Economies of scale and density in urban rail transport: effects on productivity

Daniel J. Graham a,* , Antonio Couto b,1, William E. Adeney a, Stephen Glaister a

a Railway Technology Strategy Centre, Centre for Transport Studies, Imperial College London, London SW7 2AZ, UK
b Department of Civil Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

Received 16 November 2002; accepted 4 March 2003

Abstract

This paper examines economies of scale and density in urban rail transport. It isolates the effects of constant and non-constant returns on output and productivity growth using data relating to 17 rail systems in cities around the world. Estimates reveal constant returns to scale but increasing returns to density. The productivity model shows that total factor productivity change has been of great importance in differentiating the output performance of urban rail systems. Our analysis of average labour productivity confirms the importance of shifts to other factors of production and technological change in explaining changing levels of output per worker.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Urban rail; Productivity; Scale economies; Density; Network

1. Introduction

The literature concerning the nature of scale economies in the provision of main line railway services is well known (reviews can be found in Dodgson, 1985 and Oum et al., 1999). Researchers have sought to distinguish the effects of returns to scale (RTS) from those of returns to traffic density (RTD). RTS describe the relationship between inputs and the overall scale of operations,
including both output and network size, thus given single ownership, they reflect returns to firm size. RTD describe the relationship between inputs and outputs with the rail network held fixed. The weight of evidence in the railway literature indicates that RTD are prevalent and appear to persist over a wide range of output. This is due to the prevalence of fixed costs in the railway industry and to a range of semi-fixed costs that do not vary proportionally with output. Less consensus is evident on the existence of scale economies, though the majority view is that railways operate under constant returns to scale. Much of the research has gone on to explore the influence of RTS or RTD on aspects of railway productivity, including total factor productivity (TFP) (e.g. Caves et al., 1980, 1981a; Dodgson, 1985; Freeman et al., 1985; McGeehan, 1993; Wunsch, 1996; Tretheway et al., 1997).

These themes are important. The effective regulation of national rail services by government institutions, arrangements for subsidy, fare setting, infrastructure access charging, and the like, have all been usefully informed by an understanding of the economics underpinning railway operations (see for example Foster, 1992). While previous studies have analysed cost structures and productivity for main line railways, as far as we are aware, there has been no similar exercise conducted for urban rail systems.

In this paper we concern ourselves with urban rail, that is, with underground (metro), light railway, and suburban rail systems serving cities. Our purpose is twofold. First, we seek to quantify the magnitude of RTD and RTS for urban rail operations. It should be stressed at the outset that it does not necessarily follow that the cost characteristics of these systems will correspond to those of main line railways. Dodgson (1985) notes that while there is fairly consistent evidence for increasing RTD and constant RTS across a wide range of railway outputs, this may not extend in principle to smaller railways. Second, we identify factors that give rise to productivity change in urban rail transport focusing on the role of RTD and RTS. In so doing, the paper develops a measure of TFP change.

The methodology used in the paper is conventional. However, we do develop a novel decomposition of output growth for urban rail into that due to constant returns to factor input growth, that due to returns to density and network size, and that due to technological change. It is also worth stressing that the existing data on urban rail firms is much less extensive than that typically available for use in the main line railway studies cited above. This has been a major guiding factor in the design of our methodology. Thus, an additional important contribution of this paper is to develop a tractable method of analysis that has relative modest data requirements.

In analysing urban rail labour and TFP growth we make use of data that has been collected by the Railway Technology Strategy Centre (RTSC) at Imperial College London since 1994. The data relate to 17 rail operations in 15 cities around the world. These cities are—Berlin, Glasgow, Hong Kong (2 systems), Lisbon, London, Madrid, Mexico City, Moscow, New York, Newcastle, Oslo, Paris (2 systems), Sao Paulo, Singapore, and Tokyo. Due to the sensitive commercial nature of the RTSC data we are unable to identify results with named systems.

The paper is divided into eight sections. Section 2 sets the scene by describing variation in productivity across our group of urban rail systems. A methodology to decompose output and productivity change, paying particular attention to the influence of RTS and RTD, is then developed in Section 3. Data issues and procedures for estimation are discussed in Section 4. Estimation results are presented in Section 5. Section 6 discusses our findings on RTS and RTD. Sources of productivity are analysed in Section 7. Conclusions are then drawn in Section 8.
2. Variation in the productivity of urban rail systems

As a starting point, we demonstrate below that there is indeed substantial variation in productivity across urban rail systems. Our later analysis seeks to determine the influence of some basic economic factors in giving rise to this variation.

Fig. 1 shows measures of labour productivity for the 17 systems in the RTSC database. The diagram shows the number of passenger journeys per labour hour and the number of car kilometres per labour hour for the 17 rail systems listed above. Both indicators of labour productivity show a great deal of variation. The number of passenger journeys per labour hour ranges from 23 to 110, the mean value is 50. Similarly, we find that the number of car kilometres per labour hour ranges from 4 to 18 with a mean value of just fewer than 11. Levels of labour productivity vary substantially across our systems.

