Power of Incentive Schemes, Provision of Quality and Monitoring: the case of the public transit system in Santiago de Chile

Marco Batarce* **Patricia Galilea**

Department of Transport Engineering and Logistics,
Pontificia Universidad Católica de Chile

October 2013

ABSTRACT

One important issue in procurement of public services is how to design a contract that incentivizes private firms to reduce costs and produce efficiently. In theory, fixed price contracts dominate cost-plus contracts. However, the economic efficiency gains in the fixed price contract can harm the users' welfare if the firms can reduce cost at the expenses of the delivered quality. The aim of this paper is to model and test empirically that cost-reduction effort is replaced by quality reduction when there is no monitoring and demand is almost inelastic to quality. The implementation of the transportation plan of Santiago de Chile (Transantiago) shows that unmonitored fixed price contract creates incentives to reduce costs by lowering the quality delivered. This paper measures the relative level of efficiency of the firms operating in Transantiago and the level effort to reduce costs when the intensity of monitoring changes. To do so, we specify a cost function that depends on the effort to reduce cost and an inefficiency parameter distributed across firms. Our results indicate that if the operators are monitored, they increase the cost reduction effort.

* *Corresponding author: Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul, Santiago, Chile, +56 (2) 2354 1797, mbatarce@ing.puc.cl*

1. Introduction

A major concern in procurement and contract design is how to give incentives for efficient provision of goods or services. This becomes even more relevant in the case of public goods, since their provision usually requires subsidies. Reducing such subsidies will allow a better allocation of public resources. The problem for the regulator is that the effort of the firms providing that public good is not observable; thus, he faces a moral hazard problem (Laffont and Tirole, 1991).

Contract theory indicates that when the concern of the regulator is the reduction of costs, fixed-price contracts are more efficient (Laffont and Tirole, 1991). Such type of contract consists in paying a fixed amount to the firm providing the service. This has been confirmed by Gagnepain and Ivaldi (2002a), who study empirically the case of the public transport providers in France. In their study, they show that firms providing the service under a fixed-price contract exhibit a bigger effort on the reduction of costs. These firms are the residual claimant of operational costs and, therefore, exert more effort to reduce them.

Nevertheless, Laffont and Tirole (1991) also show that when quality and effort of reducing costs are substitutes, agents may reduce costs by lowering the service level. The analysis done by Gagnepain and Ivaldi (2002a) does not consider the quality offered and they assume that operators will comply with the contracted service level. However, if the quality is not verifiable, fixed-price contracts may lead to the firm to provide low service level, as firms will only focus on lowering their costs. By contrast, if quality is verifiable, regulator may establish a monitoring scheme and force a minimum service level, recovering the incentive power of this type of contract.

This paper studies empirically the effect of the monitoring of the service level on the reduction cost effort using data from the bus public transport firms in Santiago. We address the question if the firms were producing efficiently or if they were replacing cost reducing effort with low quality provision. The question is empirically relevant as the system operates with a subsidy that reached almost 40% of the costs.

From the start of Transantiago in 2007 until year 2011, the contracts with the bus operators were, in practice, fixed-price contracts. Each firm signed a contract where they guaranteed to fulfil the operational plan established by Coordinación Transantiago, the governmental agency in charge of supervising the operators. The contracts only established a commitment on terms of following the operational plan, but no punishment in case of failing to do it. By mid-2007, Coordinación Transantiago implemented a program of compliance measures and fines to enforce the fulfilment of the contracts. Such program increased the operating fleet from 4,600 to 5,800 buses in only five months (Beltrán et al., 2013). We interpret such a change as an increment in the level of service, because the main consequence of the reduced operating fleet was low bus frequency and high crowding in buses and bus stops.

To measure the effect of monitoring on cost reduction effort, we estimate a structural cost model that takes into consideration two sources of asymmetric information: level of effort and efficiency (see Dalen and Gomez-Lobo, 1997; Gagnepain and Ivaldi, 2002b). The former source of asymmetric information is moral hazard due to a principal not capable of observing the effort exerted by the agent. The latter source is adverse selection due to a principal not capable of observing the firm's efficiency level. From an econometric point of view, the effort level is a source of endogeneity that may bias the estimates, and the efficiency is a source of unobservable heterogeneity. Under some parametric assumption, we determine the firm's optimal level of effort and introduce it in the cost function to control for endogeneity. To deal with the heterogeneity, we use a stochastic cost frontier model, where an efficiency parameter is introduced as a stochastic component of the cost function.

