Metro station operating costs: an econometric analysis

Mohammed Quddus¹, Nigel Harris² and Daniel J Graham³†

¹Transport Studies Group
Department of Civil and Building Engineering
Loughborough University
Leicestershire, LE11 3TU
UK

²Centre for Transport Studies,
Department of Civil and Environmental Engineering
Imperial College London,
London, SW7 2AZ,
UK.

E-mail: m.a.quddus@lboro.ac.uk, nigel.harris@railcons.com, and d.j.graham@imperial.ac.uk

† Corresponding author. Tel +44 (0)20 7594 6088, Fax +44 (0)20 7594 6107.
Metro station operating costs: an econometric analysis

Abstract. This paper develops an econometric analysis of metro station operating costs to identify factors that create variation in cost efficiency. Stations operating costs can be classed amongst the semi-fixed costs that a metro faces in the sense that they do not vary proportionately with metro output. They may therefore be important in determining the degree of returns to density. This paper seeks to provide an improved understanding of some of the major factor driving these costs. Empirical results show that there are strong systems specific influences on costs but over and above these we detect positive associations from a range of station characteristics including the length of passageways, the number of platforms, peak level service frequency, interchange demand and the provision of toilet facilities. In addition, we find that the presence of air conditioning has a substantial effect increasing expected station operating costs by as much as 40%.
**Introduction**

The cost structure of the mainline railway industry has received a great deal of attention in the academic literature (e.g. Caves et al 1980, Caves et al 1981a, 1981b, Freeman et al 1985, Caves et al 1985, Dodgson 1985, McGeehan 1993, Bookbinder & Qu 1993, Oum & Yu 1994, Cowie & Riddington 1996, Wunsch 1996, Tretheway et al 1997, Oum et al 1999, Cantos et al 1999, Cantos et al 2002). Research has demonstrated the very large variance in cost efficiency, or productivity, that is often present within a sample of rail firms and has developed cost and production function approaches to analyse the factors underpinning this variance.

A prominent theme in the rail efficiency literature is whether cost structures are subject to returns to scale (RTS) or returns to density (RTD). RTS describe the relationship between all inputs and the overall scale of operations, including both output and network size. RTD describe the relationship between inputs and outputs with the rail network held fixed. The evidence in the literature indicates that there are RTD due to the prevalence of fixed costs in the rail industry and to a range of semi-fixed costs that do not vary proportionally with output. There is less consistent evidence on the existence of scale economies, though the majority view is that railways operate under constant returns to scale. There are very few studies of the costs structure of urban metros though Graham et al (2003) estimates increasing RTD and constant RTS.

Station operations may provide an important source of increasing RTD in metro operations. Stations must remain staffed and functioning, with all the energy and other resources required, throughout the duration of the metro operating hours. Moreover, costs may differ quite substantially from one station to another due to the nature of engineering, the depth of station, its
size and dimensions, the technology employed, and so on. So we can conceive of station operating costs as semi-fixed costs which do not vary proportionately with system throughput and therefore may be instrumental in giving rise to increasing RTD.

In this paper we develop an econometric model to analyse variance in station operating costs. We use data on 83 stations from 13 metro systems from around the world to estimate the main drivers of cost. The paper is structured as follows. Section II describes the model specification and the data used for estimation. Section III presents our results. Conclusions are drawn in the final section.

**Model specification and data**

The data available for our analysis describe the total operating cost of each station and a range of station characteristics collected from a total of 13 metros. These are Montreal, Buenos Aires, Dublin, Glasgow, Hong Kong MTR, Hong Kong KCR, Lisbon, London, Naples, Sao Paulo, Singapore, Taipei, and Toronto. The analysis we develop below regresses the total operating costs against these station characteristics to determine their role in influencing variance in costs.

It is important to stress that we do not adopt a conventional cost function approach. We do not have data on factor prices and therefore cannot estimate the cost function. However, another important consideration in this respect is that since the operating costs of any one particular station represent only a small fraction of total metro operating costs, individual stations cannot be regarded as the appropriate units over which cost decisions are made. For instance, metro operators do not demand factor inputs at the station level in accordance with prices but make
rational decisions relating to costs and operations for the system as a whole. Furthermore, it would be wrong to ascribe any particular behavioural assumptions to individual stations, for instance, cost minimising behaviour. A metro may not seek to sustain a set level of station efficiency across the system but rather allow for disparities in efficiency to achieve some broader objectives relating to the appropriate level of system output given overall costs.

