Transportation (2011) 38:191–214
DOI 10.1007/s11116-010-9296-7

Indirect production function and the output effect of public transit subsidies

K. Obeng

Abstract This paper uses an indirect production function to decompose the effects of subsidies on output into the lump-sum, cost and inefficiency effects. Using 2006 data for U.S. transit systems it estimates an indirect production function and uses the results to calculate these effects. It finds that the lump-sum effects exceed the other effects and that the average total effect of the subsidies is a 4.72% increase in output. The range of the output change shows that in many transit systems the output increases from the subsidies are quite large. The paper suggests that reductions in allocative inefficiencies from the subsidies would result in very large increases in output.

Keywords Operating subsidy · Capital subsidy · Output effect · Indirect production function · Inefficiency

Introduction

Previous research has shown that when operating and capital subsidies are offered they create allocative inefficiencies by distorting the optimal rate of input substitution (Obeng et al. 1997). These allocative inefficiencies assume output remains unchanged. It can be argued that this assumption is unjustified because Mohring (1972) and Pederson (2003) show that based on user cost economies of scale the subsidies increase service frequencies and output. According to van Reeven (2008), this increase could make service frequency higher than its socially optimum level. If so, then the subsidies could lead to oversupply of services, and the increase in output from the higher frequencies would be due to inefficient use of resources. On the other hand Small and Gomez-Ibanez (1999) argue that the subsidies could lead to inefficiencies and reduced productivity, which if true could lead to lower levels of output. Thus, there appears to be two counteracting effects of operating and

---

A longer version of this paper is available on request to the author at obengk@ncat.edu.

K. Obeng (✉)
Department of Marketing, Transportation and Supply Chain, School of Business and Economics,
North Carolina A&T State University, Greensboro, NC 27411, USA
e-mail: obengk@ncat.edu

capital subsidies worthy of further investigation. On one hand, the user cost economies of scale argument shows that output increases with these subsidies through increased service frequency. On the other, the allocative inefficiency and reduced productivity arguments, show possible reductions in output from the subsidies. If these effects exactly offset each other then the subsidies are used to maintain existing services and do not increase output. There are no studies that attempt to bring these two effects together in the public transit economics literature. However, there are many studies that examine the impacts of the subsidies on cost and inefficiency, for example, Kim (1987), Kerstens (1996), Nolan (1996) and Karlaftis and McCarthy (2002).

While providing useful information this focus on cost can be critiqued on several fronts. First, if subsidies increase output then they must increase total cost, since it requires more inputs to produce the additional output unless there is a gain in productivity or technological improvement from the subsidies that makes it cheaper to produce each level of output. Second, since both cost and output increase we should be examining the relationship between subsidies and average cost not total cost. Alternatively, the focus should be on comparing the increase in cost from the subsidies to the increase in output. Third, focusing only on cost, past research completely ignores the possible effects of these subsidies on output by changing input prices. As developments in the public transit economics literature show, operating and capital subsidies make transit systems misperceive input prices as lower thereby making them employ more inputs than they would do otherwise (Obeng et al. 1997). In turn, this change in inputs increases the amount of output produced, a result consistent with what Cervero (1986) and Bly and Oldfield (1986) found. Therefore, the effects of the subsidies on output cannot be completely ignored, unless it is assumed they only support existing but not expanded services. Since this assumption raises empirical questions and cannot be supported in practice, ignoring the output effects of the subsidies leaves a void in the transit economics literature that requires examination.

To fill this void this paper determines the impact of operating and capital subsidies on output. It surveys the literature on public transit objectives and follows it with an indirect production function to decompose the effects of operating and capital subsidies on output into lump-sum, cost and their interaction effects. This decomposition is unique to this paper and this is the first estimation of indirect production function using public transit data. Finally, to illustrate the usefulness of the decomposition the paper specifies an empirical model and estimates it with 2006 data for U.S. bus transit systems. Using the results it calculates the proportions of output due to the lump-sum, cost and their interaction effects and sums the results to obtain the total effects of operating and capital subsidies on output. It finds that the cost and interaction effects consistently reduce output while the lump-sum effects increase output. The combined effect of these sources of output change is a 4.72% increase in output on the average. Thus, the positive lump-sum output effects of the subsidies exceed their negative cost effects. Following the results are the policy implication and conclusions respectively.

Literature review

Conceptual difficulties

Some possible reasons for the absence of focus on the output effects of operating and capital subsidies in the public transit literature are the conceptual difficulties which hinder the estimation of production functions. They include ambiguities about whether what is produced in transit systems is an intermediate or a final output (Small 1990). For example, is a vehicle mile intermediate or a final good? Those in favor of it being a final good argue that it is the output actually produced by transit systems, while others including Small (1990) argue that it is intermediate in a passenger's use of public transit service. While we do not contribute to this debate we note that all transportation outputs are heterogeneous and involve some degree of aggregation of services and trips of different qualities. This heterogeneity introduces difficulties into the estimation of production functions because standard production theory assumes aggregation of homogeneous outputs (Hanushek 1979) and some variation in inputs. However, the relatively fixed ratio between drivers and buses in some transit systems implies that for any given schedule these inputs would explain little variation in output resulting in coefficients that may be statistically insignificant. If there is no variation in inputs or input ratios, then production functions cannot be estimated to determine the output effects of public transit subsidies.

Additionally, the existence of multiple outputs in transportation such as vehicle miles, vehicles hours, passengers and passenger miles poses conceptual problems by introducing some bias into the choice of output measure as there is no consensus on a rule to use to make that choice. This lack of consensus is evident in the US National Transit Database which lists two demand and five supply measures of output (Federal Transit Administration 1998), and in a meta analysis of 33 studies on public transit efficiency by Brons et al. (2005) where 31.2, 19.4, 52.7 and 16.1% of them used output indicators that were related to passengers, seats, vehicles and revenues respectively. Even in cost studies where it may seem obvious to use service produced or what Oum and Yu (1994) call available output, there is still no consensus about whether to use vehicle miles, vehicle hours, capacity miles or seat miles. However, there is some understanding that the choice of output should be based upon the objective of the study one is conducting. For example, Oum and Yu (1994) suggest using demand-related output measures (i.e., revenue output measures) in public policy oriented studies if there is no government control over service frequency and service levels, or if that control is inconsequential. This is because government regulations on fares and level of service distort passenger demand. When these controls exist they suggest using intermediate measures of output such as vehicle miles because they will correctly reflect the service actually supplied, and it is possible to isolate the effects of regulation from them. Their results showed that the choice of output mattered in the levels and rankings of railroads based upon efficiency.

Even if there is an agreement to use available outputs such as vehicle miles, seat miles or vehicle hours in production studies, problems still arise when fleet size is the measure of capital and an exogenous variable in the production function equation. The problem is that these output measures are all products of fleet size and other variables, thus making fleet size appear on both sides of the equation. Absent other meaningful measures of capital, this problem does not favor the estimation of production functions for public transit systems.

Objectives of transit systems

Besides the conceptual difficulties above, another reason for this absence of focus is that there are many objectives of public transit systems with no consensus on which to use to model decision making (De Borger et al. 2002). The result is that a varied array of public transit objectives can be found in the transit economics literature many of which yield different outcomes when modeled explicitly. Berechman (1993) groups them into: (1) political where resource allocation is based upon political processes, (2) managerial where firms try to be efficient based upon cost per output, (3) bureaucratic in which firms try to maximize output or net earnings and (4) cost minimization. He evaluates how each applies to public transit and argues against the political and managerial objectives and notes that cost minimization is also not the main objective of transit systems. In arguing against political objectives he examines a study by Cooter and Topakian (1980) which tests and rejects the hypothesis that BART's objectives are political in the sense that its prices are set through competition by its board to win political support. For managerial objectives Berechman (1993) rejects them because they are normative and prescribe measures of efficiency, and using them does not lead to being effective. Moreover, these managerial efficiency measures do not provide information about the technology a transit system uses. Next, he examines bureaucratic objectives in terms of budget surplus maximization and argues against them too because they do not lead to optimal levels of output. He postulates that it is likely managers allow their costs to increase to meet their budgets.

A further twist about which objective to use is in the literature on technical efficiency in public transit systems. That literature shows an emerging trend favoring output maximization and input minimization as empirical objectives in public transit studies (e.g., Nolan et al. 2002). De Borger et al. (2002) explain this pattern as due to the availability of well-established methods of analyzing productivity and efficiency and the foremost objective of any public sector organization to be technically efficient (Pestieau and Tulkens 1993). Other studies suggest that transit systems pursue a modified cost minimization objective such as after-subsidy cost minimization, which could lead to after-subsidy profit (Obeng 2000), or that they pursue social objectives. Savage (2004) surmises that "managers may be motivated to maximize social welfare, number of passengers, or the amount of service provided." He notes that most empirical analyses suggest that transit systems maximize level of service. In support he cites Glaister and Collins (1978) whose work found evidence from Sydney, Australia and London, U.K. to show that indeed output maximization is pursued by some transit systems.

Besides these single objectives, Fabbri (1996) notes that transit systems may have "unconventional objectives and therefore non-standard behavioral programs." Some of these objectives are making their services universally available, serving a diversified population base and ensuring equity in service provisions, and providing high quality and environmentally friendly services. Often, these objectives are imposed by various levels of government, private agencies and political entities such as city councils as a part of their funding requirements. These varied and sometimes conflicting objectives make it difficult to develop a single empirical model to analyze the behavior of transit systems. For example, the US Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) included many social objectives for transit systems to accomplish only three of which could be modeled directly when Nolan et al. (2002) attempted to do so due to data limitations. These are pollution, safety and the provision of necessary public services.

