9TH CONFERENCE ON COMPETITION AND OWNERSHIP IN LAND TRANSPORT

HOW HAS QUALITY INCENTIVES IN STOCKHOLM BUS SERVICES WORKED? A PILOT STUDY*

Authors:
Kjell Jansson, AF-Infrastruktur AB
Roger Pyddoke, Swedish Institute for Transport and Communications Analysis

1 INTRODUCTION

This paper deals with bus contracts procured by Stockholm County Public Transport Authority (SL). It aims to find out how the incentives used over the last years have worked.

We have done this by comparing the average values of the quality variable before and after an incentive has been introduced and calculated if the differences are statistically significantly different from naught. If the difference is statistically significant and has the desired sign we consider the incentive to have worked well.

Procurement of public transport and in particular bus transport started in the beginning of the 1990-ties after re-regulation of the public transport industry. SL was among the first Public Transport Authorities to take advantage of the possibility to procure bus transport. These procurements developed into an experiment workshop where different forms of contract forms were tried.

In a previous study Jansson (2002) discussed one of the incentives applied by SL: to encourage more demand and revenue. Jansson argued that such incentives could induce adaptations of the transport supply that was not socially efficient. The paper also pointed at difficulties in providing the correct magnitude of the incentives.

The cost outcome of procurement of bus transport in Stockholm, and in Sweden in general, has been the topic in several studies, e.g., Jansson (1993), Alexandersson, Fölster and Hultén (1996) and Alexandersson and Pyddoke (2003).

* Vinnova, a Swedish research grant agency, financed this work. The Stockholm County Public Transport Authority (SL) has contributed with substantial information.

From 1996 SL has tried a set of incentives designed to improve quality. These incentives were tied to particular indicators of quality, some of which had been used even before 1996.

In this paper we will focus on these quality incentives. As far as we know this is the first study in Sweden of the effects of incentives on quality outcomes.

Since this is a pilot study we have only coded the results for a selection of three contract areas in the Stockholm County of around 25 in total.

In section 2 we will present the kind of data we have used. In section 3 the method employed is described. Results are in section 4. Conclusions are found in section 5.

2 DATA

SL has several different datasets relevant for our purpose. SL has also by itself and by commissioning various consultants conducted a number of studies of quality in bus transport. Therefore data is scattered in several sources. The primary variables we have used are presented below.

2.1 Types of data used

2.1.1 Cancelled departures

The operator in charge of a particular bus line records cancellations. For each cancellation the operator pays a fine. If SL detects that an operator fails to report a cancelled departure and SL the fine charged is significantly higher.

2.1.2 Quality monitoring

Since 2003 SL procures a consultant firm to monitor a set of quality dimensions by so called "mystery shoppers" who travel in the bus system and record the following dimensions.

- cleanliness of the vehicles,
- presence of litter,
- presence of graffiti,
- cleanliness of bus stops,
- presence of litter at bus stops,
- graffiti at bus stops,
- the service attitude of personnel,
- if the bus driver calls out bus stop names

2.1.3 Passenger complaints

Passengers can complain by telephone or e-mail. The complaints are categorized as follows

- the state of a bus stop or terminal,
- the state of the vehicle,
- the conduct of personnel,
- the quality of traffic in terms of for example cancelled or delayed trips.

A special company owned jointly by the PTA and the bus operators collects these complaints.

2.1.4 Technical punctuality

For this quality dimension automatic equipment is mounted on the vehicles. This equipment is circulated systematically among vehicles and different lines during a year. The equipment registers deviations in arrival and departure times compared to schedule as well as if the bus arrives at the terminal stop before departure time, as it should.

2.1.5 Interviews with passengers

A total of 20 000 passenger interviews are conducted during spring and autumn each year. Passengers rate the services in the following dimensions:

- punctuality,
- conduct of personnel,
- driving performance,
- cleanliness of vehicle
- cleanliness of bus stop
- quality of information about delays and cancellations

2.2 Collection of data

For this study we received data from SL for each procurement and winning operator for the years between 1993 and 2003.