In this respect, it is mainly how the station characteristics serve to influence total cost that is of interest in the present analysis. One important issue however, relates to the absence of factor price data, because this will certainly be important in determining station costs. To control for these omitted variables, which we cannot observe, we estimate the station operating costs model with a set of dummy variables for the 13 metro systems. We assume that these dummies will capture unobserved system specific effects including factor prices.

A loglinear model is used to identify the factors that influence the operating cost of a metro station. The model can be written as:

\[ \ln y_i = \alpha + \beta \ln X_i + \theta D_i + \epsilon_i \]  

(1)

where \( y_i \) is the total operating cost of a metro station \( i \), \( X_i \) is a \( k \times 1 \) vector of continuous explanatory variables describing the characteristics of station \( i \), \( D_i \) is a \( m \times 1 \) vector of dummy explanatory variables relating to metro systems, and \( \epsilon \) is white noise, \( \beta \) is a \( k \times 1 \) vector of parameters to be estimated, and \( \theta \) is a \( m \times 1 \) vector of parameters to be estimated. The log linear
model is used because it reduces the potential for multicollinearity and provides direct parameter estimates of the elasticities.

The dependent variable is the total cost of operating the station per year. This includes the costs associated with staff, utilities (for instance electricity, gas and water), the maintenance of lifts and escalators, and the maintenance of other systems such as CCTV, air-conditioning, ticketing equipment, and building.

The explanatory variables, which describe the station characteristics, and the hypotheses we seek to test with each variable are as follows.

Age of the station: The age of the station is taken as the number of years since the station opened. This figure is averaged if the station was opened in stages. Our hypothesis is that older stations will incur higher maintenance costs than new stations.

Lifts and escalators: The number of lifts and escalators within a station may influence the operating cost because this equipment needs to be operation on a daily basis and frequently maintained.

Number of ticket machines/ticket offices/ticket sales windows/entry and exit gates: The number of ticket machines includes only those machines used by the public to purchase or validate tickets. The number of ticket offices is the number of areas of the station where the ticket-selling takes place. The number of ticket sales windows relates to the number of potentially-staffed positions
used by metro staff to sell tickets to passengers. We hypothesize that these factors will influence the staff costs of the station.

*Number of opening hours per day:* This variable is taken as the average number of metro station opening hours per day. The hypothesis is, of course, that longer operating hours induce higher costs.

*Service frequency:* Two service frequency variables are considered: (1) peak frequency, (2) off-peak frequency. Frequency is calculated as the average number of trains per hour (each way) during peak periods (peak frequency) or off-peak periods (off-peak frequency). The inclusion of these variables will allow us to test whether costs are associated with frequency.

*Length of trains:* This is calculated as the total number of carriages of a train using the station. At stations with multiple lines averages are used.

*Platform dimensions:* The variables considered are the width, length, and elevation of the platform. This is to see if these are important for maintenance and cleaning costs.

*Roof length of platforms:* For underground stations, this is clearly the same as platform length, but for at-grade and elevated stations only part of platforms may be covered by a canopy, shelter or overall roof. This variable is included to understand if variation in the maintenance associated with roof lengths affects total station costs.
Length of passageways: This is measured as the total length of passageways including escalator shafts estimated by metros as an indicative proxy for the amount of cleaning and building repair which may need to be done. No account is taken of possible variations in passageway width. A better measure might have been the total floor area, but this would not have directly reflected the amount of walls and ceilings which need maintenance and cleaning; this is also discarded as a measure, as being more difficult for metros to estimate easily.

Station demand variables: The two main demand variables considered are: (1) entry demand, (2) interchange demand. Station entry demand is the total number of passengers entering the station per year. This includes passengers changing modes at the station, and entering from mainline rail or bus stations, as well as those starting their journeys locally and entering the station on foot. Interchange demand relates only to those passengers changing metro lines at the station concerned. Two secondary variables are also considered: (1) peak entry demand, and (2) peak interchange demand. Peak entry (interchange) demand is calculated as the total number of entry (interchange) passengers for the busiest hour during a standard week, and is designed to test whether peak demand (entry/interchange) drives station capacity and hence costs, or total demand drives staffing levels and hence cost.