Our basic assumption is that the monitoring affects the level of effort as an exogenous variable. This means for a given level of output, input prices and efficiency, the optimal effort exerted by the firm increases with the monitoring level. In addition, we consider the level of service and effort substitutes, and the resulting effort from the monitoring is net of quality improvements, which also increases with the monitoring. Hence, we study the net effect of monitoring on effort and quality improvement.

The paper is organized as follows. Section 2 describes briefly the Santiago bus transit system and the data available for the analysis. Section 3 presents the model that takes into account effort level and efficiency. Section 4 describes the estimation method based on stochastic frontier techniques. Section 5 presents the results of estimation and its interpretation. Section 6 concludes and discusses some policy implications.

2. Bus transit industry and data description

This section presents the bus system of Santiago and describes the data available for model estimation. For a more detailed description of the bus system see Muñoz and Gschwender (2008).

2.1 The bus system of Santiago

The bus system of Santiago (Chile) is completely regulated. The Coordinación Transantiago set the fares, which are adjusted according to some price indexes of input like fuel, labor, and dollar exchange rate. Coordinación Transantiago also defines or modifies the routes and fixes the frequencies of the bus lines. The frequencies are defined in periods of two weeks and delivered to the operators in quarterly operation plans. To monitor the operation buses, Coordinación Transantiago has a central system that collects information based on GPS and SIG continuously. This information is used to verify the fulfilment of the operation plans.

The bus system of Santiago consists of five trunk services and nine feeder services. The trunk services have exclusive operation in main roads in the city, and their routes connect the edges of the city with the Central Business District (CBD) or goes across the city from one extreme to another. The feeder services operate in exclusive zones and connect to the trunk services. Total daily demand is around of 6 millions of trips and the number of trips legs attaints 90 million per month. Because of some mistakes in the design of the system, the authority has adjusted the supply in several aspects by means of the operation plans. Table 1 displays the evolution of the driven kilometres, the fleet size, and the number of bus lines from 2007 to 2011. It should be noted that after 2010 Coordinación Transantiago started reducing the total kilometres offered as a way to reduce the subsidy for the system.

Table 1: Evolution of public transport commercial kms, fleet and bus lines

Year	Driven kms (millions)	Fleet	Bus lines
2007*	371.1	4,489	223
2008	481.4	6,399	322
2009	487.2	6,572	334
2010	512.4	6,564	357
2011	483.0	6,165	370

(Source: Coordinación Transantiago)

* Figures correspond to February 2007 when Transantiago was inaugurated

At the beginning of the period of analysis, there were four firms operating only trunks, six firms operating only feeder services (one of them operated two zones) and one firm operating two feeder zones and one trunk service. By the end of 2008, the latter firm was unable to operate the trunk and one of the feeder services; therefore two new firms entered the market at the beginning of 2009. In December 2009, the same firm became insolvent and exited the market; its zone was awarded to an incumbent firm operating another feeder zone. Thus, in the period of analysis there were thirteen firms operating in the market. We have data from ten of these thirteen firms for our estimation of the cost function. Since some trunk and feeder concessions have merged into larger firms, operating a mix of both types of services, once we estimate the operator's cost function our results should show the existence of economies of scale or spatial scope (Basso et al., 2011).

Three compliance measures were established during the time frame we focus our study of Transantiago. For more details about them and their results on Transantiago's service level, please check Beltrán et al. (2013). As we noted in the introduction, when Transantiago started in 2007, the contracts only established that the operators should fulfil the operational program designed by Coordinación Transantiago, but no punishment was established in case operators did otherwise. Therefore, the number of buses providing the service was much lower than what the operational plan established (Beltrán et al., 2013). Consequently, the authority established, in August 2007, the first compliance measure (ICPH) based on the percentage of the observed number of seat/standing places per hour (through GPS) compared with the theoretical figure indicated in the operational plan. ICPH was calculated for each company and it impacted directly on the operators' revenues. Once implemented, it had a significant impact on operation and service level: in July 2007 the number of buses on the street was 4.600, whereas the operational program established 5.600 buses. By the end of 2007, the number of buses was 5.870.

