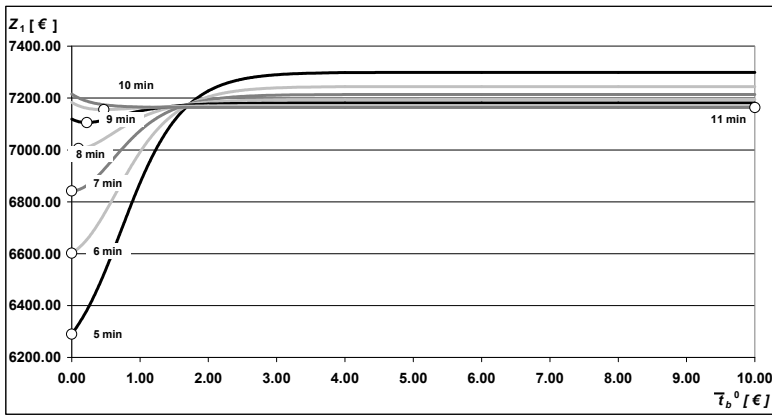


Figure 4: Objective function chart by TWT value.

A major result is that with TWT values lower than 8 min and higher than 10 min the solution is an endpoint of feasibility interval (in this case $t_b^{0,max}$ is equal to €10.00). Besides when the solution is a value that puts the first derivative equal to zero, the transit fare increases when the TWT value increases. This means that, when the transit travel time increases with respect to the road travel time, the optimum of the system (hence the minimum value of the objective

function) can be obtained by increasing transit fares and hence moving users from the transit system to the road system. Indeed, although the increase in the number of road users increases both road and transit travel times and hence objective function value, the decrease in the number of transit users decreases the number of users that incur the transit waiting time (*TWT*) and yields a decrease in the objective function that counterbalances the increase due to travel time increases.

Fig. 5 shows that eqn (20) has an asymptotic trend to endpoints of its feasibility interval (expressed by 21). In this case the *TWT* variation yields a translation of \bar{t}_b^0 and endpoints of eqn. (20) feasibility set, where the offset is equal to the product of the value δ_c , *TWT* variation and the ratio between β_a^{TIME} and β_a^{COST} .

Another important result is that, in fixing *TWT* values, an increase in road monetary costs yields a decrease in transit fares. Indeed, an increase in road monetary costs increases the objective function value while a decrease in transit fares yields a decrease in the number of road users that entails a decrease in both road and transit travel times, a decrease in users that have to support road monetary costs, and hence a decrease in the objective function value that counterbalances the increase due to the rise in road monetary costs. Finally, the dotted line, concerning the value of the road monetary cost equal to €1.50, intercepts curves in solution points of fig. 4.

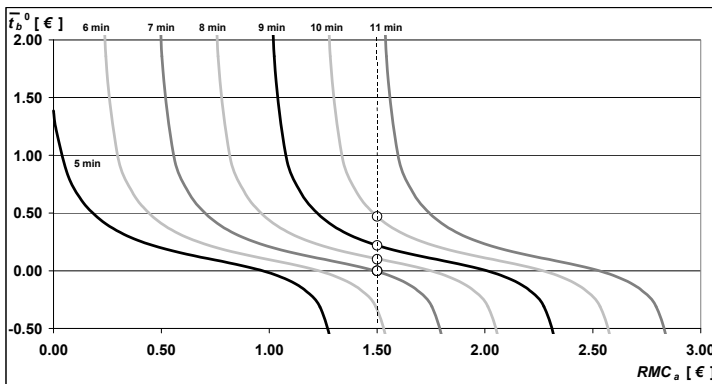


Figure 5: Transit fares by RMC_a and *TWT* values.

An increase in the ratio between β_a^{TIME} and β_a^{COST} (for instance by fixing β_a^{TIME} and decreasing β_a^{COST}), yields an increase in both width of the feasibility interval and the value of the transit fare (as shown in fig. 6).

Introducing the external costs in the objective function, assuming that:

$$EC(t_b^0, f_{c,a}, f_{b,a}) = -\beta_{TUV} d_b \tag{23}$$



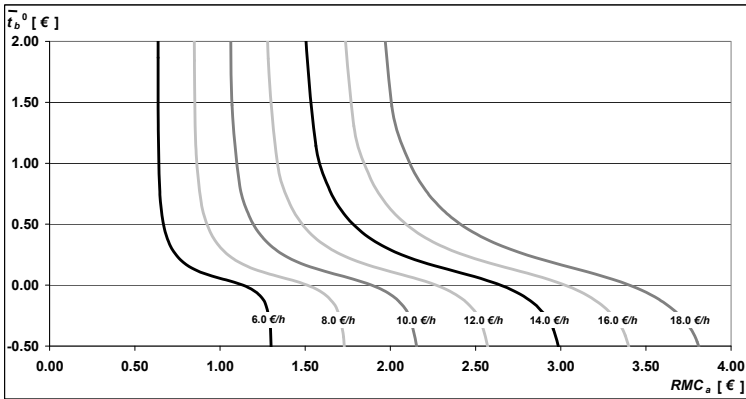


Figure 6: Transit fares by RMC_a and values of time.

