Research in Transportation Business & Management 53 (2024) 101102

Contents lists available at ScienceDirect

Research in Transportation Business & Management

journal homepage: www.elsevier.com/locate/rtbm

Benchmarking the efficiency of European metros from a production perspective

Luigi Castagna[a,*], António Lobo[b], Pierluigi Coppola[a], António Couto[b]

[a] Politecnico di Milano, Department of Mechanical Engineering, Via G. La Masa 1, 20156 Milano, Italy
[b] CITTA - Centro de Investigação do Território, Transportes e Ambiente, Faculdade de Engenharia da Universidade do Porto, Portugal

ARTICLE INFO

Keywords:
Production function
Stochastic frontier analysis
Technical efficiency
Public transport planning
transit operator performance

ABSTRACT

This paper deals with the benchmarking of the technical efficiency of 23 metro systems in Europe. Since financial data reflecting the operating costs and revenues are not often made available to the public, the aim of this work is to develop a method based on production variables to enable large-scale analysis at the European level. The methodology consists of two stages. In the first stage, a gross value of effectiveness is estimated by means of a stochastic frontier regression based on the Cobb-Douglas production function. The results show about half of the considered firms reaching scores higher than 80%. However, these gross effectiveness estimates could be influenced and constrained by long term and external factors that go beyond the control of firms' day-to-day management. For this reason, in the second stage, an exponential multiple regression is estimated to determine the effects of these factors on gross effectiveness. The elasticities obtained through a multiple regression are used for evaluating the net effectiveness, by removing positive or negative contributions to the gross effectiveness that come from the identified "long term" factors. The results show that transit firms operating smaller networks tend to have higher net effectiveness scores in the short-term compared to larger transit firms.

1. Introduction

The comparison of operators' performance is important in public transportation systems because transport firms usually operate in their local markets under concession without any competition within the market, which might lead to some inefficiencies. Therefore, it is important for regulators and stakeholders to measure and monitor the performance of a local transport operator, benchmarking it against other firms operating in the same sector. Performance evaluation is also crucial to evaluate the capacity of operators to adapt to societal or network changes across time, considering that the urban context where these firms operate is highly dynamic. Moreover, firms operating metro networks are characterized by high operation expenditures which are not fully covered by the service's farebox revenues, since in most cases fares are set by public authorities with the aim of promoting an attractive and inclusive public transport network also for the lower income groups. The need therefore arises to understand what the optimal level of public subsidies is, also in relation to the conditions of the surrounding environment (context and market).

Efficiency analyses allow us to distinguish between factors related to operators' inefficiencies (internal factors) and critical factors related to the market (external factors). Moreover, they help to identify fair levels of public subsidies for transit operation costs that are not fully covered by revenues, also in relation to different urban contexts. For these reasons, they are fundamental tools when assessing the performances of metro systems. Despite different urban transit firms having different goals (De Borger, Kerstens, & Costa, 2002), studies focused on productivity and technical efficiency as all public transport sector activities are required to operate efficiently, and this has been a matter of concern for decades for the governments, transport authorities and researchers (Lobo & Couto, 2016).

The aim of this paper is to collect and integrate data and develop a method for the benchmarking analysis of the technical efficiency of European metro systems. Financial variables were used in previous analyses (Brage-Ardao, Graham, & Anderson, 2015), (Tsai, Mulley, & Merkert, 2015). However, very often this data is not made publicly available or it is reported with different levels of detail (Tsai et al., 2015). In fact, in many European cities, the same public transport firm operates metro systems together with the other urban public transport modes (e.g., urban buses, trams, trolleybuses) eventually available in the city. The operator usually publishes only aggregated revenues and costs data in its annual reports and accounting documents, from which it is

* Corresponding author.
E-mail address: luigi.castagna@polimi.it (L. Castagna).

https://doi.org/10.1016/j.rtbm.2024.101102
Received 17 October 2022; Received in revised form 16 December 2023; Accepted 13 January 2024
Available online 23 January 2024

not possible to allocate costs and revenues related only to metro operations. Different reporting criteria and different levels of detail and data aggregation were found in the preliminary assessment of income and expense statements, even between firms in the same country. Moreover, allocative efficiency analysis implies a behavioural assumption of cost minimization or profit maximization (Coelli, Rao, O'Donnell, & Battese, 2005), which may be questionable in the case of state-owned metro operators, which usually prioritize service supply.

In contrast, most metro operators usually publish data such as network and fleet size, labour force and produced outputs (car-kilometres, passenger-kilometres, etc.). This study was developed under the perspective of production, with the aim of delivering benchmarking tools that make use of open data, to facilitate the inclusion of systems that otherwise would not be considered as well as the update of this benchmarking exercise over time. In this context, the efficiency analysis was performed through estimation of a production function that expresses a relation between the consumption of inputs and the output obtained. The inputs used to produce the output are of two types: labour and capital. Labour input is expressed by the number of employees directly involved in the operation of the network. Capital consumption is mainly related to the use of materials (fleet and infrastructure) and energy. Since there is no detailed data about consumptions and maintenance requirements, it is necessary to find proxies, and a good-one could be to use the car-kilometres to characterize the consumption of the fleet and infrastructure, also proxying the consumption of energy for train circulation. In the first stage, effectiveness scores are obtained using the number of passengers carried per year as the output variable, but these scores are affected by several factors that are beyond the managerial control in the short-medium term. For this reason, we define these first effectiveness estimates as gross effectiveness, since the effects of external factors should be statistically controlled and removed to compute a net indicator of firms' effectiveness. Therefore, in a second stage, a regression is performed to evaluate the net effectiveness based on that which the firm can control directly, removing the effects of factors that are beyond its control in the short-medium term.

This paper contributes to the existing literature by performing a uniformized benchmarking on the efficiency of European metro systems under data constraints and, at the same time, by considering the impact of external factors on the efficiency of firms, distinguishing between socio-economic factors of the urban areas where metro systems operate and factors that firms cannot control in the short term. The stochastic frontier method used in this study allows us to estimate efficiency and test hypotheses on inputs' coefficients.

2. Literature review

For decades the efficiency of transport networks has been a focus of concern because of their key role of these systems in fostering economic and social equity and development. Research related to productivity and efficiency analysis was carried out across the entire transport sector. As an example, airport operating efficiency was studied by (Oum & Yu, 2004) and (Pels, Nijkamp, & Rietveld, 2003), while (Cullinane, Song, & Gray, 2002) and (Hung, Lu, & Wang, 2010) analysed the operating efficiency and benchmarked the performance of container ports. (Lobo, Amorim, Rodrigues, & Couto, 2018; Lobo, Couto, & Rodrigues, 2016; Lobo, Rodrigues, & Couto, 2014) proposed an adaptation of efficiency measures in the context of stochastic frontier analysis to propose new operating speed frontier models for two-lane roads that allow for the estimation of any desired percentile speed; the stochastic upper speed frontier, representing the fastest drivers, is estimated based on the road conditions, and the asymmetric error accounts for speed reductions ("inefficiencies") associated with the diversity of drivers' behaviour and vehicle technology.

Focusing on the public transport sector, several studies are present in the literature related to both road and railroad public transport. (Boame, 2004; Farsi, Filippini, & Kuenzle, 2006; Karlaftis, 2004; Von Hirschhausen & Cullmann, 2010) performed efficiency analyses of bus firms and (Merkert, Mulley, & Hakim, 2017) defined a two-stage model to benchmark 58 Bus Rapid Transit (BRT) systems operating worldwide. Comprehensive reviews of studies assessing bus firms' performances can be found in (De Borger et al., 2002) and (Daraio et al., 2016). Moreover, efficiency analyses have been applied also to subunits within the same transport organisation: this is the case of (Barnum, McNeil, & Hart, 2007) who analysed the efficiency of the park-and-ride lots of the Chicago Transit Authority.

Regarding the railway sector, comprehensive reviews of studies and methodologies dealing with its operational performance can be found in (Brons, Nijkamp, Pels, & Rietveld, 2005; Catalano, Daraio, Diana, Gregori, & Matteucci, 2019; Fried, Lovell, & Schmidt, 2008; Holvad, 2020; Oum, Waters, & Yu, 1999); examples of these studies are (Alam, Xuemei, Baig, Yadong, & Shah, 2020; Bojović, Milenković, Kapetanović, & Knežević, 2016; Cantos & Maudos, 2001; Cantos, Pastor, & Serrano, 1999; Chapin & Schmidt, 1999; Couto & Graham, 2009; Cowie, 1999; Growitsch & Wetzel, 2009; Jitsuzumi & Nakamura, 2010; Kutlar, Kabasakal, & Sarikaya, 2013; Mallikarjun, Lewis, & Sexton, 2014; Oum & Yu, 1994). Also, the efficiency of public transport authorities (PTAs) was the object of several studies: (Holmgren, 2013) analysed the cost efficiency of PTAs signing gross cost contracts with operators, while (Link, 2016, 2019) assessed the efficiency in using subsidies for franchised regional services. (Sameni, Preston, & Khadem Sameni, 2016) developed a two-stage approach to assess the performance of stations: in the first stage, technical efficiency (how well stations handle train stops) is computed, while in the second stage, service effectiveness (how well stations "transform" train stops into passengers' flow) is calculated. (Smith, 2012) assessed the relative efficiency performance of rail infrastructure managers after railway deregulation to determine whether and to what extent this policy change had an impact on the efficiency of railway systems. Similar analyses conducted after the important environment and policy changes introduced by railway deregulation and liberalization were performed by (De Jorge-Moreno & Garcia-Cebrian, 1999; Lerida-Navarro, Nombela, & Tranchez-Martin, 2019; Wetzel, 2008).