Looking at factor productivity figures alone actually tells us very little about relative performance. Enterprises that are more capital intensive will of course have higher output to labour ratios. Fig. 2 looks at productivity in a different way—as a ratio of output to total operating cost. Here total operating costs is defined as the sum of all costs associated with the operations of the system including maintenance, but excluding investment and renewal. The cost data have been converted to SUS using a purchasing power parity (PPP) index published by the World Bank. Fig. 2 again shows great variation in the outputs achieved per unit of cost input. The number of passenger journeys per $1000 of cost ranges from 615 to 3900 with a mean value of 1521. The mean number of car kilometres per $1000 of cost is 342 and the differential runs from 64% below this average to 170% above. Thus, we find considerable variation when looking at the productivity of all factors of production as expressed in total operating cost.

Figs. 1 and 2 simply show that there are significant differences in levels of labour productivity and in the ratio of output to total cost across urban rail firms. They do not provide any solid explanation as to why this may be the case. However, it is worth pointing out that variations between urban rail systems in car kilometres per labour hour and car kilometres per $1000 dollars...
of total cost, are much smaller than variations in passenger journeys per labour hour or passenger journeys per $1000 of total cost. The indication may be that passenger density is a key factor affecting the performance of systems. In the remainder of this paper we develop and implement a framework that seeks to quantify how some basic aspects of production can determine and shift productivity and output growth levels.

3. Components of productivity

In this section we outline a model used to explore the production characteristics of urban rail firms. The objective is to derive a tractable means of identifying the influence of internal economies, technology, and TFP on output and productivity growth.

We assume that urban rail firms produce output $Y$ according to a variable returns to scale production technology 

$$ Y = F(X, T), \quad (1) $$

where $X$ is a vector of factor inputs with elements $X_i$ ($i = 1, \ldots, n$), and $T$ is a time index which measures technical change. Factors of production include labour, vehicles, stations, and fixed infrastructure. The production function is quasiconcave with a strictly convex input requirement structure, continuously twice differentiable, homothetic, and homogeneous of degree $\theta$.

Two components of RTS ($\theta$) can be defined—RTD and returns to network size (RTN). RTD measure the total change in output from a proportional change in inputs but with the network held fixed, while RTN measures the total change in output from a proportional change in the network

---

$^2$ We treat the output of the firm as an aggregate quantity. This is generally accepted as valid in the main line railways literature where freight–tonne miles and passenger miles are usually distinguished as separate outputs, but are each treated homogeneously with no disaggregation of traffic types. This issue of aggregation in output measures for the rail sector is discussed in Oum et al. (1999).
size, but with all other factors held fixed. Thus, for urban rail operations RTS allows us to hypothesise the existence of two distinct elements—RTD ($\delta$) and RTN ($v$)—such that $\delta + v = \theta$.

In decomposing the sources of productivity growth for urban rail firms we wish to identify the role of constant returns to scale input growth, the effects of each type of scale economy, and the impact of technological progress. To isolate these contributions we totally differentiate Eq. (1) and normalise the relevant output elasticities by the degree of RTD and RTN.

$$
\frac{d \ln Y}{d \ln X_j} = \frac{1}{\delta} \cdot \sum_j \frac{\delta \ln Y}{\delta \ln X_j} \cdot \frac{d \ln X_j}{d \ln X_j} + \frac{1}{v} \cdot \sum_k \frac{\delta \ln Y}{\delta \ln X_k} \cdot \frac{d \ln X_k}{d \ln X_k} + \left( 1 - \frac{1}{\delta} \right) \cdot \sum_j \frac{\delta \ln Y}{\delta \ln X_j} \cdot \frac{d \ln X_j}{d \ln X_j} + \left( 1 - \frac{1}{v} \right) \cdot \sum_k \frac{\delta \ln Y}{\delta \ln X_k} \cdot \frac{d \ln X_k}{d \ln X_k} + \frac{\delta \ln Y}{\delta T} \cdot dT.
$$

(2)

The subscript $j$ relates to change in all inputs not related to network size, and $k$ relates only to changes in inputs concerning the network size. Thus, the first of the five terms on the right-hand side of (2) measures constant RTD (CRD), the second measures constant RTN (CRN), the third measures non-constant RTD (NCRD), the fourth measures non-constant RTN (NCRN), and the fifth represents technical change.

TFP is defined as the sum of output growth attributable to non-constant returns and technical change: $\text{TFP} = \text{NCRD} + \text{NCRN} + \frac{\delta \ln Y}{\delta T} \cdot dT$. It is the contribution of all factors over and above that which constant returns would produce. Because Eq. (2) distinguishes the role of technology from that of scale effects it allows us to evaluate the relative importance of each factor in determining TFP growth. A discrete approximation to (2) over one year can be written as follows

$$
\ln Y_t - \ln Y_{t-1} = \left( \frac{1}{\delta} \right) \cdot \left[ \sum_j e_j (\ln X_{jt} - \ln X_{j(t-1)}) \right] + \left( \frac{1}{v} \right) \cdot \left[ \sum_k e_k (\ln X_{kt} - \ln X_{k(t-1)}) \right] + \left( 1 - \frac{1}{\delta} \right) \cdot \left[ \sum_j e_j (\ln X_{jt} - \ln X_{j(t-1)}) \right] + \left( 1 - \frac{1}{v} \right) \cdot \left[ \sum_k e_k (\ln X_{kt} - \ln X_{k(t-1)}) \right] + \hat{T},
$$

(3)

where the $e$’s are the output elasticities of factor inputs to be estimated, and where $\hat{T}$ is an estimate of the rate of technical change.