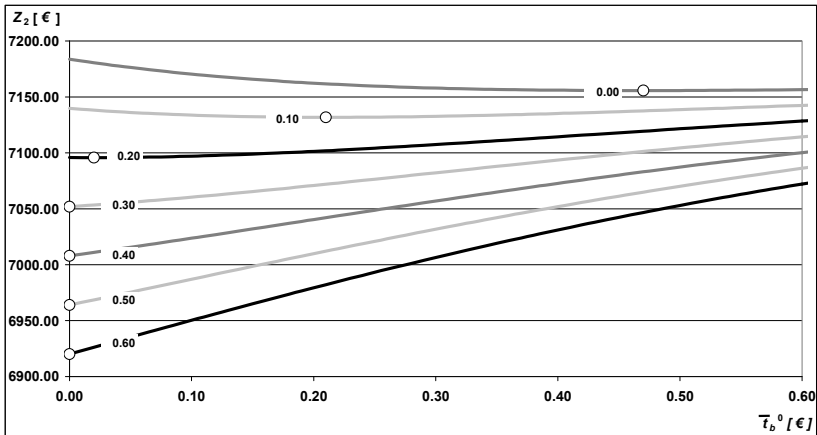


Figure 7: Objective function chart by β_{TUV} values.

where β_{TUV} is the Transit User Value that expresses the value that society associates to a user that travels on the transit system, the objective function (17) can be written as:

$$Z_2(t_b^0, f_{c,a}, f_{b,a}) = c_{oc}^l L_1 \varphi_l + d(\eta_l \varphi_l + c_p) \beta_a^{TIME} / \beta_a^{COST} - \beta_{TUV} d_b + (RMC_a \delta_c + \beta_{TUV} - (\eta_l \varphi_l + c_p) \beta_a^{TIME} / \beta_a^{COST}) f_{c,a} + d c_a(f_{c,a}) \beta_a^{TIME} / \beta_a^{COST}. \quad (24)$$

Also in this case solution $t_b^{0\wedge}$ of problem (7) is one of the points among endpoints of feasibility interval (i.e. 0 and $t_b^{0,max}$) and values \bar{t}_b^0 (of the above feasibility set) that satisfy the conditions (18) for $Z_2(\cdot)$.

Fig. 7 shows that an increase in term β_{TUV} yields a decrease in transit fares.

6 Conclusions and research prospects

The model proposed in this paper allows transit fares to be optimised, reducing the total costs of multimodal transportation system. The tests on a trial network show the importance of considering the external costs in the objective function. Research prospects will be addressed to apply the model to real networks and propose solution algorithms that can be adopted in real-scale cases.

References

- [1] Chapleau, R., Measuring impacts of transit financing policy in geopolitical context: Montreal case. *Transportation Research Record*, **1496**, pp. 52–58, 1995.
- [2] Hodge, D.C., Fiscal equity in urban mass transit systems: a geographic analysis. *Annals of the Association of American Geographers*, **78**, pp. 288–306, 1998.
- [3] Obeng, K., Expense preference behavior in public transit systems. *Transportation Research E*, **36**, pp. 249–265, 2000.
- [4] Parry, I.W.H. & Bento, A., Estimating the welfare effect of congestion taxes: the critical importance of other distortions within the transport system. *Journal of Urban Economics*, **51**, pp. 339–365, 2002.
- [5] Ballou, D.P. & Mohan, L., Decision model for evaluating transit pricing policies. *Transportation Research A*, **15**, pp. 125–138, 1981.
- [6] Ferrari, P., A model of urban transport management. *Transportation Research B*, **33**, pp. 43–61, 1999.
- [7] Huang, H.J., Fares and tolls in a competitive system with transit and highway: the case with two groups of commuters. *Transportation Research E*, **36**, pp. 267–284, 2000.
- [8] Karkaftis, M.G. & McCarthy, P., Cost structures of public transit systems: a panel data analysis. *Transportation Research E*, **38**, pp. 1–18, 2002.
- [9] Zhou, J., Lam, W.H.K. & Heydecker, B.G., The generalized nash equilibrium model for oligopolistic transit market with elastic demand. *Transportation Research B*, **39**, pp. 519–544, 2005.
- [10] Wichensin, M., Bell, M.G.H. & Yang, H., Impact of congestion charging on the transit market: An inter-modal equilibrium model. *Transportation Research A*, **41**, pp. 703–713, 2007.
- [11] Proost, S. & Van Dender, K., Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. *Transportation Research A*, **42**, pp. 1120–1230, 2008.
- [12] Montella, B., Gallo, M. & D’Acierno, L., Multimodal network design problems. *Urban Transport V - Urban Transport and the Environment for the 21st Century*, ed. L.J. Sucharov, WIT Press: Southampton, United Kingdom, pp. 405–414, 2000.
- [13] Cascetta, E., *Transportation Systems Analysis: Models and Applications, 2nd Edition*, Springer Verlag GmbH, 2009.



- [14] Osula, D.O.A., A procedure for estimating transit subsidization requirements for developing countries. *Transportation Research A*, **32**, pp. 599–606, 1998.
- [15] Cervero, R., The transit pricing evaluation model: a tool for exploring fare policy options. *Transportation Research A*, **16**, pp. 313–323, 1982.
- [16] Daskin, M.S., Schofer, J.L. & Haghani, A.E., A quadratic programming model for designing and evaluating distance-based and zone fares for urban transit. *Transportation Research B*, **22**, pp. 25–44, 1988.
- [17] Cervero, R., Flat versus differentiated transit pricing: what's a fair fare. *Transportation*, **10**, pp. 211–232, 1981.
- [18] Phillips, B. & Sanders, P., Time-based zones: an alternate pricing strategy. *Journal of Product and Brand Management*, **8**, pp. 73–82, 1999.
- [19] D'Acierno, L., Montella, B. & Gallo, M., Multimodal assignment to congested networks: fixed-point models and algorithms. *Proceedings of European Transport Conference 2002*, Cambridge, United Kingdom, 2002.