In August 2008, two new compliance measures were created: ICF and ICR. The ICF index was developed to ensure that the programmed frequency of a particular service was met, while the ICR index observes the regularity of the headway between buses for a particular service. Penalties were applied to the operators that did not comply with a high level on these indexes and they had a positive impact on operation and the quality of the service (Beltrán et al., 2013).

In order to improve the precision of the index ICPH, the amount of kilometres ran by each bus in 30 minutes was included in September 2009. The new version of this index measures the number of seat/standing places-kilometres per hour fulfilled and is called ICPKH.

2.2 Data description

To estimate the cost function, we use data from several sources because there are many different types of information involved in our estimation. We need costs, input prices, transport demand, operation variables, and fare information.

We obtained data on firms' cost from the Chilean Securities and Insurance Superintendence (Superintendencia de Valores y Seguros - SVS). The transport firms operating in Santiago must deliver financial reports quarterly, which include information operation cost, financial cost, and total cost. This information is public and available through Internet (Superintendencia de Valores y Seguros, 2012). The financial reports were submitted quarterly until March 2010, after that the regulation for this information changed to annual reports. Thus, we have quarterly information from the second quarter of 2007 to the first one of 2010. However, some firms continued reporting financial information quarterly until December 2010. Some firms exited the industry during this period (firm 5 exits in the third quarter of 2010) or few months after (firms 2 and 4 exit in the third quarter of 2011), and another firm enters (firm 10 enter in fourth quarter of 2009). Hence, we have an unbalanced panel of firms, where some firms have around 15 observations whereas others only have 2-3 observations (Table 2).

Table 2: Number of Observations for Each Firm in the Sample

Firm	Data		No. Obs.
	From (year-quarter)	To (year-quarter)	
1	2007-I	2010-I	12
2	2007-II	2010-I	12
3	2008-I	2010-I	9
4	2007-II	2010-I	12
5	2009-I	2009-III	3
6	2007-II	2010-I	12
7	2007-II	2010-IV	15
8	2007-II	2010-IV	15
9	2007-II	2010-IV	15
10	2009-IV	2010-I	2

As inputs for our cost model, we use labour, diesel, and capital. In all cases, we assume that the firms have no monopsony power in the input markets and that those markets are competitive. Thus, all firms face the same prices for labour, diesel, and capital. Usually, the labour price is calculated as the ratio between the total expenses in labour and the number of workers in the firm. However, we do not have information on such variables. Therefore, as labour price we adopt the quarterly average of the Real Cost Labour Index from the Banco Central de Chile (2012) (Central Bank of Chile).

The diesel price is obtained from the statistics collected by the National Consumer Service (Servicio Nacional del Consumidor). We use as diesel price the quarterly average price per litre of diesel in the Metropolitan Region of Santiago.

The price of capital is mainly the price of rolling stock and parts required for its maintenance. To take into account the effects of this input on the cost we consider several factors: national interest rate, international interest rate, and dollar exchange rate. All this variables should have an effect in the amortization and depreciation of rolling stock. However, after a number of preliminary estimations, the only relevant variable was the dollar exchange rate. We obtained all information on these variables from statistics from Banco Central de Chile (2012).

We obtained the operation information from Coordinación Transantiago. This is the government agency that controls the contracts with the bus service providers. The agency also defines the operation plans for the bus lines. We use the information on operation to characterize the product of the firms, because the number of passengers is not enough to describe the output of a transportation firm (Spady and Friedlaender, 1978). Therefore, we introduce the total kilometres contracted in the quarter and the average number bus lines operated by each firm.

For the number of transported passengers we use the total taps of Transantiago's smartcard recorded by the system. Users tap in to enter each bus of the network. During two hours, passengers may make up to two transfers paying a single fare as long as they do not repeat the same bus line. Thus, the total ridership in the system is estimated as the number of passengers entering the system paying their fare (even if they make more than one trip). The ridership of each firm is estimated as the number of taps registered in the buses of the firm. We are aware there is a fraction of passengers that do not pay, and consider this evasion as part of the operation cost. Indeed, if the evasion is constant in the period we analyse, it is taken into account by increasing the marginal cost.