In this paper we wish to identify the components of single-factor productivity change, as well as those of output and TFP. Chan and Mountain (1983) derive an identity that is easily modified to explain the growth in productivity of single factors. They define total change in productivity as simply the rate of technical change (i.e. $\frac{\delta \ln Y}{\delta T}$). It follows that the growth in average product of factor $g$ can be expressed in terms of the remaining $h$ factors as

$$
\Delta AP_g = (\ln Y_t - \ln Y_{t-1}) - (\ln X_{gt} - \ln X_{gt-1})
$$

$$
= \frac{1}{\theta} \sum_{h \neq g} e_h \cdot [(\ln X_{ht} - \ln X_{ht-1}) - (\ln X_{gt} - \ln X_{gt-1})] + \hat{T}
$$

$$
+ \left( 1 - \frac{1}{\theta} \right) \cdot \left[ e_g (\ln X_{gt} - \ln X_{gt-1}) + \sum_{h \neq g} e_h (\ln X_{ht} - \ln X_{ht-1}) \right].
$$

(4)
In establishing this identity we use the fact that for any function homogeneous of degree \( \theta \) Euler’s theorem states that \( \sum_i \partial Y / \partial X_i = \theta Y \) and therefore \( \sum_i \epsilon_i = \theta \) where \( \epsilon_i \) is \( \partial \ln Y / \partial \ln X_i \).

Eq. (4) decomposes change in average productivity into components that isolate respective shifts away from factor \( g \) to other factors, technological change, and the scale effects. Having already distinguished density and network effects in Eq. (4) there are no obvious benefits from attempting to do so again.

The model outlined above allows us to isolate the roles of some basic economic characteristics of urban rail firms in determining rates of output and productivity change. Our basic data includes information on inputs to production, but not on the relevant output elasticities. In the next section we outline an econometric methodology to estimate these unknown parameters.

4. Data and estimation procedures

In this section we outline the method and data used to estimate the output elasticities of factor inputs in urban rail operations. The RTSC data allow for the definition of four basic factors of production for urban rail firms; labour (number of employees) \( (L) \), fleet of vehicles \( (F) \), route length (km) \( (R) \), and the number of stations \( (S) \). Non-network factors are labour and fleet, network related factors are stations and route length. We can also define two measures of firms’ output \( (Y) \)—passenger journeys per annum and car kilometres per annum.

Through initial experimentation we found the RTSC data to be unsuitable on its own for the estimation of production function parameters. Only 17 firms currently provide data, and while we do have information for 5 years (1994–1998), the lack of variation over time in many of the variables creates substantial difficulties in statistical estimation.

For this reason we had to look elsewhere for suitable data. We decided to use the urban rail data in the Urban Public Transport Statistics produced by the RTSC on behalf of the Union Internationale des Transport Publics (UITP, 1997). These data provide information on a variety of characteristics of almost 200 urban rail systems across the world including underground, light rail, and suburban rail. The way in which they are arranged is similar to the RTSC data and allows for the definition of the same inputs and output in production. The UITP data also have the advantage that like the RTSC data, they give measures of the physical as well as monetary characteristics of firms’ inputs and outputs thus avoiding the need for exchange conversions. These data provide an extremely rich source of urban rail statistics. In some instances we found missing entries and have attempted where possible to supplement the UITP records with information from Jane’s Urban Transport Systems published by Jane’s (1996).

As mentioned previously, the UITP data allow for the same inputs and outputs to be defined. Other relevant information includes characteristics of the urban areas in which the systems operate; speed, frequency and other technical aspects of rail operations; the extent of government subsidy; labour force characteristics; and some basic indication of automation in the systems. A full description of the data can be found in UITP (1997).

One drawback of UITP data is that they are cross-sectional being currently available only for one year. For this reason, they cannot be used to evaluate the growth identities given in Eqs. (3) and (4) above. Instead, we use the UITP data only to estimate the relevant output elasticities and then apply these estimates to the time-series RTSC data to decompose output and productivity.
growth. In estimating our parameters on the basis of static cross-sectional data we are not able adequately to control for the influence of technological growth. Thus, our elasticity estimates may contain, to some degree, the effects of technology as well as the ‘pure’ elasticity effects. We return to this issue in our results section.

Given the available data we assume that the production technology of urban firms can be approximated by a variable returns-to-scale Cobb-Douglas technology. \( Y = h(\cdot)L^\alpha F^\beta S^\gamma R^\delta, \) where in additional to the variables already defined \( h(\cdot) \) is a Hicks neutral external shift parameter containing a vector of externalities that influence production.

