DESIGNING INCENTIVE CONTRACTS FOR QUALITY PROVISION IN PUBLIC TRANSPORT PROCUREMENT: THE CASE OF SANTIAGO, CHILE

Marco Batarce
Facultad de Ingeniería y Ciencias, Universidad Diego Portales
marco.batarce@udp.cl

ABSTRACT

The goal of this research is to design an optimal incentive contract for public transport procurement with microeconomic foundation and based on the model by Laffont and Tirole (1993). This contract takes into account the asymmetries of information between the operating companies and the transport authority. The needed parameters for the contract design are estimated using data on costs and demand from Santiago. The resulting optimal contract is transformed into a menu of linear contracts, which allows the authority to decentralize its implementation. The comparison of the menu of contracts and the actually implemented contract shows that the latter is a very low-powered incentive scheme for both cost reduction and quality provision.
Keywords: contracts, incentives, public transport.

RESUMEN

El objetivo de esta investigación es diseñar un contrato de incentivo óptimo para la provisión de transporte público con base microeconómica y basado en el modelo de Laffont y Tirole (1993). Este contrato toma en cuenta las asimetrías de información entre las empresas operadoras y la autoridad de transporte. Para ello se utilizan datos de costos y demanda de Santiago y se estima los parámetros necesarios para el diseño del contrato. El contrato óptimo resultante es transformado en un menú de contratos lineales que permite a la autoridad descentralizar su implementación. Al comparar el menú de contratos con los contratos actualmente implementados se muestra que estos últimos tienen bajo poder de incentivos tanto para reducir costos como para aumentar la calidad del servicio.
Palabras clave: contrato, incentivo, transporte público.

1. INTRODUCTION

A major concern in procurement and contract design is how to give incentives for efficient provision of goods or services. The problem for the regulator is that the cost reduction effort exerted by the firms providing the public good is not observable. When the regulator's concern is reduction of costs, fixed-price contracts are more efficient (Laffont and Tirole, 1993). Gagnepain and Ivaldi (2002) confirm this idea by showing that firms providing the service under fixed-price contract exhibit a bigger effort on cost reduction, since these firms are residual claimant of operational cost savings. Nevertheless, Laffont and Tirole (1991) show that when quality and cost reducing effort are substitutes, firms may reduce costs by lowering quality. If quality is not verifiable, fixed-price contracts may lead the firm to provide low quality, as firms only focus on lowering costs. Moreover, if a large fraction of population has no alternative to public transport and the demand has low elasticity to price and quality, under a fixed-price contract, the bus service operators have incentives to increase profit by reducing cost at the expense of quality. If quality is verifiable, the regulator may establish a monitoring scheme and force a minimum quality level, recovering the incentive power of this type of contract. For instance, in Santiago between 2007 and 2011, initially the contracts with bus operators were fixed-price contracts, with no penalties in case of failing to fulfil the operation plan. By mid-2007, the authority implemented a program of compliance measures and fines to enforce the fulfilment of the contracts, increasing the operating fleet from 4,600 to 5,800 and the bus frequency, a dimension of service quality.

Consequently, the regulator faces a trade-off between developing costly quality performance measures and giving the firm monetary incentives. The extreme case is the implementation of cost-plus contracts. If quality is verifiable, the firm has no incentives for cost reduction by reducing quality. Notwithstanding, if quality is unverifiable and the regulated price is divorced from realized costs, the firm have incentive to increase its margin by lowering quality. In general, if quality is not verifiable, and it is not possible to base rewards and penalties explicitly on the realized quality, a strong conflict can arise between creating incentives for cost reduction and creating incentives to increase quality. The optimal resolution of this conflict depends upon the relative importance of reducing costs and increasing quality as shown by Laffont and Tirole (1993).

The goal of this research is to design an optimal contract for public transport procurement with microeconomic foundation and based on the model by Laffont and Tirole (1991). This contract takes into account the asymmetries of information between the operating companies and the transport authority. For this I use data on cost and demand from Santiago and estimate the parameters needed for the contract design. The methodology includes the estimation of a structural cost model that takes into consideration three sources of asymmetric information: level of effort, level of quality and efficiency (Dalen and Gomez-Lobo, 1997; Gagnepain and Ivaldi, 2002, Batarce and Galilea 2014). Data correspond to an unbalanced panel of firms operating between 2007 and 2010 in Santiago, Chile. The resulting contract is compared with the level of fines set by the authority to enforce the fulfilment of the contracts in 2007.

2. MARKET DESCRIPTION AND DATA

The bus system was designed as a bus network of feeder and trunk services, and a metro system with four lines. The operation of the bus system was organized into fourteen groups of services.

Five of them consisted of trunk services (hereafter identified by T1 to T5) and nine consist of feeder services (hereafter identified by F1 to F9). The trunk services had exclusive operation in main roads in the city, and their routes connect the edges of the city with the Central Business District (CBD) or go across the city from one extreme to another. The feeder services operate in exclusive zones (mainly residential or with low density of land use) and connect to the trunk services and the metro system. Because of the new feeder-trunk structure, most passengers transfer between buses and metro lines, with 1.6 legs per trip on average. The only means of paying the bus fare is a smartcard, which permits fare integration over the entire trip (among buses and metro).