The level of monitoring is represented by dummy variables that equal one, starting in the quarter when the measure is implemented. We do so, because there is no consistent measure of the intensity of monitoring in the data. We hypothesize that each time the authority introduces a new compliance measure the level of monitoring increases.

Table 3 shows the averages of the main variables used in estimation for each firm in the period for which we have data. There are differences among the firms' averages because the data used correspond to different length periods. Notice that the total taps are different from the total number of passengers because of the evasion.

Table 3: Average of main variables used in the estimation

Firm	Cost (millions $)	Taps (millions)	Bus lines	Driven kms	Labor Index	Diesel price	Dollar Ex. Rate
1	7.2	12.3	42.6	10,000	102.5	81.0	533.0
2	14.9	24.4	74.9	19,277	102.5	81.0	533.0
3	6.1	11.3	36.1	8,350	103.0	82.3	538.6
4	37.4	40.2	102.5	44,185	102.5	81.0	533.0
5	11.8	8.0	36.3	8,054	104.5	55.7	573.4
6	29.1	32.0	48.3	20,812	102.5	81.0	533.0
7	44.7	45.8	81.8	36,773	103.8	81.0	527.9
8	41.8	50.3	63.0	33,194	103.8	81.0	527.9
9	23.7	31.1	36.8	20,701	103.8	81.0	527.9
10	8.2	5.2	46.0	2,385	106.8	78.7	518.6
Total	26.4	31.3	60.5	24,256	103.2	80.4	532.2

3. Modelling approach

This section presents the modelling approach used to represent the behaviour of the firms. The objective is to obtain an expression relating the unobservable (for the econometrician) variables with the data. These unobservable variables are of two types: due to asymmetric information (effort and quality) and technological heterogeneity (efficiency level). The former are endogenous variables and its value is the results of the firm maximizing profit behaviour. Thus, in what follows, we express these variables as a function of the observables variables and technological heterogeneity.

In order to model the firm's behaviour, we follow the Dalen and Gomez-Lobo (1997)'s structural modelling approach to introduce efficiency level and manager's effort in the cost function of the bus firms. This approach takes into account information asymmetries that characterize regulated industries.

The basic modelling assumption is that firm faces some degree of productive inefficiency and the manager of the firm may reduce it by exerting effort. As the effort is costly for the manager, there is a trade-off between inefficiency and effort when manager maximizes firm's profits. A second assumption is that the level of service is chosen by the manager and is costly. As a result of these assumptions, effort and quality are substitutes and used by the manager to reduce cost.

The model assumes that the firm's cost function depends on the price of inputs (*w*), the level of capital (*K*), output (*Q*), efficiency (θ), effort (e_0), and a quality index (*q*) such that total cost equals $C(w, K, Q, \theta - e_0 + q)$. Notice the cost depends on the net level of efficiency, $\theta - e_0$. The inefficiency parameter, θ, is positive and such that the greater it is, the less efficient the firm is.

Also, it is assumed that effort net of quality (e_0-q) has a cost for the manager which can be expressed in monetary terms by the function $\psi(e_0-q)$. The cost effort function satisfies $\psi'(e_0-q)>0$, $\psi''(e_0-q)>0$, and $\psi'''(e_0-q)\geq 0$. Notice that improving the quality increases the costs, but such cost rise may be compensated by the manager's effort. The net effort is positive, if the manager is able of compensating the impact of quality improvement on cost, and negative, otherwise. In what follows, we denote net effort as *e*.

Then, for a fixed-price contract with transfer *T*, the firm chooses the level of net effort that maximizes its profits. The firm's problem is

$$\max_{e}\left\{T - C(w, K, Q, \theta - e) - \psi(e)\right\}. \tag{1}$$

The first order condition for problem (1) is

$$-\frac{\partial C}{\partial e} = \psi'(e). \tag{2}$$

Therefore, optimal level of effort is a function of input prices, capital, output, and efficiency levels, $e^*(w, K, Q, \theta)$. By replacing the optimal level of effort in the cost function, we get a function depending only on observable variables and the efficiency parameter.