Efficiency and productivity studies have been applied, albeit to a lesser extent, also to urban rail transit efficiency assessment. As previously defined, labour and capital are the inputs used by transit operators to produce the outputs. Labour input is used to measure the human effort in the production process (Jain, Cullinane, & Cullinane, 2008), while capital expresses a stock from which a flow of services is derived (Oum et al., 1999). (Jain et al., 2008) used the number of staff as labour and the number of train cars as capital inputs, while the total length of tracks run by trains expresses the route length. A similar choice was made by (Lobo & Couto, 2016) and (Xue & Zhao, 2021). The former used network length, number of stations and number of train cars to characterize firms' capital and the number of employees to represent the labour force. The latter used similar inputs and added the total traction energy consumptions. Urban rail transit uses several production factors for the provision of the services and, as defined in (Graham, Couto, Adeney, & Glaister, 2003), four basic factors for the production of urban rail firms are labour, fleet of vehicles, number of stations and fixed infrastructure. (Graham, 2008) used the number of employees and the route length plus fleet capacity as another input variable which represents the standing and seat capacity of the fleet. Even when the analyses are concerned with assessing the cost efficiency of urban rail firms, the inputs reflect the factors listed above. (Tsai et al., 2015) determined the input prices by dividing the corresponding costs by the quantities of inputs: labour price is the staff cost per employee; rolling stock price, which is a proxy of capital price, is the ratio between non-staff costs and the total number of cars. While the production factors of metro systems are the same as those of railway firms, outputs are not the same. Indeed, as stated in (Jain et al., 2008), when analysing the performance of the former, only outputs related to the provision of passenger service can be included, such as passengers' trips per year or car-kilometres per year. (Novaes,

2001) used the total passengers carried per year as an output variable, while (Xue & Zhao, 2021) considered the passenger turnover volume and the ticket revenues. Differently, (Graham, 2008) decided to use the car-kilometres per year to characterize the output of the system as this data reflects that which a urban rail firm is able of producing, while measures such as passengers and passenger-kilometres are demand-related. This is the reason that both output measures are usually included in the analysis: the first one (car- kilometres) to measure systems' efficiency, while the second one (passengers) to assess systems' effectiveness. This is the case of (Lobo & Couto, 2016) who performed two different measurements: they measured the number of car-kilometres produced per year for estimating efficiency and measured the number of passenger carried per year for evaluating effectiveness. Also (Jain et al., 2008) and (Tsai et al., 2015) considered passenger trips and car-kilometres as output variables in their analyses. (Wey, Kang, & Khan, 2020) proposed a two-stage framework for the performance assessment of Taipei metro transit in which they decomposed the overall rail transit efficiency into production efficiency and service effectiveness, considering two different high-capacity and medium-capacity transit systems that perform services using shared inputs (i.e., labour force) and dedicated inputs (i.e., the specific route length and the dedicated fleet of each system). In particular, operating route length, fleet vehicles and shared factors (such as staff) are inputs and vehicle-kilometres are the output for the production process. Vehicle-kilometres are also an input for the service process, together with the number of stations and other shared factors. Finally, the outputs of the service process are passengers and passenger-kilometres. When instead dealing with costs, (Brage-Ardao et al., 2015) used the train service costs as the dependent variable, rationalized by the car-kilometres, passenger journeys and train hour outputs.

In the evaluation of the efficiency of both railway and urban rail transit firms, it is important to stress that scores are influenced by variations in the market, operating, institutional and regulatory policy environments (Oum & Yu, 1994), which cannot be fully controlled or managed by firms. Among these factors, some are totally uncontrollable, while others can be controlled by firms, but only in the long term (Gathon & Pestieau, 1995). For these reasons (Oum & Yu, 1994) applied a regression to a railway gross efficiency index previously evaluated to identify a residual index as a closer indicator of managerial and technical efficiency. The gross efficiency index represents the combined outcome of true managerial and operational efficiency and the effects of constraints imposed by external conditions. The variables used in the regression capture the effects of government policy and other uncontrollable variables on the gross efficiency indices, and the results showed that to properly compare and make inferences about management and operation efficiency measures among firms, it is necessary to control for these differences in operating and market environment. Focusing on urban rail transit, (Tsai et al., 2015) performed Tobit regressions to assess the effects of external factors on the urban rail cost efficiency, while (Lobo & Couto, 2016) considered, in a second stage of their stochastic frontier analysis, the socio-economic factors of cities where metro systems operate to assess their effects on the systems' performance. The effectiveness scores obtained in this second stage were compared with the ones previously obtained (without external variables), and conclusions were drawn on whether metro systems operate in a favourable or unfavourable surrounding environment. Finally, (Wey et al., 2020) introduced environmental and transport policy factors to evaluate their impacts on the efficiency and effectiveness of the Taipei metro system and its associated subsystems. A review of the variables used to characterize railway firms' inputs and outputs as well as the characteristics of the network, service, policy, and external environment can be found in (Canavan, 2015) and (Catalano et al., 2019).

Since the degree of technical efficiency can only be measured in relation to "best practices", an efficiency frontier must be constructed (Brons et al., 2005) that can be either a production or a cost frontier. The production frontier represents the maximum output attainable from each input level, hence it represents the current state of the technology in the industry (Coelli et al., 2005). In other words, the production frontier defines the maximum output that can be produced from a specified set of inputs, given the existing technology available to the firms involved (Batiese, 1992). Existing approaches to define production frontiers can be distinguished as parametric versus non-parametric and deterministic versus stochastic. The most widely used methods are the deterministic non-parametric Data Envelopment Analysis (DEA) and the Stochastic Frontier Approach (SFA), which are both able to address the complexity of measuring railway efficiency (Makovsek, Benezech, & Perkins, 2015). A comparative study can be found in (Cullinane, Wang, Song, & Ji, 2006), where DEA and SFA were applied to the same dataset to assess the technical efficiency and the scale properties of container ports. The obtained efficiency scores and the relative methods' strengths and weaknesses were compared and analysed. The main advantages of the DEA technique are that it does not require strong a priori assumptions regarding production technology (Couto, 2004) and it can be used with much smaller data samples compared to SFA (Tsai et al., 2015). The drawbacks instead are that DEA results are greatly dependent on the observed best practices in the sample (Couto, 2004) and the inclusion of additional firms may change efficiency scores (Coelli et al., 2005). Moreover, measurement errors and other noise may influence the shape and position of the frontier, and outliers may influence the results (Coelli et al., 2005; Holvad, 2020). Finally, since this technique does not involve the estimation of a statistical model, it cannot be used to test whether the efficiency index for a specific observation is statistically significant or not (Couto, 2004). The main advantages of the stochastic frontier methodology are that it accounts for noise and can be used to conduct conventional test of hypothesis; however, it requires to specify the functional form of the inefficiency component and of the production function (Coelli et al., 2005). Additionally, a large dataset with panel data is needed to estimate robust results (Tsai et al., 2015). In this way, our study follows in the footsteps of other studies using SFA for cross-country production efficiency benchmarking in the transport sector, such as (Gathon, 1989) for urban buses, (Cullinane et al., 2002) for ports, (Pels et al., 2003) for airports, (Coto-Millán, Inglada, Legidos, & Rodríguez-Álvarez, 2004) for airlines, and (Lobo & Couto, 2016) for metros.

2.1. Data collection and pre-processing

To evaluate the efficiency of the European metro systems, the first step was to collect data on the operation of the metros and the socio-economic factors of the surrounding urban context. The collected data allowed us to build up a database that includes 25 cities and 328 panel data observations, covering the period from 2000 to 2020. The cities and operators whose metro systems are included in the database are: Barcelona (TMB), Berlin (BVG), Bilbao (Metro Bilbao), Brescia (MetroBS), Brussels (STIB), Bucharest (Metrorex S.A.), Budapest (BKV), Glasgow (SPT), Hamburg (Hamburger Hochbahn), Helsinki (HKL), Lausanne (TL), Lisbon (Metropolitano de Lisboa), London (TfL), Madrid (Metro de Madrid), Milan (ATM), Munich (MVG), Oslo (Sporveien Oslo AS), Paris (RATP), Porto (Metro do Porto), Prague (DPP), Rome (ATAC), Toulouse (Tisséo), Turin (GTT), Warsaw (Metro Warszawskie) and Wien (Wiener Linien GmbH & Co KG). These metro systems are the ones for which it was possible to find publicly available data in official sources, such as annual reports of transit operators or authorities.

The collected variables characterizing metro systems are the network length (*NL*), the number of stations (*NS*), the number of train sets (*NT*), the number of coaches (*NC*) and the number of employees (*NE*), which were collected from official reports released by public transport operators or transport authorities. In case of few missing values, a linear regression was performed among the available data to fill with the necessary information.

The number of employees of the metro systems required further calculations, especially in the case of public transport firms operating several transport systems. In this case, when data on workers directly employed in metro's operations were not available in the report, the following approximation was made: for each firm, the total fleet was calculated as the sum of the number of trains, buses, trams and any other public transport vehicle operated by the firm. The percentage of metro trains in the fleet was then calculated and it was used to derive the workforce directly involved in the operations of the metro system by multiplying the total number of employees by this percentage.

Two additional variables dealing with offer and demand were collected: car-kilometres (*CARKM*) and passengers carried per year (*PASS*). The published data did not always report car-kilometres: in about half of the cases, the available information referred to train-kilometres. In these situations, the procedure applied requires calculating the average number of coaches per train for each year. Then, the value found is multiplied by the train-kilometres value of the same year to obtain a proxy of the car-kilometres. The summary statistics of metro systems included in the database are presented in Tables 1 and 2.