Each group of services, both feeder and trunk, was tendered and awarded to the highest bidder that fulfilled the requirements. The administration of the fare transactions generated by the system was also tendered and awarded to a consortium of banks and an information technology firm.

Every group of services was awarded to only one firm, but one firm can operate more than one group of services. At the beginning of the period of analysis, there were ten firms. Four firms operated only trunks (T1, T2, T4 and T5), five firms operating only feeder services (one of them operated two groups F2 and F4) and one firm (Firm 4, see Table 1) operating two feeder groups and one trunk group of service (F5, F8 and T3). By the end of 2008, in a new tender, the Firm 4 lost the concession of the feeder group F5, which was awarded to an entrant firm. In the last quarter of 2009, the Firm 4 could not continue with the operation of the trunk group T3, which was awarded to another entrant firm. In October 2011, the Firm 4 became insolvent and exited the market; its feeder group was awarded to an incumbent firm operating another feeder group. Thus, in the period of analysis, there are twelve firms operating in the market. Data from nine of these firms is used for the estimation of the cost function.

When Transantiago started in February 2007, the contract established an operation fleet of 5,600 buses. In the second quarter of 2007, however, the actual observed operating fleet was 4,600 buses (Beltran et al., 2012). The problem was that the contracts only established a commitment on terms of following the operation plan, but no punishment in case of failing to follow it. By mid-2007, the transport authority implemented a program of compliance measures and fines to enforce the fulfillment of the contracts. Such a program increased the operating fleet from 4,600 to 5,800 buses in only five months (Beltran et al., 2012). This implies that the actual total driven kilometers differs from those established in the contracts, mainly for the three first quarters of 2007.

Although Transantiago was designed to operate without subsidies, the new conditions on the firms along with the new operation scheme and the fare integration led to a higher cost than the estimated one. For instance, formal contracts for drivers increased labor costs, and the integrated fare implied additional collection and management costs. In addition, fare integration reduced the revenues of the system. Users tap in to enter each bus of the network and may make to two transfers paying a single fare during two hours as long as they do not repeat the same bus line. Moreover, the Government decided to not increase the fare during the first two year of operation and to use the subsidy to cover the operational deficit. As a consequence, the total subsidy has incremented gradually since 2007 and reached around US$50 million per month (40% of the total costs) in 2012.

Besides the cost increment because of the increase of driven kilometers and bus fleet, there has been a significant increment of fare evasion in bus services. By the end of 2011, the fraction of passengers evading the fare payment reached 23%. There are many reasons for this evasion. One of them is that fares increased by 50% between 2010 and 2012, and paying the fare twice on every workday represents around 15% of their income for a significant fraction of users. Moreover, most public transport users have no alternative for travelling and the bus travel demand exhibits low price elasticity; in consequence, the high fares may increase the propensity to evade the payment.

To estimate the cost function, I use the same data used by Batarce and Galilea (2017), which comprise many different types of information involved in our estimation: costs, input prices, transport demand, operation variables and level of fines imposed because of the nonfulfillment of contracted kilometers, which is associated to quality level. Detailed description of the available information is reported in Batarce and Galilea (2017). Data corresponds to an unbalanced panel of firms, where some firms have 13 observations whereas others have as few as 3 observations (Table 1). Table 2 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. The cost is obtained from firms' financial reports, the variables related to bus operation from the transport authority, and the input prices (labor, diesel and dollar) from Chilean Central Bank's statistics.

Table 1: Number of observations for each firm in the sample

Firm	Available data		No. Obs.	Operated group of services
	From (year-quarter)	To (year-quarter)		
1	2007-4	2010-1	10	F1
2	2007-4	2010-1	10	F2, F4
3	2008-1	2010-1	9	F3
4	2007-4	2010-1	10	F5, F8, T3
5	2009-1	2009-3	3	F9
6	2007-4	2010-1	10	T1
7	2007-4	2010-4	13	T2
8	2007-4	2010-4	13	T4
9	2007-4	2010-4	13	T5

Table 2: Quarterly average of main variables used in the cost function estimation

Firm	Cost (million CL$)	Smartcard transactions (million)	Fare evasion rate (%)	Bus lines	Driven kilometers (million)	Labor Index	Diesel Price (CL$/lt)	Dollar Ex. Rate (CL$/US$)
1	7.2	12.3	14.6	42.6	10.0	102.5	81.0	533.0
2	14.9	24.4	14.6	74.9	19.3	102.5	81.0	533.0
3	6.1	11.3	15.0	36.1	8.4	103.0	82.3	538.6
4	37.4	40.2	14.6	102.5	44.2	102.5	81.0	533.0
5	11.8	8.0	15.0	36.3	8.1	104.5	55.7	573.4
6	29.1	32.0	14.6	48.3	20.8	102.5	81.0	533.0
7	44.7	45.8	15.3	81.8	36.8	103.8	81.0	527.9
8	41.8	50.3	15.3	63.0	33.2	103.8	81.0	527.9
9	23.7	31.1	15.3	36.8	20.7	103.8	81.0	527.9
Full sample	26.4	31.3	14.9	60.5	24.3	103.2	80.4	532.2

3. OPTIMAL INCENTIVE CONTRACT FOR COST REDUCTION AND QUALITY PROVISION

When designing an incentive contract, the regulator must find a combination of output *q* and net monetary transfer *t*, in addition to the reimbursement of the cost, such that the firm agree on trade. The contract is represented by a pair (*q*, *t*), but it may include other verifiable variables (this idea is clarified below). The contract is determined according to the regulator's objective and the available verifiable information.