For any input, for instance route length \((R)\), we can specify a simple relationship that explains factor productivity in terms of the relative intensity of other factors and returns to scale:

\[
\frac{Y}{R} = h(\cdot) \left( \frac{L}{R} \right)^\alpha \left( \frac{F}{R} \right)^\beta \left( \frac{S}{R} \right)^\gamma R^{(\alpha + \beta + \gamma + \delta - 1)}.
\]

From Eq. (6) it is clear that RTN \((\nu)\) is equal to \(\gamma + \chi\), while RTD \((\delta)\) are equal to \(\alpha + \beta\).

Taking logs of (6) gives

\[
\ln\frac{Y}{R} = \ln h(\cdot) + \alpha \ln \left( \frac{L}{R} \right) + \beta \ln \left( \frac{F}{R} \right) + \gamma \ln \left( \frac{S}{R} \right) + (\alpha + \beta + \gamma + \chi - 1) \ln R.
\]

All the production function parameters relevant to our study of productivity growth can identified by the estimation of (7) above.

As mentioned previously, the UITP data define two outputs—annual car kilometres and annual passenger journeys. Regressions based on passenger journeys tended to produce much better results than those based on car kilometres and in this paper we use passenger journeys as the dependent variable. In transportation analyses passenger journeys are referred to as final outputs while car kilometres are referred to as intermediate outputs (e.g. Small, 1992; Berechman, 1993). For the transit firm, it is often argued that final outputs are not under the same degree of control as intermediate outputs. However, it is the case that urban rail systems employ labour and capital (for instance automated ticket machines and station assistants) to maximise passenger journeys rather than simply car kilometres.

Eq. (7) includes a vector of Hicks neutral effects in urban rail operations. The UITP data allowed us to examine the impact of a variety of influences on production. Initial regressions considered the effects of local urban characteristics (population, population density, and economic activity rates), labour force characteristics (proportions of employees in administration, maintenance, and operations), firm specific operation characteristics (average speed, hours of operation,
capacity/utilisation), and the degree of automation. We found that with the exception of automation these variables performed poorly. The estimating equation for the Cobb-Douglas function is

$$
\ln \frac{Y}{R} = \delta D_A + \sigma \text{SUB} + \alpha \ln \left( \frac{L}{R} \right) + \beta \ln \left( \frac{F}{R} \right) + \gamma \ln \left( \frac{S}{R} \right) + \left( \alpha + \beta + \gamma + \chi - 1 \right) \ln R + \mu,
$$

(8)

where in addition to the variables already defined $D_A$ is a dummy variable which takes a value of 1.0 if the system has automatic ticket machines and 0 otherwise. It is hoped that the inclusion of this effect will provide some control for the degree of automation of the system. The SUB variable records the proportion of operating revenue that is derived from public subsidy. Given the existence of zero values we represented this variable in the regressions using a Box–Cox transformation.

5. Estimation results

Estimates from the Cobb-Douglas production function are shown in Table 1. As discussed in the previous section, the data used for estimation are taken from UITP (1997). They are annual cross-sectional data with 99 observations including 38 metros, 41 light railways, and 20 suburban rail systems. In addition to the variables included in Table 1 the Cobb-Douglas function was also estimated with a set of dummy variables corresponding to the type of system—i.e. underground, light rail, or suburban rail. The motivating logic was to capture differences that might arise through system type by specifying different intercept coefficients. An $F$-test between the relevant restricted and unrestricted model was calculated at 2.96 with 2 and 90 degrees of freedom. This $F$-ratio is not statistically significant at an acceptable level, and thus on this basis, we could not reject the hypotheses that there is no strong dimension in the data according to system type that requires differentiation.

The explanatory power of the Cobb-Douglas model is high as indicated by the $R^2$ value of 0.87. The associated $F$-ratio of 104 with 6 and 92 degrees of freedom allows us to reject the hypothesis that the model has no explanatory power at the 1% level of significance.

---

4 Strictly speaking a production function is defined over real inputs and outputs and should therefore not include subsidy. It may be reasonable to suppose, however, that the output we observe is different than the maximum output available because of subsidy. In this way we view the subsidy variable as behaving like technical change.

5 The Box–Cox transformation is an available metric that not only permits zero values but also contains the natural logarithm as a limiting case. This metric, proposed by Box and Cox (1964), is defined for a variable $x_i$, as:

$$
f_i(x_i) = \begin{cases} 
\frac{(x_i^\lambda - 1)}{\lambda} & \text{if } \lambda \neq 0, \\
\ln x_i & \text{if } \lambda = 0.
\end{cases}
$$

Provided that $\lambda$ is strictly positive the Box–Cox metric is well defined for zero levels: $f_i(0) = -1/\lambda$. Furthermore, the natural logarithm metric is a limiting case of the Box–Cox:

$$
\lim_{\lambda \to 0} \frac{(x_i^\lambda - 1)}{\lambda} = \ln x_i.
$$
The elasticity of output with respect to labour is estimated at 0.969 and with respect to the fleet at 0.373, both of which are statistically significant at the 1% level. The elasticity of output with respect to the number of stations is estimated at 0.133 but is significant only with 90% confidence. The coefficient on route length, which effectively measures internal firm economies of scale, is positive but insignificant, and thus the hypothesis of CRS cannot be rejected. We therefore assume that returns to scale are in fact constant.