Cost minimization

Despite these various objectives and their limitations as pointed out by Berechman (1993), most recent research show that cost minimization is a favored empirical objective in the public transit economics literature. As in the private unregulated sector, this objective assumes substitution possibilities between the inputs required to produce the same level of output but not how changes in input prices affect output. The vast literature on transit cost functions beginning with Viton (1980) and numerous subsequent works such as Karlaftis and McCarthy (2002) and Obeng and Sakano (2009) attests to this objective's popularity. In its use, cost is a function of competitive input prices, output, firm and environmental characteristics. Among its advantages, it minimizes some of the conceptual problems noted earlier such as fixed input proportions since we no longer must use the physical quantities of inputs but their prices. In addition, it overcomes endogeneity problems because output is now an exogenous variable. Another advantage is that when it is formulated as a frontier it permits the estimation of allocative and technical inefficiencies, whereas only technical inefficiency can be estimated in single equation production frontiers. And when cost functions include subsidies as variables their signs allow inferences to be made about their impacts on cost but not on output.

The cost function approach has clearly enriched our knowledge about transit system economics. It has allowed researchers to answer questions about returns to scale, input demand elasticities, input substitution, inefficiencies and total factor productivity among others. Yet, they have been critiqued on the grounds that they do not allow (1) the effects of input price changes on output to be studied because firms minimize costs for given levels of output, (2) the impact of total expenditures on output to be determined and, (3) cost may not be at a minimum level. According to Berechman (1993) some earlier cost functions tended to be very simplistic. This is because their exogenous variables were only input prices and output and did not include external variables (reflecting differences in urban area spatial structure and population density) and those about a transit system's decision environments (regulation, demand and subsidy formula) all of which affect resource allocation. Another limitation is that the input demand elasticities derived from cost functions are output constant input elasticities of demand instead of Marshallian demand elasticities needed for policies. Garofalo and Malhotra (1990) write that the main drawback of demand elasticities from cost functions is their assumption that a change in input use occurs only through substitution effects, thus ignoring the output effect. The implication is that the values of these elasticities are smaller and lead to underestimation when used in policies. These limitations, of course, do not negate the results from cost functions but suggest finding ways to enhance the information they provide.

Output maximization

In an attempt to do so, indirect production function has been suggested as suitable in public sector studies instead of cost functions. This approach builds upon the fact that an indirect production function is the dual of the production function and provides the same information about technology as do cost functions and more. In fact, it is obtained by solving a minimum cost function for output and does not result in any loss of information about cost. Thus, in this function, output is endogenous and cost and input prices are exogenous and this clearly avoids many of the conceptual problems discussed. A similar function has been used by Chambers (1982) to derive the output effect of a change in factor prices. He showed that if the indirect production function is homothetic the output effect of a change in the price of an input is the negative of its corresponding optimal share in cost.

In instances where cost minimization may not reflect firm behavior, output maximization may be an alternative objective. As Fare et al. (1988) note, output maximization is most "appropriate to producer performance evaluation when resource usage can be reliably compared on the basis of cost but benefits in terms of outputs produced or services provided cannot be priced reliably enough to allow revenue comparisons" (p. 73). They continue that it is most appropriate where regulation prevails, or where actual prices may not be observed or are not exogenous. For example, although there is readily available information about the passenger revenues of US public transit systems, that information does not reflect the value of service because fares are subsidized. Shephard (1973) adds that output maximizatiom is most appropriate in public and service sector organizations where decision makers are concerned with how much benefit in terms of output would be obtained for given levels of expenditures.

Though both cost minimization and output maximization have their merits, the use of the latter as an objective and the estimation of an indirect production function instead of a cost function rest on some advantages besides the aforementioned. Kim (1987) and Hilmer and Holt (2005) discuss these advantages including the fact that they allow us to determine (1) the impacts of changes in input prices on output, (2) the effects of changes in budget on output and (3) they avoid simultaneity problems in estimating single equation production functions. Also, as we have noted, the elasticities of demand from estimating an indirect production function are not as restrictive as those from cost functions. And, being the dual of the production function it offers a way to estimate its parameters. However, it can be critiqued because it could lead to actual output levels which are not Pareto optimal (Nash 1978) due to production inefficiencies, and it leaves unanswered questions about whether indeed transit output is endogenous or how to deal with outputs such as vehicle miles and passenger miles which are linked because passenger demand is a function of service level.

These critiques notwithstanding, this paper assumes output maximization. It shows through derivation that the indirect production function is flexible enough to permit a decomposition of the effects of operating and capital subsidies on output into (i) the lump-sum effect from the subsidies received being dependent on output; (ii) the cost effects from the subsidies making transit systems misperceive their input prices and (iii) the interactions between these two effects. The first two of these output effects parallel Schmidt's (2001, p. 242) assertion that "Any subsidy program that makes the subsidy dependent on the amount of output the firm produces will give the firm an incentive to increase output by raising the firm's perceived marginal revenue, as will a subsidy on inputs that lowers perceived marginal cost." He calculates that because of this incentive US transit systems increase their outputs by 6–8% above what they would have produced without the subsidies but does not decompose this output change among its sources. The third, interaction effect, is an additional effect of allocative distortion on output.

In addition our use of the indirect production function is based upon the premise that US transit systems are often given subsidies to increase and improve their services and make them generally available to the populations they serve. The ISTEA, for example, lists the expansion of the consumer base of public transit systems as one of its objectives (Nolan et al. 2002), and the federal formula for allocating Section 5307 grants to transit systems is partially based on vehicle miles and passenger miles. Because these subsidies are based upon it, decision makers are interested in knowing how much they increase output. In the section below we add to the transit economics literature by deriving a decomposition formula to show that output maximization allows us to calculate these changes in output from offering public transit subsidies.

Derivation: output decomposition

Two types of public transit subsidies can be distinguished in the US; they are operating subsidies and capital subsidies. Operating subsidies account for the inability of public transit operating revenues, particularly fare revenues, to cover operating expenses. They cover the costs of labor, fuel, materials and supplies as well as vehicle maintenance. Capital subsidies are for equipment purchases, right-of-way acquisition, corridor development and the construction of new facilities, and labor to supervise these activities. These subsidies are offered by federal, state and local governments. At the federal level, operating subsidies are apportioned to transit systems based on a legislative formula. That formula allocates 9.32% of the Section 5307 funds to urbanized areas with populations between 50,000 and 200,000 based upon population (50%) and population times population density (50%). The rest, 90.68%, goes to transit systems in cities with populations of 200,000 or more. For transit systems in these latter cities, the bus tier is allocated to them based upon bus vehicle revenue miles (45.4%), population (22.7%) and population times density (22.7%). The rest, 9.2%, is an incentive tier allocated based upon passenger miles squared over operating cost.[1]

Given these percentages, the formula favors both output maximization and cost minimization. But, because the incentive tier of the subsidy is very small, cost minimization has a lesser impact on the amount of federal subsidy a transit system receives than the maximization of vehicle revenue miles and passenger miles. By rewarding output increases and penalizing high operating cost firms, however, the formula encourages intensive use of some inputs (Schmidt 2001) particularly overuse of capital relative to labor and fuel (Obeng and Azam 1995). Recently, this formula has been used to apportion US federal capital subsidies to transit systems under the American Recovery and Reinvestment Act of 2009.[2] Similar formulae for allocating subsidies are found across various states while cities do not have such a formula.[3]

Irrespective of the formulae used in their allocation, subsidies cover expenditures by being spent on inputs. Their disproportionate use on a particular input can cause allocative distortion and become a source of inefficiencies in public transit systems. For example, capital subsidies affect how transit management perceives its costs of capital inputs and operating subsidies affect how it perceives the costs of its non-capital inputs and this could cause allocative distortion. Lately, federal restrictions on the use of capital subsidies have been relaxed. Under both the Transportation Equity Act for the twenty-first century and the Consolidated Appropriations Act of 2005 transit systems operating in cities with 200,000 populations or more are allowed the flexibility to use federal capital subsidies to cover operating expenditures if they no longer receive federal operating subsidies.

Given this background consider a typical transit system's annual budgetary decision, which we conceptualize as a four-step process with a possibility that the first two steps could be reversed. In the first step the transit system determines its cost of production. Next, it estimates the subsidies it will receive from all sources and spends them on inputs with most going to the inputs with the largest shares in cost and targeted by the subsidies. Then, the transit system calculates its after-subsidy cost (i.e., cost net of the subsidies) and establishes the maximum B for it. This ensures that it has enough passenger revenue to pay for after-subsidy cost.[4] If this maximum is equal to passenger revenue the transit system makes zero after-subsidy profit; if it is less than passenger revenue it makes after-subsidy profit. And if it is more than passenger revenue more subsidies would be needed suggesting strict inequality as the binding constraint. A reason for after-subsidy profit could be that US transit systems have excessive subsidies particularly from dedicated local sources or they

---

[1] For the FY 2009 apportionment formula see, http://www.fta.dot.gov/documents/2009fullyear_-_Table_4_-_sec_5307__Apportioment_Formula.xls.

[2] The formula for capital subsidies can be found in http://www.fta.dot.gov/documents/ARRA_Table_3_-_sec_5307__Apportioment_Formula-HBS.xls.