Types of metro stations: Dummy variables are used to reflect the overall type of metros in terms of being at-grade, elevated, sub-surface (typically constructed by cut & cover, and 5-6m below ground) or deep tube. The idea for this categorisation is that at-grade and sub-surface stations can be managed without lifts or escalators for passengers to travel vertically, whereas elevated and deep tube stations normally need this equipment, which adds significantly to (e.g. electricity) costs.
Other variables: The presence of air conditioning, toilets for public use, platform screen doors and shops are all included in modelling through a dummy variable. Each is thought to generate costs (respectively electricity, cleaning, maintenance and management time).

Results

Prior to model fitting, a number of statistical tests were performed to determine the nature of the data. For example, it is possible that the explanatory variables may be correlated with each other (the effect of multicollinearity) or that the data exhibits heteroskedasticity (the effect of non-constant variance).

Although imperfect multicollinearity does not violate the assumptions of the classical model, if its presence is sufficiently acute it can lead to biased inefficient and even wrongly signed estimates. If the overall goodness-of-fit, $R^2$, is relatively high (say more than 0.8) but only few explanatory variables are significantly different from zero or there are high pair-wise correlations among the regressors, then it is possible that multicollinearity may be present. Here, we use the variance inflation factor (VIF) proposed by Chatterjee et al. (2000) to determine the presence of multicollinearity. The number of ticket gates of a station, for example, is found to be highly correlated with the entry demand of the station, and the length of the platform of a station is correlated with the length of the longest train passing the station. Based on the VIF test, the highly correlated variables are excluded from the explanatory variables used in the final model.
Data from the London metro stations are not included in the model as operating costs are not obtainable at the station level for the categories which are consistent with the other metros. This reduces the total number of observations to 83. However, we still have to estimate more than 30 parameters which are found to be uncorrelated with each other. Some of the explanatory variables such as entry and interchange demand, lifts and escalators are then combined to minimize the number of parameters to be estimated. A dummy variable is used to represent the presence of lifts or escalators within a station in the model. This variable takes on a value of 1 if there are any lifts or escalators in a metro station and a value of 0 otherwise. Summary statistics (observations, mean, standard deviation, minimum, and maximum) of the final explanatory variables used in the model are shown in Table 1.

**Table 1: Summary statistics of explanatory variables used in the model**

Another important assumption of the classical linear regression model is that the disturbances appearing in the regression function are homoskedastic. The problem of heteroskedasticity is common in cross-sectional analysis because the data usually involves observations from heterogeneous units (i.e. stations from different metros), and therefore heteroskedasticity may be expected if data from small, medium and large stations are sampled together. Conducting the Park Test (Park, 1966) we find that our data are not characterised by heteroskedasticity. This may be due to the use of the log linear model which reduces the variances among the variables.

**Table 2: Model estimation results for the operating cost of a metro station**
Table 2 presents our results. Two models are considered: (1) without metro dummies, (2) with metro dummies. The second model includes the metro specific effects to control for heterogeneous environments. Ramsey’s RESET test (an $F$-test) is used to select the better model (Ramsey, 1969) and this shows that the addition of metro station dummies significantly increases the goodness-of-fit of the model. Therefore, the model with the metro station dummies is used for the interpretation of the results.

The model goodness-of-fit, the adjusted $R^2$ is 0.88 which shows a good degree of explanatory power for a cross-sectional model. The comparison between the observed cost and the predicted cost is shown in Figure 1. The mean prediction error is found to be only 2.3%. It is noticeable that the names of the metros are omitted to preserve confidentiality.

**Figure 1 here - Figure 1: Observed and predicted costs**

Table 2 shows a number of statistically significant effects on metro station operating costs that arise having controlled for unobservable system specific effects.

The age of the station is found to be negatively associated with the operating cost of a metro station at the 90% confidence level. This is unexpected as we would expect an older station to require more maintenance and hence be associated with higher costs. The explanation of this counterintuitive finding may be due to the fact that more recent stations (e.g., KCR, Hong Kong)
tend to be larger and to have higher-quality facilities which also require a relatively high maintenance treatment.

The length of passageways, the total number of platforms, the peak hour service frequency, and the entry and interchange demand are found to be statistically significant at the 95% confidence level and positively associated with the operating cost. These results confirm our hypotheses. The elasticity associated with the peak period service frequency is higher compared to others. The result suggests that a 10% increase in peak period service frequency (each way, per hour) is associated with a 4.8% increase in the operating cost and a 10% increase in the number of platforms leads to a 2.7% increase in the operating cost. The length of the roof is also found to be positively associated with the cost but only at the 90% confidence interval.