The second category of collected data concerns the socioeconomic attributes of the cities where the metro systems operate: metropolitan area total land area (*AREA*), total population of the metropolitan area (*POP*), age dependency ratio (*DEPDEM*), average household size (*AHS*), unemployment rate (*UR*), population density (*PD*), number of cars registered per 1000 inhabitants (*CREG*), gross domestic product per capita (*GDP*) and diesel pump price (*GAS*).

The average household size and the number of cars registered per 1000 inhabitants were collected from the Urban Audit Database (Eurostat, 2021). The metropolitan land area, total population, unemployment rate, population density, and GDP per capita were collected from the Regions and Cities Statistical Atlas (OECD, 2021a) of the Organisation for Economic Co-operation and Development (OECD). The age dependency ratio was calculated as the ratio between the population not included in the labour force (aged 0 to 14 years and over 65 years) and the population included in labour force (aged between 15 and 64) available in the same database. The values of the GDP per capita are expressed in 2015 constant prices (USD) and constant purchasing power parities. Since the OECD database only includes data up to the year 2018, for systems with observations related to 2019 and 2020, a regression was performed to find these values as well as to complete the missing values.

The diesel pump prices are national average values obtained by averaging all price observations related to the year under consideration reported in the Oil Bulletin Price History database (European Commission, 2021). These values are then adjusted to the reference year 2015 by means of the GDP deflator values of the different countries (World Bank, 2021). In the World Bank database, Poland and Romania's deflators were referred to 2010 and 2005 base years, respectively. For this reason, Poland's results were additionally adjusted to the reference year 2015 by dividing them by the respective deflator published in the GeoBook dataset (OECD, 2021b). On the other hand, for Romania's values the prices were adjusted using the European Union (EU) average, due to data unavailability.

Finally, as a general procedure, any data not available in the OECD Statistical Atlas was sought in the Eurostat Urban Audit database or in

Table 2
Mean values of CARKM and PASS.

System	Number	CARKM		PASS	
		(thousands)	St. Dev	(thousands)	St. Dev
Barcelona	1	77,467.4	12,392.7	362,120.0	32,070.3
Berlin	2	125,616.7	6138.5	516,586.7	42,445.7
Bilbao	3	18,008.3	3466.2	82,618.0	9412.0
Brescia	4	5051.7	661.3	15,924.7	2203.3
Brussels	5	23,997.4	7062.7	135,471.4	11,709.9
Bucharest	6	36,320.9	8162.7	175,452.3	4190.8
Budapest	7	29,766.6	582.6	294,298.8	22,018.9
Glasgow	8	3525.0	138.9	12,462.5	776.3
Hamburg	9	80,752.6	7884.3	206,960.6	29,629.9
Helsinki	10	–	–	62,166.7	10,995.8
Lausanne	11	3576.4	316.1	43,862.0	2518.2
Lisbon	12	22,940.2	2779.1	169,918.7	15,397.9
London	13	530,338.6	40,437.2	1,194,898.1	158,219.1
Madrid	14	169,293.7	26,727.6	613,696.5	48,992.7
Milan	15	56,848.6	5200.2	324,525.8	21,764.4
Munich	16	62,428.8	3824.8	350,444.4	34,975.4
Oslo	17	42,460.0	4505.3	112,000.0	11,291.6
Paris	18	237,931.7	17,414.4	1,431,457.9	110,281.3
Porto	19	17,991.9	5782.6	46,301.2	18,282.2
Prague	20	51,258.4	6878.1	500,266.6	61,405.8
Rome	21	37,878.0	6417.2	269,748.5	53,698.5
Toulouse	22	–	–	111,975.0	4273.5
Turin	23	9570.8	1857.3	36,727.1	11,302.8
Warsaw	24	36,034.5	1097.9	185,416.0	8984.9
Wien	25	82,783.3	3150.5	445,266.7	13,302.0

Table 1
Mean values of the variables characterizing metro systems.

System	Number	Years	NL	NS	NT	NC	NE
Barcelona	1	2000–2019	97	126	148	734	3323
Berlin	2	2002–2020	147	172	259	1294	5267
Bilbao	3	2001–2019	40	37	42	179	674
Brescia	4	2013–2018	13	17	18	54	131
Brussels	5	2006–2019	40	69	75	320	1792
Bucharest	6	2008–2019	69	51	71	308	4192
Budapest	7	2000–2011	31	40	78	389	1671
Glasgow	8	2012–2019	11	15	14	41	277
Hamburg	9	2003–2019	102	90	233	800	2905
Helsinki	10	2001–2020	25	18	32	129	277
Lausanne	11	2015–2019	14	29	38	76	386
Lisbon	12	2000–2019	38	49	110	335	1605
London	13	2002–2006, 2010–2019	405	271	607	4174	17,848
Madrid	14	2000–2019	260	272	411	2037	6539
Milan	15	2000–2005, 2012–2019	83	96	169	861	2904
Munich	16	2003–2011, 2019	91	94	99	595	2470
Oslo	17	2015–2019	85	101	115	345	614
Paris	18	2000–2011,2013–2019	218	300	701	3573	12,181
Porto	19	2003–2018	56	68	79	238	387
Prague	20	2002–2011, 2013–2019	59	56	144	718	2762
Rome	21	2001–2011, 2013–2019	43	55	85	512	2508
Toulouse	22	2016–2019	27	38	116	232	840
Turin	23	2006, 2008–2019	11	19	28	56	176
Warsaw	24	2015–2019	29	28	75	450	2441
Wien	25	2013–2016, 2018–2019	80	106	135	882	4127

the publications of national statistic offices.

Moreover, the presence of other metro-like and tram systems in the same metropolitan area is depicted by the dummy variable OUR which is set equal to 1 when at least one of these systems exists.

3. Model description

3.1. Gross effectiveness

In the first modelling stage, a gross effectiveness measurement to account for factors that are beyond the short-medium term management control was carried out using a stochastic frontier regression based on the Cobb-Douglas production function. The stochastic frontier methodology approach, introduced by (Aigner, Lovell, & Schmidt, 1977) and (Meeusen & Van Den Broeck, 1977), allows us to distinguish between efficient and inefficient production and to estimate the degree of inefficiency (Brons et al., 2005). Differently from the deterministic ones, stochastic production frontiers state that some firms fail to achieve their production frontier, and the inefficiencies cannot be fully explained by measurable variables (Oum et al., 1999). For this study, the stochastic frontier approach has been chosen because it allows us both to estimate technical efficiency and to test hypothesis on inputs' coefficients.

Following the principles highlighted in (Coelli et al., 2005), the functional form of the stochastic production frontier is given by:

$$ln\ y = \beta X + v - u \tag{1}$$

where:

- y is the output produced;
- X is the vector containing the logarithms of inputs;
- β is the vector of unknown parameters, the inputs' coefficients;
- v is the symmetric random error that accounts for statistical noise (can be positive or negative);
- u is the one-sided distribution error.

As stated in (Coelli et al., 2005), the statistical noise arises from the inadvertent omission of relevant variables from the vector X, as well as from measurements errors and approximation errors associated with the choice of the functional form. The noise term has the same probability of being favourable or not to the production, thus it takes the form of a normal and symmetric distribution, giving the random (i.e., stochastic) nature to the production frontier $\exp(\beta X + v)$ (Lobo & Couto, 2016). As a result, the stochastic frontier output can lie above or below the deterministic frontier defined by $\exp(\beta X)$.

u is a non-negative random variable associated with technical inefficiency. Specific distributions for u explored in the literature are exponential, half-normal, truncated normal and gamma (De Borger et al., 2002). In this study, it is assumed that error u follows an exponential distribution. Finally, the model is estimated using the Maximum Likelihood (ML) method.

Fig. 1 clarifies the concept of the stochastic production frontier for two firms A and B that use a given input level x_i to produce the output y_i ($i = A, B$); the same figure makes it possible to specify the concept of technical efficiency (*TE*) (Eq. (2)), which is defined as the ratio between the observed output for each firm i and the relative stochastic frontier output. It measures the actual output produced by firm i with a given input level, compared with the output that a fully efficient firm could produce with the same input vector.

$$TE_i = \frac{y_i}{exp(\beta X_i + v_i)} = \frac{exp(\beta X_i + v_i - u_i)}{exp(\beta X_i + v_i)} = exp(-u_i) \tag{2}$$

So far, we have discussed about technical efficiency, although this concept assumes different meanings depending on whether the output variable expresses supply or demand. Since in this paper, as it will be described below, the number of passengers carried per year, i.e., a demand variable, is used as an output, it is appropriate to refer to effectiveness rather than efficiency.

In this study, the Cobb-Douglas production function is the functional form used to express the relationship between inputs and outputs, as required by stochastic frontier methodologies.