[3] Even though we have discussed the federal subsidy allocation formula it is noteworthy that we are not modeling the formula but how the funds from the formula are used.

[4] It is possible to argue that the sequence of decisions is the reverse of that described herein. For example, the subsidy decision could be made first followed by a determination of the level of transit output level to be produced. Furthermore, the sequence may be different in contracted services where the transit system determines the service to be provided (output) and specifies the subsidy to be paid to the contractor. Despite these possible differences in sequencing, they do not alter the results in this paper.

pursue profit maximization as a goal, which is doubtful because many of them can hardly cover their costs and rely on subsidies. Yet, another and one underlying this paper is that they maximize output by providing more services to attract passengers. As long as the marginal cost of the additional service is less than the marginal revenue from it after-subsidy profits result. Therefore, in the final step of the process the transit system maximizes its production of services subject to the after-subsidy cost constraint, B.

This four-step process envisions input demand as a function of subsidies. However, it can be shown that the reverse is also true. Assume the transit system just described produces vehicle miles of service (Q) with labor (L), capital (K) in terms of vehicles, and all other inputs proxied by fuel (F). Thus, $Q = Q(L, K, F)$is the transit system's production function and it incurs total actual resource cost $C = w_L L + w_K K + w_F F$ and total operating cost of $C_o = w_L L + w_F F$ in producing this level of output where w_L, w_K, w_F are the respective market prices of the inputs. The operating subsidy (A_o) that this transit system receives is $A_o = w_L L + w_F F - pQ(L, K, F)$ where, p is the price per vehicle mile. From this, operating subsidy depends upon input levels. Similarly, based upon the US federal subsidy allocation formula previously discussed, Obeng (2010) shows that subsidies are functions of input levels and derives the functional form of the operating subsidy a US transit system receives as $A_o = h_o(L, K, F, D, M)$where M is urbanized area size in terms of square miles and D is population density.

A similar function can also be derived for capital subsidy. Here, the amount of capital subsidy that a US transit system receives depends, for example, upon how many vehicles are bought, with the federal government paying 80% of it and state and local governments paying the rest. And, as noted earlier, the Transportation Equity Act for the twenty-first century and the Consolidated Appropriations Act of 2005 both allow federal capital subsidies to be used to cover operating expenditures. As well, the American Recovery and Reinvestment Act of 2009 uses the federal operating subsidy allocation formula to distribute its capital subsidies. Based upon this information the capital subsidy function is also $A_{\hat{K}} = h_{\hat{K}}(L, K, F, D, M)$.[5]

These functions are expanded to include the sources of the subsidies, transit system and local area characteristics. Since population density and urbanized area size are very highly correlated the latter is deleted from the subsidy functions. More specifically, the operating subsidies and capital subsidies functions are specified as $A_o(L, K, F, N, D, Y_{\mathrm{UAF}})$ and $A_{\hat{K}}(L, K, F, D, Y_{\mathrm{GEN}}, Y_{\mathrm{CAP}}, Z)$ respectively where $\partial A/\partial F > 0$, $\partial A/\partial L > 0$ and $\partial A/\partial K > 0$, $\partial A/\partial(D) > 0$. Y_{UAF} and Y_{CAP} are respectively binary variables showing receipt of funds from the federal urban area formula grant and capital subsidy programs. Y_{GEN} is a binary variable showing receipt of funds from state and local general revenue sources, average vehicle age is z, and network size in terms of route miles is N.

The Lagrangian of this constrained optimization is,

$$\begin{aligned} \text{Max}\ Q(L, K, F) + \lambda\Big[\Big\{ & w_L L + w_K K + w_F F - A_{\hat{K}}(L, K, F, D, Y_{\mathrm{PM}}, Y_{\mathrm{GEN}}, Y_{\mathrm{CAP}}, z) \\ & - A_o(L, K, F, N, D, Y_{\mathrm{UAF}})\Big\} - B\Big] \end{aligned} \tag{1}$$

---

[5] For example, in buying buses, the amount of capital subsidies a transit system receives from federal sources depends upon the number bought. At the margin, this subsidy is $0.8(w_K K)$ after suppressing the recent changes that have been made to allow these subsidies to be used for non-capital purposes. Similarly, because the federal share in operating losses is 50% at the margin, the total federal operating subsidies a transit system receives is $0.5(pQ - w_L L - w_F F)$ where, pQ is fare revenue and p output price. In both cases the amounts of the subsidies clearly depend upon input levels.

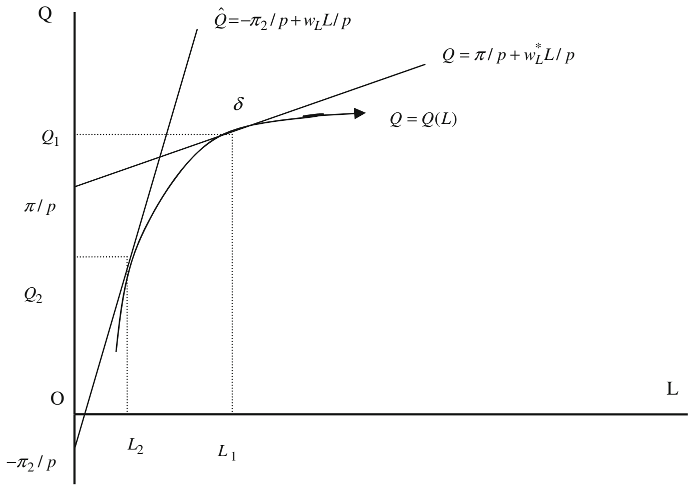

Fig. 1 Output, loss and after-subsidy profit

Where, λ is a Lagrangian multiplier associated with the constraint and the decision variables are L, K, F, λ. Figure 1 shows this optimization problem, and it is drawn by applying the approach in Kumbhakar and Bokusheva (2009). Here, $Q(L)$ is the production function, C^* the terms in braces in Eq. 1 or after-subsidy cost, and it is assumed that $B = pQ$. Since in Fig. 1 the production function depends only on labor, $C^* = w_L^* L$ where w_L^* is the after-subsidy wage (or implied wage) of labor and $w_L^* < w_L$. Also, we assume that the transit system earns after-subsidy profit $\pi_1 = p(Q) - w_L^* L$.[6] This assumption is consistent with the data in Obeng (2010) which shows real after-subsidy profit for US transit systems between 1995 and 2006 after accounting for operating subsidies. Solving this equation for output gives $Q = (\pi_1/p) + (C^*/p) = (\pi_1/p) + \left(w_L^*/p\right)L$ as the profit function. Using this function, output is maximized at δ in Fig. 1 with the transit system producing Q_1 with L_1 units of labor.

Without subsidies, US transit systems make losses. These losses make a transit system perceive its wage rate correctly as w_L, produce $\hat{Q}$ and incur a loss of $-\pi_2 = p\hat{Q} - w_L L$. Solving this equation gives a loss function $\hat{Q} = (w_L L/p) - (\pi_2/p)$ whose slope of w_L/p is steeper than w_L^*/p. With this function, the transit system maximizes output by producing Q_2 with L_2 units of labor. Comparing Q_1 toQ_2, the latter is smaller suggesting that indeed the potential to make a profit when a transit system receives these subsidies is an incentive enough for it to hire more inputs and increase its output.

To calculate this increase in output we first differentiate Eq. 1 with respect to input quantities to obtain first order conditions. From these conditions the ratio of the marginal products f_i, f_j of any input pair (i, j) is,

$$\frac{f_i}{f_j} = \frac{w_i\left(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki}\right)}{w_j\left(1-\mu_{oj}H_{oj}-\mu_{Kj}H_{Kj}\right)} = \frac{w_i^*}{w_K^*} = \sigma_{ij}\frac{w_i}{w_j} \quad \text{for } i = L,\ K,\ F \text{ and } i \neq j \tag{2}$$

Where, for any input x_i, the ratios of the subsidies to input cost are $H_{oi} = A_o/w_i x_i$, $H_{Ki} = A_{\hat{K}}/w_i x_i$ and $\sigma_{ij} = (1-\mu_{oi}H_{oi}-\mu_{Ki}H_{ki})/(1-\mu_{oj}H_{oj}-\mu_{Kj}H_{kj})$ measures allocative

---

[6] This assumption can also be after-subsidy loss.

distortion or allocative inefficiency between input pairs. If σ_{ij} is less than one then the subsidies make the perceived (or implied) price of input i very low relative to the perceived (or implied) price of input j resulting in the substitution of i for j; the reverse being true also. The result is that transit systems overuse a less productive input j relative to input i. Also in this equation μH is input subsidy as a share in input price and if it is zero for one input the subsidies cause distortions only in the other input.

From the dual of this constrained optimization problem, the after-subsidy minimum cost function can be written as $C^* = C^*\left(w_L^*, w_K^*, w_F^*, Q\right)$ and the after-subsidy total cost as $C^* = w_L^*L + w_K^*K + w_F^*F$. Substituting $w_i^* = w_i(1 - \mu_{oi}H_{oi} - \mu_{Ki}H_{Ki})$ into the after-subsidy total cost equation gives $C^* = \sum{}_i w_i x_i(1 - \mu_{oi}H_{oi} - \mu_{Ki}H_{Ki})$. Since S_i is each input's share in actual total cost, that is $S_i = w_i x_i/C$, the actual cost of each input is $CS_i = w_i x_i$ and its substitution into the after-subsidy total cost equation gives $C^* = C\sum{}_i S_i(1 - \mu_{oi}H_{oi} - \mu_{Ki}H_{Ki})$. This after-subsidy cost is also the implied cost of the transit system, and it is what influences a transit system's production plans.[7] Thus, though total resource cost is C, a firm receiving these subsidies behaves based upon the cost C^*.