3.1 Incentive contracts

Following Laffont and Tirole (1993), suppose that the cost of public transport provision depends on output level q, firm's efficiency level θ, and effort to reduce the inefficiency e. Then, the firm's cost C may be represented by a function such that $C=C(q,\theta-e)$. The effort exerted by the firm to reduce the inefficiency is also unobservable by the regulator, and induces a disutility for the firm. Suppose the disutility of effort that the firm incurs (in monetary terms) is $\psi(e)$, with $\psi' > 0$, $\psi'' > 0$, and $\psi(0) = 0$. In general, the efficiency parameter θ and the exerted effort e are firm's private information, therefore the regulator or transport authority cannot design a contract contingent on them.

Suppose also that an output level of q values for consumer $S(q)$, with $S' > 0$ and $S'' < 0$, and an utilitarian regulator that wants to maximize total welfare when designing the contract. Total welfare comprises the net surplus of consumer/taxpayer and the firm's utility. The latter is given by

$$U = t - \psi(e). \tag{1}$$

Let $\lambda > 0$ denote the shadow cost of public funds. This means that taxation inflicts disutility \$ ($1+\lambda$) on taxpayers to levy \$1 for the state. The net consumer surplus is $S(q) - (1+\lambda)(t - C(q,\theta - e))$. Then, for the utilitarian regulator, the social welfares is

$$\begin{aligned} W &= S(q) - (1+\lambda)(t - C(q,\theta - e)) + t - \psi(e) \\ &= S(q) - (1+\lambda)[C(q,\theta - e) + \psi(e)] - \lambda U. \end{aligned} \tag{2}$$

The firm accepts the contract with the regulator only if it obtains at least as much utility as outside the relationship. If the outside opportunity level of utility or reservation utility is normalized to zero, the firm's individual rationality constraint is

$$U = t - \psi(e) \geq 0. \tag{3}$$

Suppose that the efficiency parameter and the effort are observable, therefore the regulator may design a contract that implements the first-best solution under complete information. This means, to find optimal levels of output and effort, and a net monetary transfer. Indeed, if effort is contractible and efficiency observable, the regulator solves

$$\max_{\{q,e,U\}} \left\{ S(q) - (1+\lambda)[C(q,\theta-e) + \psi(e)] - \lambda U \right\}$$
$$s.t. \quad U \geq 0$$

The solution of this problem is

$$S'(q^*) = (1+\lambda)\frac{\partial C}{\partial q}(q^*, \theta - e^*),$$
$$\psi'(e^*) = -\frac{\partial C}{\partial \theta}(q^*, \theta - e^*),$$
$$U = t - \psi(e^*) = 0.$$

The first equation is the usual condition for efficient production: marginal cost equal to marginal benefit. The second equation implies that the marginal disutility of effort must be equal to marginal cost savings by improved efficiency. The third equation implies that the firm receives no rent because of the shadow cost of public funds.

According to Laffont and Tirole (1993), the first-best solution is implementable with a fixed-price contract. To see this, suppose that the regulator offers a fixed gross transfer T for a output level q^*, then the firm solves the problem

$$\max_{e} \left\{ T - [C(q^*, \theta - e) + \psi(e)] \right\}.$$

The solution is

$$\psi'(e) = -\frac{\partial C}{\partial \theta}(q^*, \theta - e^*).$$

To leave the firm with no rent, the regulator needs to set the gross transfer equal to $T = C(q^*, \theta - e^*) + \psi(e^*)$.

When regulator cannot observe the efficiency parameter neither the effort exerted by the firm, the solution is to use incentive contracts (Laffont and Tirole, 1993, Laffont and Martimort, 2001). In practice, however, contracts for public transport provision are fixed-price or cost-plus. As shown, the former type of contract induces the firm to exert the optimal effort level. It is said the contract is a high-powered incentive scheme. As the regulator, however, does not have information on the firm efficiency, the firm may obtain positive rents because of its informational advantage. It should be noted that if there is not shadow cost of public funds, this rent will not be a problem for the regulator.

The cost-plus contract, on the contrary, induces no effort from the firm. Indeed, if the monetary transfer is the realized cost plus a fee t, the firm chooses the effort by maximizing $C + t - [C + \psi(e)] = t - \psi(e)$, which leads to zero effort. It is said that the contract is a low-powered incentive scheme. Therefore, if the regulator's concern is to produce efficiency, the best choice is a fixed-price contract.