This function is expressed, for n inputs, as:

$$y = \beta_0 \cdot \prod_{n=1}^{N} x_n^{\beta_n} \tag{3}$$

The econometric software NLOGIT 5 used for the estimations

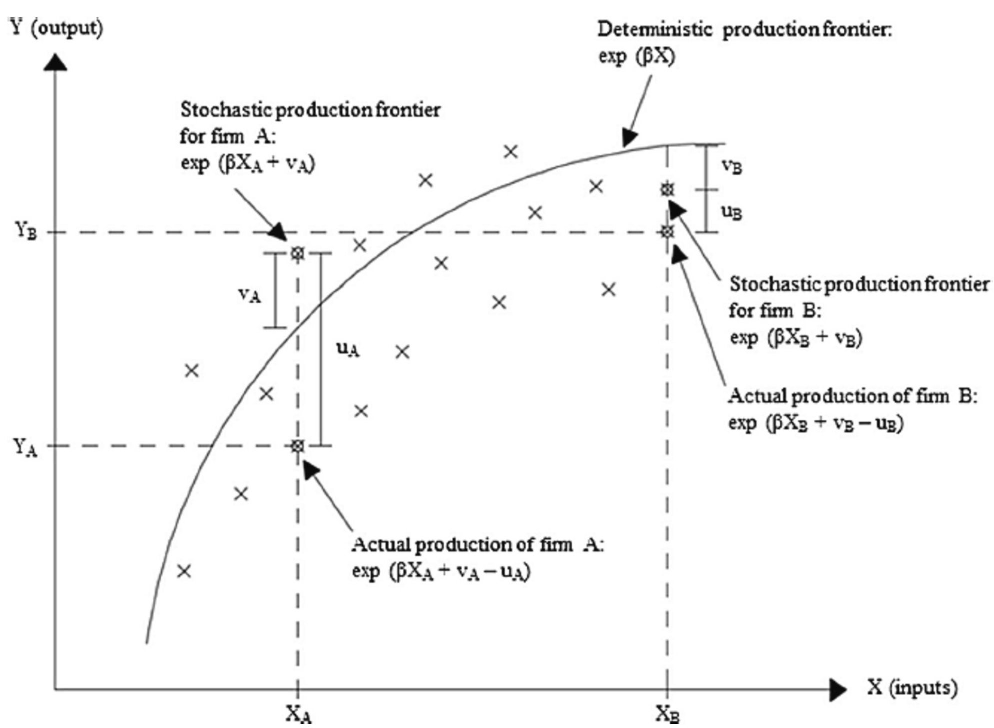

Fig. 1. Stochastic production frontier from (Lobo & Couto, 2016) based on the work of (Coelli et al., 2005).

requires a linearized form of the Cobb-Douglas equation (Eq. (4)) and returns, for each observation, the expected value $E(u_i|\varepsilon_i)$.

$$lny = ln\beta_0 + \sum_{n=1}^{N} \beta_n \cdot lnx_n \tag{4}$$

$E(u_i|\varepsilon_i)$ corresponds to the mean of the conditional distribution f $(u|\varepsilon)$, estimated following the approach described by (Jondrow, Lovell, Materov, & Schmidt, 1982). It corresponds to the punctual estimate of the error u for each observation, allowing us to apply Eq. (2) to obtain the gross effectiveness (GE) of each firm i (Eq. (5)).

$$GE_i = e^{-E(u_i|\varepsilon_i)} \tag{5}$$

3.2. Net effectiveness

To understand how the day-to-day management affects the production, in this stage the net effectiveness is derived by considering the factors associated to long-term (strategic) management decisions and the socio-economic context. In this sense, an exponential multiple regression is performed starting from the gross effectiveness results of the first stage, which may be expressed as:

$$ln\ GE_i = ln\beta_0^{'} + \sum_{j=1}^{J} \beta_j \cdot lnx_{j,i} + \sum_{k=1}^{K} \beta_k \cdot lnx_{k,i} + \epsilon_i^{'} \tag{6}$$

where GE_i are the gross effectiveness values obtained in the previous stage, $x_{j,i}$ are factors related to long-term decisions about the network (network length, number of stations and coaches), $x_{k,i}$ are factors related to the socio-economic context of the cities where metro systems operate, β are coefficients that have to be estimated and $\epsilon_i^{'}$ is the random error term of the regression.

The elasticities β_j obtained using the exponential multiple regression are used for evaluating the net effectiveness, by removing positive or negative contributions to gross effectiveness due to the identified "long term" factors. The effectiveness attributed to the long-term factors is removed from the gross effectiveness, such as the net effectiveness (*NE*) is given by:

$$NE = exp\left(\ln GE_i - \sum_{j=1}^{J} \beta_j \cdot lnx_{j,i}\right) \tag{7}$$

The socio-economic context variables can be considered as inherent to the definition of effectiveness, and for this reason their effects are not removed. Otherwise, the obtained measure would be representative of a net efficiency, i.e., it would stem exclusively from inward company effects.

4. Model application and results

4.1. Gross effectiveness

In a first stage, a gross value of effectiveness is calculated by regressing the output passengers carried per year (*PASS*) against proxies that represent the yearly consumption of the inputs to produce that output.

The Cobb-Douglas function's input variables are the number of employees, the network wear, the fleet wear and tear and a time-trend variable. The number of employees (*NE*) is introduced as a proxy of the labour costs that each firm has to face to produce the service. The network wear (*NETW*) is introduced as the ratio between the car-kilometres produced in each year and the total length of the network. It represents the total number of coaches that are crossing a 1-km section of the network in one year. This ratio is a proxy for the consumptions related to the wear and tear of the infrastructure: the higher this ratio, the higher is the wear and tear of the network. The fleet wear and tear (*FLEETW*) is defined as the ratio between the car-kilometres produced in each year and number of coaches in the fleet. This ratio is a proxy of consumptions related to the rolling stock and it represents the total number of kilometres travelled every year by each coach in the fleet. In this sense, the *NETW* and *FLEETW* variables are proxies of consumptions by representing the wear and tear of the infrastructure and fleet, and therefore the maintenance, depreciation and energy expenses. The time trend variable (*YR*) aims to capture potential gains of expertise and know-how, as well as the technological progress throughout the years.

As the reports on Helsinki and Toulouse metro systems do not include the car-kilometres production, the respective observations were discarded from the database. The same was done for all observations with missing values for the *CARKM* and *PASS* variables. As a result, the final database consists of 264 observations.

Based on Eq. (4), the econometric software NLOGIT 5 estimates the optimal frontier and returns the input elasticities β_n and $E(u_i|\varepsilon_i)$ for each observation. The estimated input elasticities are presented in Table 4.

Eq. (5) makes it possible to obtain the effectiveness for each observation starting from the value of $E(u_i|\varepsilon_i)$. Fig. 2 shows the average gross effectiveness' values for each firm, ordered from the most to the less effective. About half of the firms are characterized by a gross effectiveness score that is higher than 80%, with Oslo, Munich and Prague achieving values higher than 90%. Instead, only the Warsaw and Glasgow systems reach gross effectiveness values lower than 50%.

4.2. Net effectiveness

As previously stated, the effectiveness obtained through the previous method could be influenced and constrained by context effects, as well as by long-term strategic decisions that metro operators are not able to change, at least in the short term, or that are, in many cases, political decisions. Following a similar approach to the one defined by (Oum & Yu, 2004), a regression analysis was performed to evaluate the effects of each input on the gross effectiveness and compute the net effectiveness, i.e., the effectiveness that is related with short- to medium-term management decisions, after removing the long-term factors. The considered factors depending on long-term decisions are related to the fleet size (*NC*) and to the network size, namely the network length (*NL*) and the number of stations (*NS*). Due to the significant differences in size of the metro systems considered in this analysis, the *NS* and *NC* variables were introduced in the model divided by the network length *NL* (variables *ns* and *nc*). The context effects are represented by the socio-economic factors of the cities where systems operate (see Table 3), which may impact the effectiveness by providing a favourable or unfavourable operational context.

The exponential multiple regression is defined as in Eq. (6), where in its linearized form, the natural logarithm is applied to all input variables except for the dummy variable *OUR*. Table 5 shows the β coefficients of the un-discarded (significant) variables included in the model: *NL*, *ns*, *nc*, *OUR* and *PD*. Since the *NS* and *NC* variables were divided by *NL*, the elasticity of the network length β_{NL} is given by $\beta_{NL} - (\beta_{ns} + \beta_{nc})$.

The R-Squared value obtained is 0.285, and following the principles highlighted in (Oum & Yu, 2004), it is possible to conclude that only 28.5% of the total variation in the gross effectiveness index can be explained by the variables included in the regression. This, in turns, implies that the remaining 71,5% of the variation can be assigned to differences in technical efficiency and to additional factors that are not included in this analysis.

Eq. (7) makes it possible to calculate the net effectiveness values, whose average results are reported in Fig. 3. The results show that, taking apart the long-term constraints, the most effective systems are Prague, Budapest and Bilbao and, in general, smaller networks have higher net effectiveness scores than larger ones.

