ARTICLE IN PRESS

Case Studies on Transport Policy xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Case Studies on Transport Policy

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

Accounting for factors affecting cost and technical inefficiency

K. Obeng

Department of Marketing and Supply Chain, College of Business and Economics, North Carolina A&T State University, 1601 E. Market St., Greensboro, NC 27411, United States

ARTICLE INFO

Keywords:
Heterogeneity
Technical inefficiency
Cost
Stochastic frontier
Public transit
Regulations
Environmental variables

ABSTRACT

This paper addresses heterogeneity in the deterministic portion of a cost frontier and in the parameterisation of technical inefficiency using an unbalanced panel of bus transit systems. It conceptualises six groups of variables which affect heterogeneity in cost and technical inefficiency: organisational size, ownership, service delivery methods, regulations, external factors and trend. Then, it specifies and estimates a translog cost frontier and from its results identifies those variables that affect technical inefficiency. It finds that purchasing some passenger services and providing others directly and bus useful-life regulation increase technical inefficiency. It also finds that MPO-owned single mode bus systems have less technical inefficiency and while network size increases technical inefficiency it reduces cost. From this result it cautions making inferences about how a variable impacts cost from its effect on technical inefficiency. It examines the managerial implications of the results and suggests using variables in the parameterisation of technical inefficiency whose effects on costs have been established empirically or through conceptualisation.

1. Introduction

Many researches on technical inefficiency using panel data such as Battese and Coelli (1995), Matas and Raymond (1998), Cornwell, Schmidt and Sickles (1990), Lee and Schmidt (1993) and Cuesta (2000) variously address heterogeneity in the context of maximum likelihood estimation. In addition Caudill, Ford and Gropper (1995) showed that when maximum likelihood estimation methods are used heterogeneity leads to overestimation of the constant term, underestimation of the slope coefficients as well as biases in the estimation of production frontier models. To address this problem, they suggested a stochastic frontier model which parameterised the variance of inefficiency in terms of internal organisational variables, arguing as in Schmidt (1986) that they are responsible for differences in firms. Concurring with this view, Greene (2005) notes that in maximum likelihood estimation heterogeneity confounds technical inefficiency estimates if not accounted for properly. Since then, subsequent stochastic frontier models have been developed which allow heteroscedasticity in the variance of technical inefficiency to be studied by including firm-level variables in the variance of the one-sided error term and factors outside the control of management in the variance of the two-sided random error. For example Amsler, Schmidt and Tsay (2014) developed a post-truncation approach that allowed environmental variables to affect the mean of technical inefficiency, its variance or both.

Instead of parametrising the variance of technical inefficiency some recent extensions of these models decompose technical inefficiency while still accounting for heterogeneity. The Ahn et al., (2007) model for example has multiple components of inefficiency and Abrate et al. (2008) used time-invariant variables in their deterministic equation and time-varying factors in the stochastic component of their function, while Kumbhakar et al. (2012) decomposed technical inefficiency into persistent and residual components and found persistent (time-invariant) inefficiency to be larger than residual inefficiency. Also, Colombi et al. (2014) decomposed technical inefficiency into individual effect (latent heterogeneity), long term persistent inefficiency (time invariant due to input rigidities which inhibit input substitution), short run inefficiency (time varying) and random inefficiency. They conclude that their decomposition is appropriate when firms are heterogeneous, implying a large cross-section of firms to give cross-firm data variation, and a long panel to give within-firm data variation.[1]

This within-firm data variation has been noted by Greene (2003, 2005) as essential when fitting stochastic frontier models using panel data and that short panels could give inconsistent results. Further, as he notes, short panels may be unsuitable to many of the stochastic frontier models and cannot be fitted to the true fixed effect model because of the incidental parameters problem. In such instances, he suggests using

E-mail address: obengk@ncat.edu.

[1] A recent application of this decomposition to Italian airlines attributed their cost inefficiency to temporary or short-term inefficiency (Martini, Scotti, Viola and Vittadini 2020).

https://doi.org/10.1016/j.cstp.2020.04.010
Received 27 May 2019; Received in revised form 12 March 2020; Accepted 28 April 2020
2213-624X/

advanced numerical algorithms to solve this problem while Chen, Schmidt and Wang (2014) propose a simulated likelihood method to solve it. In addition, Wang and Ho (2010) propose a methodology based on first differencing to remove individual fixed effects, and within-transformation by subtracting the sample mean from each observation to remove time invariant individual effects before estimation. Further, they propose an approach to recover the individual effects from first order conditions after the estimation. While this method resolves the problem, it requires within-firm data variation absent which the value of each variable becomes a zero after the differencing.

Recognising the importance of heterogeneity and the advances made to decompose it, the extant literature in public transit economics has focussed mostly on addressing it by parameterising technical inefficiency with very few applications of the types of decomposition proposed by KL and CKMV. This literature can be grouped into two. The first which assumes that heterogeneity resides in technical inefficiency is a two-step approach with two streams of research: the post hoc, which calculates technical efficiency by parametric or non-parametric methods and relates it to organizational and environmental variables using a second-stage regression (e.g., Kerstens 1999); and the latent class approach which considers heterogeneity in its classification of transit systems and then in its decomposition of the resulting technical inefficiency by Tobit regression (Obeng 2013). The second and the most prevalent is the one-step approach which again assumes that heterogeneity resides in technical inefficiency and parameterises and estimates it using stochastic frontier models.[2] Using this latter approach and panel data, Farsi et al. (2006) conclude that ignoring heterogeneity could lead to an upward bias in the calculation of inefficiency scores because unobserved heterogeneity, output and network size may be correlated. A similar parameterisation by Vigren (2016) found low technical efficiency in Swedish transit systems operating in high density cities, where competitive tendering was used, and differences in cost efficiency between the transit systems that received incentives and those that did not. Additionally, Zhang, Juan and Xiao (2015) found that gross cost contracts provided incentives for Chinese transit systems to be technically efficient compared to net cost and management contracts, while Jarboui et al. (2013, 2014a, 2014b) found inverse relationships between board size and managerial optimism and technical efficiency, and a positive relationship between CEO tenure and technical efficiency. Similarly, Piacenza (2006) using this approach rejected the absence of cost inefficiency and found that transit systems receiving fixed-price subsidies had lower distortions from minimum cost, and that speed accounted for firm differences in cost inefficiency.