To identify and estimate the model, we assume that the cost is given by a Cobb-Douglas function, and that net effort (e) and efficiency parameter (θ) enter the cost function as a global effect on productivity. In contrast, several authors assume that labour is the factor that induces heterogeneity in firms' productivity, and that, at the same time, can be changed by the manager's effort (Wolak, 1994; Dalen and Gomez-Lobo, 1997; Gagnepain and Ivaldi, 2002a). This differentiation is only a matter of identification and does not imply substantial differences in the interpretation of the results. Therefore, if there are two inputs, labour (L) and materials (M), and a fixed capital level (K), the cost function is

$$C(w_L, w_M, K, Q, \theta - e) = \beta w_L^{\beta_L} w_M^{\beta_M} K^{\beta_K} Q^{\beta_Q} \exp(\theta - e) , \tag{3}$$

where w_j is the price of factor j, and β, β_L, β_M, β_K, β_Q are parameters of the model.

For estimation it is also needed a functional form for the cost of net effort. We assume the following function

$$\psi(e) = \alpha\left[\exp(e) - 1\right], \quad \alpha > 0 . \tag{4}$$

Therefore, the optimal net effort level is

$$e^* = \frac{1}{2}\left(\ln \beta + \beta_L \ln w_L + \beta_M \ln w_M + \beta_K \ln K + \beta_Q \ln Q + \theta - \ln \alpha\right), \tag{5}$$

and the firm's cost evaluated on it is

$$C(w_L, w_M, K, Q, \theta - e^*) = \left[\alpha\beta w_L^{\beta_L} w_M^{\beta_M} K^{\beta_K} Q^{\beta_Q} \exp(\theta)\right]^{1/2} . \tag{6}$$

Equation (6) relates the data on cost, prices, capital and output with the parameters of the cost function without the effect of the endogenous effort. The left hand side of equation (6) corresponds to the observed cost in the sample of firms.

To model the effect of monitoring on the effort, we assume that the optimal level of net effort change as a function of the level of monitoring. We also assume the level of monitoring increases every time the regulator introduce a change in the index used to monitor the firms, but do not assume any functional form. This allows us to avoid the definition of a continuous variable representing the level of monitoring and depend on any functional form. Therefore, the monitoring effects on net effort are modelled by means of dummy variables as follows

$$e_t = e^* + \sum_m \delta_m d_{mt} , \tag{7}$$

where δ_m is the effect of the change in monitoring level m to be estimated, and d_{mt} is a dummy variable that equals one if the monitoring level changes at or before period t, and zero otherwise. This way, a positive value for d_m means the cost reducing effort is greater than the quality increase led by the monitoring level increment.

Replacing the equation (7) in the cost function (6), we get the following expression for the cost (in logarithm) in any period t.

$$\ln C_t = \frac{1}{2}\left[\ln \alpha\beta + \beta_L \ln w_{Lt} + \beta_M \ln w_{Mt} + \beta_K \ln K_t + \beta_Q \ln Q_t + \theta_t - \sum_m \delta_m d_{mt} \right]. \quad (8)$$

4. Estimation method

For estimation, we add a random term to the cost function presented in equation (8). This way, our model turns out to be a stochastic cost frontier (Gagnepain and Ivaldi, 2002b). Hence, the efficiency parameter is estimated as an exogenous effect. The net effort is an endogenous effect, but the structural approach allows us to get rid of it as described in previous section.

For the firm i in period t, we model the cost frontier as a true fixed effect model (Greene, 2005). By doing so, the firms' (systematic) heterogeneity due to differences in route structure, geographical conditions of operation areas or another technological condition can be represented by the model and does not bias the estimation. If there no such systematic heterogeneity, fixed effects should be not statistically different each other.