SL prepared the necessary data about when and where incentives were introduced in the different contract areas. These data were coded as variables 0 and respectively 1 for the indication of presence of incentives.

3 METHOD

3.1 Statistical analysis

Explained variables are the various kinds of quality measures. Explanatory variables, on the other hand, may be both different kinds of incentives as well as other potential explanatory variables as business cycles or cost development.

The data available from SL includes both dependent variables and potentially explanatory variables for the effects, for each contract area for each period. Each month in the table represents a separate observation.

All kinds of data are not available for all kinds of incentives. Punctuality is a quality dimension for which there are data for the longest period of time, since 1996. For the other dimensions measurements commenced 1998. The various incentives in turn have been introduced between 1994 and 2002. The following table gives an overview of starting points for measurements and incentives.

Table 3.1.1 Measurements and incentives for various quality variables

Quality factor/ method	First measurement, year	Measurement frequency, times per year	First Incentive, year
Passenger rating	2000	2 (20 000 interviews)	2001
Cancelled departures	1998	Continuous	1994
Monitoring (MSS)	2003	4	2001
Punctuality	1996	12	1994
Complaints	1998	Continuous	-
Bus at stop in advance	1996	Continuous	1999

The measure of punctuality for example, is available for each departure, each line and contract area for the sample of buses that carried the equipment for automatic registration according to a fixed scheme. For the present project these data have to be aggregated in some way. A problem that we discussed in the study was that the bus stops for which punctuality was measured were not stable over time. The purpose was to measure at the bus stops with the largest flows in both directions. But because the largest flows have shifted so has the bus stops, and the changing flows have also redefined the functions of some bus stops. In this study we have used measurements from departure stops even when these bus stops are not the largest in terms of flow. The purpose of this choice is twofold, firstly being that the flows at departure stops have been more stable, secondly that fewer external factors as congestion may influence punctuality at the departure stop.

3.2 Tests of incentives - method

For the test we looked for sequences where we had data on the quality outcomes *before* as well as *after* the incentive was introduced. We found a number of such sequences in the three selected contract areas Råsta-Solna, Råsta-Sollentuna and Söderort. Thereafter we calculated the following statistical tests.

The method is simple. We assume that the variables are normally distributed and we use a statistical standard test for comparison of two means from two different samples. These samples are assumed drawn from two different stochastic variables X_1 and X_2, from distributions with unknown means and variances. We want to test the hypothesis that the mean of the first stochastic variable X_1 that is μ_1 equals the mean of the other stochastic variable X_2 that is μ_2 against the hypothesis that the mean of X_1 that is μ_1 is larger than or less than the mean of X_2 that is μ_2. We do this by observing the mean of the sample $\bar{x}_1$ is larger or less than the mean of the second sample $\bar{x}_2$. We calculate the difference $\bar{x}_1 - \bar{x}_2$. But if we take in to account that there is a positive variance for both μ_1 and for μ_2, then the difference between $\bar{x}_1$ and $\bar{x}_2$ may vary between different samples from X_1 and X_2. We therefore calculate the likelihood that the null hypothesis is true given that we observe $\bar{x}_1 - \bar{x}_2$. To do this we create the test statistic T that is t-distributed and compare the obtained value with the t-distribution.

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

If the observed value of T is sufficiently far from zero we conclude that the null hypothesis is false at a certain degree of significance. We may for example calculate a value of T such that the likelihood that we observe has a value of T that is larger than this, if μ_1 and μ_2 are equal, is less than 1 percent. If we then observe a value of T that is larger than this value we say that we may reject the null hypothesis at the 1 per cent significance level.

3.3 Available data in Råsta-Solna and Råsta-Sollentuna

For both these contract areas incentives had been introduced for making the bus available to passengers before departure time. At the terminal the drivers should advance till the stop and let passengers board before time of departure. This is measured as the number of minutes before departure time that the bus was made available to the passengers per a certain number of departures.