While the knowledge gained from these parameterisations of public transit technical inefficiency has been insightful, many studies do not include a cross-section of variables reasonably expected to contribute to managerial inefficiency especially those that affect cost and technical inefficiency simultaneously.[3] For example, the received public transit literature above on one-step estimation has focussed individually on the effects of regulations, organisational and environmental contextual variables on technical inefficiency (not cost) and not in aggregate. Thus, it is unknown what the results would be if variables representing different aspects of transit operations are considered together especially if they are correlated with cost and technical inefficiency. This approach in the public transit economics literature also limits understanding the true effects of some of these variables on cost by leaving researchers to infer the cost effect of a variable that affects technical inefficiency from the duality between cost and production functions, which may not always be correct if for example regulations prevent full adjustment to cost minimisation input levels. For example consider an incentive regulation that rewards transit systems when their services are effective in terms of passenger miles and reduces operating cost as found in US transit systems. Such a regulation distorts input use in favour of capital whose costs are high and increases total cost as Obeng and Azam (1995) found, as well as increases technical efficiency by rewarding transit systems that produce large outputs. If in the same stochastic frontier model this regulation is included only in the parameterisation of technical inefficiency alone and not in the deterministic cost part, probably its true effects cannot be determined leading to misdirected policies. There are other variables yet to be identified whose cost effects cannot be inferred from their effects on technical inefficiency. Research is therefore needed to begin identifying some of these variables and this is where this paper makes its contribution to the public transit economics literature. To do so it studies cost and technical inefficiency by including some variables describing the regulatory and operating environments of transit systems in the stochastic and deterministic parts of the cost function and estimating the resulting stochastic frontier model. The results allow us to compare the coefficients to identify network size as that whose cost effect cannot be inferred from its effect on technical inefficiency, and to surmise that input regulations do not allow the true effect of partial contracting on technical inefficiency to be realised. The rest of the paper is organised as follows. The next section is the methodology and it is followed by the data, results and conclusion respectively.

2. Methodology

The treatment of heterogeneity by parameterising the technical inefficiency term in a stochastic frontier model in this paper begins by considering a transit system$n$ that uses the set of inputs $x_{n1t}\cdots x_{nit,}$to produce the output $Q_{nit}$where $n = 1, 2, 3, \ldots N$ indexes transit systems, $i = 1, 2, 3\cdots j$ indexes the inputs, and the trend and technology variable$t = 1, 2, 3, \cdots .T$ shifts the production frontier. We begin by assuming that this transit system has no allocative distortion from regulations, funding and administrative restrictions and market imperfections, thus allowing full adjustments of inputs to cost minimising levels in response to input price changes. This adjustment makes the transit system minimize its total cost $C_{nt} = \sum_1^i w_{nit}x_{nit}$subject to a production technology where, $w_{nit}$is each input's observed actual price. With this assumption the transit system's cost minimization problem is,

$$\min_x L = \sum\nolimits_1^i w_{nit}x_{nit} + \gamma(Q_{nt} - Q_{nt}[x_{n1t}.\,..x_{nit}z_{n1t,}\cdots z_{njt}t]e^{v_{nt}-u_{nt}}) \tag{1}$$

where$\gamma$is the Lagrangian multiplier and $Q[\bullet]$ is the production function. Also, assume that this transit system has technical inefficiency in producing its output and let$Q_{nt}[\cdot]e^{v_{nt}-u_{nt}}$be the production frontier capturing this inefficiency, and $z_{n1t,}\cdots z_{njt}$each period's set of heterogeneity variables for this transit system. This technology assumes a composed error $\varepsilon_{nt} = v_{nt} - u_{nt}$ where $v$is a normally distributed random error, v_{nt} $N(0, \sigma_v^2)$ and the technical inefficiency term u_{nt} follows a truncated normal distribution $N(\mu, \sigma_u^2)$ with a mean of μ and a varianceσ_u^2. Thus $-u_{nt}$ is technical inefficiency or the proportion by which this transit system's actual output falls short of the maximum output it can produce with its inputs.

Using Eq. (1) the minimum cost frontier for this transit system is derived as $C_{nt} = C_{min}(w_{n1t,}w_{n2t}, ..w_{nit}, z_{n1t,}\cdots z_{njt}, Q_{nt}, t)e^{v_{nt}+u_{nt}}$ where $u_{nt}$is now the proportion by which its actual cost exceeds minimum cost due to technical inefficiency and$v_{nt}$is the proportion of its excess cost due to random error. Imposing linear homogeneity restrictions on $C_{nt}(\bullet)$ by using w_{n1t} as the price of the reference input, we approximate the transit system's empirical linearly homogeneous minimum cost frontier by the translog technology below:

[2] This is an alternative to treating the inefficiency determinants as heteroscedastic variables in technical inefficiency by parameterizing the variance of technical inefficiency in terms of these determinants as noted earlier.

[3] Though not considered in this paper Battese and Coelli (1995) note that the variables in the parameterization can include input variables provided their inefficiency effects are stochastic, and interaction terms between input variables and firm specific variables. This is also true in the Huang and Liu (1994) model where the variables appearing in the parameterization included inputs, government and institutional regulations and their interactions.