We discuss in more detail the influence of RTD and RTN in the next section. There are two additional results given above that are of interest. First, we see that the elasticity of output with respect to public subsidy is negative. This is consistent with previous results for mainline railways (for example Caves et al., 1981b; Friedlander et al., 1993). The estimated elasticity is, however small predicting a 0.04% decline in passenger journeys if the proportion of public subsidy increased by 10%. The indication is that public subsidy may not be increasing mobility for the urban populations being served by rail.

Second, the model suggests a powerful effect associated with the dummy variable for automated ticketing. This result suggests that much higher output is achieved on systems with greater automation, and intuitively this seems to make sense. However, the direction of causality is not clear—it may well be that it is the busiest and most heavily used systems that most require automation. The data did not allow us to explore this issue in greater detail. But given the productivity framework outlined above, the most important aspect for us is to ensure that we have provided some controls for this effect in estimating our basic elasticities. Our estimates of technological change, being based on residual quantities, will then incorporate automation as well as other firm-specific effects.

### 6. Returns to scale and density in urban rail operations

Table 2 expresses the results estimated above in terms of output elasticities and estimates of RTS and RTD.

In the first instance we can see that there are constant RTS, as directly estimated from our Cobb-Douglas specification. This result means that we have not found a systematic relationship...
between larger urban rail systems and output levels. In other words, it indicates that doubling the total size of an entire urban rail operation would leave unit costs unchanged.

However, we do find that there are substantial economies of density. The estimated elasticity of output with respect to density is 1.34. Thus, we find on average, that if the use of factors associated with density increases by 10%, the average output of urban rail operations increases by 13.4%. This of course implies that systems with higher density have lower unit costs. The sources of RTD are related to the fixed and semi-fixed costs that are prevalent in urban rail operations. For instance costs such as minimum track maintenance, station staffing, and simply keeping the stations and system in operation, do not vary proportionately with system throughput.

Our estimate of RTN is negative, −0.342. This result we found to be consistent across different specifications of the Cobb-Douglas function, and not just those reported. Thus, larger urban rail networks, holding staff and fleet constant, are associated with fewer passenger journeys per annum, because they have a lower frequency of service.

The results given above are consistent with those typically found in analyses of the cost structures of main line railways (e.g. Keeler, 1983; Dodgson, 1985; Caves et al., 1980, 1981a,b, 1985). It is not the size of the urban rail firm, with respect to the overall scale of operations, that differentiates output performance but the density of traffic on the network. That these density effects exist means that the productivity and cost efficiency of urban rail firms is not associated only with the managerial and organisational abilities of the firm, but is rooted in some basic economic characteristics of operations that may fall outside the immediate control of the system operator. In the following section we go on to look at how these economies of density affect productivity.

7. Sources of urban rail productivity growth

Table 3 shows average annual output and TFP growth for urban rail systems over the period 1994–1998 broken down into the components identified in Eq. (3) above. Due to the quality of the data it has been possible to calculate our figures for only 15 of the 17 systems for which we have information. Two points are worth stressing before we proceed to interpret these figures. First, the length of time covered in our data is relatively short for an analysis of TFP and this necessarily tempers the degree of confidence that can be attached to the results. Second, in calculating the figures in Table 3 we have applied the elasticities estimated using a cross-section of world urban

Table 2
Urban rail output elasticities and returns to scale and density

| Elasticity of output with respect to labour | 0.969 |
| Elasticity of output with respect to fleet | 0.373 |
| Returns to density | 1.342 |
| Elasticity of output with respect to stations | 0.133 |
| Elasticity of output with respect to network length | −0.475 |
| Returns to network size | −0.342 |
| Returns to scale | 1 |
rail systems. This is particularly relevant in the interpretation of the residual technology component. Here we have to adopt a fairly broad definition of technology, which in addition to capturing the pure efficiency gains induced through the passage of time as technological progress takes place, also includes a range of unobservable system-specific effects. These will include the degree of allocative and technical efficiency that arises other than through scale effects (i.e. from managerial and organisational decisions). In essence then, this component measures the deviation of each firm in efficiency terms from the industry standard less the non-constant returns effect, and is similar in concept to a measure of X-(in)efficiency (e.g Leibenstein, 1966; Stigler, 1976).

The second column on the left-hand side of Table 3 shows average annual rates of output growth (passenger journeys) for the urban rail firms. Seven of the systems we examine have expanded passenger journeys while the remaining eight have experienced output decline and in some cases substantial contraction. The mean value for all our systems shows a decline in the average annual number of passenger journeys of 0.8%.

Columns 3–5 in Table 3 show the constant returns contribution of input change to output change. Seven of the 15 systems have reduced factors associated with density (labour and fleet) over the five-year period and in some cases this is associated with substantial declines in output. Systems III, X, XII, and XV have all increased density substantially over the period and our model predicts fairly strong output growth on this basis assuming constant returns. Note that the variation in output effects due to CRN is much smaller than that due to CRD as there are less extreme changes in these factors of production.