3.4 Available data in Söderort

For Söderort we found analysable data for:

3.4.1 Incentive for late departure

Late departure is measured as the number of minutes after departure time that the bus departed per a certain number of departures. We have data on late departures in this contract area from January 1996 to December 2003. We have 69 observations before the incentive was introduced and 27 observations after.

3.4.2 Incentive for the total number of complaints

Complaints are measured as the total number of complaints per a certain number of departures. We have data on complaints from April 1998 to December 2003. We have 43 observations before the incentive was introduced and in October 2001 and 26 observations after.

3.4.3 Incentive for passenger quality ratings

Besides the direct measurements SL also used interviews with passengers to capture the subjective satisfaction with the services. In the interviews passengers rate their perceptions of quality scale from not satisfied over indifferent to satisfied. The variable we use is the proportion of the passengers that have responded with satisfied.

The quality dimensions are:
- total rating,
- cleanliness
- quality of information about delays and cancellations,
- conduct of personnel,

4 RESULTS

4.1 Results contract areas Råsta-Solna and Råsta-Sollentuna

The table below shows results for the variable "bus at stop in advance" for these areas.

Table 4.1.1 Results for Råsta-Solna. Bus available in advance at the stop

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Bus at stop in advance	Sep. 96 to Dec 03	Aug. 99 to Dec 04	35	53	2.95	3.77	-8.45	improvement 1 % level

Thus, the time during which the buses are available to passengers before departure has increased since the incentive was introduced.

Table 4.1.2 Results for Råsta-Sollentuna. Bus available in advance at the stop

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Bus at stop in advance	Sep. 96 to Dec 03	Aug. 99 to Dec 04	35	53	3.91	2.99	7.22	worsening 1 % level

The time during which the buses are available to passengers before departure has decreased since the incentive was introduced.

4.2 Conclusions on incentives in Råsta-Solna and Råsta-Sollentuna

For these contract areas we find the surprising result that the incentive for bus available at stop in advance has opposite effects in Sollentuna and Solna. In addition these effects are statistical significant for both contract areas.

4.3 Results contract area Söderort

The table below shows results for the variables:

- late departure,
- passenger complaints,
- conduct of personnel according to complaints,
- total quality rating,
- cleanliness in vehicle,
- information on delays and cancelled departures,
- conduct of personnel according to interviews.

Table 4.3.1 Results for Söderort. Late departures

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Late departure	Jan. 96 to Dec 03	Jan. 02 to Dec 04	69	27	38.57	36.11	-3.56	improvement

The late departures have decreased significantly after the incentives were introduced.

Table 4.3.2 Results for Söderort. Total number of complaints

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Passenger complaints	Apr. 98 to Sep. 03	Oct. 01 to Dec 04	43	26	87.62	111.1	-3.46	worsening

The total number of complaints has increased significantly since the incentives were introduced.

Table 4.3.3 Results for Söderort. Conduct of personnel according to complaints

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Conduct of personnel Complaints	Apr. 98 to Dec 03	Oct. 01 to Dec 04	42	27	15.71	21.44	-3.33	worsening

The number of complaints for conduct has increased significantly since the incentives were introduced.

Table 4.3.4 Results for Söderort. Total rating

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Total quality rating	May. 00 to Dec 03	Oct. 01 to Dec 04	4	4	58.75	66.25	-2.42	Improvement 10 % level

The total rating has there fore increased significantly since the incentives were introduced.

Table 4.3.5 Results for Söderort. Cleanliness in vehicles

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Cleanliness in vehicle	May. 00 to Dec 03	Oct. 01 to Dec 04	4	4	53.50	62.00	-2.04	not significant

The rating for cleanliness in vehicles has not increased significantly since the incentives were introduced.

Table 4.3.6 Results for Söderort. Information about delays and cancellations

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Information on delays and cancellations	May. 00 to Dec 03	Oct. 01 to Dec 04	4	4	19.75	21.00	-0.56	not significant

The rating for quality of information about delays and cancellations has not increased significantly since the incentives were introduced.