$$\ln\left(\frac{C_{nt}}{w_{1t}}\right) = \beta_0 + \beta_1\ln\left(\frac{w_{n2t}}{w_{n1t}}\right) + \beta_2\ln\left(\frac{w_{n3t}}{w_{n1t}}\right) + \beta_Q\ln(Q_{nt}) + \beta_{row}\ln(RW_{nt}) + 0.5\beta_{11}\left(\ln\left(\frac{w_{n2t}}{w_{n1t}}\right)\right)^2 + \beta_{12}\ln\left(\frac{w_{n2t}}{w_{n1t}}\right)\times\ln\left(\frac{w_{n3t}}{w_{n1t}}\right) + \beta_{1q}\ln\left(\frac{w_{n2t}}{w_{n1t}}\right)\times ln(Q_{nt}) + \beta_{r1}\ln\left(\frac{w_{n2t}}{w_{n1t}}\right)\times\ln(RW_{nt}) + 0.5\beta_{22}\left(\ln\left(\frac{w_{n3t}}{w_{n1t}}\right)\right)^2 + \beta_{2q}\ln\left(\frac{w_{n3t}}{w_{n1t}}\right)\times ln(Q_{nt}) + \beta_{r2}\ln\left(\frac{w_{n3t}}{w_{n1t}}\right)\times\ln(RW_{nt}) + 0.5\beta_{qq}(\ln(Q_{nt}))^2 + \beta_{rq}\ln(Q_{nt})\times\ln(RW_{nt}) + 0.5\beta_{rr}(\ln(RW_{nt}))^2 + \beta_{den}\ln(DEN)_{nt} + \sum_t \varphi_t(D_{nt}) + v_{nt} + u_{nt} \quad (2)$$

For heterogeneity in cost we include RW for the transit system's network size in terms of miles of right-of-way and DEN for population density. $D_t$is a binary time variable for each $t = 1, 2, \cdots\cdots.T$measuring neutral technical change whose rate is the difference between the values of $\varphi_t$in successive periods as in Cuesta (2000).[4] Since this paper's objective is to add to understanding the factors affecting cost and technical inefficiency in public transit systems to inform policy better, we follow Holmgren (2013) and use the panel formulation of the cost frontier in Battese and Coelli (1995) for time varying technical inefficiency. So, we parameterise technical inefficiency in terms of producer time-specific (or environmental time-specific) variables (y)that potentially affect it instead of just time only.[5]

The choice of variables to include in this parameterisation is informed by our earlier discussion and the results of recent work by Obeng et al. (2016) which examined the effects of government regulations on technical inefficiency in US public transit systems. That study found technical inefficiency to be higher in transit systems that met the spare-bus ratio regulation and in transit systems that operated < 150 vehicles. Margari, Erbetta, Petraglia and Piacenza (2007) found that high-powered incentive contracts (incentive regulation) improved efficiency in their study of Italian bus transit systems and Nolan (1996) found in his US study that federal and state subsidies and fleet age increased technical efficiency. Using these findings and those from the aforementioned studies as a guide, this paper conceptualizes six types of heterogeneity variables that potentially affect public transit technical inefficiency. They are organisational contextual variables including types of ownership (e.g., MPO-owned), methods of service delivery (e.g., partial contracting whereby direct provision and contracting-out of some passenger services are combined), government regulations, external environmental factors (population density and service area), organisational size (e.g., network size in terms of route miles) and trend.[6,7]

With this parameterisation the logarithm of technical inefficiency in Eq. (2) is,

$$u_{nt} = \vartheta_0 + \sum_n \vartheta_{nt}y_{nt} + W_{nt} \quad (3)$$

$$fory = f(organizationalsize, ownership, servicedelivery, regulation, externalfactors, trend)$$

where, W_{nt}follows a truncated normal distribution and the point of truncation is $-(\vartheta_0 + \sum_n \vartheta_{nt}y_{nt})$;that is, $W_{nt} \geq -(\vartheta_0 + \sum_n \vartheta_{nt}y_{nt})$. If $\vartheta_{nt} = 0$for each variable then technical inefficiency is the same for all transit systems and equal to the constant termϑ_0. If $\vartheta_{nt} \neq 0$then technical inefficiency depends on the values of y_{nt}each period and also varies across observations. The first term accounts for the contributions of missing variables (unobserved heterogeneity) to technical inefficiency and the second shows the contributions of observed heterogeneity to time-varying technical inefficiency.[8] Since we use the Battese and Coelli (1995) stochastic frontier model for panel data the effects of unobserved heterogeneity resides in the constant term but cannot be extracted. By this model the technical inefficiency effects are the product of an exponential function of time and the non-negative firm specific term given by,

$$u_{nt} = \{\exp[-\eta(t-T)]\}\left|\sum_n \vartheta_0 + \vartheta_{nt}y_{nt} + W_{nt}\right| \quad (4)$$

where, $\eta$is the time varying parameter. Substituting this equation into Eq. (2) the resulting equation can be estimated by maximum likelihood methods.

However, before the estimation and following Greene (2007) the log-likelihood function of the truncated normal model is specified as,

$$\text{Log}L_n = -\frac{1}{2}log2\pi - log\sigma - \frac{1}{2}\left(\frac{d\varepsilon_{nt}}{\sigma} + \alpha\lambda\right)^2 - log\Phi(\alpha\sqrt{(1+\lambda^2)}) + log\Phi\left(\alpha - \frac{d\varepsilon_{nt}\lambda}{\sigma}\right)$$

$$where\alpha = \frac{\mu}{\sigma\lambda}, \sigma_u = \frac{\sigma\lambda}{\sqrt{1+\lambda^2}}, \lambda = \frac{\lambda_u}{\lambda_v}, \sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$$

Where the Φterm is the cumulative density function of the standard normal distribution. This likelihood function is maximized with respect to β, σ, $\lambda$and α and the parameter μ obtained from $\mu = \alpha\sigma\lambda$.Then, we use the Jondrow et al. (1982) estimator below to calculate stochastic technical inefficiency (u) as

$$E[u|\varepsilon] = \frac{\sigma\lambda}{1+\lambda}\left[\frac{\phi(z)}{1-\Phi(z)}\right], \varepsilon = v + u, z = \varepsilon\lambda/\sigma \quad (6)$$

where, $\phi(\bullet)$is the probability density function. The value of each transit system's technical efficiency (TE) is then calculated as$TE = \exp(-u)$.