The effect of air conditioning is captured by a dummy variable. This variable is found to be positively associated with the operating cost and is statistically significantly different from zero at the 95% confidence level. This is an indication that average operating cost is high in a station with the air-conditioning facility if all other factors remain constant. The coefficient ($\theta$) of the effect of the air-conditioning is 0.35 indicating that the relative effect on the average operating cost due to the presence of air-conditioning is $100 \times \{\exp(\theta) - 1\}$, or 41%. In other words, a station with the air-conditioning facility has an extremely large impact on costs increasing the expected operating cost by 41%, holding all other factors included in the model constant.

The presence of toilets within a station is also found to be positively associated with the operating cost. This is expected as there are some costs associated with the maintenance and staffing of
toilets. However, the coefficient of this variable is unexpectedly high. Perhaps this is because this variable represents the effects of some other factors which are not included in the model.

Interestingly, the type of metro station has little effect on the operating cost. As explained previously, a categorical variable (grade, sub-surface, elevated and tube) is used to reflect the overall type of metro station. None of the coefficient estimates are statistically significant at the 95% confidence level. The tube-type metro station shows a positive coefficient relative to the at-grade-type station but only at the 87% confidence level.

The system specific dummy variables are expressed relative to and intercept for the Metro-1. The result suggests that the Metro-3, Metro-5, Metro-6, and Metro-7 metros are costlier compared to the Metro-1. The operating cost associated with the Metro-5, for instance, is about 93% higher relative to the Metro-1 if all other factors included in the model remain constant.

The number of ticket offices in a station, the total number of entrances, the operating hours per day and the presence of lifts or escalators, the width of the platforms, and the length of the longest train are found to be statistically insignificant. This is perhaps because the metro specific dummies included in the model pick up the effects hypothesised from these factors.

The models are re-estimated without the statistically insignificant variables (below 90% confidence level) of the models presented in Table 2 (with metro dummies). The results are shown in Table 3. It is interesting to note that the model goodness-of-fit remains the same after excluding five insignificant explanatory variables. The age of the station now becomes insignificant. As expected, the metro specific dummies are now picked up most of the effects.
The operating cost of the Metro-2, Metro-8, Metro-9, and Metro-10 metros are now lower relative to the Metro-1. The effects of all other factors remain invariable.

Table 3 is about here  - Table 3: Re-estimated models with the significant variables of the models shown in Table 2

Conclusions

In this paper we have developed an econometric model to investigate variance in metro station operating costs. The model regresses total metro station operating costs on a series of station characteristics and a set of metro systems specific dummy variables. The results show that there are strong unobserved system specific effects, confirming the need to differentiate the data in this way. Over and above the system specific effects we have identified some factors that appear to have an important influence on the levels of station costs. These include the length of passageways, the number of platforms, peak level service frequency, interchange demand and the provision of toilet facilities. In addition, we find a very strong effect from the existence of air conditioning which raises the expected station operating costs by as much as 40%.

Stations operating costs can be classed amongst those semi-fixed costs that do not vary proportionately with metro output. For this reason, they may be very important in determining the magnitude of RTD on the costs structure and productive efficiency of the firms. This paper has provided an improved understanding of some of the major factor driving these costs.
References


Biographical Sketch

Dr Mohammed A Quddus is a Lecturer in transport studies at Loughborough University in the UK. He received his PhD from Imperial College London in 2006 in the area of map matching algorithms for transport telematics applications. His main research interests include transport planning and policy, transport risk and safety, intelligent transport systems, and geographic information science (GI Science).

Nigel Harris is a railway planner with research interests in fares, network modelling, service planning, demand forecasting, scheme appraisal and railway business planning. He has co-authored/edited key texts on the privatisation of British Rail and on planning passenger and freight railways, as well as having published over 50 other papers. He is a visiting lecturer at the Universities of Newcastle and Birmingham.

Dr Daniel Graham is a Senior Research Fellow in the Centre for Transport at Imperial College London. He was previously at the London School of Economics where he received his PhD in 1996. He currently works on a range of themes in transport economics and policy and in urban and regional economics.