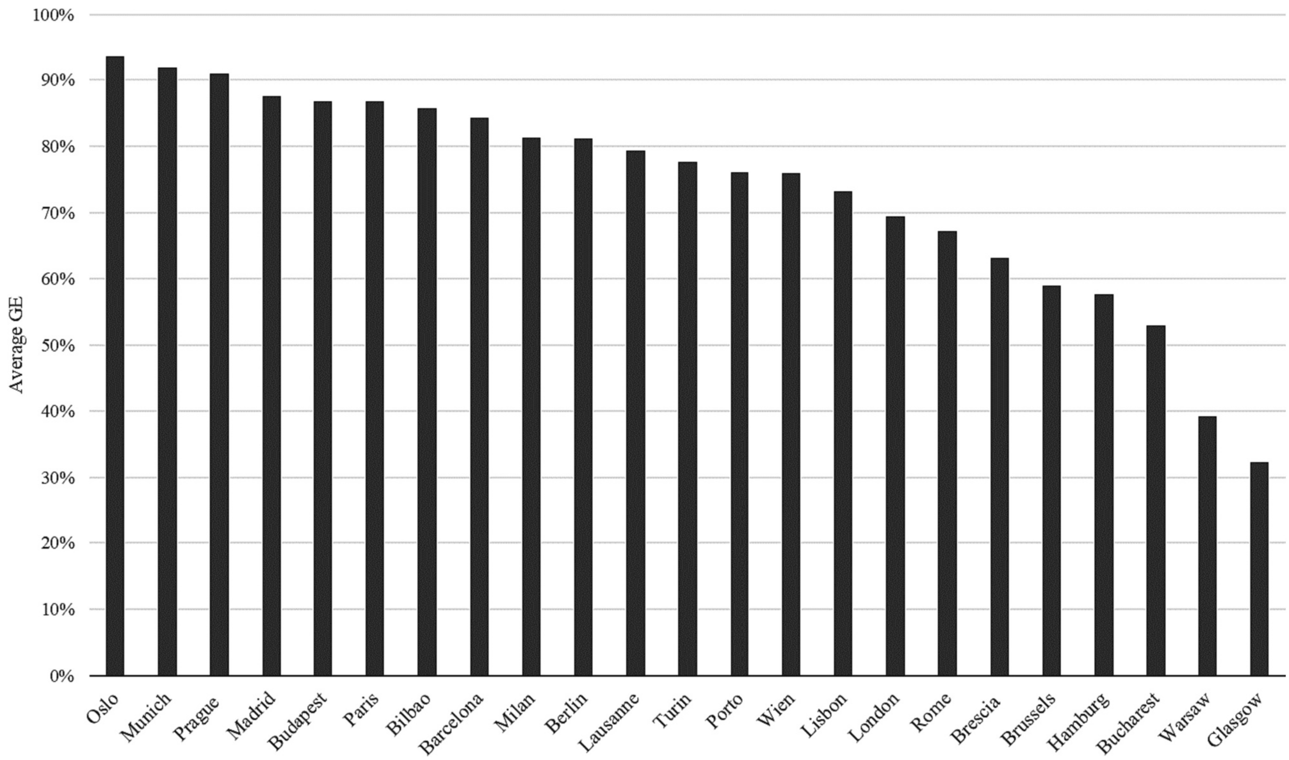

Fig. 2. Average gross effectiveness scores.

Table 3
Socioeconomic factors of the cities (mean values).

System	Number	AREA [km^2]	POP	DEPDEM	AHS	UR	PD [inh. /km^2]	CREG	GDP (USD)	DPP (EUR)	OUR
Barcelona	1	2626	4,747,969	0.47	3	0.13	1808	418	41,039.61	1.069	1
Berlin	2	17,453	5,051,211	0.46	2	0.11	289	359	39,680.74	1.244	1
Bilbao	3	1345	1,004,164	0.49	3	0.12	747	425	42,115.74	1.145	1
Brescia	4	603	478,193	0.57	2	0.08	793	613	35,898.41	1.472	0
Brussels	5	4818	2,532,996	0.52	2	0.09	526	529	72,000.07	1.276	1
Bucharest	6	1754	2,278,718	0.39	3	0.05	1299	461	49,644.29	1.038	1
Budapest	7	6395	2,855,789	0.44	2	0.04	447	353	37,271.75	1.397	1
Glasgow	8	3365	1,810,745	0.49	2	0.07	538	374	34,991.88	1.544	0
Hamburg	9	7192	3,188,037	0.49	2	0.06	443	422	57,503.14	1.238	0
Helsinki	10	4688	1,368,375	0.46	2	0.07	292	398	57,645.39	1.241	1
Lausanne	11	775	412,921	0.46	2	0.07	539	364	69,417.67	1.623	0
Lisbon	12	4321	2,893,187	0.52	2	0.11	670	543	39,086.30	1.166	1
London	13	6968	11,538,565	0.47	3	0.06	1656	374	62,948.53	1.538	1
Madrid	14	7883	6,328,381	0.44	3	0.16	803	515	46,806.10	1.150	1
Milan	15	3115	4,755,904	0.52	2	0.06	1527	572	59,044.43	1.348	1
Munich	16	5495	2,620,845	0.46	2	0.05	477	517	75,051.78	1.263	1
Oslo	17	7403	1,362,465	0.48	2	0.04	184	493	66,937.30	1.839	1
Paris	18	17,584	12,390,631	0.49	2	0.10	705	391	63,327.21	1.181	1
Porto	19	953	1,303,434	0.46	3	0.14	1368	559	27,515.75	1.312	1
Prague	20	5757	2,064,350	0.44	2	0.03	359	547	51,588.65	1.167	1
Rome	21	6162	4,093,786	0.51	2	0.09	664	657	52,764.67	1.337	1
Toulouse	22	6216	1,408,199	0.53	2	0.08	227	587	48,312.00	1.302	1
Turin	23	1702	1,745,045	0.57	2	0.09	1025	634	43,141.92	1.447	1
Warsaw	24	8599	3,156,590	0.52	3	0.03	367	672	61,198.00	1.077	1
Wien	25	9617	2,881,827	0.47	2	0.08	300	376	54,433.37	1.207	1

5. Discussion

5.1. Input effects

The results from the stochastic frontier model used to estimate the gross effectiveness (Table 4) show that increasing the number of employees (*NE*) has a positive impact on the demand: this could be explained considering that having a higher number of employees makes it possible to increase the quality level of system service and customer care, thus improving user experience. Also, the *NETW* variable has a positive impact on the demand and a possible interpretation could be that a higher value of *NETW* implies a higher number of trains running on the network and therefore a higher level of service and attractiveness for the users, who can benefit from higher service frequency. Finally, *FLEETW* has a negative effect on the demand: a higher wear and tear of the vehicles could result in a reduction of the level of service due to increased need for extraordinary maintenance, service disruption and lower levels of comfort and user experience. The time-trend variable was

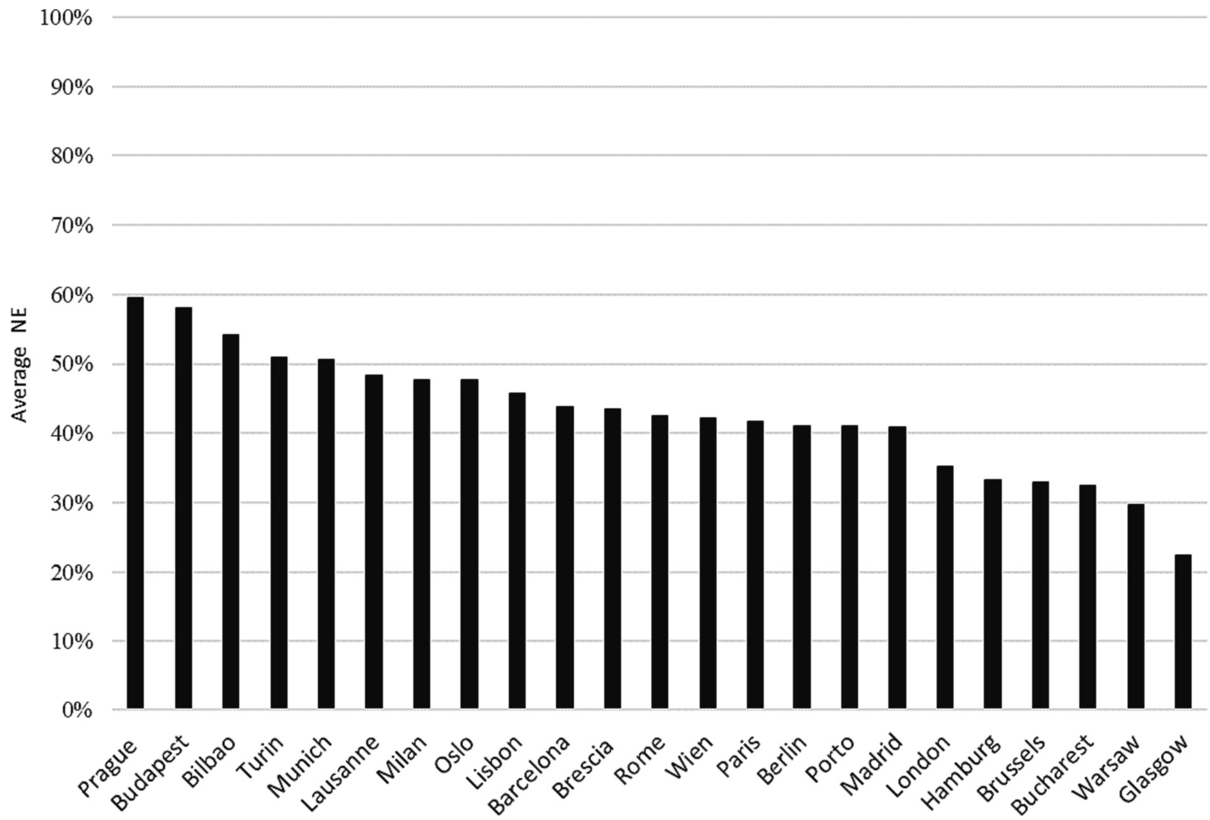

Fig. 3. Average net effectiveness scores.

Table 4
Results of the stochastic frontier regression.

Variable	β	Standard Error	*p*-value
CONSTANT	4.309	0.386	0.000
NE	0.711	0.241	0.000
NETW	0.685	0.680	0.000
FLEETW	−0.361	0.074	0.000
	No. of observations = 264		
	Log-likelihood = −76.886		
	$\sigma_u = 0.324$		
	$\sigma_v = 0.168$		

not statistically significant at the 5% level, thus it was dropped from the model. This may be due to the fact that the urban rail sector has consolidated fairly well in terms of technological evolution over the last 20 years. It is important to highlight that these results have to be interpreted in relation to the number of passengers carried per year, considered as output in the analysis, i.e., in relation to their contribution to the service effectiveness. Different results could be obtained if the analysis was performed under a production efficiency (e.g., considering car-kilometres as output) or cost perspectives.