3.2 Concern for quality provision

If the regulator is concerned for the quality of public transport services, besides efficient production, the fixed-price contracts are not the right instrument. As the firm is residual claimant for its cost savings, it has incentives to reduce costs by lowering quality. When quality is observable, but unverifiable, Laffont and Tirole (1991, 1993) propose an incentive contract with monetary transfers depending on an index of quality, corrected by price, which is measured by the demand. As suggested by Sappington (2005), with this mechanism firm's revenue reflects realized cost, which induces the firm to increase demand by increasing quality, without incurring in loses because of higher cost of quality provision. The performance-based contract proposes by Hensher and Wallis (2005) is an incentive contract for quality provision in the same spirit as the proposed one by Laffont and Tirole (1991, 1993).

However, if the quality is unobservable for the consumer or demand is inelastic to quality, the mechanism based on incentives does not work because firm's cost and revenue are divorced. In this case the firm has no incentive to increase quality because demand is not responsive. With an incentive contract, the monetary transfer to the firm becomes independent of the realized cost of quality provision, and the only incentive is for cost savings. As quality is unverifiable, the firm reduces cost by lowering quality.

A way to ensure a minimum quality level is monitoring it, which may be carried out by the regulator by means of specialized staff and equipment. If quality is unobservable even for the regulator, control measures may be based on other variables correlated with quality or on variables that measure only some dimensions of service quality. In this case, the regulator receives a noisy signal of quality. For instance, in the case of public transportation, bus frequency is an observable and verifiable variable related to several dimensions of quality, such as waiting time and crowding. In practice, the problem is how to implement rewards and punishments regarding the level of this noisy signal of quality, because the uncertainty prevents the regulator from setting a single threshold for verifying the fulfillment of the standard[1]. For instance, small deviations under the minimum quality standard may be neglected, which implies no punishment for the firm. In turn, deviations above the minimum may not be rewarded. One solution is to implement fines and rewards as function of the difference between the signal of quality and a minimum level of it set by the regulator.

To model the contract when quality is a concern for the regulator, suppose that the cost depends also on a variable *s* that is either the provided quality or another variable used to measure it. Then, the cost is a function $C = C(q, s, q - e)$ increasing on *s*. Suppose also that an output *q* with quality *s* values for consumers $S(q,s)$, with $\partial S/\partial q > 0$, $\partial^2 S/\partial q^2 < 0$, $\partial S/\partial s > 0$ and $\partial^2 S/\partial s^2 < 0$.

Suppose again that the efficiency parameter and the effort are observable, along with the provided quality. The regulator may design a contract that implements the first-best solution under complete information by maximizing the social welfare $\left\{S(q,s) - (1+\lambda)[C(q,s,\theta - e) + \psi(e)] - \lambda U\right\}$ on (*q*, *s*, *e*, *U*), subject to the firm's rationality constraint $U \geq 0$. The first-order conditions for this problem imply

[1] This is a moral hazard problem and may be analyzed, for instance, as shown by Laffont and Martimort (2001).

$$
\begin{aligned}
&S_q(q^*,s^*) = (1+\lambda)C_q(q^*,s^*,\theta-e^*),\\
&S_s(q^*,s^*) = (1+\lambda)C_s(q^*,s^*,\theta-e^*),\\
&\psi'(e^*) = -\frac{\partial C}{\partial \theta}(q^*,s^*,\theta-e^*),\\
&U^* = t^* - \psi(e^*) = 0.
\end{aligned}
$$

When quality is not observable, a fixed-price contract does not implement the optimal effort neither the optimal quality. Indeed, under fixed-price the firm solves

$$
\begin{aligned}
&\max_{(e,s)} \left\{T - [C(q^*,s,\theta-e)+\psi(e)]\right\}\\
&\text{s.t. } s \geq s_0 ,
\end{aligned}
$$

where s_0 is a minimum quality level. If s is lower s_0 than the contract is terminated. The idea of this lower bound of quality is that the firm cannot degrade quality to unacceptable levels for consumers. For instance, in the case of water utilities, the minimum quality may be non-turbid tap water and better levels of quality may be related to concentration of minerals or pollutants invisible to the naked eye. In the case of frequency-based public transport, the lower bound of quality may be a minimum frequency under which the firm can blame on not providing the service at all, rather than a poor quality service. For instance, a bus line with contracted bus headway of 5 minutes may provide a headway of 7 minutes, which implies a reduction of 25% of frequency, without user notices this degradation because of high irregularity caused by bus traffic interaction. However, an increment of headway until 15 minutes can be detected despite the traffic congestion and irregularity, leading to an unacceptable quality for users.

Assuming that the cost is strictly increasing on quality ($\partial C/\partial s > 0$), the first order conditions for the firm´s problem are

$$
\begin{aligned}
-\frac{\partial C}{\partial \theta}(q^*,s,\theta-e) &= \psi'(e),\\
s &= s_0.
\end{aligned}
$$

If cost is not separable on quality, the firm's optimal level of effort will not be the first-best effort obtained under complete information.