3. Data

Context: This study uses data for US public transit systems that operate bus-only modes in applying these equations. Thus, the data do not include those that operate light rail, subways, trolleys, inclined plane and ferries. A rationale for this choice is to avoid mixing modes with different operating characteristics and technologies. Another is that the data source, the National Transit Database (NTD), already uses this classification and contains detailed information on single-mode bus transit systems. These bus-only systems generally operate in cities with a million population or less, and unlike those in larger cities receive federal operating subsidy. Initially, we considered and included all such systems in this database for the period 2007–2011 and to increase

[4] In the Battese and Coelli (1995) model a trend variable appeared in the deterministic and stochastic parts of the model and they noted that the distributional assumptions of inefficiency allow technical change, the time-varying inefficiency and the constant terms in the deterministic and stochastic equations to be identified.

[5] The rationale for this model choice is that the data did not fit fixed and random effects models perhaps because it is a short panel with less within-firm data variation.

[6] In an initial formulation we included some of these variables particularly those representing regulations in the deterministic portion of the frontier and did not get good results.

[7] Interactions terms were also considered and discarded for lack of fit.

[8] In the absence of this parameterization $U_{nt} = \vartheta_0 + v_{nt}$ after adding the error term to it and if there are individual effects $\alpha_n$they can be added to $\vartheta_0$to obtain $U_{nt} = \omega_n + v_{nt}$ where $\omega_n = \vartheta_0 + \alpha_n$and the equation estimated as a fixed or random effect model.

within-firm data variation selected those that had at least 4 years data in the period studied. This resulted in data for 77 transit systems and 359 observations.

Output and inputs: Previous research guided the selection of output and inputs for this study. In three recent papers JFB used operating cost and the number of employees as measures of transit inputs in their production function, arguing that their approach was robust and avoided problems of input substitution. They also reported studies that used a single measure of input (e.g., Jha and Singh 2000). Apart from these studies, many others have used two inputs which are labour and fleet size (Singh and Vanketesh, 2003) or three inputs which are labour, fleet size and energy in their production and cost functions as in OSN and Agarwal, Yadav and Singh (2009) and Nagadevara and Ramanayya (2010). Following these latter studies, we use labour (x_{n1t}), non-labour inputs (x_{n2t}) broadly defined to represent all other inputs which are neither labour nor capital and whose costs are included in public transit operating expenses (e.g., fuel, tires, supplies, and utilities), and capital (x_{n3t})mwasured by fleet size. We measure these inputs as follows: labour worked-hours measure the quantity of labour employed, fleet size is the quantity of capital, and the total gallons of all types of fuel are a proxy for all other inputs. Because this study's focus is on technical efficiency, we use the produced output, vehicle miles, as output as opposed to passenger miles which are demand-related and reflect service effectiveness.

Input prices: All prices and cost are in base year 1984 prices. Labour's price(w_{n1t}) is real total labour compensation and benefits per labour worked hour; the price of the proxy input(w_{n2t}) is real total non-labour operating cost per gallon of fuel. For the rental price of capital(w_{n3t}) we use the formula in OSN to calculate it. This formula is $w_{n3t} = r_t(R_t + d)e^{-d(AGE_{nt})}$ where, r_t is the weighted average real price of a new public transit bus in each year $t$as reported by APTA (2007, 2008, 2009, 2010 and 2011). $R_t$is each year's rate on a high yield municipal-bond for the city or county where the transit system is located; $d$is a straight line rate of depreciation assuming a bus useful life of 20 years; and AGE_{nt} is each transit system's average fleet age. From these prices the real yearly total cost of each transit system is$C_{nt} = w_{n1t}x_{1t} + w_{n2t}x_{2t} + w_{n3t}x_{3t}$.

Regulation variables: Regulations in US transit systems are different from what prevails elsewhere because there is not a central body overseeing and monitoring the types of services transit systems provide in terms of quantity, quality and fares. Rather they are in the form of administrative and funding requirements, executive instruments and laws passed by congress. In this study we consider five federal regulations eventually using three in the models because not all transit systems met them, thus making it possible to code each of them using a binary system. For the labour protection regulation (CAP) which transit systems must certify they meet before receiving federal capital subsidy, a transit system was coded one (yes = 1) and having met the regulation if it received capital subsidy and a zero (no = 0) otherwise. Similarly, a transit system receiving Federal Urbanized Are Formula Grant (IN) was coded one (yes = 1) and having met the incentive regulation and zero (no = 0) otherwise. The federal government regulates that transit systems cannot keep more than 20% of the buses they buy with federal subsidy as spare (SP). To determine if a transit system met this regulation, we subtracted the yearly percentage of its fleet of buses it operated in maximum service from 100, and considered a transit system as having met it if the resulting percentage was 20% or less. And for the requirement that buses bought with federal subsidy must be used for at most twelve years (FA), a transit system was coded as having met it if its average fleet age was less than or equal to 6 years (yes = 1) and not having met it otherwise. Lastly, the federal requirement that transit systems receiving operating subsidy must purchase some passenger services from the private sector (DOPT) was treated as a service delivery method and its coding discussed under that heading.

Service delivery method, ownership and others: To account for ownership differences, we distinguish city-owned (CITY) and MPO-owned transit systems (MPO) since they are the most dominant in our data and use a binary system to code each of them (yes = 1, no = 0). Regardless their types of ownership and coding, by regulation U.S. transit systems that receive federal operating subsidy are required to contract-out portions of their passenger services to private sector companies to provide (DOPT) and those that do so are coded one (yes = 1) and zero (no = 0) otherwise. Similarly, we used a binary coding (yes = 1, no = 0) to identify those that provide their services directly (DP), recognizing that those are their core areas. That is, the areas of their expertise and in focusing on them they are able to provide the quantity and quality of service they desire at a cost they can afford, which might lead to lower technical inefficiency and lower cost. While other measures such as purchasing the entire passenger service from private sector sources (PT) could have been used, an initial estimation of the model that included them did not yield good results. Besides these variables, the data include population per square mile (DEN) reflecting congestion, service area in terms of square miles (SQ) describing the external operating environments of the transit systems, and network size in terms of miles of right of way (RW) all of which are observed heterogeneity variables.

