Workshop 1 Report: Developing an effective performance regime

Wijnand Veeneman a, *, Andrew Smith b

a Faculty of Technology Policy and Management, Delft University of Technology, The Netherlands
b Institute for Transport Studies, University of Leeds, UK

A R T I C L E   I N F O

Article history:
Available online 30 October 2014

Keywords:
Contracts
Performance measurement
Tendering
Productivity and efficiency
KPIs

JEL classification:
L14
D24

A B S T R A C T

This workshop discussed the challenges faced in developing performance regimes: in particular, the way in which public transport authorities secure the performance of their operator(s). Earlier Thredbo workshops focused mostly on setting and measuring performance standards and incentivizing performance. This year’s workshop also looked more widely. The first additional topic was the context in which the performance regime is operating: how well is the market developed and what consequences does that have for the regime? The second was the maturity of the regime. Which conditions have to be fulfilled to have a fully-fledged and mature performance regime? These questions were addressed based on papers (and workshop participants) discussing performance in Australia, New Zealand, Japan, Greece, France, Ireland, Sweden, The Netherlands, Chile and Latin America more widely, and The United States. Key findings are that a wider set of conditions has to be in place to make a performance regime work. Appropriate technology is needed to capture good quality data. Mature institutions — that is, with the necessary legal powers to enforce contracts, guard against capture by the operators, and with appropriate staffing and resources — are also crucial. Maturity differs widely in the countries covered in the workshop, and thus different solutions are needed in different contexts. In particular, in situations of “low maturity”, regimes that place greater emphasis on passenger/demand metrics are likely to be more appropriate. The distinction between enforcing and incentivizing is also important in developing an appropriate performance regime. A suggested analytical framework for an effective performance regime which takes account of the above factors is set out, together with areas for future research. Obtaining greater information on the marginal costs and benefits of improving performance and also how better to benchmark complex and diverse operations against each other are key areas for future research. Other key research needs identified include: how to strike the right balance between enforcement versus seeking improvement; operationalizing KPIs (e.g. targeting frequency versus punctuality); and understanding real as opposed to assumed behavior by authorities and public and private operators.

© 2014 Elsevier Ltd. All rights reserved.

1. Overview of the papers presented

Based on evidence from France, Faivre d’Arcier and Bouf show that tendering of entire networks has not brought about the efficiency gains expected. They point out that operating costs per trip have risen substantially from 1980, far more than the growth rate of fare box revenues. Often, the high level of subsidies is legitimized on the basis of other public values (Veeneman, Van de Velde, & Lutje Schipholt, 2006). However, the real performance with respect to these values is often unclear when making alterations to the network. Faivre d’Arcier suggests evaluating the performance of public transport in more detail, taking account of a wide range of public values (e.g. environmental performance). In the presented approach the performance of each line in a network on various public values is modeled and presented to decision-makers. This allows the authority and operator to change the services, incorporating the performance of the various lines on patronage, cost and revenue, but also on CO2 emissions or car kilometers avoided, all on a line-by-line level. It represents one of the examples of the broadening of performance regimes seen throughout the workshop. A key challenge though is whether measuring line by line performance conflicts with the need to consider overall system performance, given that ridership on one line affects usage on others.

Batarce and Galilea, Gomes-Lobo and Briones, and Melo all look at the development of the performance regime in Santiago. They show how a contracting strategy and its performance rely heavily on monitoring. They see the most effect from the first
steps to introduce stronger management of compliance, with later steps having less effect. Batarce and Galilea show the dependency of gross-cost contracts on sound monitoring and monetizing. Gómes-Lobo and Briones note the positive effect of putting some demand risk on to the operator, even in gross-cost contracts. Melo shows that the increased focus on compliance in Santiago has indeed increased the quality of service by reducing waiting times for travelers.