Table 5
Results of the exponential multiple regression.

Variable	β	St. Error	p-value
$\beta_0{}'$	−0.429	0.223	0.0555
NL	−0.012	0.023	0.0000
ns	0.297	0.071	0.0000
nc	−0.109	0.043	0.0116
OUR	0.240	0.045	0.0000
PD	−0.099	0.032	0.0021
	R-Squared = 0.285		
	No. of observations = 264		

Regarding the net effectiveness model estimations (Table 5), the network length (*NL*) has a slightly negative coefficient, reflecting the fact that increasing the network length while keeping all the other variables (i.e., *NS*, *NC*) constant brings no benefits in terms of effectiveness. This would result in a drop in the frequency and therefore decrease the attractiveness of the service. The coefficient of the number of stations (*ns*) is positive, since networks with a higher number of stations are supposed to be more effective in attracting more users as they increase the accessibility of the network. The number of coaches (*nc*) has a negative coefficient. This could be explained by the fact that increasing the number of coaches in the fleet, and so the number of trains, while keeping the other variables constant could result in a congestion of the network with a consequent decrease in frequency and speed. Moreover, an increase in the number of coaches without interventions in other variables could result in lack of personnel to operate them. Both situations lead to a reduction in the attractiveness of the service and, as a consequence, to a reduction in the number of passengers. Focusing instead on the factors related to the socio-economic environment, the dummy variable *OUR* significantly has a positive coefficient. The reason may be that, as this variable deals with effectiveness and effectiveness is demand-related, the presence of different systems, which are usually complementary to each other and favour multimodality and accessibility to the different parts of a city, increases the attractiveness of the whole public transport network, metro included.

The coefficient of the population density of the metropolitan area (*PD*) is slightly negative. An increase of the population in the area may lead to an increase in demand which could exceed the capacity of the service, thus leading to overcrowded vehicles and congestion. When the *AREA* and *POP* variables are introduced instead of *PD*, they are both found not to be statistically significant. Metro systems are high-frequency transport systems tailored for high density urban areas, and for this reason, what is relevant for effectiveness is how dense its population is and not the absolute size of the urban area.

The results show that the other socio-economic variables introduced are not statistically significant. This could be due to two main reasons. The first is related to database limitations or unobserved external factors that are not related to the collected ones and may still have an important impact on effectiveness. The second is that European metro systems are generically well adjusted to the socio-economic environment where they operate, since the context variables do not seem to have a relevant impact on their effectiveness.

The *DEPDEM*, *AHS*, *UR* and *GDP* variables are not statistically significant because they may be capturing contradictory effects with respect to the usage of metro systems. *DEPDEM* reflects the percentage of the total population of two different age groups that may be characterized by different mobility patterns: the youngers tend to move more, they do not have a driving license and therefore they have to rely on public transport. Elderly groups tend to move less, especially as they get older, and this is also reflected in their usage of public transport. In turn, higher values of *AHS* may imply a reduction in the car availability per person and usually imply a higher percentage of non-drivers (e.g., children) who have to rely on public transport. Lower *AHS* values may lead to an increase in the car availability per person. *UR* may also be capturing contradictory effects between the fact that unemployed people have less possibilities to afford a car on their own but still need to move for their daily routines and seeking a job, and the fact that, by definition, unemployed people do not need to commute. In countries or cities with higher GDP per capita, people may be wealthier and afford to buy and maintain cars, but at the same time public transport is usually more efficient and attractive in these cities. The *CREG* variable includes both people that exclusively use private cars for their trips and people that can actually use it as a complementary transport mode to access the public transport network (e.g., park-and-ride). Finally, *DPP* mostly affects car owners (although it may also affect public transport fares). Those who already own a car may be less sensitive to fuel price changes, which may explain a negligible impact of *DPP* on travel mode choice.

5.2. Effectiveness scores

In this study, the average gross effectiveness values range between 32% and 94%, with all firms having scores higher than 50%, except for Warsaw and Glasgow which reach average values of 39% and 32%, respectively. Oslo, Munich and Prague are the top three systems in terms of gross effectiveness, reaching 94%, 92% and 91%, respectively. These findings are in line with the ones of (Lobo & Couto, 2016), where effectiveness values ranged between 41% and 95% (without considering the Turin metro, which at the time obtained a very low score because data was only available for its opening year). Despite using a non-parametric method, (Tsai et al., 2015) determined values of efficiency ranging between 43% and 88% using a bootstrapped DEA methodology to assess the relative (cost) efficiency of 20 urban rail transit operators in Asia, Australia, Europe and North America between 2009 and 2011. Continuing with non-parametric methods, (Jain et al., 2008) used DEA to assess the technical and scale efficiency of 15 urban rail transit systems and found technical efficiency scores ranging between 35% to 100%. As reported by (De Borger et al., 2002), among studies that performed international comparison, (Gathon, 1989) used a deterministic translog production frontier to benchmark 60 urban transit firms in Europe for the year 1984. The results highlighted technical efficiency scores ranging between 58% to 100%. (Wunsch, 1994, 1996) performed an international comparison of European transit firms between 1988 and 1993, finding technical efficiency scores ranging from 43% to 100% using a Free Disposal Hull method and from 26% to 100% using DEA. Fig. 4 aggregates the average gross and net effectiveness scores estimated in this analysis.

As previously mentioned, the net effectiveness results highlight that transit firms operating smaller networks tend to have higher net effectiveness values compared to larger ones. However, this should be not taken as a proof that these firms are better than the ones operating larger networks, as the net effectiveness does not consider, for example, technical advancement, rolling stock quality and economies of scale, which are mainly affected by large, long-term investments. The net

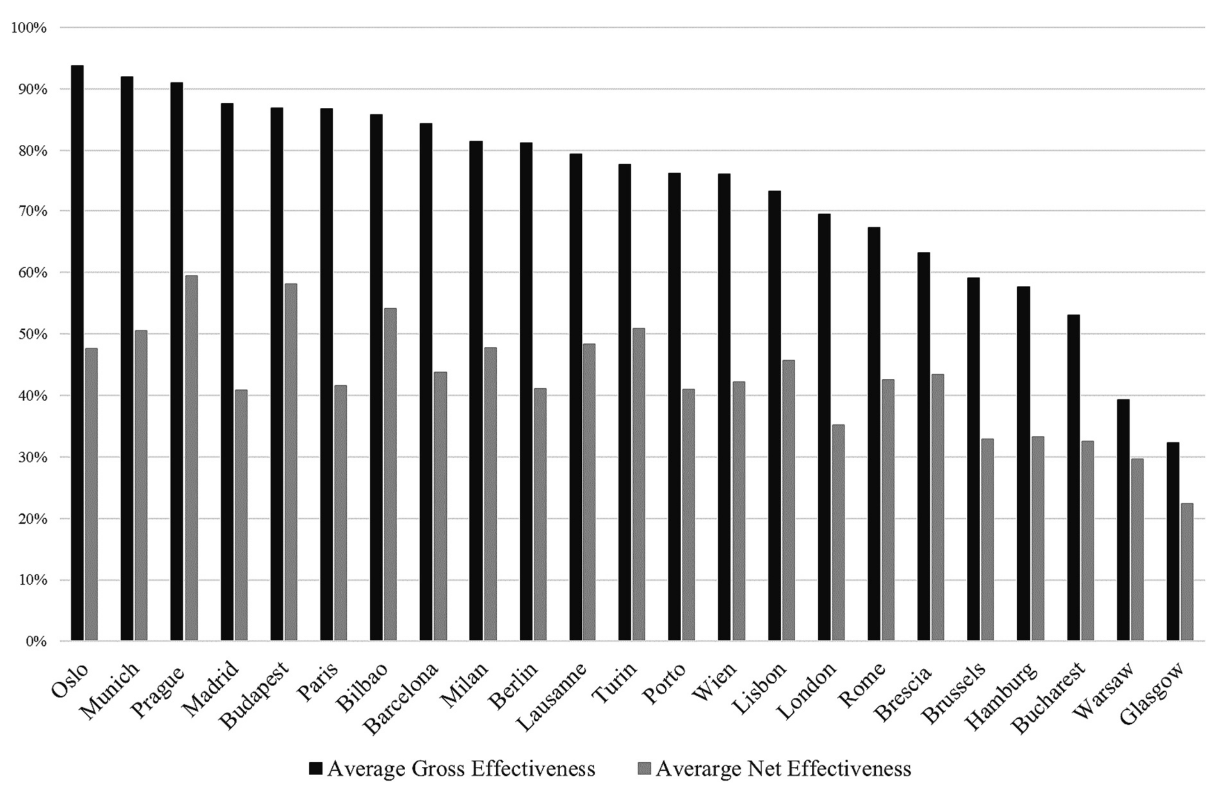

Fig. 4. Average gross and net effectiveness scores.

effectiveness results should be interpreted as a ranking of technical efficiency not considering the long-term decisions that still have an influence in the short-term production outcomes. The good performance of smaller networks in the net effectiveness could be explained by the fact that the day-to-day management of a smaller network is simpler and more flexible than that of a larger network. Adjusting the operating schedule and the service frequency according to the demand could be easier in an automated and single-line system like Turin than in a dense system like Madrid. It should also be added that, for the firms at the top of the net effectiveness ranking, which is dominated by smaller systems like Prague, Budapest and Bilbao (see Fig. 3), the network, fleet and labour consumption in the short-term are better proportioned to their output.