Descriptive statistics: Table 1 shows descriptive statistics about the data that this study used. From the table all the transit systems we studied met the federal labour protection requirement, 93.6% received incentive subsidy and 39.3% met the bus useful-life regulation. Also, 41% had dedicated local subsidy sources, 28.4% met the spare bus ratio regulation, 39.6% operated their services entirely by themselves, 6.1% used the mixed approach whereby they provided some of their passenger services themselves and contracted out the rest to private sector companies and 54.3% contracted out their entire service to private sector companies. Thus, overall, 60.7% of the transit systems studied involved private sector companies in their operations through partial- or full-service contracting. Also, the table shows that 60.7% of these transit systems were city-owned, 34.3% were MPO-owned, and other agencies such as universities, state and human services organizations owned the rest (5.0%).

4. Estimation and results

The empirical equation to be estimated with this data is Eq. (2) with Eq. (4) substituted into it and incorporating the assumed truncated normal distribution of inefficiency. Before the estimation some modifications were made to the model. First, because all the transit systems on one hand and 93.6% on the other met the federal labour protection and the incentive regulations respectively they were deleted from the final model for lack of variation. Second, initial estimation showed that none of the coefficients of the binary time variables was statistically significant in the deterministic model so one year (t_{2008}) was randomly selected and included in the final model. After these modifications the resulting model showing time-varying technical inefficiency was estimated by the Battese and Coelli (9 9 5) stochastic frontier method for panel data in LIMDEP (Greene 2007). The results of the estimation are in Table 2. The fit statistics at the top of this table clearly show that the source of the variance in the errors in the stochastic frontier is the truncated inefficiency term (72.4%) and not the normally distributed random error term. Next, considering the deterministic component of the stochastic cost frontier, many of the first order coefficients which show long run elasticities are statistically significant and give 14.2% and 20.1% as the respective cost shares of non-labour and capital inputs. Adding these cost shares and subtracting the result from one, gives labour cost share of 65.7%. Besides these cost shares, the elasticity of cost with respect to output being 0.78 and statistically different from one (t = (0.79–1.00)/0.0221) = -9.50) indicates economies of scale. Other first order coefficients that are statistically significant are those of network size, population density and service area which are −0.0397, 0.4319 and 0.0982 respectively. Of the second order terms, many have statistically significant coefficients except the capital-labour relative

Table 1
Descriptive Statistics.

Cost, Prices and Output	Mean	Std. Dev.	Minimum	Maximum	Cases
$\ln(C_{nt}: Totalcost)$	15.9093	1.0419	13.7361	18.7746	359
$\ln(w_{n2t}: Priceofnon - laborinput)$	0.6035	0.4102	−1.0505	2.8014	359
$\ln(w_{n2t}: Rentalpriceofcapitalinput)$	7.1393	0.3480	6.2971	8.1754	359
$\ln(w_{n3t}: Priceoflabour)$	2.5319	0.2961	1.5598	3.2281	359
$\ln(Q_{nt}: Outputinvehiclemiles)$	14.9219	0.9972	12.8639	17.6690	359
Regulations					
SP: Meets spare-bus ratio regulation (yes = 1, no = 0)	0.2841	0.4516	0.0	1.0	359
IN: Meets incentive regulation by receiving incentive-tier subsidy (yes = 1, no = 0)	0.9359	0.2452	0.0	1.0	359
CAP: Meets federal labour protection regulation (yes = 1, no = 0)	1.0000	0.0000	1.0	1.0	539
FA: Meets bus useful-life regulation (yes = 1, no = 0)	0.3928	0.4890	0.0	1.0	359
DOPT: Meets regulation to purchase some passenger service from private companies (yes = 1, no = 0)	0.0613	0.2402	0.0	1.0	359
Service Delivery Method					
DP: Directly provides service (yes = 1, no = 0)	0.3955	0.4896	0.0	1.0	359
PT: Purchases entire service from private sector (yes = 1, no = 0)	0.5432	0.4988	0.0	1.0	359
Ownership Type					
MPO: MPO-owned (yes = 1, no = 0)	0.3426	0.4752	0.0	1.0	359
CITY: City-owned (yes = 1, no = 0)	0.6072	0.4890	0.0	1.0	359
Organisational Size					
$\ln(RW_{nt}: Milesofright - of - way)$	5.7666	0.7913	4.0943	8.2699	348
Others					
Local dedicated subsidy (yes = 1, no = 0)	0.4095	0.4924	0	1	359
$\ln(SQ_{nt}: Serviceareinsquaremiles)$	4.9624	0.9780	3.0445	7.4593	359
$\ln(DEN_{nt}: Populationdensity)$	7.6729	0.3205	7.0148	8.8633	359

price term (-0.0202), the second order terms of output (0.0062) and network size (-0.0219), and the output-network interacting term (-0.0183). Additionally, the coefficient of time is statistically insignificant suggesting no technical change in the transit systems studied during the period considered.

Turning our attention now to technical inefficiency, the variance parameters $\lambda$and $\sigma_u$of the compound error and the parameter of the time-varying inefficiency $\eta$are all statistically significant. The latter result being negative shows a statistically significant decline in technical efficiency over time in the transit systems studied. Further, the results show a mean technical inefficiency of 0.74 with a standard deviation of 0.12. Since we assume no allocative distortion, this technical efficiency is also cost efficiency. While this is noteworthy, the sources of technical inefficiency are useful in identifying the factors which increase or decrease it. The information near the bottom of the table shows these sources. A statistically significant negative coefficient of a variable shows it is a source of a decrease in technical inefficiency, and a statistically significant positive coefficient shows it is a source of an increase in technical inefficiency.

Network size: From the table, a source of increase in technical inefficiency is network size in terms of miles of right-of-way whose coefficient is 0.1338 and statistically significant at < 0.0005 probability level. This implies that an increase in a transit system's network size increases technical inefficiency, or that technical inefficiency is larger in large transit systems compared to small ones. A possible reason for this finding is increases in organisational complexity as system size increases which lead to lower productivity and higher cost. Contrary to this finding, we noted above that the first order coefficient of network size is -0.0397 and statistically weak ($p = 0.068$) in the deterministic part of the cost equation. Combining that finding with that regarding network size in the inefficiency equation, it cannot be said that the higher technical inefficiency in transit systems with extensive networks that we found translates into higher cost. For there may be some cost advantages of transit systems with larger networks which we have not captured in this study. Furthermore, we find a statistically significant coefficient of population density (0.4319, $p < 0.0001$) in the cost equation, indicating that an increase in population density increases cost less proportionately and providing an advantage to transit systems operating in areas with high densities.