Like Batarce and Galilea, Georgiadis et al. use data envelopment techniques to come to their conclusions. Data envelopment analysis can be used to perform a more integrated analysis of the performance of various operators or lines. Georgiadis et al. were able to incorporate the effect of population density and traffic conditions in the evaluation of lines and point at the possibility to reduce vehicle size as a promising efficiency improvement in the bus network of Thessaloniki. Together with the approach presented by Faivre d’Arcier, this could provide an integrated line-by-line evaluation tool to optimize the overall performance of the network.

Tribone et al. show how the set up of data analysis and visualization has to align closely to the needs of those intended to use it. On the basis of a case study of the punctuality on the Boston Red-line metro, they show how the translation from data to performance metrics is best developed in close cooperation with those operationally involved. Van Oort also looks at punctuality; however he focuses on the way in which this is set into the contract. His conclusion is that the performance management of punctuality focuses too often on the vehicles, rather than on the effects on the passengers. Van Oort proposes ways to improve on that, for example by focusing on regularity rather than punctuality on short headway services. Like Van Oort, Wretstrand et al. show how a limited system boundary of a performance regime can lead to sub-optimization (for example, it is important to consider the overall safety performance of a mode, including incidents that occur when walking to a bus stop). Looking at safety for the traveler, they show how widening the analysis towards sidewalks and stop design could optimize overall system performance in a more efficient way.

Veeneman et al. show the effects of tendering in the Netherlands and the performance of various contracts, also using a data envelopment approach. The conclusion is that non-tendered concessions have improved their performance faster than tendered contracts. This is not an effect of the direct awarding approach, but of the higher potential for improvement these concessions had, being in the major Dutch cities, which traditionally had low cost efficiency. The realistic threat of tendering was as effective in terms of improving performance as tendering itself. Hewitt also shows how tendered and non-tendered concessions interact in Wellington and how the possibility to tender can be used to improve performance. The Wellington authority ranks its operators, to allow the best performers on cost-recovery to keep their concessions. This case, and experience from the workshop more widely, emphasizes the greater importance that performance ought to play in the re-award process — this important incentive mechanism is generally seen as being largely absent in many cases around the world.

Finally, Tiznado et al. presented a paper focusing on the incentives of bus drivers, rather than bus operators. They found that drivers preferred to be incentivized on operational performance rather than passenger satisfaction — though it is suggested that such measures be incorporated into driver incentives (recognizing that part of the problem is that drivers do not fully understand the implications of such measures for their remuneration).

2. Discussion

This section is structured as follows. The workshop identified the context and degree of maturity as being very important in determining the optimal performance regime to implement. These are discussed first. We then discuss the relationship between the operator and authority, highlighting a number of trade-offs that need to be considered. We then set out an analytical framework for an effective performance regime, before setting out some important future research questions in this area.

2.1. Context

In the discussion several aspects of the performance regime were highlighted. Also from the plenary papers (Seftel & Rikhotso, 2013) it became clear that the contexts vary widely and that variation can have strong impacts on the ability to implement a successful performance regime, independent of the regime itself.

In the discussion, four key factors came forward. First of all, the level of development of the market is essential. When the absence of competition does not allow for the ultimate incentive, being replaced by another operator, performance regimes are less effective.

The variation on this was high for the cases discussed in the workshop, from 10 possible competitors in Helsinki, 5 in Wellington, 8 in Santiago, 4 in Utrecht, 2 in Lyon and only one operator in Boston. In Australia and New Zealand negotiation and tendering are combined, with the possibility to switch between both approaches, allowing for long-term incentivizing of operators that perform well, with the threat of tendering being a useful discipline on operator performance. Pure negotiation with an existing set of operators as in Santiago allows for some form of performance management by shifting market share between them, but this seems to be less effective. In case of the integrated single operator of Boston, competition and choosing a potentially better performing operator is not an option to incentivize the operator. Here the performance regime only exists to stimulate the existing operator into better services and efficiency.

The way in which the operators are chosen (when there in no free market entry), also plays an important role in securing quality. Competitive tendering in Europe for example does not allow for the use of past performance in the choice, as it could hinder market entry of operators with no prior performance in the tendered area. Using past performance in awarding concessions could provide a strong quality incentive, like it does in some other examples in the workshop (e.g. Australia and New Zealand).