Similar equations as the after-subsidy cost have been derived in the shadow pricing literature by Atkinson and Halvorsen (1986), De Borger (1993) and Kumbhakar (1997) to show relationships between shadow cost and actual cost. Because $\sum{}_i S_i(1 - \mu_{oi}H_{oi} - \mu_{Ki}H_{Ki})$ is less than one, it is the proportion of total resource cost transit systems misperceive to be their own and must pay with their own internally generated funds from passenger revenues, rentals, advertisements and investments. Further, because transit systems misperceive their costs as low they overuse some of their inputs resulting in overproduction and an increase in overall resource cost. This increase in output also results from the subsidies increasing service frequency (van Reeven 2008; Tistato 2007; Mohring 1972) and it is calculated in the next step using an indirect production function.

Using the duality between cost and production functions, if a transit system's minimum implied cost function is $C^*\left(w_L^*, w_K^*, w_F^*, Q\right)$ then under output maximization there exists an indirect production function $Q\left(w_L^*, w_K^*, w_F^*, C^*\right)$ which is the solution to solving for output at the cost minimization point.[8] This function is non-decreasing in implied cost C^*, non-increasing in implied input prices w_L^*, w_K^*, w_F^*, homogeneous of degree zero in C^* and w_i^* and quasi-convex in input prices, i.e. $\partial^2 C^*/\partial w_i^{*2} > 0$.[9] Assume a flexible technology of the translog type. Then, expanding $Q\left(w_L^*, w_K^*, w_F^*, C^*\right)$ up to the second order using Taylor's series gives the translog indirect production function,

$$\ln Q = \beta_0 + \sum\nolimits_i \beta_i\left(\ln C^* - \ln w_i^*\right) + 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}\left(\ln C^* - \ln w_i^*\right)\ln\left(\ln C^* - \ln w_j^*\right) \tag{3}$$

---

[7] These authors showed that this relationship is exact if the production function underlying the cost function is Cobb-Douglas. In another context, Kumbhakar (1997) generalized the relationship between implied and actual cost to situations where the cost function is translog.

[8] These implied prices can be obtained by maximizing output subject to a net cost constraint where net cost is total cost less the amounts of operating subsidies and capital subsidies expended. These subsidies are functions of all inputs. They can also be obtained by minimizing net cost subject to a production function constraint.

[9] If the implied cost function is Cobb-Douglas of the form $C^* = (1/\eta_0)^\theta w_L^{*\beta_L\theta} w_K^{*\beta_K\theta} w_F^{*\beta_F\theta} Q^\theta$ which is homogeneous of degree one in input prices implying that $\theta(\beta_L + \beta_K + \beta_F) = 1$ then the indirect production function is, $Q = \eta_0\left(C^*/w_L^*\right)^{\beta_L}\left(C^*/w_K^*\right)^{\beta_K}\left(C^*/w_F^*\right)^{\beta_F}$.

Substituting the implied cost $C^{*} = C\sum{}_{i}S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ and the implied input price $w_i^{*} = w_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ provided earlier into Eq. 3 and adding an error term ε, gives,

$$\begin{aligned}\ln Q &= \beta_0 + \sum\nolimits_i \beta_i(\ln C - \ln w_i + \ln U_i)\\ &\quad + 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}(\ln C - \ln w_i + \ln U_i)\left(\ln C - \ln w_j + \ln U_j\right) \;+\; \varepsilon.\\ &= \beta_0 + \sum\nolimits_i \beta_i(\ln C - \ln w_i + \ln U_i) + 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}(\ln C - \ln w_i)\left(\ln C - \ln w_j\right)\\ &\quad + \sum\nolimits_i \beta_i \ln U_i + \; 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}\left([\ln C - \ln w_i + \ln U_i]\ln[U_j]\right.\\ &\quad \left.+ \ln[\ln C - \ln w_i + \ln U_j]\ln[U_i] + \ln[U_j]\ln[U_i]\right) + \varepsilon.\end{aligned}$$

$$\text{where}\quad U_i = \left\{\sum\nolimits_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\right\}/(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki}) \tag{4}$$

U_i is the weighted change in the optimal rate of input substitution as a result of having the subsidies, where the weight is the actual share of an input in cost. Expanding Eq. 4 through substitution of U_i into the term $\sum_i\beta_i \ln U_i$ we have,

$$\begin{aligned}\ln(Q) &= \ln\hat{Q} - \sum_i \beta_i \ln(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\\ &\quad + \left(\sum_i \beta_i \ln\left[\sum\nolimits_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\right]\right)\\ &\quad + 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}\left([\ln C - \ln w_i]\ln[U_j] + \left[\ln C - \ln w_j\right]\ln[U_i] + \ln[U_j]\ln[U_i]\right) + \varepsilon.\end{aligned} \tag{5}$$

Where, $\ln(\hat{Q}) = \beta_0 + \sum_i \beta_i(\ln C - \ln w_i) + 0.5\sum_i\sum_j \beta_{ij}(\ln C - \ln w_i)\ln(\ln C - \ln w_i)$ and consistent with Fig. 1 it is the logarithm of the output $\hat{Q}$ that would have been produced had the transit systems not received the subsidies. In Eq. 5, because the second term $-\sum_i \beta_i \ln(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ does not affect input shares or input proportions it is the lump-sum effect of the subsidies on output by changing implied input prices. Further, because $(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ is positive and less than one, its logarithm is negative and this makes $-\sum_i \beta_i \ln(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ always positive. Therefore, operating and capital subsidies increase output through their lump-sum effects.

Comparatively $\sum_i \beta_i\{\ln\sum_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\}$ is the weighted effect of the cost impacts of the subsidies on output as a result of increased use of inputs. Specifically, $\sum_i \beta_i$ is the weight and $\ln\sum_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})$ shows how much the subsidies make minimum implied cost less than actual total cost. Thus $\sum_i \beta_i$ is a factor that converts the cost impacts of the subsidies into output impacts. The value of this third term is always negative and, again, it shows the opportunity cost of the subsidies in terms of forgone output.

The term $0.5\sum_i\sum_j \beta_{ij}([\ln C - \ln w_i][\ln U_j] + [\ln C - \ln w_j][\ln U_i] + \ln[U_j]\ln[U_i])$ adds to or reduces output. It is the interaction of the lump-sum and the opportunity cost effects of the subsidies. Because as we have noted U is the weighted change of the optimal rate of input substitution this term captures allocative distortion. It may also be considered a relic of the translog model because it cannot be obtained from a linear or a Cobb-Douglas model. Since both $(\ln C - \ln w_i)$ and $\ln(U_j)$ are positive, the sign of β_{ij} determines the direction of the contribution of $0.5\sum_i\sum_j \beta_{ij}([\ln C - \ln w_i]\ln[U_j] + [\ln C - \ln w_j]\ln[U_i] + \ln[U_i]\ln[U_j])$ to output. If this sign is positive then the whole term is positive and the result is a further increase in output. If β_{ij} is negative an additional decrease in output would result from this term. Finally, if β_{ij} is zero, this term disappears and the decomposition reduces into the opportunity cost and the lump-sum effects of the subsidies. This is the result that would have been obtained if we were to use a Cobb-Douglas technology.[10]

Adding these three terms, the total effect of the subsidies on output denoted by ln (ξ), can be written as below.

$$\begin{aligned}\ln(\xi) = &-\sum_i \beta_i \ln(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki}) + \left(\sum_i \beta_i \ln\left[\sum\nolimits_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\right]\right)\\ &+0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}\left([\ln C-\ln w_i]\ln\left[U_j\right]+\left[\ln C-\ln w_j\right]\ln[U_i]+\left[\ln U_j\right][\ln U_i]\right)\end{aligned} \tag{6}$$

Rewriting this equation gives,

$$\begin{aligned}\ln(\xi) = &-\sum_i \beta_i \ln(U_i) + 0.5\sum\nolimits_i\sum\nolimits_j \beta_{ij}\left([\ln C-\ln w_i]\ln U_j+[\ln C-\ln w_j]\right.\\ &\left.\ln U_i+[\ln U_i][\ln U_j]\right).\end{aligned} \tag{7}$$

If $\ln(\xi_{ij})$ is positive, the subsidies increase output more than they decrease it. Alternatively if it is negative, the subsidies reduce output more than they increase it, while if it is zero then both the negative and positive output effects of the subsidies exactly offset each other, thus leaving output unchanged. In the latter case, the subsidies are used to maintain existing services without increasing them and this makes it appropriate to examine the impacts of the subsidies on costs only. Thus, it is in the latter case where the subsidies maintain existing services, but not increase them, that there is support for studies that use cost functions to examine the impact of the subsidies. It is important to note that the first term of Eq. 7 is a combination of the lump-sum and cost effects of the subsidies and shows efficiency gain (loss) if U_i is less (greater) than one.