Regulations and Service Delivery Method: For the regulation variables, the coefficient of the bus useful life regulation is positive (0.11) and statistically significant at the 0.088 probability level. Though statistically weak, this finding suggests a higher level of technical inefficiency in transit systems that meet the bus useful-life regulation. It is found from the results also that meeting the spare-bus ratio regulation reduces technical inefficiency but this effect is not statistically significant. As regards the contracting regulation, one of its objectives is to control cost and allow transit systems to shed the portions of their services such as specialized transportation which they deem themselves incapable of providing efficiently to private sector companies. If successful, then contracting-out some passenger services to private transit systems should reduce technical inefficiency and reduce cost. In Table 2, with a coefficient of 0.29 whose level of statistical significance is 0.0000, technical inefficiency is higher in transit systems that meet the contracting regulation by providing some passenger services themselves and contracting-out others to private sector companies. Though the reverse may be expected, it is the presence of technical inefficiency in these transit systems that drives the regulation for them to contract-out portions of their services to private sector companies in the first place. Notwithstanding, another explanation for this finding is that because such transit systems receive federal capital subsidy, they cannot reduce their levels of employment without violating the labour protection clause and so are left with many employees to produce less output when services are contracted out to private sector companies. Similarly, if a contractor must provide own vehicles, the contracting agency is left with many buses and employees to produce a less amount of output which also leads to technical inefficiency. Thus, our finding reflects the fact that contracting out some services while complying with restrictive input level regulations increases technical inefficiency.

Type of ownership: In addition to the above results, the table shows that the coefficient of MPO-owned transit systems in the technical inefficiency equation is negative (-0.31) and statistically significant ($p = 0.0499$). This suggests that technical inefficiency is lower in MPO-owned transit systems than in other forms of ownership, or that they are more efficient compared to other types of transit ownership. Because MPO-owned transit systems account for 34.5% of the transit systems studied, these findings suggest that technical inefficiency is higher in the remaining 65.5%. These latter transit systems are operated by cities, agencies, universities and other private companies, some with more and others with less resources that may be less concerned about efficiently

Table 2
Stochastic Cost Frontier Model.

Log likelihood function			374.62	
Estimation based on			N = 359, K = 27	
Inf. Cr.			AIC = -691.7, AIC/N = -1.94	
Stochastic frontier based on panel data (77 transit systems)				
σ_v^2			0.0034	
σ_v			0.0580	
σ_u^2			0.0243	
σ_u			0.1558	
$\sigma = \sqrt{\sigma_v^2 + \sigma_u^2}$			0.1663	
$\Gamma = \sigma_u^2/\sigma_v^2$			0.8784	
$\sigma_u^2/(\sigma_u^2 + \sigma_v^2)$			0.7241	
Battese-Coelli Models: Time Varyingu_{it}				
Time dependent u_{it} = exp[-η(t-T)]*|U(i)|				
	Coefficient	Standard Error	z	Prob. |z| > Z*
Deterministic Component of Stochastic Frontier Model				
Constant	−0.3043*	0.0514	−5.92	0.0000
$\ln(w_{n3t}) - \ln(w_{n1t})$	0.2009*	0.0200	10.04	0.0000
$\ln(w_{n2t}) - \ln(w_{n1t})$	0.1423*	0.0290	4.91	0.0000
lnQ_{nt}	0.7854*	0.0221	33.48	0.0000
$0.5 \times [\ln(w_{n2t}) - \ln(w_{n1t})]^2$	−0.0801*	0.0256	−3.28	0.0010
$[\ln(w_{n2t}) - \ln(w_{n1t})] \times [\ln(w_{n3t}) - \ln(w_{n1t})]$	0.0830**	0.0388	2.14	0.0326
$[\ln(w_{n2t}) - \ln(w_{n1t})] \times lnQ_{nt}$	−0.1417*	0.0491	−2.89	0.0039
$0.5 \times [\ln(w_{n3t}) - \ln(w_{n1t})]^2$	−0.0202	0.0215	−0.94	0.3471
$[\ln(w_{n3t}) - \ln(w_{n1t})] \times lnQ_{nt}$	0.1033*	0.0302	3.32	0.0009
$0.5 \times [\ln(Q_{nt})]^2$	0.0062	0.0242	0.25	0.7997
$\ln(RW_{nt})$	−0.0397***	0.0217	−1.83	0.0680
$[\ln(w_{n2t}) - \ln(w_{n1t})] \times \ln(RW_{nt})$	0.1296**	0.0580	2.50	0.0124
$[\ln(w_{n3t}) - \ln(w_{n1t})] \times \ln(RW_{nt})$	−0.1411*	0.0389	−3.63	0.0003
$\ln(Q_{nt}) \times \ln(RW_{nt})$	−0.0183	0.0407	−0.45	0.6539
$0.5 \times [\ln(RW_{nt})]^2$	−0.0058	0.0447	−0.13	0.8975
$\ln(SQ_{nt})$	0.0982*	0.0180	5.45	0.0000
$\ln(PDEN_{nt})$	0.4319*	0.0397	10.88	0.0000
t_{2008}	−0.0092	0.0166	−0.55	0.5808
Offset [mean = μ_i] parameters in one sided error				
Constant	0.2946*	0.0670	4.39	0.0000
Meets bus useful life regulation (yes = 1, no = 0)	0.1055***	0.0619	1.71	0.0880
Purchases some services (yes = 1, no = 0)	0.2655*	0.0935	2.84	0.0045
MPO owned (yes = 1, no = 0)	−0.3090**	0.1576	−1.96	0.0499
Meets spare bus ratio regulation (yes = 1, no = 0)	−0.0255	0.0679	−0.38	0.7075
$\ln(RW_{nt})$	0.1338*	0.0382	3.50	0.0005
Variance parameters for compound error				
λ	2.6878*	0.0568	47.34	0.0000
o_u	0.1558*	0.0062	250.88	0.0000
ηparameter for time varying inefficiency				
Constant	−0.0394*	0.0125	−3.15	0.0016
	Mean	Std. Deviation	Minimum	Maximum
Technical Efficiency	0.7402	0.1244	0.3562	0.9804

Notes: * p < 0.01, ** p < 0.05, *** p < 0.10.

managing their limited resources, than being effective in providing services to their populations.