In the cases discussed, there was a variety of choices made in terms of demand risk: putting it entirely with the operator, authority or some form a shared demand risk by limiting the downside and up-side risk to the operator, giving a bonus for patronage growth, or fare-box revenue sharing. The discussion showed that putting demand risk with the operator is a robust way of incentivizing for performance in terms of quality of services. Obviously, this has to be combined with some level of freedom of the operator on the tactical level. A key issue though is whether some aspects of demand risk are outside the control of operators (such as the state of the economy). In rail in Britain, for example, there has been considerable debate about how to insulate rail franchisees from exogenous demand risk — that is, relating to the growth in GDP or employment in the economy.

These aspects of the context set the stage for the effectiveness of a performance regime. The remainder of the discussion on performance regimes was focused on improving performance once an operator is chosen and contracted.

2.2. Degree of maturity

As well as differences in context, the cases discussed showed different levels of maturity of the performance regime. In the
discussion three key aspects of maturity emerged. First, history will affect the degree of maturity of the regime. For example, how experienced are the operators? Where operators have been running services for many years they will have a good degree of understanding of demand and supply conditions. They will also potentially have considerable power and there is a danger that they could “capture” the authority to the detriment of passengers.

Second, a performance regime needs a sound institutional basis. Clearly, a good contract is part of that, with a well described system of KPIs and related incentives. This has been the major focus of earlier Thredbo discussions on performance regimes (see for example Nelson & Merkert, 2013). However, contracts are only as strong as the institutional environment in which they are functioning. Consequently, a performance regime is also dependent on the institutional maturity of a country, mostly of its legal system. Adequate resourcing, with staff that are capable of managing complex contracts (from an economic, operational and legal perspective) is also key, as is continuity of experience within the authority over time.

Related to this, the staff of the authority have to act on the data and ensure that the incentive mechanism in the contract is enforced. The discussion in the workshop showed that a mechanistic approach to invoking penalties or awarding bonuses (enforcing) is not always the most productive use of the performance regime. A mature KPI and incentive system needs to be followed up by strong negotiations from the authority staff with the operator. The staff of the authority needs understanding of the difference between situations where it is appropriate to directly impose penalties and award bonuses (enforce) and when to see them as a starting point for further discussions (improvement).

Finally, KPIs should be defined in a way that technology is available to objectively and reliably harvest the data, and communicate that data to the authority in a trustworthy way. When contract-level KPIs are set that do not take into account the operational level of data harvesting and communication and the quality of that data available, the KPIs do not reach their potential in the performance regime. Also the technology itself has to be mature.

A mature system needs all three of the above ingredients. The maturity of the system will affect the ability of the authority to produce sensible and sophisticated KPIs that are not easily gamed. In the contract, the KPIs are described and operationalized. Clearly, they should be objective, robust, and with limited possibilities to be gamed. Examples from the Netherlands rail showed how KPIs of punctuality incentivized the operator to leave the station early, only to stop at the next block with a red signal. Passengers could have made that connection if the train had waited the extra minutes at the platform, with no penalty to the overall travel time. The KPI was not robust in improving quality of service. The operator gamed the KPI, to the detriment of the service.

The above discussion has an impact on the type of performance regime that will be appropriate in different contexts. In situations of low maturity performance metrics that are linked to demand and the passenger are likely to be optimal, given the difficulties in less mature regimes of specifying and enforcing a sophisticated performance regime based on operational measures. More mature environments would then look to implement a regime based on such operational measures, but in the longer term it will be desirable to harness natural incentives, based on demand, alongside operational measures (thus completing the cycle of maturity — see Fig. 1).

2.3. Interaction between operator and authority

The workshop discussed the role of KPIs in the interaction between authority and operator. First of all, the use of the term KPI was contested. The essence of indicators of operators’ performance is not that they show key performance, but that they show those performance indicators on which compliance is needed. For example, patronage generally is a key performance aspect. However, in the case of net cost contracts, an additional system of KPIs penalties and bonuses may not be necessary. When the authority’s main goal is to incentivize patronage, the net cost contract (possibly with supplement of fare box revenues) could provide enough of an incentive. However, when compliance is needed with more and other factors, other KPI’s have to be added.