Substituting Eq. 7 into Eq. 5 gives $\ln(Q)=\ln(\hat{Q})+\ln(\xi)\ +\ \varepsilon$, which links the form of the decomposition to that derived by Kumbhakar (1997) for translog cost functions and fills the void in the transit economics literature mentioned earlier in terms of absence of focus on the effect of subsidies on output. Notice that whether or not output increases from the subsidies does not mean cost behaves likewise; the change in cost from the subsidies is $\ln\left(\sum_i S_i(1-\mu_{oi}H_{oi}-\mu_{Ki}H_{Ki})\right)$. Since the impacts of subsidies on cost and output can be both determined from Eq. 5, it is advantageous to estimate indirect production functions instead of cost functions when studying the impacts of operating and capital subsidies.

The empirical model

To apply Eq. 6 using cross-sectional data requires modifying it to account for heterogeneity. We do so by adding variables reflecting the characteristics of the observations (i.e., transit systems) to capture heterogeneity. Many previous studies use a similar approach to examine the relationships between organizational and environmental characteristics and public transit performance. Among them, Pina and Torres (2001) consider a city's

---

[10] The decomposition is most apparent when the underlying technology is characterized as translog, Cobb-Douglas or Lewbel. If, for example, the technology is generalized quadratic, $Q=\sum_i\sum_j\alpha_{ij}\left\{\left(\frac{C^*}{w_i^*}\right)^{\alpha\beta}\left(\frac{C^*}{w_j^*}\right)^{\alpha(1-\beta)}\right\}^{1/\beta}$ the decomposition cannot be obtained.

industrial characteristics, geographical extent, population density, income per capita and the age of the population as exogenous variables in their study of transit performance. They find that the population of an urban area, environmental variables and type of management do not have significant impacts on efficiency. Obeng (1987) finds that the variables that affect cost, partial measures of productivity and the measure of economies of scale are average vehicle speed, the ratio of employer to employee paid benefits, subsidies, capacity utilization, route miles, the peak-base ratio, average fleet age, number of modes operated, and the ratio of supervisors, professionals and executives to total employment. Guiliano (1980) identified market conditions (e.g., hours of service availability and the peak-base ratio), system size (e.g., service area), age of the firm and unionization as affecting efficiency. Kerstens (1996) classified the variables affecting public transit performance into competition (e.g., the extent of privatization or contracting), organizational differences (e.g., ownership) and operating environment (e.g., network length, number of lines, peak–base ratio, number of stops, vehicle speed), vehicle age, and method of financing (e.g., subsidies).

Similar variables as those listed above are used in this paper to account for heterogeneity. In particular, we follow Pina and Torres (2001), Obeng (1987) and Kerstens (1996) and use population density (D), average vehicle speed (V), average vehicle age (z), and network size in terms of route miles (N) as measures of heterogeneity and include them in the indirect production function. Thus, the empirical indirect production function to be estimated is,

$$\begin{aligned}\ln Q = \beta_0 + \sum\nolimits_i \beta_i(\ln C - \ln w_i + \ln U_i) + 0.5 \sum\nolimits_i \sum\nolimits_j \beta_{ij}(\ln C - \ln w_i + \ln U_i)\\ \left(\ln C - \ln w_j + \ln U_j\right) + \eta_n \ln(N) + \eta_Z \ln(z) + \eta_V \ln(V) + \eta_D \ln(D) + \varepsilon\end{aligned} \tag{8}$$

Imposing the symmetry constraints $\beta_{ij} = \beta_{ji}$, employing Roy's (1943) identity and Shephard's lemma, the observed share (S_i) of an input in cost is,[11]

$$S_i = \frac{(-\partial \ln Q/\partial \ln w_i)}{(\partial \ln Q/\partial \ln C)} = \frac{\left(\beta_i + \sum_j \beta_{ij}\left(\ln(C/w_j) + \ln(U_i)\right)\right)}{\left(\sum\limits_i \beta_i + \sum_i \sum_j \beta_{ij}\left\{\ln\left(C/w_j\right) + \ln\left(U_j\right)\right\}\right)} \tag{9}$$

Both Eq. (8) and i - 1 of the share equations from (9) form a system to be estimated jointly. This system is homogeneous of degree zero in the parameters and for unique identification of their parameters we follow Gajanan and Ramaiah (1996) and impose the following restrictions on the coefficients.

$$\sum\nolimits_i \beta_i = 1, \quad \beta_{ij} = \beta_{ji} \tag{10}$$

Additionally to improve convergence we estimate Eqs. 8 and 9 jointly with the hypothesized subsidy functions below. These functions include all the variables discussed in the previous section of this paper.

---

[11] A reviewer suggests that Eq. 9 is from Shephard's lemma. Our check shows early references attribute it to Roy (1943). Of course it can also be derived from Shephard's lemma.

$$\left.\begin{array}{r}\ln(A_o) = \varphi_0 + \sum_i \mu_{oi}\ln(x_i) + \varphi_n \ln(N) + \varphi_D \ln(D) + \varphi_{\mathrm{UAF}}(Y_{\mathrm{UAF}}) \\ \ln\left(A_{\hat{K}}\right) = \upsilon_0 + \sum_i \mu_{ki}\ln(x_i) + \upsilon_Z \ln(z) + \upsilon_{\mathrm{CAP}}(Y_{\mathrm{CAP}}) \ + \upsilon_{\mathrm{GEN}}(Y_{\mathrm{GEN}}) \ + \upsilon_{\mathrm{PM}}\ln(Y_{\mathrm{PM}}) \\ \text{where } \ x = L,\ K,\ F,\end{array}\right\} \tag{11}$$

Where φ and υ are parameters to be estimated and $Y_{\mathrm{PM}} = h(Q)$ shows passenger miles. The signs of the coefficients of all the variables are expected to be positive.

Data

The data to estimate the set of equations are for the single mode bus transit systems included in the 2006 U.S. National Transportation Statistics (NTS) database. Initially all such transit systems reporting their data were included in the sample providing 100% enumeration.[12] Later, observations missing relevant data on operating subsidies, labor hours, and gallons of fuel, vehicle miles and route miles were deleted. Similarly, transit systems whose data on key variables (e.g., ratio of operating subsidies to capital cost, ratio of capital subsidy to labor cost) were judged by the author to be unreasonable or whose data were listed by the US Department of Transportation as questionable in the NTS database were deleted.[13] These deletions left 227 observations to be used in this study.

The data for these observations include operating cost, total annual vehicle miles of service, total annual hours worked by labor, gallons of fuel, fare revenue, total capital subsidies, total operating subsidies, fleet age, fleet size, transit background data such as population density, service area and average speed, and the shares of labor and fuel in total operating cost. Other variables are labor cost calculated as the sum of wages, salaries and fringe benefits, fuel cost which is total operating cost less labor cost, population density, service area and capital user costs. Capital cost is calculated as $w_K K = Kr(R + d)e^{-d(z)}$ where K is fleet size, r is the weighted average price of a new public transit bus in 2006. This price was calculated from awarded bus purchase contracts reported in various issues of METRO magazine by dividing the contract amount by the number of vehicles bought.[14] R is the average prime rate for 2006, d is a straight line rate of depreciation assuming a bus useful life of 20 years and w_K is bus user cost. Following Nadiri and Schankerman (1981) and the fact that $C = w_L L + w_F F + w_K K$ total capital cost calculated as above was added to total operating cost to obtain total cost and the shares of labor, fuel and capital in total cost calculated as each input's cost divided by total cost. After that, the cost of purchased transportation was allocated to the inputs according to their shares in cost and this cost was also added to total cost. Finally, input shares were recalculated and input prices calculated as input cost divided by input quantity. Table 1 shows descriptive statistics for the transit systems used.

---

[12] Notice that these are the transit systems submitting their annual data to the Federal Transit Administration and that not all transit systems do so. Therefore, they do not represent all the transit systems in the U.S.

[13] Some of the ratios of operating subsidies to capital cost were 100 or higher, and the ratios of capital subsidies to labor cost were in some case greater than 50.

[14] This is comparable to the average 2007 and 2008 new bus price of $424,880 reported by APTA (2008).

Table 1 Descriptive statistics

Variable	N	Mean	Std. Dev.	Minimum	Maximum
Total cost ($)	227	18,946,934.26	29,553,144.59	983,648.73	276,868,066.00
Passenger miles		16,111,996.22	30,280,491.94	8,025.00	289,297,904
Vehicle miles	227	2,959,289	4,348,777	25,950	39,504,428
Labor wage ($)	227	18.21	58.77	6.12	894.96
Capital user cost per vehicle ($)	227	44,422.73	6,075.01	3,636.42	56,646.74
Fuel price per gallon ($)	227	8.45	9.64	3.38	99.87
Labor hours	227	6,20,217.04	843,732.79	9,152.00	6,810,714.00
Fleet size	227	93.9736	119.0029	7.0000	905.0000
Gallons of fuel	227	583,620.75	1,004,167.93	21,609.00	10,091,084.00
Capital subsidy ($)	227	2,608,230.23	4,156,500.35	1,239.00	30,114,012.00
Operating subsidy ($)	227	11,211,794.83	17,813,901.89	6,017.00	154,588,939.00
Received funds from capital program (yes = 1, No = 0)	227	0.4846	0.5009	0.0000	1.0000
Received funds from local dedicated subsidy sources (Yes = 1, No = 0)	227	0.3700	0.4839	0.0000	1.0000
Received subsidy from state dedicated subsidy sources (Yes = 1, No = 0)	227	0.4405	0.4975	0.0000	1.0000
Received funds from federal urban area formula funds (Yes = 1, No = 0)	227	0.8899	0.3137	0.0000	1.0000
Received funds from local and state general revenues (Yes = 1, No = 0)	227	0.3524	0.4788	0.0000	1.0000
Route miles	227	328.61	349.16	7.00	2,674.76
Average fleet age	227	5.50	4.23	0.33	55.25
Service area (square miles)	227	293.83	510.00	14.00	3,353.00
Population density	227	2,327	1,182	1,055	7,068
Average vehicle speed (mph)	227	14.37	3.55	9.24	47.73