5. Conclusion

This paper's objective was to understand the factors affecting public transit cost and technical inefficiency by parameterising the latter in a stochastic cost frontier model in terms of variables expected to affect it and including some of these variables in the deterministic cost. A translog cost frontier which assumed the one-sided inefficiency term followed a truncated normal distribution was then specified and estimated using an unbalanced panel consisting of 77 US transit systems and 359 observations distributed variously across five years. The results showed 0.74 technical efficiency on the average and an increase in network size in terms of route miles increasing technical inefficiency. This means that transit systems with extensive networks have higher levels of technical inefficiency when compared to smaller transit systems, a finding inconsistent with that of JRB.

Also, the results showed that meeting the spare-bus ratio regulation had no effect on technical inefficiency but meeting the bus-useful life regulation increased technical inefficiency. Further, we found that transit systems that contracted out some of their passenger services to private sector companies while providing others themselves had higher levels of technical inefficiency. While this finding is inconsistent with what we expected and counterintuitive, it is the presence of technical inefficiency and some services falling outside the core competencies of some transit systems that drive contracting-out efforts. Thus, the services which transit systems provide themselves are those they have competencies providing, and those they contract out are the ones outside their core competencies which they perceive contractors as having the capability and expertise to provide at a lower cost and at a level of service the bus transit systems desire. That our finding contradicts this expectation suggests that there must be systemic inefficiencies that cannot be addressed with partial contracting.

The paper also examined the effects of different types of public transit ownership on technical inefficiency, finding that being MPO-owned reduced technical inefficiency. That is, such transit systems have lower levels of technical inefficiency when compared to other types of ownership such as being city-, privately- or university-owned. This shows an advantage of MPO owned transit systems and may be explained by their abilities to leverage their planning resources and expertise to select and design better routes responsive to passenger demand as well as being larger operations with large budgets which allow them to use specialised resources.[9] Moreover, MPO-owned transit systems cover wider areas traversing municipal boundaries and generally have longer routes so this finding shows that they are able to use their existing resources to produce large outputs measured in vehicle miles.

Comparing the long run cost elasticity of network size, which is its first order coefficient, to the effect of network size on technical inefficiency, no consistent pattern emerges; that is, network size has different effects on technical inefficiency and cost. This inconsistency does not allow us to make strong inferences about how this variable affects cost and technical inefficiency. If the finding of network size increasing technical inefficiency is what to expect, then there may be cost-increasing effects of this variable which this study did not capture. Also, this finding suggests that making inferences that because a variable reduces (increases) technical inefficiency it would reduce (increase) cost would be erroneous and policies based on such inferences misdirected since we found an opposite effect, which may be true for other variables we did not examine. It also suggests caution when deciding which variables to include in the deterministic part of the frontier and which to include in the parameterisation of technical inefficiency. Where there are inconsistencies as we found for network size, theory- and empirical-driven decisions should guide which to include in each of them. Certainly, if the coefficient of a variable is statistically significant in say cost and not in technical inefficiency, then it seems reasonable to remove it from the latter and retain it in the former. We would like to extend this caution to all variables used in parameterising technical inefficiency whose effects on cost have not been established in previous research or through conceptualisation.

Managerial Implications: The above results have managerial implications. For example, the finding that being an MPO owned transit system reduces technical inefficiency means that such a transit system uses fewer inputs to produce a given level of output or that it is more technically efficient relative to other transit systems. However, given the range of technical efficiency for the transit systems studied (0.3562 to 0.9804) there is not a transit system which is technically efficient among them. Thus, all will benefit from actions which increase technical efficiency such as reducing inputs or increasing output. While increasing output appears to be a simpler strategy, it must be done carefully by considering demand; that is, management increasing passenger services only when justified by high demand potential.

If the input reduction approach is used some strategies for management to consider would include longer headways on routes with less demand and during off-peak periods, using attrition and retirement to reduce labour inputs, larger and longer vehicles to reduce fleet size and labour employment, and energy efficient vehicles to reduce fuel use. Because the study also found that transit systems with extensive route networks have higher technical inefficiency, it suggests that they do not use their resources optimally and should consider the same suggestions above to reduce their inputs or increase output. Doing so would reduce their costs more than that reported in this paper. Another consideration for management is whole service contracting of passenger service to meet the federal contracting-out regulation. This is because we found higher levels of technical inefficiency in transit systems that used the mixed service delivery system by which they contracted-out portions of their services and provided others themselves with higher levels of inputs they were required to keep to meet federal regulations regarding transit inputs especially labour and capital.

Limitations: As instructive as perhaps the above results may appear, there are limitations of this study worth considering. One is the rather short panel (5 years) used in this analysis and another is the use of unbalanced panel data. While the unbalanced panel allowed us to obtain cross-firm data variation the short panel limited within-firm data variation thus preventing us from using fixed and random effects models. Future work that uses a longer panel should provide results that can be compared to those in this study for them to be generalized.

6. Author's Contribution

I attest that I am the sole author of this paper and wrote it myself.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abrate G., Erbette F. and Fraquelli G. 2008. Cost inefficiency or just heterogeneity? An application of stochastic frontier models to the Italian water industry. Higher Education and Research on Mobility Regulation and the Economics of Local Services, Working Paper # 3.

Agarwal, S., Yadav, S.P., Singh, S.P., 2009. Total factor productivity growth in the state road transport undertakings of India: an assessment through MPI approach. Indian Economic Review 44 (2), 203–223.

Ahn, S.C., Lee, Y.H., Schmidt, P., 2007. Stochastic frontier models with multi time-varying individual effects. J. Prod. Anal. 27, 1–12.

American Public Transit Association. 2011. Public transportation fact book, 61st Edition. American Public Transit Association, Washington D.C.