In the discussion a number of dilemmas became apparent that are relevant for the way in which KPIs are defined and used. The first dilemma is that between service related and task related KPIs. Service related KPIs (like customer satisfaction) focus the operator more on integrated service quality; however, they are harder to objectively measure and less actionable to the operator than task related KPIs (like buses on time).

A second dilemma is between improvement and punishment. Is the performance regime put in place to improve services, or to punish the operator in case of underperformance? For example, a major penalty could reduce innovation potential at the operator level or even lead to bankruptcy. This eventually has negative effects on service quality. On the other hand, if penalties are never administered after failing to meet KPI standards, the credibility of the performance regime is at stake.

A third dilemma is between partnership and “whack-a-mole”. In a trusted partnership, the authority is generally codeciding and as such is also partly responsible for missing KPI targets. The role as independent judge in administrating penalties is compromised. Staying independent, not getting involved, and just administering bonuses and penalties, denies the fact that authorities control many key ingredients of service quality delivery. The last dilemma not only focuses on the relation between authority and operator, but also between operators. KPIs can incentivize the operators to cooperate to offer quality or to optimize on their own performance, which are not necessarily the same.

A performance regime can be set up to provide the authority a great deal of power over the operator. That power has to be used carefully, as inappropriate use can stifle the possibilities of the operator to provide attractive services. With failure by the operator to reach set targets comes the inclination of the authority to increase penalties and strengthen interventions. On the other hand, there are many different areas in which governments are active that influence the possibilities of the operator of these services. The obvious areas are infrastructure development and urban planning. However, many others are relevant for the success of the operator, from fuel tax, working conditions regulation, to school times, parking policies, etc. Striking the right balance between building a trusted partnership on the one hand and enforcement by

![Fig. 1. The cycle of maturity.](image-url)
transparency acting on (un)realized KPI thresholds on the other is
the broader and key dilemma of a performance regime. Trans-
parently acting on (un)realized KPIs should include understanding
of what the data is saying. The data has to be translated into
meaningful information and knowledge about the situation to
effectively improve the performance. In the workshop we saw ex-
amples of this in the field of punctuality.

2.4. An analytical framework for an effective performance regime

The workshop gathered a number of robust lessons for the
development and operation of a performance regime. First of all,
KPIs have to be distinguished between aspirations and targets.
Aspirational KPIs can be related to patronage and system wide
quality. Incentives can be natural, simply by putting demand risk
with the operator. Target KPIs have to be realistic and based on
historic data, preferably from the region in which services are
provided. The incentives also have to be realistic and related to the
marginal costs, market price and societal gains of the attainment of
the KPIs thresholds. As noted below, new research on under-
standing the marginal costs and benefits of improvements in
performance is a key area for future research.

Incentives and KPIs work on three levels (see Fig. 2). First, some
goals can be secured in the contract award, with incentives to the
operator to promise (and consequently realize) goal attainment on
specific KPIs. The incentive is the awarding of the contract itself. An
example is the environmental performance of vehicles, which op-
erators include in their bid to get the contract.

Second, additional goals can be secured in the contract as nat-
ural incentives. The clearest example is the use of net cost contracts
in which demand risk is placed on the operator. The incentive is the
awarding of the contract itself. An example is the environmental performance of vehicles, which operators include in their bid to get the contract.

Finally, other public values might not align with those of the
operator, thus requiring these to be robustly measured and incor-
porated into the core of the performance regime. For example, the

\[
\text{Fig. 2. The interaction between authority and operator on three levels of a performance regime.}
\]
exposing operators to demand risk, the discussion in the workshop emphasized the importance of performance being linked to re-award of the contract. This mechanism was deemed not to play a significant enough role at present, though clearly this varied across different countries as noted above. Another important factor is the extent to which benchmarking can be used to compare the performance of operators. It was noted that yardstick competition, as used by economic regulators, could be a powerful tool, but that developing a credible regime using appropriate techniques and dealing with heterogeneity between operators is a major challenge.