Fuel is a proxy for all non-labor and non-capital inputs. Therefore, its costs include the costs of materials, tires and all types of liquid fuels, and a portion of the cost of purchased service

Results

Estimation

Before estimating Eqs. 8, 9 and 11 it must be determined if the equations are identified. The condition for identification in our system of non-linear equations is that the number of endogenous variables appearing on the right-hand-side of the output equation must be less than the number of the predetermined and additional endogenous variables appearing in the subsidy and share equations but not in the output equation (see Kelejian and Oates 1989). Because H_{oi} and H_{Ki} are independent variables there are no endogenous variables on the right-hand-side of the output equation. Hence, the equation is identified. Even if we use A_o/w_ix_i and $A_{\hat{K}}/w_ix_i$ instead of H_{oi} and H_{Ki} respectively, still the output equation will be identified because it will have two right-hand-side endogenous variables ($A_o,\ A_{\hat{K}}$) and there will be four predetermined variables (Y_{UAF}, Y_{CAP}, $Y_{\mathrm{GEN,}}Y_{\mathrm{PM}}$) in the subsidy equations with non-zero coefficients excluded from the output equation. Furthermore, the subsidy and share equations are indentified because they do not contain endogenous variables.

Examining the output and cost share equations, the coefficients of H_{oi} and H_{Ki} are from the subsidy equations. Therefore, the cost, share and subsidy equations form a system of non-linear seemingly unrelated equations. Consequently, they are estimated jointly by iterative methods using the Marquardt optimization technique after imposing the restrictions and the non-negativity constraint, $\beta_i > 0$ on the coefficients.[15] The choice of this method is because upon convergence it gives similar results as would be obtained from maximum likelihood methods. Table 2 shows the results of the estimation. Convergence was achieved in 35 iterations and at that point 210 observations were used and 17 rejected.[16] For those rejected their implied input prices were negative and the model did not fit their data well. From the adjusted coefficients of determination the indirect production function explains 90.48% of the variation in output while 51.15 and 75.31% of the variation in capital subsidy and operating subsidy respectively are explained by their equations. Additionally, the equations explain 77.44 and 64.65% of the variation in the labor and fuel cost shares respectively. Most of the estimated coefficients are highly significant statistically as can be seen in Table 3.

Examining the signs of the coefficients, those of the subsidy equations are consistent with prior expectation. Surprisingly, with a zero coefficient, fleet size does not affect operating subsidy. In comparison, receipt of funds from the federal urban area formula grant has a positive and statistically weak coefficient in the operating subsidy equation. All the coefficients of the capital subsidy equation are statistically significant including the coefficients of receipt of funds from capital subsidy programs and local and state general revenues. And, contrary to our expectation, the coefficient of fleet age is negative and statistically significant in the capital subsidy equation showing that transit systems that keep their buses longer generally receive less capital subsidies. This could be a reflection of the years-of-use regulation in place for transit vehicles purchased with federal money. That regulation requires that vehicles bought with federal subsidies must be used for at least 12 years.

Regarding the estimated coefficients of the indirect production function, the sign of average bus speed is positive and statistically significant while the coefficient of population density, average fleet age and route miles are non-significant. These results show that transit systems that maintain relatively high average speeds produce large outputs. Using these results, on the average, a transit system's share in total cost that it must pay with its non-subsidy funds is calculated to be 36.04% leaving 63.96% to be accounted for by subsidies.

Similarly, using the results and the equation below, there are slight economies of scale in the transit systems studied. These economies show that a percentage increase

$$\partial \ln C/\partial \ln Q = 1/\left(\sum_i \beta_i + \sum\nolimits_i \sum\nolimits_j \beta_{ij}\{\ln(C/w_i) + \ln(U_i)\}\right) \tag{12}$$

in output increases cost by 0.9965% (std. dev. = 0.0321). This cost increase is statistically not different from the value of one for constant returns to scale. Therefore, we cannot dismiss constant returns to scale in the transit systems studied.

---

[15] Except the ratios of subsidies to input costs each variable is normalized by subtracting its mean from its value. This allows us to calculate allocative distortion for the mean transit system.

[16] This result was obtained after many trials using different starting values. In all cases the values of the coefficients at convergence were very close suggesting a global convergence point had been reached.

Table 2 Nonlinear iterative seemingly unrelated regression estimation results

Variables	Parameter	Std. error	t-value	Probability
Operating subsidy				
Constant	−0.2072	0.1277	−1.6200	0.1064
Labor (logarithm)	0.3949	0.0217	18.1700	<0.0001
Fuel (logarithm)	0.2134	0.0053	40.1200	<0.0001
Capital (logarithm)	0.0000	0.0000	–	–
Population density (logarithm)	0.6861	0.1256	5.4600	<0.0001
Route miles (logarithm)	0.3034	0.0490	6.1900	<0.0001
Allocation from urban area formula grant (Yes = 1, No = 0)	0.2266	0.1345	1.6800	0.0936
Capital subsidy				
Constant	−0.4606	0.1185	−3.8900	0.0001
Capital (logarithm)	0.4460	0.0714	6.2500	<0.0001
Labor (logarithm)	0.1354	0.0288	4.6900	<0.0001
Fuel (logarithm)	0.0649	0.0116	5.6000	<0.0001
Population density (logarithm)	0.9232	0.2462	3.7500	0.0002
Average fleet age (logarithm)	−0.3496	0.1528	−2.2900	0.0232
Allocation from federal capital program (Yes = 1, No = 0)	0.6087	0.1665	3.6600	0.0003
Funds allocated out of general revenue (Yes = 1, No = 0)	0.0008	0.0002	3.8000	0.0002
Output				
Constant	−0.0214	0.0254	−0.8400	0.4005
log $(CU_L)/w_L)$	0.6635	0.0047	140.4700	<0.0001
log $(CU_F)/w_F)$	0.2767	0.0035	80.1200	<0.0001
log $(CU_K)/w_K)$	0.0598	0.0025	24.0800	<0.0001
0.5log $(CU_L)/w_L)$log $(CU_L)/w_L)$	−0.0679	0.0088	−7.6800	<0.0001
log $(CU_F)/w_F)$log $(CU_L)/w_L)$	0.0522	0.0057	9.1900	<0.0001
log $(CU_L)/w_L)$log $(CU_K)/w_K)$	−0.0052	0.0032	−1.6000	0.1116
0.5log $(C(U_F)/w_F)$log $(C(U_F)/w_F)$	−0.0583	0.0043	−13.7000	<0.0001
log $(C(U_F)/w_F)$log $(C(U_K)/w_K)$	−0.0046	0.0017	−2.6200	0.0095
0.5log $(CU_K)/w_K)$log $(CU_K)/w_K)$	0.0091	0.0025	3.6400	0.0004
Population density (logarithm)	0.0706	0.0677	1.0400	0.2978
Average fleet age (logarithm)	−0.0560	0.0403	−1.3900	0.1661
Average speed (logarithm)	0.6983	0.1232	5.6700	<0.0001
Route miles (logarithm)	0.0021	0.0265	0.0800	0.9375

Model uses 210 observations. The R^2 for the output, capital subsidy, operating subsidy, labor share in cost and fuel share in cost are respectively 0.9048, 0.5115, 0.7531, 0.7744 and 0.6465

Effects of both operating and capital subsidies on output

Based upon Eq. 6 we calculate the effects of the subsidies on output first by considering both subsidies together and then separately. In Table 3 when we consider a transit system that receives both subsidies, the lump-sum effects of these subsidies are positive and increase output by 77.36% per transit system. This shows that the subsidies by reducing