Laffont and Tirole (1991) propose an incentive contract to induce the regulated firm to exert cost reducing effort and provide quality. Their model assumes that the quality is observable but unverifiable. Therefore, the sources of asymmetric information are four: the efficiency level of the firm (θ), the cost reducing effort (e), the provided quality level (s) and a demand parameter (ε) known to the firm, but unobservable to the regulator.

As the demand and the market price are observable, the regulator can infer the unobservable quality from them and reduce the asymmetric information to three variables. In addition, the functional form of cost and demand allows them to group the parameters θ and ε into one unknown parameter ξ. As the distributions of θ and ε are assumed common knowledge, the distribution of ξ is given by the convolution of these distributions.

Laffont and Tirole (1991) find the optimal incentive contract by maximizing the social welfare subject to the rationality constraint and the incentive compatibility constraint. The solution implies that, for a given level of effort, the optimal quantity, quality and price are the same as those in complete information case. The optimal contract inducing cost reducing effort only distorts the optimal effort with respect to the first-best case.

The incentive contract can be implemented by means of a menu of linear contracts as the following:

$$t(\xi, z) = a(\xi) + b(\xi)(z(\xi) - z) \tag{4}$$

where the net transfer to the firm t depends on the firm's type ξ and on a performance index z that summarizes the cost and the provided quality. $z(\xi)$ is the announced value of the index and z is the observed ex post value. This performance index depends on the functional forms of costs and demand models. For instance, Laffont and Tirole (1991) use a linear demand $p = a - bq + ks$ and a linear cost function $C = (\theta + s - e)q$, then z has the following form:

$$z = \frac{C}{q} - \frac{p + bq}{k}.$$

The parameter $b(\xi)$ is the share of overrun in the performance index and $a(\xi)$ a fixed payment. These parameters are given by the following conditions:

$$b(\xi) = \psi'(e^*(\xi)), \tag{5}$$

$$a'(\xi) = \psi'(e^*(\xi)) \cdot z'(\xi), \tag{6}$$

where $e^*(\xi)$ is the optimal effort from the solution of the regulator's problem with the incentive compatibility constraint. Next section presents an application of this optimal contract.

4. APPLICATION TO SANTIAGO BUS TRANSIT SYSTEM

To design an incentive contract according to Laffont and Tirole (1991), it is needed to know some parameters describing the firms' demand and costs. To this end, I use the demand function estimated by Batarce and Galilea (2017) and the cost function estimated by Batarce (2017).

Batarce and Galilea (2017) assume that demand depends on the fare, the frequency and the supplied capacity. They use the total driven kilometers as a proxy for frequency, since the driven kilometers are the only available aggregate information and, whenever the routes do not change significantly, kilometers are proportional to frequency. Batarce and Galilea (2017) measure capacity as the number of seats/standing places supplied by the bus fleet. In addition, they take into account the effect of monthly seasonality and firm heterogeneity (fixed effects). Both effects are modeled by means of dummy variables. The functional form that Batarce and Galilea (2017) estimate is a linear demand as follows:

$$q_{it} = \eta p_t + \gamma_1 s_{it} + \gamma_2 k_{it} + \gamma_3 t + \gamma_0 + \sum_{j=2}^{14} \phi_j f_{ji} + \sum_{m=1}^{11} \delta_m d_{mt} + \varepsilon_{it}, \qquad (7)$$

where q_{it} is the total smartcard transactions corrected by the fare evasion rate for the group of services i in the period t; p_t is the bus fare in period t; s_{it} is the driven kilometers for the group of services i in the period t; k_{it} is the number of supplied seats/standing places; t is a trend; f_{ji} is a dummy variable equal to 1 if the group of services i corresponds to j and 0 otherwise; d_{mt} is a dummy variable that is 1 if the period t corresponds to the month m (from January to November) and 0 otherwise. The term ε_{it} is an independent random error for the group of services i in the period t, which is interpreted as asymmetric information known by the operator and unknown by the regulator (and the econometrician). The estimated parameters η, γ_0, γ_1, γ_2, γ_3, ϕ_2,…, ϕ_{15}, are shown in Table 3, parameters δ_1,…, δ_{11} are not shown.

Table 3: Estimated parameters of the demand function (Batarce and Galilea, 2017).