Therefore, the model to estimate is

$$c_{i,t} = \frac{1}{2}\left[\beta_{0i} + \beta' z_{i,t} + \theta_{i,t} \right] - \delta' d_t + \varepsilon_{i,t}\ , \quad (9)$$

where $c_{i,t}$ is the logarithm of firm's cost in period t, $z_{i,t}$ is the vector of input prices, capital and output, β_{0i} is the fixed effect, β is the vector of parameters. As mention above, δ is the vector of parameters representing the effects of monitoring on net effort. $\theta_{i,t}$ is the exogenous efficiency level, which is a random term distributed half-normal $N^+(0,\sigma_\theta)$. $\varepsilon_{i,t}$ is the random symmetric disturbance iid normal $N(0,\sigma_\varepsilon)$. Notice also that the parameter β_{0i} comprises α and β from the structural model ($\beta_{0i} = ln(\alpha\ \beta)$), since they are not identified separately.

We impose homogeneity of degree one on input prices. This is a standard procedure for cost function estimation and guaranties theoretical consistency.

To specify the cost function, it should be considered that transport output is multidimensional. Indeed, it is recognized that transport output is defined by a flow vector with components identified by origin, destination, period and commodity type (Jara-Díaz, 1982). In the case of Transantiago, such description of the product is unfeasible because it implies an output vector of high dimension because a trunk service can have around 70 stops along its route. Hence, the output in the cost function is specified as an aggregated measure. In this case, the aggregated product is defined as the total passenger flow.

In addition, to take into account some characteristics of the multi-product nature of the problem, we adopt a hedonic output approach (Spady and Friedlaender, 1978). We define the output as a vector of aggregates with components describing different characteristics and technological factors of the output. These components are useful to calculate economies of scale unambiguously. On this respect, as the aggregated output is the total passenger flow, the calculation of economies of scale is straightforward. Indeed, the degree of multiproduct economies of scale (*S*) is equal to the inverse of the cost elasticity to aggregated output (Jara-Díaz and Cortés, 1996). Therefore, the additional variables used to describe the output are the kilometres contracted during the period and the number of bus lines operated by each firm.

To estimate the parameters of the model we use the Chen et al. (2011)'s estimator, which is based on the within estimator used for linear panel data models. Chen et al. (2011) show the joint distribution of the symmetric random disturbance and the efficiency parameter belongs to the skew normal family. The authors derivate the corresponding likelihood function based only on the joint distribution of the deviations from the means. This estimator is not subject to the incidental parameter problem (Arellano and Honoré, 2001). The fixed effects are estimated using the estimates of the other parameters of the model as corrected residuals by the expectation of θ_{it} (see Chen et al, 2011).

The estimation method allows us to compute the average technical efficiency level among the firms by using the expectation of $\theta_{i,t}$. In addition, if we assume the fixed effects represent efficiency differences, and not only heterogeneity, we can compute the relative efficiency by using the expression $\eta_i = \beta_{0i} - min_j\{\beta_{0j}\} + \theta_i$. Also, we can compute relative efficiency grouping the firms according to their type: trunk or feeder. We can do the same with the cost distortion given by the expression $exp(\theta_i)$ or $exp(\eta_i)$.

5. Results

This section presents the results of the estimation of the structural model and the standard cost frontier model. We compare the results of both approaches. Also, this section presents the relative efficiency level of the firms.

After a number of preliminary estimations, the variables to include in the model are ridership and number of bus lines operated by firm to describe the output, price of labour, diesel and dollar exchange rate as price of inputs, and number buses as capital.

In both models, all parameters are significant with 95% of confidence (Table 4), and the goodness of fit is high with R^2 equal to 0.97.

Cost function parameters exhibit important differences between models. The most relevant difference between the models is the degree of economies of scope. The structural model exhibits constant returns to scale, whereas the standard model exhibits significant economies of scale (2.06). The latter effect is consistent with findings of Dalen and Gomez-Lobo (1997), and Gagnepain and Ivaldi (2002b). Theory also suggests down bias in the standard model due to the endogeneity of effort, which is controlled in the structural model.

The elasticity of diesel price in structural model is more than twice as that in standard model. The elasticity to labour is 0.08 in the structural model and 0.48 in the standard model. Elasticity to dollar exchange rate varies little from one model to another.