American Public Transit Association. 2010. Public transportation fact book, 60th Edition. American Public Transit Association, Washington D.C.

American Public Transit Association. 2009. Public transportation fact book, 59th Edition. American Public Transit Association, Washington D.C.

Association, American Public Transit, 2008. Public transportation fact book, 58th Edition. American Public Transit Association, Washington D.C.

American Public Transit Association. 2007. Public transportation fact book, 57th Edition. American Public Transit Association, Washington D.C.

Amsler, C., Schmidt, P., Tsay, W., 2014. A post-truncation parameterization of truncated normal technical efficiency. J. Prod. Anal. 44 (2), DOI. https://doi.org/10.1007/s11123-014-0409-8.

Battese, G.E., Coelli, T.J., 1995. A model of technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.

Caudill, S.B., Ford, J.M., Gropper, D.M., 1995. Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. Journal of Business and Economic Statistics 13 (1), 105–111.

Chen, Y., Schmidt, P., Wang, H., 2014. Consistent estimation of the fixed effects stochastic frontier model. (2014). Journal of Econometrics 181, 65–76.

Colombi, R., Kumbhakar, S.C., Martini, G., Vittadini, G., 2014. Closed-skew normality in stochastic frontier with individual and long/ short-run efficiency. J. Prod. Anal. 42, 123–136.

Cornwell, C., Schmidt, P., Sickles, R., 1990. Production frontiers with cross-sectional and time series variation in efficiency level. Journal of Econometrics 46, 185–200.

Cuesta, R., 2000. A production model with firm-specific temporal variation in technical inefficiency: With application to Spanish dairy farms. J. Prod. Anal. 13, 139–158.

Farsi, M., Filippini, M., Kuenzle, M., 2006. Journal of Transport Economics and Policy 40 (1), 95–118.

Greene, W., 2003. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. Stern School of Business, New York University, New York, Department of Economics https://archive.nyu.edu/jspui/bitstream/26162/2/2-10.pdf.

Greene W. 2007. IMDEP version 9.0, Econometric Modeling Guide vol. 2. Econometric Software Inc., Plainview NY.

Greene, W., 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126, 269–303.

Holmgren, J., 2013. The efficiency of public transit operation – An evaluation using stochastic frontier analysis. Research in Transportation Economics 39, 50–57.

Huang, C.J., Liu, J.T., 1994. Estimation of non-neutral stochastic frontier production function. J. Prod. Anal. 5, 171–180.

Jarboui, S., Forget, P., Boujelbene, Y., 2014a. Inefficiency of public road transport and internal corporate governance. Case Studies on Transport Policy 2 (3), 153–167.

Jarboui, S., Forget, P., Boujelbene, Y., 2013. Public road transport efficiency: a stochastic frontier analysis. Journal of Transport Systems Engineering 17 (5), 64–71.

[9] The data show that the logarithm of total cost for an MPO-owned transit system is 16.37 compared to 15.59 of city-owned transit systems. It also shows the logarithms of output of 15.37 for MPO- and 14.61 for City-owned, network sizes of 6.06 and 5.60 for MPO- and City-owned transit systems respectively.

Jarboui, S., Forget, P., Boujelbene, Y., 2014b. Transport firms' inefficiency and managerial optimism: a stochastic frontier analysis. Journal of Behavioral and Experimental Finance 3, 41–51.

Jha, R., Singh, S.K., 2000. Small is efficient: a frontier approach to cost inefficiencies in Indian state road transport undertakings. International Journal of Transport Economics 28 (1), 95–114.

Kerstens, K., 1999. Decomposing technical efficiency and effectiveness in French urban transport. Annales d'economie et de Statistique 54 (54), 129–155.

Kumbhakar, S.C., Lien, G., Hardaker, J.B., 2012. Technical efficiency in competing data models: a study of Norwegian grain farming. J. Prod. Anal. https://doi.org/10.1007/s11123-012-0303-1.

Lee, Y.H., Schmidt, P., 1993. A production frontier model with flexible temporal variation in technical inefficiency. In: Fried, H., Lowell, C.A.K., Schmidt, S. (Eds.), The measurement of productive efficiency: techniques and applications. Oxford University Press, Oxford.

Margari, B.B., Erbetta, F., Petraglia, C., Piacenza, M., 2007. Regulatory and environmental effects on public transit efficiency: a mixed DEA-SFA approach. J. Regul. Econ. 32 (2), 131–151.

Martini, G., Scotti, D., Viola, D., Vittadini, G., 2020. Persistent and temporary inefficiency in airport cost function: An application to Italy. Transp. Res. Part A 132, 999–1019.

Matas, A., Raymond, J., 1998. Technical characteristics and efficiency of urban bus companies: The case of Spain. Transportation 25, 243–263.

Nolan, J.F., 1996. Determinants of productive efficiency in urban transit. Logistics and Transportation Review 32 (3), 319–342.

Obeng, K., 2013. Bus transit technical inefficiency using latent class indirect production frontier. Appl. Econ. 42 (928), 3933–3942.

Obeng, K., Azam, G., 1995. The intended relationship between federal operating subsidies and cost. Public Finance Quarterly 23 (1), 72–94.

Obeng, K., Sakano, R., Naanwaab, C., 2016. Understanding overall output efficiency in public transit systems: the roles of input regulations, perceived budget and input subsidies. Transp. Res. Part E 89, 137–150.

Schmidt, P., 1986. Frontier production function. Econometric Review 4, 289–328.

Singh, S.K., Vanketesh, A., 2003. Comparing efficiency across state transport undertakings: a production frontier approach. Indian Journal of Transport Management 27 (3), 374–391.

Vigren, A., 2016. Cost efficiency in Swedish public transport. Research in Transportation Economics 59, 123–132.

Wang, H.J., Ho, C., 2010. Estimating fixed effect panel stochastic frontier by model transformation. Journal of Econometrics 157 (2), 286–296.

Zhang, C., Juan, Z., Xiao, G., 2015. 20 contractual practices affect technical efficiency? Evidence from public transport operators in China. Transp. Res. Part E 80, 39–55.