Parameter	Variable	Estimate	t-test
η	Fare	-0.0036	-2.91
γ_1	Kilometers (million)	0.4551	7.11
γ_2	Seats/standing places (thousand)	0.0936	8.12
γ_3	Trend	0.0169	3.04
γ_0	Constant	2.8296	6.21
ϕ_2	Feeder 2	-0.1233	-0.79
ϕ_3	Feeder 3	1.0782	6.80
ϕ_4	Feeder 4	1.9382	11.65
ϕ_5	Feeder 5	-0.2746	-1.74
ϕ_6	Feeder 6	0.9012	4.86
ϕ_7	Feeder 7	-0.0485	-0.30
ϕ_8	Feeder 8	-0.0486	-0.30
ϕ_9	Feeder 9	-0.0084	-0.06
ϕ_{10}	Trunk 1	1.5829	3.31
ϕ_{11}	Trunk 2	0.5831	0.63
ϕ_{12}	Trunk 3	-1.2945	-3.57
ϕ_{13}	Trunk 4	1.3954	1.57
ϕ_{14}	Trunk 5	1.7697	3.82
σ_ε	Demand standard deviation	0.8333	

The cost function is estimated using the same methodology and data as Batarce (2017). This methodology overcomes the estimation problems caused by the endogeneity of cost reducing effort and quality level, both variables are unobservable. To solve the problem the method uses the first order conditions of the firm's profit maximization including the disutility of effort and the cost of fines due to under-provision of quality (contracted kilometers). The cost structure is modeled using the following function:

$$C_{it} = A_i w_{1t}^{\beta_1} w_{2t}^{\beta_2} w_{3t}^{\beta_3} \exp(\alpha_1 q_{it} + \alpha_2 s_{it} + \alpha_3 \ln b_{it} + \theta - e), \qquad (8)$$

where C_{it} is the firm i's total cost in period t, q_{it} is the number of passengers (smartcard transactions corrected by the fare evasion rate), s_{it} is driven kilometers, b_{it} is the bus fleet during the period, and w_{1t}, w_{2t}, w_{3t} are input prices of diesel, material for maintenance (measured with the dollar exchange rate) and labor (measure with the real labor cost index) during period t. Technology is represented by A_i and is specific for each firm, since it includes differences related to the geographical area where each firm operates. θ_{it} is the firm's efficiency level, which is private information. The firm may exert an effort e to reduce inefficiency and costs. The parameters to estimate are A_i, α_1, α_2, β_1, β_2, β_3 and the distribution of θ. As the cost function should be homogeneous of degree one in the input prices, the restriction $\beta_1 + \beta_2 + \beta_3 = 1$ is imposed. The distribution of θ is half-normal with standard deviation σ_θ. Estimated parameters are presented in Table 4.

Table 4: Estimated parameters of the cost function.

Parameter	Variable	Estimate	t-test
α_1	Passengers (million)	0.0205	9.51
α_2	Kilometers (million)	0.0169	5.49
α_3	Buses	1.0636	8.04
β_1	Fuel	0.4561	9.78
β_2	Dollar exchange rate	0.2974	4.01
β_3	Labor	0.2465	-
A_1	Firm 1's fixed effect	5.0704	31.45
A_2	Firm 2's fixed effect	5.9333	44.80
A_3	Firm 3's fixed effect	4.7616	24.82
A_4	Firm 4's fixed effect	6.0880	21.92
A_5	Firm 5's fixed effect	6.5607	12.05
A_6	Firm 6's fixed effect	5.5065	19.01
A_7	Firm 7's fixed effect	4.8857	23.51
A_8	Firm 8's fixed effect	4.9548	24.44
A_9	Firm 9's fixed effect	5.0352	25.16
σ_θ	Efficiency standard deviation	0.0853	-11.58

In addition, to consistently estimate the cost function in Eq.(8), the disutility of effort is specified as:

$$\psi(e) = K[\exp(e) - 1] \qquad (9)$$

The parameter K is not identified separately form the rest of parameters in Eq.(8), but this lack of identification does not prevent us from designing the optimal contract.

To compute the linear contract according to Laffont and Tilrole (1991), it is necessary to determine the functional form of the performance index z, which depends on the functional form of demand

and costs. The index satisfies the relationship $z = \xi - e$, where ξ is the convolution of the firm's cost and demand private information represented by the random variables θ in Eq. (8) and a linear transformation of ε in Eq. (7) respectively. This relationship is obtained after writing the provided quality as function of the remained terms of the demand and replacing it into the cost function. In this way the two sources of asymmetric information can be reduced into one. The variable ξ is given by:

$$\xi = \theta - \frac{\alpha_2}{\gamma_1} u \tag{10}$$

Therefore, the performance index z as a function of parameters of demand and costs is given by the following equations:

$$z = \ln C - \ln C_0 - B_1 q - B_2 p + B_3 \tag{11}$$

$$C_0 = A\, w_1^{\beta_1} w_2^{\beta_2} w_3^{\beta_3} \tag{12}$$

$$B_1 = \alpha_1 + \frac{\alpha_2}{\gamma_1} \tag{13}$$

$$B_2 = -\frac{\eta \alpha_2}{\gamma_1} \tag{14}$$

$$B_3 = -\frac{\alpha_2}{\gamma_1}\left(\gamma_0 + \gamma_2 k + \gamma_3 t\right) \tag{15}$$

By assuming that ε is normally distributed with zero mean and standard deviation σ_ε and θ follows a half-normal distribution with standard deviation σ_θ, then the distribution of ξ turns to be skewed normal with shape parameter $\gamma_1 \sigma_\theta / (\alpha_2 \sigma_\varepsilon)$ (Azzalini, 1985). In what follows, F denotes the probability distribution of ξ and f denotes its density.