Table 4: Results of stochastic frontier cost model estimation

Parameters	Structural Model		Standard Model	
	Estimates	Std. err.	Estimates	Std. err.
Diesel	0.6979	0.0507	0.2875	0.0247
Dollar Ex. Rate	0.2180	0.1331	0.2353	0.0447
Labor	0.0841	-	0.4772	-
Buses	0.7038	0.1467	0.7141	0.0626
Passengers	1.0105	0.0679	0.4859	0.0351
Services	0.1727	0.0468	0.1187	0.0241
δ_1 (ICPH)	0.1735	0.0368		
δ_2 (ICFR)	-0.1871	0.0315		
δ_3 (ICPKH)	-0.1464	0.0291		
σ_e	0.1639	0.0064	0.0840	0.0030
σ_θ	0.2750	0.0095	0.1485	0.0045

The effect of monitoring in the cost through the net effort is significant. The parameters point out the introduction of the first compliance measure (ICPH) resulted in a significant increase of the cost reducing effort that compensate the increment on cost due to the increment on level of service. Indeed, the net cost reduction is 8% when the ICPH is implemented. By contrast, the following compliance measures do not exhibit the same effect and increase the cost. When ICF and ICR are implemented, the net effect on cost is an increment of 1%. In the case of ICPKH, the most demanding measure, costs increase

9%. This means that for the managers is not profitable to increase the effort and offset the cost increment due to required higher quality.

Regarding the average level of efficiency in the period of analysis, Table 5 shows the relative levels for different assumptions on the source of heterogeneity, based on the structural model. As we cannot identify the parameter α, we cannot compute the total cost distortion ($exp(\theta_i - e_i)$), but we present the cost distortion due to technical inefficiency ($exp(\theta_i)$). Also, we present the cost distortion assuming the heterogeneity represented by the fixed effects is due to technical inefficiency. This means deviations from the more efficient firms are a measure of the relative technical inefficiency ($\beta_i = \beta_{0i} - min_j\{\beta_{0j}\}$), which leads to relative technical cost distortion ($exp(\theta_i + \beta_i)$). In addition, we group the firms according to the type of service for compute relative cost distortion.

Table 5: Fixed effects estimates, technical inefficiency, cost distortion and relative cost distortions

Firm	Fixed effect (b_{0i})	Technical inefficiency	Technical cost distortion	Relative cost distortion		
				Global	Feeder	Trunk
1	3.8266	0.125	1.136	1.374	1.374	
2	4.4133	0.134	1.151	2.503	2.503	
3	3.6367	0.119	1.128	1.128	1.128	
4	5.1694	0.130	1.142	5.289	5.289	3.069
5	5.5450	0.180	1.214	8.186	8.186	
6	4.4746	0.152	1.177	2.721		1.579
7	4.4642	0.129	1.139	2.606		1.512
8	4.2987	0.136	1.152	2.233		1.296
9	4.1809	0.143	1.157	1.994		1.157
10	4.1406	0.190	1.222	2.023	2.023	

Technical cost distortion is 16% on average (Table 5). Trunk services exhibit higher level of cost distortion than the feeders do. Firms 10 and 5 have the higher levels of cost distortion, 22% and 21% respectively. The former is the newest operator in the sample with only two quarters of operation, which could explain the level inefficiency, if we consider a period of initial adjustment. The latter was insolvent in 2010-III, which is consistent with its low level of efficiency.

When we consider the fixed effect as technical heterogeneity and include the differences as relative inefficiency the average cost distortion attaint 300%. Firms 4 and 5 drive this increment as they go bankrupt in 2011-III and 2010-III respectively (Table 5). Excluding these firms, the average cost distortion for feeder services is 76% and for trunk services is 39%. Notice that firm 4 operates two units of feeder services and one unit of trunk services, thus the cost distortion is computed for both cases.

We remark that the cost distortions presented in Table 5 are not net of cost reduction effort. This means the firms may attain this level of cost only if they do not exert any effort to compensate the technical inefficiencies. Since the effort is costly for managers, we conjecture the ranking of efficient firms remain unchanged after include effort, particularly for the less efficient ones.

6. Conclusions and policy implications

Our model allows us to obtain cost function parameters unbiased because of we take into account the endogeneity of the effort and quality chosen by the firms' managers. One important result is the industry exhibits constant returns to scale. In contrast, the model that does not consider the effort lead to wrong conclusion of significant economies of scale.