Table 3 Output effects of operating and capital subsidies

Effects of both subsidies	*N*	Mean	Std. deviation	Minimum	Maximum
Implied system share in cost	209	0.4986	0.0896	0.2798	0.9973
Total effect of subsidies on output	209	0.0472	0.0488	−0.0018	0.3524
Effect of cost impact of the subsidies	209	−0.7104	0.1677	−1.2738	−0.0027
Interaction effect	209	−0.0150	0.0279	−0.1600	0.0755
Lump-sum effects of subsidies on output	211	0.7736	0.1940	0.0030	1.4439
Labor-capital allocative distortion	211	0.7214	0.3721	0.2825	3.2581
Labor-fuel allocative distortion	211	1.6208	1.2003	0.4242	13.6007
Capital-fuel allocative distortion	211	2.4662	1.9646	0.4237	23.2709
Effect of capital subsidies					
Implied system share in cost	217	0.9521	0.0449	0.7823	0.9999
Total effect of subsidies on output	217	0.0070	0.0149	−0.0125	0.0840
Effect of cost impacts of the subsidies	217	-0.0502	0.0490	−0.2456	−0.0001
Interaction	217	0.0006	0.0023	−0.0037	0.0136
Lump-sum effects of subsidies on output	219	0.0566	0.0570	0.0001	0.2963
Labor-capital allocative distortion	219	1.4085	0.7283	0.9898	5.6200
Labor-fuel allocative distortion	219	1.0050	0.0203	0.9349	1.1021
Capital-fuel allocative distortion	219	0.8040	0.1964	0.1816	1.0036
Effect of only operating subsidies					
Implied system share in cost	220	0.5468	0.0792	0.3958	0.9996
Total effect of subsidies on output	220	0.0487	0.0574	−0.0020	0.5355
Effect of cost impacts of he subsidies	220	−0.6128	0.1309	−0.9269	−0.0004
Interaction effect	220	−0.0143	0.0433	−0.5136	0.0660
Lump-sum effects of subsidies on output	222	0.6766	0.1759	0.0004	1.8077
Labor-capital allocative distortion	222	0.5431	0.0930	0.3088	0.9996
Labor-fuel allocative distortion	222	1.9523	6.1655	0.4544	91.9273
Capital-fuel allocative distortion	222	3.5620	11.1134	1.0006	166.1716

implied input prices increase the quantities of inputs demanded and make transit systems almost double their outputs. The size of this increase is affected by the effects of the cost impacts of the subsidies which reduce output by 71.04% per transit system when both subsidies are received. The fourth row of Table 3 shows that the interaction effects of the lump-sum and cost impacts add 1.50% on the average to the output reduction. Combining these results, the subsidies increase output by 4.72% per transit system with a range of −0.18 to 35.24%. This range shows that while in some transit systems the effects of the subsidies on output are quite small or that the subsidies actually reduce output by a very small proportion, in others the effects are quite large. Surprisingly though, the total effect of the subsidies on output is negatively related to transit system size measured in terms of fleet operated. The correlation between them is −0.1795 with a standard error of 0.0076.

Overall, the results suggest that the subsidies increase output except in four transit systems. These four transit systems are North East Transportation Authority, Central Florida Regional transportation Authority (LYNX), Pee Dee Regional Transportation Authority (PDRTA) and Clarksville Transit System (CTS). Because output increases in most transit systems and in few it decreases, it is inappropriate to assume that the outputs of the transit systems studied remain constant when capital and operating subsidies are offered as cost studies assume. For the transit systems studied, a proper accounting of the effects of operating and capital subsidies would be obtained by estimating indirect production functions as in this study. Though there are potentials for output to change by a large proportion in all the transit systems studied, inefficiencies in input overuse reduce that change.

Effects of either subsidy on output

Surprisingly enough, when the individual impacts of both subsidies are considered and compared, Table 3 shows that the total effects of operating subsidies on output are far larger than the total effects of capital subsidies on output. When only operating subsidies are considered, on the average, their lump-sum effects result in 67.66% increase in output on the average while the cost impacts and the interaction term reduce output by 61.28 and 1.43% on the average respectively resulting in a 4.87% increase in output overall. Comparatively, when only capital subsidies are considered their lump-sum effects increase output by 5.66% while the cost impacts reduce output by 5.02% and the interaction effect increases output by 0.06% respectively. The net result is a 0.70% output gain from capital subsidies on the average. Thus, while both types of subsidies are important, operating subsidies have larger impacts on output than do capital subsidies, at least in the transit systems studied. In fact, in this study capital subsidies have very little effect on increasing output suggesting that they support current services.

This finding may be because capital subsidies are mainly for equipment replacement. To operate bus services more intensively to increase output requires the same capital but more labor and fuel whose costs are partially supported by operating subsidies. It could also be that operating subsidies cover those costs that heavily influence short run production decisions. For example, decisions to purchase or replace capital are made quite infrequently and involve large expenditures which increase the scale of transit operations. Once such decisions are made what influences how much service to produce is a transit system's ability to cover its short run costs, and this makes it important to have operating subsidies.

Sources of inefficiencies

Equation 7 shows that all the increases in output could also be considered as due to allocative inefficiency because the subsidies affect the optimal rate of input substitution. That interpretation shows that operating subsidies cause more allocative inefficiencies than do capital subsidies. Given this result it is important to examine the sources of the allocative inefficiencies. To do so labor-capital, capital-fuel and labor-fuel allocative distortions from the subsidies are calculated using σ_{ij} from Eq. 2 and the results are also shown in Table 3. Focusing only on when both subsidies are received, the value of the labor-capital allocative distortion is 0.7214 and it shows distortions in the optimal rate of substitution between these inputs. Since the value of this distortion is less than one it shows that the subsidies have made labor relatively cheap leading to its overuse relative to capital. For capital-fuel allocative distortion its value of 2.4662 shows that because of subsidies fuel is overused relative to capital. That is, the subsidies have reduced the cost of fuel so much that transit management misperceives it as relatively cheap compared to other inputs therefore, leading to its substantial overuse relative to capital. This overuse could take the form of buying and running less fuel efficient vehicles, routing services through congested routes, excessive idling of vehicles, extended service, improper vehicle maintenance, and possibly wrong engine choices during the bus purchase decision-making process. Finally, the labor-fuel allocative distortion is 1.6208 showing that fuel is also overused relative to labor. This could take the form of operating larger and longer buses that increase fuel use and reduce the number of drivers per shift. From these results, input overuse from the subsidies particularly the overuse of labor and fuel relative to capital has led to overproduction of transit services.

Policy implications

The results above have policy implications that are discussed in this section. Perhaps the most important finding is that absent revenue considerations in the single mode US transit systems studied, a justification for capital and operating subsidies exists in terms of a modest 4.72% increase in output on the average. This increase is smaller than, yet very close to the 6–8% reported by Schmidt (2001). Based upon this result it is inappropriate to assume output remains unchanged when transit systems receive operating and capital subsidies. Decision makers and those who control financial resources to transit systems, therefore, can use this increase to compare the performance of the transit systems they fund. Furthermore, this small increase suggests that most of the subsidies support existing services.

Another result is the negative and significant correlation of −0.1795 between the increase in output and fleet size. Because this correlation is statistically significant large single mode bus transit systems do not enjoy as much an increase in output as do small systems. A possible explanation is the difficulty in increasing output in large single mode bus transit systems because many may be operating near capacity and unable to expand their services by adding routes or extending them. In such transit systems responses to increases in demand could be achieved through intensive operation of their vehicles or by changing schedules. In terms of policy this finding suggests that large output increases cannot be used to justify subsidies to large single mode bus transit systems; the justification for the subsidies to them is to maintain and improve quality of service.

A breakdown of the output increase shows that operating subsidies increase output by 4.87% on the average and capital subsidies by 0.70%. In both cases the lump-sum effects of the subsidies are the reasons output increases and they occur because the subsidies make transit systems misperceive their input prices as having fallen. In turn, this misperception of price makes transit systems behave as if their incomes have risen resulting in their production of more output than they did before. Thus, the output increase is not because transit systems switch to more productive inputs, nor are they because they extend their routes. For, the data shows that capital subsidies to these transit systems are mainly for replacing old vehicles and modernizing their facilities.

Comparatively, these transit systems use their operating subsidies to cover the costs of essential resources they need daily to operate their vehicles intensively especially when demand increases. This makes the relationship between operating subsidies and output stronger resulting in more increases in output than would be obtained from capital subsidies. In some transit systems these output increases are modest, as we found, and in others they are quite substantial. Regardless of the amounts, however, these increases support a policy of providing more operating subsidies than capital subsidies to single mode bus transit systems. While this suggestion does not argue against capital subsidies to these firms, it results from the fact that they are used mainly for facilities and capital replacement which are activities that do not impact output by much. Moreover, such a policy should be tampered with judgments about the possible effects of the subsidies resulting in more deficits.

An explanation for the increase in output is that both subsidies change the optimal rates of input substitution in favor of labor. From the calculated values of σ_{ij} they make the implied price of labor very low relative to the implied price of capital resulting in a labor-capital allocative distortion of 0.7214 when operating and capital subsidies are considered jointly, and distortions of 0.5431 and 1.4085 respectively when these subsidies are considered separately. Both subsidies, therefore, have opposite effects with operating subsidies encouraging more use of labor relative to capital and capital subsidies doing the reverse. These results are expected. This capital overuse takes the form of operating (1) more buses, (2) longer and larger buses, and (3) improving and building facilities to attract customers. Additionally, the subsidies have encouraged more use of fuel evidenced by the values of 2.4662 and 1.6208 for capital-fuel and labor-fuel allocative distortions. These distortions mean that both capital and fuel have very low marginal productivities requiring more of them to be used to produce any given level of output. They also mean that the increase in output would have been large if the subsidies had encouraged transit systems to increase their use of more instead of less productive inputs.

Both the overuse of capital and fuel result in wasted resources and high costs. Therefore, policies are needed to control these inputs and reduce cost. Regarding capital, its overuse can be controlled by enforcing the federal years-of-use regulation (of 12 years) for vehicles bought with federal money and the 20% spare ratio required by the Federal Transit Administration. For fuel, its overuse can be controlled by operational changes such as reducing idling, avoiding congested and circuitous routes and by switching to other types of fuel.