Table 4.3.7 Results for Söderort. Conduct of personnel according to interviews

	Measurements during	Incentives during	No. of observations		Average without incentive	Average with incentive	Test variable T	Quality change
			before incentive	after incentive				
Conduct of personnel Interviews	May. 00 to Dec 03	Oct. 01 to Dec 04	4	4	67.75	71.50	-1.71	not significant

The rating for conduct of personnel has not increased significantly since the incentives were introduced.

4.4 Conclusions on incentives in Söderort

The incentives seem to have influenced the number of late departure in the desired direction. The total rating of quality has also gone the right way. Cleanliness, information about delays and cancellations and conduct has not improved significantly. The number of complaints has increased significantly both totally and for conduct. The development of different quality ratings is therefore equivocal.

5 CONCLUSIONS FROM THE STUDY

The primary purpose of this pilot study preliminary study was to find out if an analysis of quality incentives for bus services was at all possible with the kind of data that exists in SL: s databases. The answer is that available data can be used both for evaluation of how quality has evolved and how incentives have worked.

We have also been able to make some further observations that may be useful for further evaluation and development of the incentives.

- The introduction of incentives seems to have proceeded in a somewhat *ad hoc* fashion. In some cases measurements of the quality variables has been commenced in good time before the incentive and sometimes not. There does not seem to have been any conscious plan to evaluate incentives.

- The number of incentives is large and varying and it is not always clear why a particular incentive has been chosen.

- Not all incentives have been paired with quality measurements.

- There are quality measurements for variables that lack a corresponding incentive.

The reason for this situation – a lack of structure – appears to be a result of ambitions on the part of SL to try out new incentive forms.

The statistical tests that we have conducted in the present pilot study for three contract areas suggests that four incentives worked well and had the desired effects and that three incentives did not work well, whereas two incentives did not seem to have any significant effect. From this one may not conclude that the full set of incentives for the whole set of bus contracts has not produced quality improvements and a positive net welfare. That would clearly be a premature conclusion from these limited calculations. The paper does however show that both data and methods allow for meaningful evaluation of the design of individual incentives and that these methods could be applied to the whole set of contracts.

Since the method of analysis seems to work well and since the data collected by SL is comprehensive and well structured it would be most interesting to carry out a full study by use of all data for all contract areas. One could then also use both cross-series and time-series data, and combinations of the two.

REFERENCES

Alexandersson, G., and Pyddoke, R, (2003). Leder upphandling till lägre kostnader för lokal busstrafik?, (Will procuremenet of local bus services imply lower costs?). *Working paper*. The Swedish Road and Research Institute (VTI).

Eriksson, M. and Jansson, K. (2003), Bedömning av några incitament i upphandling av SL:s kollektivtrafik (Assessment of some incentives in the procurements by SL). Commissioned by Stockholm County Public Transport Authority.

Jansson, K. (1993). Swedish Competitive Tendering in Local and Regional Public Transport - Overview and comparative case studies. International Conference on Competition and Ownership in Land Passenger Transport, Toronto.

Jansson, K., (2002). Incitament i upphandling av SL:s kollektivtrafik – Principer för bedömning (Incentives in procurement of SL: s public transport – Principles for assessment*)*. Commissioned by Stockholm County Public Transport Authority.

Pyddoke, R. (2003). Erfarenheter av upphandling i transportsektorn och tänkbara rekommendationer ? Översikt av effekter på kostnader, kvalitet samt utveckling av ny teknik (Experience of procurement in the transport sector and possible recommendations? Overview on impact on costs, quality and technique development*)*. *Working paper*. The Swedish Road and Research Institute (VTI).

Stockholm County Public Transport Authority (2004). , All references to contracts and operators for the years 1993-2003 and the used incentives for all contract areas.

ÅF-Infrastruktur AB (2004). Measurements of a number of quality indicators for the contract areas Råsta-Solna, Råsta-Sollentuna and Söderort.