Thus, regarding the constant returns contribution of input growth (CRIG), Table 3 demonstrates how decisions to expand and contract different input combinations leads to output change in the absence of non-constant returns. For instance, system V has reduced network size over the

Table 3
Average annual output and TFP change for urban rail systems, 1994–1998 (%)

<table>
<thead>
<tr>
<th></th>
<th>dY/Y</th>
<th>CRD</th>
<th>CRN</th>
<th>CRIG</th>
<th>NCRD</th>
<th>NCRN</th>
<th>NCRIG</th>
<th>TECH</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>13.90</td>
<td>10.14</td>
<td>2.00</td>
<td>-1.15</td>
<td>-3.47</td>
<td>-2.68</td>
<td>-6.15</td>
<td>0.40</td>
<td>-5.75</td>
</tr>
<tr>
<td>II</td>
<td>3.55</td>
<td>-3.27</td>
<td>-0.05</td>
<td>-3.32</td>
<td>-1.12</td>
<td>0.07</td>
<td>-1.05</td>
<td>7.92</td>
<td>6.87</td>
</tr>
<tr>
<td>III</td>
<td>-1.42</td>
<td>9.74</td>
<td>0.00</td>
<td>9.74</td>
<td>3.33</td>
<td>0.00</td>
<td>3.33</td>
<td>-14.49</td>
<td>-11.16</td>
</tr>
<tr>
<td>IV</td>
<td>1.33</td>
<td>0.95</td>
<td>-0.30</td>
<td>0.65</td>
<td>0.33</td>
<td>0.40</td>
<td>0.73</td>
<td>-2.71</td>
<td>-1.98</td>
</tr>
<tr>
<td>V</td>
<td>6.09</td>
<td>0.82</td>
<td>3.74</td>
<td>4.56</td>
<td>0.28</td>
<td>-5.01</td>
<td>-4.73</td>
<td>6.26</td>
<td>1.53</td>
</tr>
<tr>
<td>VI</td>
<td>1.99</td>
<td>-0.79</td>
<td>0.14</td>
<td>-0.65</td>
<td>-0.27</td>
<td>-0.18</td>
<td>-0.45</td>
<td>3.10</td>
<td>2.65</td>
</tr>
<tr>
<td>VII</td>
<td>0.23</td>
<td>0.16</td>
<td>0.00</td>
<td>0.16</td>
<td>0.05</td>
<td>0.00</td>
<td>0.05</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>VIII</td>
<td>1.44</td>
<td>-5.03</td>
<td>-0.38</td>
<td>-5.41</td>
<td>-1.72</td>
<td>0.51</td>
<td>-1.21</td>
<td>5.17</td>
<td>3.97</td>
</tr>
<tr>
<td>IX</td>
<td>-0.21</td>
<td>-5.68</td>
<td>0.07</td>
<td>-5.61</td>
<td>-1.94</td>
<td>-0.09</td>
<td>-2.03</td>
<td>7.43</td>
<td>5.40</td>
</tr>
<tr>
<td>X</td>
<td>5.23</td>
<td>5.59</td>
<td>-0.05</td>
<td>5.54</td>
<td>1.91</td>
<td>0.06</td>
<td>1.98</td>
<td>-2.29</td>
<td>-0.31</td>
</tr>
<tr>
<td>XI</td>
<td>-8.08</td>
<td>-4.49</td>
<td>-0.89</td>
<td>-5.38</td>
<td>-1.54</td>
<td>1.19</td>
<td>-0.35</td>
<td>-2.35</td>
<td>-2.70</td>
</tr>
<tr>
<td>XII</td>
<td>-8.41</td>
<td>4.22</td>
<td>-0.82</td>
<td>3.40</td>
<td>1.44</td>
<td>1.10</td>
<td>2.55</td>
<td>-14.36</td>
<td>-11.81</td>
</tr>
<tr>
<td>XIII</td>
<td>-2.55</td>
<td>0.07</td>
<td>0.00</td>
<td>0.07</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
<td>-2.65</td>
<td>-2.62</td>
</tr>
<tr>
<td>XIV</td>
<td>6.58</td>
<td>-3.20</td>
<td>0.55</td>
<td>-2.65</td>
<td>-1.09</td>
<td>-0.73</td>
<td>-1.83</td>
<td>11.06</td>
<td>9.23</td>
</tr>
<tr>
<td>XV</td>
<td>1.92</td>
<td>3.41</td>
<td>-0.04</td>
<td>3.38</td>
<td>1.17</td>
<td>0.05</td>
<td>1.21</td>
<td>-2.67</td>
<td>-1.45</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.79</td>
<td>-0.51</td>
<td>0.26</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.35</td>
<td>-0.53</td>
<td>-0.01</td>
<td>-0.54</td>
</tr>
</tbody>
</table>

Notes: CRIG—constant returns to input growth, NCRIG—non-constant returns to input growth.
period being examined while increasing the density of other factors giving rise to strong CRIG. Systems III, VII and XIII have retained a constant network size and achieved positive CRIG by increasing density. Other systems such as I and XI have reduced density and as a result have contracted output.