Conclusion

This paper's purpose is to fill a gap in the public transit economics literature by estimating the effects of subsidies on output, recognizing that previous studies fail to do so though such studies have greatly improved our understanding of public transit cost structure, technology, performance and efficiency. Particularly, the paper's purpose is to add to the body of knowledge on the effects of public transit subsidies on cost by extending those effects to output. Its main contributions are first, it provides a decomposition of the effects of capital and operating subsidies on public transit output into the lump-sum, cost and their interaction or allocative inefficiency effects. This decomposition is unique to this paper and as shown it is the production counterpart of what Kumbhakar (1997) derived in his work on translog cost functions. Because this decomposition is not possible with cost functions we conclude that additional information is gained when indirect production functions are used in studies on public transit subsidies. Second, it finds that the positive lump-sum effects of the subsidies on output are larger than the negative effects of the cost impacts and allocative inefficiencies from the subsidies. This implies that the subsidies together increase output beyond maintaining existing output. Overall, this increase in output is 4.72% on the average mainly due to the effects of operating subsidies. Thus, we conclude that a modest justification exists for operating and capital subsidies in terms of output increase. However, we also find that there is an undesirable outcome of the subsidies in that they could lead to lower output in some transit systems as found in four cases. A third contribution is that the cost effects of the subsidies on output are negative and large and could negatively affect the overall outcome of the subsidy. This finding leads to the conclusion that greater focus needs to be given to the mechanisms used to pay subsidies in the US to ensure that their negative effects on output are minimized. In particular, containing the cost effects of the subsidies could lead to large output gains.

Limitation

The results of this paper are limited by the functional form of the indirect production function used. Although we used a flexible functional form specification the ability to separate the lump-sum and the cost effects rests with the translog, Cobb-Douglas and Lewbel functional forms. The decomposition is not apparent with a generalized quadratic specification; only the combined effects of the cost and lump-sum effects can be obtained from it. Also, the results may have been affected by the choice of heterogeneity variables. Future research using different measures of heterogeneity should add to the results in this paper especially the size of the output effect.

Acknowledgments I am grateful to the anonymous referees and Prof. R. Sakano of the Department of Economics and Finance, North Carolina A&T State University, Greensboro NC, for their comments.

References

American Public Transit Association: 2008 Public Transportation Fact Book, 59th edn. APTA, Washington DC (2008)

Atkinson, S.E., Halvorsen, R.: The relative efficiency of public and private firms in a regulated environment: the case of public utilities. J. Public Econ. **29**, 281–294 (1986)

Berechman, J.: Public Transit Economics and Deregulation Policy. North Holland, Amsterdam (1993)

Bly, P.H., Oldfield, R.H.: The effects of public transit subsidies on demand and supply. Transp. Res. A **20**(6), 415–427 (1986)

Brons, M., Nijkamp, P., Pels, E., Rietveld, P.: Efficiency of urban public transit: a meta analysis. Transportation **32**, 1–21 (2005)

Cervero, R.: Examining the performance impacts of transit operating subsidies. ASCE J. Transp. Eng. **115**(5), 467–480 (1986)

Chambers, R.G.: Duality, the output effect and applied comparative statics. Am. J. Agric. Econ. **64**(1), 152–156 (1982)

Cooter, R., Topakian, G.: The political economy of a public corporation: pricing objectives of BART. J. Public Econ. **13**, 299–318 (1980)

De Borger, B., Kerstens, K., Costa, A.: Public transit performance: what does one learn from frontier studies? Transp. Rev. **22**, 1–38 (2002)

De Borger, B.: The economic environment and public enterprise behaviour: Belgian railroads, 1950–86. Economica **60**, 443–463 (1993)

Fabbri D. (1996). Public transit subsidy: from the economics of welfare to the theory of incentives. Working Paper. Department of Economics, University of Bologna, Italy

Fare, R., Grosskopf, S., Lovell, C.A.K.: An indirect approach to the evaluation of producer performance. J. Public Econ. **37**, 71–89 (1988)

Federal Transit Administration: 1998 National Transportation Statistics: National Transit Database Section 15 Report Year. U.S. Department of Transportation, Washington DC (1998)

Gajanan, S.N., Ramaiah, K. (1996). An econometric estimation of Hicksian and Marshallian elasticities in Indian manufacturing. South. J. Econ. 63:406–417

Garofalo, G.A., Malhotra, D.M.: The demand for inputs in the traditional manufacturing region. Appl. Econ. **22**, 961–972 (1990)

Glaister, S., Collins, R.: Maximization of passenger miles in theory and practice. J. Transp. Econ. Policy **12**, 304–321 (1978)

Guiliano, G. (1980). The effect of environmental factors on the efficiency of public transit service. In: Presented at the 60th Annual Meeting of the Transportation Research Board, Washington, DC

Hanushek, E.A.: Conceptual and empirical issues in the estimation of educational production functions. J. Hum. Resour. **14**(3), 351–388 (1979)

Hilmer, C.E., Holt, M.T.: Estimating indirect production functions with a more general specification: an application of the Lewbel model. J. Agric. Appl. Econ. **37**(3), 619–634 (2005)

Karlaftis, M.G., McCarthy, P.: Cost structures of public transit systems: a panel data analysis. Transp. Res. E **38**, 1–18 (2002)

Kelejian, H.H., Oates, W.E.: Introduction to Econometrics: Principles and Applications, 3rd edn. Harper and Row Publishers Inc., New York (1989)

Kerstens, K.: Technical efficiency measurement and explanation of French urban transit companies. Transp. Res. A **36**(6), 431–452 (1996)

Kim, H.Y.: Analyzing the indirect production function for U.S. manufacturing. South. Econ. J. **55**(2), 494–504 (1987)

Kumbhakar, S.C.: Modeling allocative inefficiency in a translog cost function and cost share equations: an exact relationship. J. Econ. **56**, 351–356 (1997)

Kumbhakar, S.C., Bokusheva, R.: Modeling farm production decisions under an expenditure constraint. Eur. Rev. Agric. Econ. **36**(3), 343–367 (2009)

Mohring, H.: Optimization and scale economies in urban bus transportation. Am. Econ. Rev. **62**, 591–604 (1972)

Nadiri, M.I., Schankerman, M.A.: The structure of production, technological change and the rate of growth of total factor productivity in the U.S. Bell system. In: Cowing, T.G., Stevenson, R.E. (eds.) Productivity Measurement in Regulated Industries. Academic Press, New York (1981)

Nash, A.C.: Management objectives, fare and service levels in bus transport. J. Transp. Econ. Policy **12**, 70–85 (1978)

Nolan, J.F.: Determinants of productive efficiency in urban transit. Logist. Transp. Rev. **32**(3), 319–342 (1996)

Nolan, J.F., Ritchie, P.C., Rowcroft, J.E.: Identifying and measuring public policy goals: ISTEA and the US bus transit industry. J. Econ. Behav. Organ. **48**, 291–304 (2002)

Obeng, K.: The deadweight costs of operating and capital subsidies. J. Transp. Res. Forum **49**(1), 37–57 (2010)

Obeng, K.: Classification of Bus transit policy variables. Transp. Plan. Technol. **11**, 257–272 (1987)

Obeng, K.: Expense preference behavior in public transit systems. Transp. Res. E **36**, 249–265 (2000)

Obeng, K., Azam, G.: The intended effects of federal transit operating subsidies. Public Finan. Q. **23**(1), 72–94 (1995)

Obeng, K., Azam, G., Sakano, R.: Modeling Economic Inefficiency Caused by Public Transit Subsidies. Praeger, Westport (1997)

Obeng, K., Sakano, R.: Public transit subsidies, output effect and total factor productivity. Res. Transp. Econ. **23**, 85–98 (2009)

Oum, T., Yu, C.: Economic efficiency of railways and implications for public policy. J. Transp. Econ. Policy **28**(2), 121–138 (1994)

Pederson, P.: On the optimal fare policies in urban transportation. Transp. Res. B **37**, 423–435 (2003)

Pestieau, P., Tulkens, H.: Assessing and explaining the performance of public enterprises. Finanzarchiv **50**, 293–323 (1993)

Pina, V., Torres, L.: Analysis of the efficiency of local government service delivery. An application to urban public transport. Transp. Res. A **35**, 929–944 (2001)

Roy, R.: l'Utilité: Contribution à la Théorie des Choix. Hermann, Paris (1943)

Savage, I.: Management objectives and the causes of mass transit deficits. Transp. Res. A **38**, 181–199 (2004)

Schmidt, S.: Incentive effects of expanding federal mass transit formula grants. J. Policy Anal. Manag. **20**(2), 239–261 (2001)

Shephard, R.W.: Indirect production functions. Research Report, Operations Research Center, University of California, Berkeley (1973)

Small, K., Gomez-Ibanez, J.: Urban transportation. In Mills, E., Cheshire, P. (eds.) Handbook of Regional and Urban Economics. Elsevier Science, Amsterdam (1999)

Small, K.: Urban Transit Economics. Harwood Academic Publishers, Reading (1990)

Tistato, P.: User economies of scale: Bus subsidy in Adelaide. Econ. Rec. **73**(223), 329–347 (2007)

Van Reeven, P.: Subsidisation of urban transport and the Mohring effect. J. Transp. Econ. Policy **42**(2), 349–359 (2008)

Viton, P.A.: On the economics of rapid-transit operations. Transp. Res. **14A**, 247–253 (1980)

Author Biography

K. Obeng is a Professor of Transportation at North Carolina A&T State University, Greensboro NC. He has PhD in City and Regional Planning with concentration in transportation planning from the University of Pennsylvania and has published widely in urban transportation. His research interests are in the areas of public transit cost, subsidy, productivity, performance, and public transit management and transportation safety.