3. Future questions

The workshop has developed several key research questions. A performance regime needs to strike the right balance between enforcing and improving. From the discussions it became clear that both extremes seem rather unproductive. Also it became clear that the right balance is dependent on maturity of the institutional environment and the contract type. However, where the right balance is, would be an excellent question for further research. The same question applies to the number of KPIs and the balance between system related KPIs (aimed at overall system performance) and task related KPIs (aimed at task execution).

A second important area highlighted for future research concerns valuation. What is the marginal cost and indeed benefit of improving reliability for example? The discussion on punctuality triggered further questions on how specific values can best be operationalized into KPIs, in a given context. For high frequency services moving away from schedule compliance to headway compliance makes a great deal of sense. Also, for other KPI’s their formulation is context dependent. More understanding of that context dependency is needed.

The workshop discussed the need for further research on methodologies for benchmarking performance between operators. In particular it was noted that statistical and or data envelopment analysis (DEA) approaches could be utilized as a comparator method to normalized unit cost approaches. Such statistical methods are widely used by economic regulators in the utility sector, for example.

Finally, performance regimes have long been seen as a set combination of KPIs and incentives. However, a wider range of factors comes into play when evaluating the effectiveness of a performance regime. A shift is needed in the research from understanding the performance regime under assumptions of operator and authority behavior to understanding the performance regime, including real-life behavior of operator and authority. Related to this point is the question as to how behavior is linked to ownership — how do private operators differ from state owned ones or intermediate non-profit bodies such as Network Rail in Britain?

Papers presented in the workshop

The looming crisis in French public transit Dominique Bouf and Bruno Faire D’Arcier LET — University of Lyon, France

Measuring the performance of urban public transport in relation to public policy objectives Bruno Faire D’Arcier LET — University of Lyon, France

Incentive schemes, provision of quality and monitoring: the case of the public transit system in Santiago de Chile Patricia Galiliea and Marco Batare Pontificia Universidad Católica de Chile

Incentives in bus concession contracts: the Latin American experience Andres Gomez-Lobo University of Chile and Julio Briones Pontificia Universidad Católica de Chile

Implementing New Zealand’s new public transport operating model: a description of the challenges and progress to date Rhona Hewitt Greater Wellington Regional Council, New Zealand Rachel Drew NZ Bus limited, New Zealand

Can regulatory policies improve the performance of a bus system? A statistical analysis for the case of Transantiago Carlos Meio Industrial Engineering School, Diego Portales University, Chile

Measuring and improving the efficiency and effectiveness of bus public transport systems Georgios Georgiadis, Ioannis Politis and Panagiotis Papaioannou Department of Civil Engineering, Aristotle University of Thessaloniki, Greece

Incorporating service reliability in public transport design and performance requirements: international survey results and recommendations Niels van Oort Department of Transport and Planning, Delft University of Technology, The Netherlands and Goudappel Coffeng Mobility Consultants, The Hague, The Netherlands

Efficient frontier analysis of Dutch public transport tendering: a first analysis Wijnand Veeneman, Janneke Wilshut, Thijis Urlings and Jos Blank Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands Didier van de Velde Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands and inno-V, The Netherlands

An automated data driven performance regime for operations management, planning, and control Dominick Tribune Massachusetts Bay Transportation Authority, Boston, MA, USA David Block-Schachter, John Attanucci and Nigel H.M. Wilson Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, USA

Safety as a key performance indicator: creating a safety culture for enhanced passenger safety and comfort Anders Wretstrand, Bengt Holmberg and Monica Berntman Department of Technology and Society, Lund University, Sweden

Incentive schemes for bus drivers: the case of the public transit system in Santiago de Chile Ignacio Tiznado, Patricia Galiliea, Felipe Delgado and Markus Niehaus, Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile

Additional literature