The parameters for the optimal linear contract are defined by the following equations:

$$b(\xi) = \left[K C_0 \exp(B_1 q + B_2 p - B_3 + \xi) G(\xi) \right]^{1/2} \tag{16}$$

$$a'(\xi) = \frac{1}{2}\left[1 - \frac{\lambda}{1+\lambda} H'(\xi) G(\xi)\right] b(\xi) \tag{17}$$

$$G(\xi) = \left(1 + \frac{\lambda}{1+\lambda} H(\xi)\right)^{-1} \tag{18}$$

$$H(\xi) = \frac{F(\xi)}{f(\xi)} \tag{19}$$

The contract is a function of the total cost C and the demand q, since the fare is fixed by the regulator, and the other terms in z do not depend on the performance of the firm (effort and provided quality). The transfer is:

$$t(\xi) = a(\xi) + b(\xi)\ln\left(\frac{C(\xi)}{C}\right) - b(\xi)B_1(q(\xi) - q) \tag{20}$$

In Eq.(20), $C(\xi)$ and $q(\xi)$ correspond to the value announced by the firm when choosing the contract, and C and q are the realized values. The first term in Eq.(20) is the fixed transfer to the firm. The second term is the sharing of cost overruns, which is in logarithm because of the exponential form of effort in the cost function. When the realized cost is higher than the announced cost, the fixed transfer is reduced. The objective of this reduction is to induce cost reducing effort. By the contrast, if the realized cost is lower that the announced one, the firm get a higher transfer. The third term is to induce quality provision by increasing the transfer when the firm increases demand (the demand increment is due to higher quality). This part of the contract is similar to the performance-based contract (Hensher and Wallis, 2005), where the operator is rewarded if the realized demand is higher than a demand base level. It is worth noticing that the demand increment leads to cost increment; therefore, the contract incentives lead the firm to report the true cost and the true quality (measured by the demand) that it can deliver. Alternatively, the contract may be written in terms of the average cost, instead of total cost and demand, which is a figure usually known by managers of transport firms.

To interpret the parameter $b(\xi)$ is useful to approximate the logarithm around one by $\ln(x) \approx (x-1)$. The approximate transfer is:

$$t(\xi) \approx a(\xi) - b(\xi)\left(\frac{C - C(\xi)}{C(\xi)}\right) - b(\xi)B_1(q(\xi) - q), \tag{21}$$

which shows that $b(\xi)$ is the reward or penalty by one percent of cost savings or overruns respectively. Eq. (21) also shows that $b(\xi)/C(\xi)$ is risk sharing of cost overruns. As $B_1 q$ is the demand-elasticity of the cost, including the effect of quality (kilometers) s on the demand, the third term in Eq.(22) compensates the increments of cost due to demand increments.

For the chosen functional form of costs and effort disutility, it is possible to show that $b(\xi) = C(\xi)G(\xi)$. If the transfer $\tilde{t}$ is defined as follows:

$$\tilde{t}(\xi) = a(\xi) - G(\xi)\big(C - C(\xi)\big) - \tilde{c}(\xi)(q(\xi) - q), \tag{22}$$

where $\tilde{b}(\xi) = b(\xi)/C(\xi) = G(\xi)$ and $\tilde{c}(\xi) = b(\xi)B_1 = C(\xi)G(\xi)B_1$, then the gross transfer T is:

$$
\begin{aligned}
T(\xi) &= C + \tilde{t}(\xi), \\
&= C + a(\xi) - G(\xi)\left(C - C(\xi)\right) - \tilde{c}(\xi)(q(\xi) - q), \\
&= a(\xi) + G(\xi)C(\xi) + (1 - G(\xi))C - \tilde{c}(\xi)(q(\xi) - q), \\
&= \tilde{a}(\xi) + \tilde{b}(\xi)C - \tilde{c}(\xi)(q(\xi) - q).
\end{aligned} \tag{23}
$$

As $G(\xi)$ is a monotone, decreasing function and satisfies $G(\underline{\xi}) = 1$ and $G(\bar{\xi}) = 0$, Eq.(23) shows that the efficient firm chooses a gross transfer without sharing the cost overruns, and the contract becomes a fixed-price contract with incentives for demand increments. On the contrary, less efficient firms choose a higher fraction of overruns ($G(\bar{\xi}) < 1$) and the contract becomes like a cost-plus contract.

Table 6 summarizes the menu of linear optimal contracts. To compute these parameters it is assumed that the firm's fixed technological effect (A_i) is the average across the firms in the data. It is also assumed that the values for input prices are those in December 2016, such that the fuel is 462 CL$, the dollar exchange rate is 667 CL$/US$, and the labor is 130, and the bus fleet is equal to 971, which is approximately the average fleet of trunks services in December 2012. For the values coming from the demand model, the constant term of demand is computed as the average fixed effect of trunk services plus the average monthly effect and the constant of the model, the trend is equal to 114, and the seat/standing places is equal to 87,440 (the average in December 2011).