Also, we are capable to measure the effect of monitoring the quality on the effort exerted by the managers. We observe that the introduction of the first compliance measure (ICPH) lead to the firms to increase quality to fulfil the contracts and to increase the effort to reduce cost. The net effect is a reduction of cost. The following control measures do not have the same effect and the result is a rise in costs. This shows that the cost of effort is too high and for firms' managers is optimal do not exert more effort to compensate the increase in quality.

Our results on measures of efficiency are consistent with observed. The most inefficient firms go bankrupt in the period of analysis or just before. Regarding the technical inefficiency, the average cost distortion is 16% above the frontier.

This paper suggests that when contract are incomplete and the service level is not adequately incorporated in the contract, fixed-price contracts may not be efficient. So, authorities are advised to design contracts that allow certain level of quality monitoring when deciding for fixed-price contracts. Also, monitoring proves to be an effective tool to incentivise cost reduction efforts without hindering the service level.

The results of this paper allow us to do further research on the design of optimal incentive contracts to apply in the case of Transantiago. In particular, a topic for research is the amount of fines given the provided level of service below the contracted level.

References

Arellano, M., Honoré, B., 2001. Panel data models: some recent developments. In: J.J. Heckman & E.E. Leamer (Eds.), *Handbook of Econometrics*, Vol. 5, pp. 3229–3296, Elsevier.

Banco Central de Chile, 2012. Estadísticas, Series de Indicadores, http://www.bcentral.cl/estadisticas-economicas/series-indicadores/index_erd.htm. Accessed March 20, 2012.

Basso, L.J., Jara-Díaz, S.R., and Waters II, W.G., 2011. Cost Functions for Transport Firms. In: de Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), *A Handbook of Transport Economics*, Edward Elgar Publishing Ltd., Nothampton.

Beltrán, P., Gschwender, A., Palma, C., 2013. The Impact of Compliance Measures on the Operation of a Bus System: the Case of Transantiago. *Research in Transportation Economics*, 2013, 39(1), pp. 79-89. http://dx.doi.org/10.1016/j.retrec.2012.05.026. Accessed Jul. 15, 2013.

Chen, Y., Schmidt, P, Wang, H.J, 2011. Consistent estimation of fixed effects stochastic frontier model. Working paper, Michigan State University.

Coordinación de Transportes de Santiago, 2012. Informe Financiero de Transantiago, Periodo Enero 2010 – Abril 2012, http://www.coordinaciontransantiago.cl/corporativo/. Accessed May 30, 2012.

Dalen, M.D., Gomez-Lobo, A., 1997. Estiamting cost functions in regulated industries characterized by asymmetric information. *European Economic Review* 41, pp. 935–942.

Gagnepain, P., Ivaldi, M., 2002a. Incentive Regulatory Policies: the Case of Public Transit Systems in France. *RAND Journal of Economics* 33(4), pp. 605-629.

Gagnepain, P., Ivaldi, M., 2002b. Stochastic frontiers and asymmetric information models. *Journal of Productivity Analysis* 18, pp. 145–159.

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

Jara-Díaz, S.R., 1982. The Estimation of Transport Cost Functions: A Methodological Review. *Transportation Review* 2, pp. 257–278.

Jara-Díaz, S.R., Cortés, C., 1996. On the Calculation of Scale Economies from Transport Cost Functions. *Journal of Transport Economics and Policy* 30, pp. 157–170.

Laffont, J.J., Tirole, J., 1991. Provision of quality and power of incentive schemes in regulated industries. In: Barnett, W.A., Cornet, B., d'Aspremont, C., Gabszewicz, J.J., Mas-Colell, A. (Eds.), *Equilibrium theory and applications: Proceedings of the sixth international symposium in economic theory and econometrics*, Cambridge University Press.

Muñoz, J.C., Gschwender, A., 2008. Transantiago: A Tale of Two Cities. *Research in Transportation Economics* 22, pp. 45-53.

Superintendencia de Valores y Seguros, 2012. http://www.svs.cl/sitio/index.php. Accessed several times between March and June 2012.

Spady, R., Friedlaender, A.F., 1978. Hedonic Cost Functions for the Regulated Trucking Industry. *Bell Journal of Economics* 9, pp. 159–79.

Wolak, F.A., 1994. An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction. *Annals of Economics and Statistics* 34, pp. 13–69.