Looking at the gross effectiveness ranking, it is observed that it is more heterogeneous in the sense that it alternates between larger and smaller systems. In fact, regarding the gross effectiveness, there are much more factors to consider, including the favourable or unfavourable socio-economic context, the abovementioned potential advantages of operating a large network, but also the possibility that large networks are less flexible to cope with inefficiencies due to operational, legal or policy issues.

5.3. Closing remarks

The goal of this study was the benchmarking of the technical efficiency of European metro systems in a production perspective. The final database used for the analysis is a panel data from 23 metro systems covering the period from 2000 to 2020. The developed model consists of two stages: in the first one, gross values of effectiveness are estimated for each firm considering the number of carried passengers per year as output, while in the second stage the net effectiveness is calculated. The gross effectiveness analysis allows us to assess the impact of factors beyond firm's control on the short-medium term. Then, the effectiveness attributed to the long-term factors is removed from the gross effectiveness to obtain the net effectiveness. Additionally, the developed models allow us to rank metro systems in terms of gross and net effectiveness.

Gross effectiveness estimates, obtained through a stochastic frontier regression, found that about half of the metro systems reaches scores higher than 80%, with Oslo, Munich and Prague achieving values higher than 90%. These gross effectiveness estimates could be influenced by external factors that are beyond the control of transit operators in the short- and medium-term, but which have an impact on operations. However, the results from the second stage, in which the obtained effectiveness scores were regressed against the long-term and context factors, show that most of those context factors are not significant, except for the population density and the presence of other urban rail system in the same metropolitan area. This could be due to several reasons, such as errors in the database, unobserved external factors that may still have an impact on the effectiveness, or the possibility that European metro systems are well suited to the urban socio-economic context where they operate. Finally, the obtained results highlight that transit firms operating smaller networks tend to have higher net effectiveness values compared to larger ones, with Prague, Budapest and Bilbao reaching the best performances. The better performance of small networks in relation to net effectiveness can be related to an increased flexibility of small firms to cope with inefficiencies through simple day-to-day management decisions compared to larger firms, provided that their network is well suited to the size and context of the cities where they operate.

The main limitation of this study stems from the fact that, under the scope of this work, it was not feasible to collect costs data, and for this reason, it was decided to use a production function instead of a cost function for this benchmarking exercise. Nevertheless, data scarcity on the expenses of metro systems, different reporting criteria and the aggregation of costs with other transport modes operated by the same firms are severe constraints for a benchmarking of this transport sector. By considering proxies for the firms' consumptions derived from inputs' quantities, this work provides a valuable effort to overcome these issues.

Focusing on future research directions, a further investigation is planned about the factors beyond current management control that currently affect the performances of metro operators. This additional investigation aims to assess if there are other external factors currently not included in our analysis that affect the performance of metro systems and try to understand whether and to what extent they influence the system's effectiveness.

Moreover, the same initiative also envisages the further development of this analysis through the collection and integration of data related to revenues and expenses which will allow to integrate an important aspect of firms' operations. The described issues concerning different reporting criteria and levels of detail will have to be addressed, which could lead to the use of a more straightforward methodology.

Author statement

During the preparation of this work the authors have not used any tool or service of generative AI and AI-assisted technologies in the writing process.

CRediT authorship contribution statement

Luigi Castagna: Data curation, Formal analysis, Methodology, Writing – original draft, Conceptualization, Visualization. **António Lobo:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **Pierluigi Coppola:** Conceptualization, Supervision, Validation, Writing – review & editing. **António Couto:** Conceptualization, Supervision, Validation.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgement

The contributions of António Lobo and António Couto are a result of project DynamiCITY: Fostering Dynamic Adaptation of Smart Cities to Cope with Crises and Disruptions, with reference NORTE-01-0145-FEDER-000073, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This research was funded in part by the Fundação para a Ciência e a Tecnologia, I.P. (FCT, Funder ID = 50110000187) under the grant with DOI 10.54499/CEECINST/00010/2021/CP1770/CT0003.

References

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21–37. https://doi.org/10.1016/0304-4076(77)90052-5

Alam, K. M., Xuemei, L., Baig, S., Yadong, L., & Shah, A. A. (2020). Analysis of technical, pure technical and scale efficiencies of Pakistan railways using data envelopment analysis and Tobit regression model. *Networks and Spatial Economics, 20*(4), 989–1014. https://doi.org/10.1007/s11067-020-09510-9

Barnum, D., McNeil, S., & Hart, J. (2007). Comparing the efficiency of public transportation subunits using data envelopment analysis. *Journal of Public Transportation, 10*(2), 1–16. https://doi.org/10.5038/2375-0901.10.2.1

Batiese, G. E. (1992). Frontier production functions and technical efficiency: A survey of empirical applications in agricultural economics. *Agricultural Economics, 7*(3–4), 185–208. https://doi.org/10.1111/j.1574-0862.1992.tb00213.x

Boame, A. K. (2004). The technical efficiency of Canadian urban transit systems. *Transportation Research Part E: Logistics and Transportation Review, 40*(5), 401–416. https://doi.org/10.1016/j.tre.2003.09.002

Bojović, N., Milenković, M., Kapetanović, M., & Knežević, N. (2016). Innovations impact on efficiency of European railway companies. *Management, 21*(79), 13–26.

Brage-Ardao, R., Graham, D. J., & Anderson, R. J. (2015). Determinants of train service costs in metro operations. *Transportation Research Record: Journal of the Transportation Research Board, 2534*(1), 31–37. https://doi.org/10.3141/2534-05

Brons, M., Nijkamp, P., Pels, E., & Rietveld, P. (2005). Efficiency of urban public transit: A meta analysis. *Transportation, 32*(1), 1–21. https://doi.org/10.1007/s11116-004-0939-4

Canavan, S. (2015). *Performance modelling of urban metro rail systems: An application of frontier, regression, and causal inference techniques*. Ph. D thesis. United Kingdom: Department of Civil and Environmental Engineering, Imperial College London.

Cantos, P., & Maudos, J. (2001). Regulation and efficiency: The case of European railways. *Transportation Research Part A: Policy and Practice, 35*(5), 459–472. https://doi.org/10.1016/S0965-8564(00)00007-0

Cantos, P., Pastor, J. M., & Serrano, L. (1999). Productivity, efficiency and technical change in the European railways: A non-parametric approach. *Transportation, 26*(4), 337–357. https://doi.org/10.1023/A:1005127513206

Catalano, G., Daraio, C., Diana, M., Gregori, M., & Matteucci, G. (2019). Efficiency, effectiveness, and impacts assessment in the rail transport sector: A state-of-the-art critical analysis of current research. *International Transactions in Operational Research, 26*(1), 5–40. https://doi.org/10.1111/itor.12551

Chapin, A., & Schmidt, S. (1999). Do mergers improve efficiency? Evidence from deregulated rail freight. *Journal of Transport Economics and Policy, 33*(2), 147–162. http://www.jstor.org/stable/20053802.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). Springer.

Coto-Millán, P., Inglada, V., Legidos, B. R., & Rodríguez-Álvarez, A. (2004). Changes in the world air industry: An analysis of technical efficiency. *International Journal of Transport Economics, 31*(3), 341–354. http://www.jstor.org/stable/42748282.

Couto, A. (2004). *The impact of high speed technology on demand and productivity in European railways: An econometric analysis*. Faculty of Engineering of the University of Porto (Unpublished PhD dissertation).

Couto, A., & Graham, D. J. (2009). The determinants of efficiency and productivity in European railways. *Applied Economics, 41*(22), 2827–2851. https://doi.org/10.1080/00036840801949782

Cowie, J. (1999). The technical efficiency of public and private ownership in the rail industry: The case of Swiss private railways. *Journal of Transport Economics and Policy, 33*(3), 241–251. http://www.jstor.org/stable/20053814.

Cullinane, K., Song, D.-W., & Gray, R. (2002). A stochastic frontier model of the efficiency of major container terminals in Asia: Assessing the influence of administrative and ownership structures. *Transportation Research Part A: Policy and Practice, 36*(8), 743–762. https://doi.org/10.1016/S0965-8564(01)00035-0

Cullinane, K., Wang, T.-F., Song, D.-W., & Ji, P. (2006). The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis. *Transportation Research Part A: Policy and Practice, 40*(4), 354–374. https://doi.org/10.1016/j.tra.2005.07.003

Daraio, C., Diana, M., Di Costa, F., Leporelli, C., Matteucci, G., & Nastasi, A. (2016). Efficiency and effectiveness in the urban public transport sector: A critical review with directions for future research. *European Journal of Operational Research, 248*(1), 1–20. https://doi.org/10.1016/j.ejor.2015.05.059

De Borger, B., Kerstens, K., & Costa, Á. (2002). Public transit performance: What does one learn from frontier studies? *Transport Reviews, 22*(1), 1–38. https://doi.org/10.1080/01441640010020313

De Jorge-Moreno, J., & Garcia-Cebrian, L. I. (1999). Measuring of production efficiency in the European railways. *European Business Review, 99*(5), 332–344. https://doi.org/10.1108/09555349910288219

Eurostat, Urban Audit Database. Available online https://ec.europa.eu/eurostat/web/cities/data/database (accessed on 30/09/2021).

European Commission, Weekly Oil Bulletin. Available online https://ec.europa.eu/energy/data-analysis/weekly-oil-bulletin_en (accessed on 30/09/2021).

Farsi, M., Filippini, M., & Kuenzle, M. (2006). Cost efficiency in regional bus companies: An application of alternative stochastic frontier models. *Journal of Transport Economics and Policy, 40*(1), 95–118.