Table 6: Parameters for the menu of linear contracts

$F(\xi)$	$\tilde{a}(\xi)$	$\tilde{b}(\xi)$	$\tilde{c}(\xi)$	$q(\xi)$
0.00	63,360	0.00	1883	55.8
0.05	55,075	0.01	1873	51.4
0.40	48,604	0.02	1855	47.9
0.60	45,396	0.03	1834	46.2
0.70	43,397	0.04	1816	45.1
0.80	40,724	0.05	1782	43.6
0.90	36,261	0.10	1698	41.0
0.95	31,439	0.17	1561	38.1
1.00	14,703	0.55	847	24.0

Finally, the optimal menu of contracts comprises a fixed payment, a fraction of overruns sharing and a reward or punishment per passenger over or below the reference demand respectively. These contracts can be compared with a proposed contract for Santiago, in which the authority wants to induce high quality by a payment per passenger and a discount for shrinking driven kilometers. The firm that fulfill with the expected demand and the contracted kilometers gets a total transfer composed by a payment per passenger t_p equal to 560 CL$/pax and a payment per kilometer t_s equal to 420$/km, approximately. Moreover, if the firm provide more kilometers than the contracted ones, it does not obtain any reward. Suppose that the contract sets a gross transfer equal to a_0, with expected demand is $q_0 = q(\xi_m)$, where ξ_m is the mean of ξ, and contracted kilometers

s_0. The efficient cost for this case is $C_0 = C(\xi_m) = 32{,}677$ million CL$. It is possible to write the same contract in term of costs instead of kilometers by using the approximation $\Delta C = (dC/ds)\Delta s = (\alpha_1\gamma_1 + \alpha_2)C\Delta s$. Then, the actual transfer may be written as follows:

$$\begin{aligned}\hat{T} &= a_0 - t_s(s_0 - s) - t_p(q_0 - q) \\ &= a_0 + \frac{t_s}{(\alpha_1\gamma_1 + \alpha_2)C_0}(C - C_0) - t_p(q_0 - q) \\ &= a_0 - \frac{t_s}{(\alpha_1\gamma_1 + \alpha_2)} + \frac{t_s}{(\alpha_1\gamma_1 + \alpha_2)C_0}C - t_p(q_0 - q) \\ &= \hat{a} + \hat{b}C + \hat{c}(q_0 - q)\end{aligned} \tag{24}$$

The parameters in Eq. (24) are equivalent to those in the menu of linear contracts; however, their values are very different. On the one hand, the cost overrun sharing parameter $\hat{b}$ is 0.49, which implies low power of incentives for cost reduction. On the other hand, the parameter to reward or punish for quality provision $\hat{c}$ is 560, which is also lower than the parameter for the menu of linear contracts and implies low power of incentives for quality.

6. CONCLUSIONS

This paper compute a menu of linear contracts to induce cost reduction effort and quality provision for the bus transit services in Santiago. This menu of contracts allow the authority to decentralize the implementation of the optimal regulatory mechanism. The comparison of the linear contracts with the actually implemented contract shows that the transport authority has implemented a low-powered incentive scheme for both cost reduction and quality provision.

The recommendation of using a menu of contracts stems from the information asymmetry, which prevent the regulator from knowing the firm’s costs, demand and effort. This means, the authority needs not only to increase the contract parameters (payment per kilometer and per passenger), but also to allow the firm selects the contract from the menu, so to reduce the information asymmetry. In this way, the firm self-selection induces cost reduction, quality provision and information rent reduction. All these incentives may lead to a reduction of the subsidies required by the bus system of Santiago.

ACKNOWLEDGMENTS

This research was supported by CONICYT through the FONDECYT Project N° 11160857.

REFERENCES

Azzalini, A. (1985). A class of distributions which includes the normal ones. **Scandinavian journal of statistics**, 12(2), 171-178.

Batarce, M. (2017). Fixed-price Contracts and Enforcement of Minimum Quality in Public Transport: The Case of Santiago, Chile. Paper presented in **ITEA Annual Conference and School on Transportation Economics**, Barcelona, 21-23 June, 2017.

Batarce, M. and Galilea, P. (2017). Cost and fare estimation for the bus transit system of Santiago. **Transport Policy**, *forthcoming*.

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**, 39(1), 79-89.

Dalen, D.M. and A. Gomez-Lobo (1997) Estimating cost functions in regulated industries under asymmetric information. **European Economic Review**, 31, pp. 935-942.

Gagnepain, P. and M. Ivaldi (2002) Incentive regulatory policies: The case of public transit systems in France. **RAND Journal of Economics**, 33(4), 605–629.

Hensher, D. A., and Wallis, I. P. (2005). Competitive tendering as a contracting mechanism for subsidising transport: the bus experience. **Journal of Transport Economics and Policy**, 39(3), 295-322.

Laffont, J.J. and Martimort, D. (2002). **The Theory of Incentives: The Principal-Agent Model.** Princeton University Press, Princeton, New Jersey.

Laffont, J.J. and 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.

Laffont, J.J. and Tirole, J. (1993) **A theory of incentives in procurement and regulation**. The MIT Press, Cambridge.