Columns 6–8 in Table 3 are again concerned with the impact of input growth but this time they capture the amount of output change that can be explained by non-constant returns. Our elasticity estimates indicate that RTD are increasing and thus the expansion of factors of production related to density raises output with network effects held fixed. Of the eight systems that employed more labour and fleet over the period we find output effects from NCRD ranging from 0.02% to 3.3%. Non-constant output effects of decreasing density range from −0.3% to −3.5%. The mean value over all systems is −0.17%.

The elasticity associated with RTN is a negative fraction and the interpretation of the NCRN figures in Table 3 needs careful explanation in light of this. In considering CRN we model what would happen if output changed in constant proportion to network size. That the elasticity is in fact a negative fraction means that in our model NCRN effects becomes a kind of compensating factor that corrects for the constant returns assumption. As the figures in Table 3 show, the NCRN effects are consistently of greater magnitude than the CRN effects. The mean NCRN value over all our systems is −0.35% and values range from −5.0% to 1.2%. In essence, this indicates that our systems are not working with constant network effects and that changes in these factors of production do have strong implications for output growth due to non-constant returns. To state the importance of the result more clearly, we can say that because of the existence of CRS, firms that expand network inputs, without a corresponding expansion in the inputs associated with density, will find that the rate of output growth lags behind the rate of growth in network size.

Our model thus shows that NCRIG have a crucial part to play in differentiating the output performance of urban rail systems. The most important issue relates to the extent to which firms are aware of the fact that non-constant returns are very influential within the industry. Growth in the density of operation has increasing returns, growth in the network size has decreasing returns, and we identify a clear impact from these basic economic characteristics in differentiating firms’ output performance.

The next effect identified in Table 3 relates to the role of ‘technology’, calculated as a residual given the other effects. As mentioned in the introduction to this section, the figures in Table 3 can be regarded as comparing our systems against a measure of the industry standard. This interpretation is particularly appropriate with respect to efficiency effects because here we are looking at how the growth in efficiency in each system compares with what would be expected given the general production characteristics of the industry based on cross-section data.

Our mean figure over all systems indicates a per annum decline in the output contribution of technology of −0.01% over the period being examined. The table also shows a great deal of variation amongst the systems in the output contribution of technological growth. Some systems, such as III and XII, deviate enormously from the industry average showing high negative output effects of over −14% from the technology that they employ. Other systems including II, IX, and XIV have achieved large expansions of output by using productive technologies. In short, our model suggests that the technological characteristics of operations, which include aspects of managerial and organisational practice, can explain output change to a large extent and the magnitude of this effect is not trivial.
Finally in Table 3 we have calculated the combined effects of NCRIG and technology in a measure of TFP. TFP here represents all output change that cannot be explained by constant returns input growth. Over the period being examined TFP has declined in eight systems and has increased in the other seven. As we have seen both NCRIG and technology have made substantial contributions to TFP, but they differ in the way they affect systems.

Table 4 shows our analysis of labour productivity growth based on the implementation of Eq. (4) above. Since our elasticity estimates determine CRS over all inputs the scale effects on labour productivity drop out (i.e. \( \theta = 1 \)). We are therefore left with explanations based on shifts away from labour to other factors and technological change.

The second column on the left-hand side of Table 4 shows actual rates of labour productivity growth for our systems between 1994 and 1998. Eight of the systems included in the table have experienced labour productivity growth while seven have experienced decline. Change in labour productivity has varied enormously across our systems.

Before going on to consider the other figures shown in table it is worth making an observation about the nature of urban rail operations. In many industries, for instance manufacturing sectors, factors of production can be truly substitutable. For instance, machines can be employed to do the jobs of workers. In the urban rail sector we have a slightly different situation, although one which affects all industries in detail, in which many factors cannot really perform the same tasks. Labour cannot perform the functions of fleet or route, stations cannot provide a substitute for workers and so on. However, that is not to say that from the firm’s point of view the factors are not substitutes with respect to the objective of revenue maximisation. For instance, passenger journeys can be expanded by increasing the route length and number of stations, by investing in and running more trains, or by employing more labour and increasing the density of use of the