Fried, H. O., Lovell, C. K., & Schmidt, S. S. (2008). *The measurement of productive efficiency and productivity growth*. Oxford University Press.

Gathon, H.-J. (1989). Indicators of partial productivity and technical efficiency in the European urban transit sector. *Annals of Public and Cooperative Economics, 60*(1), 43–60. https://doi.org/10.1111/j.1467-8292.1989.tb02008.x

Gathon, H.-J., & Pestieau, P. (1995). Decomposing efficiency into its managerial and its regulatory components: The case of European railways. *European Journal of Operational Research, 80*(3), 500–507. https://doi.org/10.1016/0377-2217(94)00133-W

Graham, D. J. (2008). Productivity and efficiency in urban railways: Parametric and non-parametric estimates. *Transportation Research Part E: Logistics and Transportation Review, 44*(1), 84–99. https://doi.org/10.1016/j.tre.2006.04.001

Graham, D. J., Couto, A., Adeney, W. E., & Glaister, S. (2003). Economies of scale and density in urban rail transport: Effects on productivity. *Transportation Research Part E: Logistics and Transportation Review, 39*(6), 443–458. https://doi.org/10.1016/S1366-5545(03)00017-6

Growitsch, C., & Wetzel, H. (2009). Testing for economies of scope in European railways: An efficiency analysis. *Journal of Transport Economics and Policy, 43*(1), 1–24.

Holmgren, J. (2013). The efficiency of public transport operations – An evaluation using stochastic frontier analysis. *Research in Transportation Economics, 39*(1), 50–57. https://doi.org/10.1016/j.retrec.2012.05.023

Holvad, T. (2020). Efficiency analyses for the railway sector: An overview of key issues. *Research in Transportation Economics, 82*, Article 100877. https://doi.org/10.1016/j.retrec.2020.100877

Hung, S.-W., Lu, W.-M., & Wang, T.-P. (2010). Benchmarking the operating efficiency of Asia container ports. *European Journal of Operational Research, 203*(3), 706–713. https://doi.org/10.1016/j.ejor.2009.09.005

Jain, P., Cullinane, S., & Cullinane, K. (2008). The impact of governance development models on urban rail efficiency. *Transportation Research Part A: Policy and Practice, 42*(9), 1238–1250. https://doi.org/10.1016/j.tra.2008.03.012

Jitsuzumi, T., & Nakamura, A. (2010). Causes of inefficiency in Japanese railways: Application of DEA for managers and policymakers. *Socio-Economic Planning Sciences, 44*(3), 161–173. https://doi.org/10.1016/j.seps.2009.12.002

Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2–3), 233–238. https://doi.org/10.1016/0304-4076(82)90004-5

Karlaftis, M. G. (2004). A DEA approach for evaluating the efficiency and effectiveness of urban transit systems. *European Journal of Operational Research, 152*(2), 354–364. https://doi.org/10.1016/S0377-2217(03)00029-8

Kutlar, A., Kabasakal, A., & Sarikaya, M. (2013). Determination of the efficiency of the world railway companies by method of DEA and comparison of their efficiency by Tobit analysis. *Quality & Quantity, 47*(6), 3575–3602. https://doi.org/10.1007/s11135-012-9741-0

Lerida-Navarro, C., Nombela, G., & Tranchez-Martin, J. M. (2019). European railways: Liberalization and productive efficiency. *Transport Policy, 83*, 57–67. https://doi.org/10.1016/j.tranpol.2019.09.002

Link, H. (2016). A two-stage efficiency analysis of rail passenger franchising in Germany. *Journal of Transport Economics and Policy, 50*(1), 76–92.

Link, H. (2019). The impact of including service quality into efficiency analysis: The case of franchising regional rail passenger serves in Germany. *Transportation Research Part A: Policy and Practice, 119*, 284–300. https://doi.org/10.1016/j.tra.2018.11.019

Lobo, A., Amorim, M., Rodrigues, C., & Couto, A. (2018). Modelling the operating speed in segments of two-lane highways from probe vehicle data: A stochastic frontier approach. *Journal of Advanced Transportation, 2018*, 1–10. https://doi.org/10.1155/2018/3540785

Lobo, A., & Couto, A. (2016). Technical efficiency of European metro systems: The effects of operational management and socioeconomic environment. *Networks and Spatial Economics, 16*(3), 723–742. https://doi.org/10.1007/s11067-015-9295-5

Lobo, A., Couto, A., & Rodrigues, C. (2016). Flexible stochastic frontier approach to predict spot speed in two-lane highways. *Journal of Transportation Engineering, 142*(8), 04016032. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000862

Lobo, A., Rodrigues, C., & Couto, A. (2014). Estimating percentile speeds from maximum operating speed frontier. *Transportation Research Record: Journal of the Transportation Research Board, 2404*(1), 1–8. https://doi.org/10.3141/2404-01

Makovsek, D., Benezech, V., & Perkins, S. (2015). *Efficiency in Railway operations and infrastructure management, International transport forum discussion papers. 2015/12*. Paris: OECD Publishing. https://doi.org/10.1787/5jrvzrnmhx7k-en

Mallikarjun, S., Lewis, H. F., & Sexton, T. R. (2014). Operational performance of U.S. public rail transit and implications for public policy. *Socio-Economic Planning Sciences, 48*(1), 74–88. https://doi.org/10.1016/j.seps.2013.08.001

Meeusen, W., & Van Den Broeck, J. (1977). Efficiency estimation from cobb-Douglas production functions with composed error. *International Economic Review, 18*(2), 435–444. https://doi.org/10.2307/2525757

Merkert, R., Mulley, C., & Hakim, M. M. (2017). Determinants of bus rapid transit (BRT) system revenue and effectiveness – A global benchmarking exercise. *Transportation Research Part A: Policy and Practice, 106*, 75–88. https://doi.org/10.1016/j.tra.2017.09.010

Novaes, A. G. (2001). Rapid-transit efficiency analysis with the assurance-region DEA method. *Pesquisa Operacional, 21*(2), 179–197. https://doi.org/10.1590/S0101-74382001000200004

OECD, Regions and Cities Statistical Atlas. Available online: https://regions-cities-atlas.oecd.org/ (accessed on 30/09/2021).

OECD, OECD.Stat. Available online at https://stats.oecd.org/Index.aspx?DataSetCode=DACDEFL, (Accessed on 30/09/2021).

Oum, T. H., Waters, W. G., & Yu, C. (1999). A survey of productivity and efficiency measurement in rail transport. *Journal of Transport Economics and Policy, 33*(1), 9–42. http://www.jstor.org/stable/20053789.

Oum, T. H., & Yu, C. (1994). Economic efficiency of railways and implications for public policy: A comparative study of the OECD Countries' railways. *Journal of Transport Economics and Policy, 28*(2), 121–138. http://www.jstor.org/stable/20053031.

Oum, T. H., & Yu, C. (2004). Measuring airports' operating efficiency: A summary of the 2003 ATRS global airport benchmarking report. *Transportation Research Part E: Logistics and Transportation Review, 40*(6), 515–532. https://doi.org/10.1016/j.tre.2004.08.002

Pels, E., Nijkamp, P., & Rietveld, P. (2003). Inefficiencies and scale economies of European airport operations. *Transportation Research Part E: Logistics and Transportation Review, 39*(5), 341–361. https://doi.org/10.1016/S1366-5545(03)00016-4

Sameni, M. K., Preston, J., & Khadem Sameni, M. (2016). Evaluating efficiency of passenger railway stations: A DEA approach. *Research in Transportation Business & Management, 20*, 33–38. https://doi.org/10.1016/j.rtbm.2016.06.001

Smith, A. S. J. (2012). The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector. *Transportation Research Part E: Logistics and Transportation Review, 48*(2), 503–515. https://doi.org/10.1016/j.tre.2011.10.003

Tsai, C. H. P., Mulley, C., & Merkert, R. (2015). Measuring the cost efficiency of urban rail systems an international comparison using DEA and tobit models. *Journal of Transport Economics and Policy, 49*(1), 17–34.

Von Hirschhausen, C., & Cullmann, A. (2010). A nonparametric efficiency analysis of German public transport companies. *Transportation Research Part E: Logistics and Transportation Review, 46*(3), 436–445. https://doi.org/10.1016/j.tre.2009.11.005

Wetzel, H. (2008). Productivity growth in European railways: Technological progress, efficiency change and scale effects. In *, Vol. No. 101. Working paper series in economics*. Lüneburg: Leuphana Universität Lüneburg, Institut für Volkswirtschaftslehre.

Wey, W.-M., Kang, C.-C., & Khan, H. A. (2020). Evaluating the effects of environmental factors and a transfer fare discount policy on the performance of an urban metro system. *Transport Policy, 97*, 172–185. https://doi.org/10.1016/j.tranpol.2020.05.004

World Bank, GDP Deflators. Available online at https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS.AD (accessed 30/09/2021).

Wunsch, P. (1994). *Costing busses: Back to the basics, Brussels, FUSL (SMASH Cahier 9405)*.

Wunsch, P. (1996). Cost and productivity of major urban transit Systems in Europe: An exploratory analysis. *Journal of Transport Economics and Policy, 30*(2), 171–186. http://www.jstor.org/stable/20053108.

Xue, L., & Zhao, S. (2021). Evaluating and analyzing the operation efficiency of urban rail transit Systems in China Using an integrated approach of DEA model, Malmquist productivity index, and Tobit regression model. *Journal of Transportation Engineering, Part A: Systems, 147*(10), 04021061. https://doi.org/10.1061/JTEPBS.0000561