<table>
<thead>
<tr>
<th></th>
<th>( \Delta A P L )</th>
<th>Shift to other factors</th>
<th>Total shift</th>
<th>Tech.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fleet</td>
<td>Route</td>
<td>Stations</td>
</tr>
<tr>
<td>I</td>
<td>-3.38</td>
<td>0.51</td>
<td>1.26</td>
<td>0.01</td>
</tr>
<tr>
<td>II</td>
<td>8.20</td>
<td>1.85</td>
<td>0.24</td>
<td>0.63</td>
</tr>
<tr>
<td>III</td>
<td>-14.60</td>
<td>-4.62</td>
<td>-0.69</td>
<td>-1.75</td>
</tr>
<tr>
<td>IV</td>
<td>-3.06</td>
<td>-1.05</td>
<td>0.00</td>
<td>-0.22</td>
</tr>
<tr>
<td>V</td>
<td>4.91</td>
<td>-0.48</td>
<td>-0.06</td>
<td>-1.43</td>
</tr>
<tr>
<td>VI</td>
<td>2.92</td>
<td>0.18</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>VII</td>
<td>0.66</td>
<td>0.79</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>VIII</td>
<td>4.41</td>
<td>1.12</td>
<td>0.47</td>
<td>0.74</td>
</tr>
<tr>
<td>IX</td>
<td>7.54</td>
<td>2.78</td>
<td>0.42</td>
<td>0.99</td>
</tr>
<tr>
<td>X</td>
<td>-2.43</td>
<td>-2.78</td>
<td>-0.41</td>
<td>-0.99</td>
</tr>
<tr>
<td>XI</td>
<td>2.15</td>
<td>7.71</td>
<td>1.04</td>
<td>1.16</td>
</tr>
<tr>
<td>XII</td>
<td>-14.76</td>
<td>-2.87</td>
<td>0.28</td>
<td>-1.17</td>
</tr>
<tr>
<td>XIII</td>
<td>-2.65</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>XIV</td>
<td>10.06</td>
<td>0.38</td>
<td>0.26</td>
<td>0.20</td>
</tr>
<tr>
<td>XV</td>
<td>-2.43</td>
<td>-1.26</td>
<td>0.05</td>
<td>-0.84</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.16</td>
<td>0.15</td>
<td>0.20</td>
<td>-0.17</td>
</tr>
</tbody>
</table>
Of the systems shown in Table 4, seven have experienced shifts from other factors to labour while eight have shifted investment into other factors from labour. The mean value over all systems for factor substitution is 0.17% indicating a small shift from labour into other factors. For three of the systems shown in Table 4, VII, X, and XV, our model suggests that factor substitution has had a greater impact on labour productivity than technology. While for five others, III, IV, VIII, IX, and XI, factor substitution and technological change have been equally influential in labour productivity change.

Labour productivity change in the remaining seven systems has been more strongly affected by technology than factor substitution. For systems II, IV, V and XIV technological growth has served to reduce labour demand per unit of output, and for I, XII, and XIII technology has been labour augmenting.

Our analysis of annual average labour productivity change confirms that both factor substitution and technological growth are important in determining variation in labour productivity growth. There is no consistency across the systems in the direction of influence of these components of change. As regards factor substitution, it would seem that labour shift from and to fleet are the most prominent.

In essence, the analysis presented above has allowed us to consider how firm efficiency, as measured with respect to TFP or as ‘technology’ in the labour productivity analysis, enters into the output and labour productivity performance of firms. Our analysis reveals some important general characteristics of urban rail firms. First, and perhaps most crucial, is the importance of strong increasing returns to density in operations with constant returns to scale. The output performance of urban rail firms is strongly affected by the extent to which they exploit this characteristic of the industry, being aware that there are increasing RTD but decreasing RTN. Second, we have shown that TFP has a real influence on rates of output change between firms. Rates of TFP are formed both by differences in the extent of NCRIG and in the state of technological progress and other firm-specific ‘efficiencies’ as captured by our technology effect. Third, we have shown that technology and factor substitution both contribute to differences in labour productivity between firms.

8. Conclusions

This paper has examined output, TFP, and labour productivity change in urban rail transport. It has shown that the urban rail industry is characterised by constant returns to scale but also by powerful increasing returns to density. We have shown that the level of output performance of these firms is dependent upon the extent to which they are able to exploit these economies of density. Our results also indicate that TFP growth is created by both non-constant returns and technology effects and is extremely important in differentiating the growth rates of firms. Factor substitution and technology are both evident in explaining the variation between firms in rates of labour productivity growth.

The results tell us a great deal about the cost characteristics of urban rail firms. Costs are characterised by large fixed components. And this is true not only in the long term given ‘sunk’
costs in ‘fixed’ capital, but even in the short term as certain constant elements in the nature of costs prove invariant to specific aspects of urban rail output, for instance, in relation to passenger density or minimum track maintenance. The fact that fixed costs are prevalent in short and long run costs implies that average short and long run costs will be generally higher than marginal costs, which of course has important implications for fares policy, subsidy, and regulation.

But our results have also proven useful in comparing productivity performance across urban rail systems. We have shown that RTD do have an important influence on TFP and labour productivity. Some firms may not be able to increase the density of use of their system, or may not desire to do so because they wish to provide services to urban areas that will not be densely used. Under these circumstances, the operator should perhaps be less concerned that their productivity figures appear to be below average because it may not reflect managerial or organisation inefficiencies. In this paper we have developed a methodology that make inter-firm comparisons on a more even basis by isolating the sources that give rise to variation in productivity.

Thinking about production and cost characteristics in this way may also give urban rail firms additional information in reaching decisions about aspects of network expansion and frequency of service. Those under financial pressure, or those wishing to achieve maximum output growth from investment, will find that allocating resources to achieve greater density of the core system is more productive than reducing density by extending the network to outlying areas. If contraction is required, cutting the extremities of the network is likely to be better than reducing the density of service on the existing system. In practice many metros may already be operating at, or near, the maximum frequency of service. In these cases, if sufficient latent demand exists, then density increases could be achieved by expanding the network or by building new lines.

References


