Contract design and performance of railway maintenance: effects of incentive intensity and performance incentive schemes

Kristofer Odolinski – VTI

CTS Working Paper 2016:20

Abstract

In this paper we study the effect of contract design on the performance of railway maintenance in Sweden, using a panel data set over the period 2003-2013. The marginal effect of incentive intensity is estimated, showing that the power of incentive schemes improve performance as measured by the number of infrastructure failures. In addition, the performance incentive schemes result in a reallocation of effort from failures not causing train delays to failures causing train delays.

Keywords: contract design, incentive intensity, maintenance, railways

JEL Codes: D82, L92

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
www.cts.kth.se

Contract design and performance of railway maintenance: effects of incentive intensity and performance incentive schemes

Kristofer Odolinski

The Swedish National Road and Transport Research Institute (VTI), Department of Transport Economics, Box 55685, 102 15 Stockholm, Sweden (kristofer.odolinski@vti.se)

Abstract

In this paper we study the effect of contract design on the performance of railway maintenance in Sweden, using a panel data set over the period 2003-2013. The marginal effect of incentive intensity is estimated, showing that the power of incentive schemes improve performance as measured by the number of infrastructure failures. In addition, the performance incentive schemes result in a reallocation of effort from failures not causing train delays to failures causing train delays.

1.0 Introduction

Government agencies often procure goods and services instead of producing it in-house. This procurement accounts for a significant part of national economies, with estimates at 19 per cent of the gross domestic product (GDP) in the European Union (European Commission 2011). Cutting costs and improving quality are frequently stated goals when introducing competitive tendering and contracting of services previously offered by a state-owned monopoly. However, careful contract design is required in order to achieve the goals of such reform, with appropriate specification and monitoring of quality along with incentive schemes to deal with moral hazard and adverse selection. Whether or not different contract designs have the desired effects needs to be tested empirically, both for policy reasons and to assess if theoretic arguments for certain designs are valid in the current case.

This paper contributes to this line of research by studying the incentive structures in railway maintenance contracts in Sweden. More specifically, the purpose with this paper is to provide evidence on the effect of incentive intensity on infrastructure performance as well as the effect of tilted performance incentive schemes.

Sweden chose to gradually expose its maintenance of railways to competitive tendering in 2002. One objective of the transfer from in-house to tendered production of rail maintenance was to provide scope for innovation (Banverket 2000). To do so, firms (contractors) are given degrees of freedom by the contracts: most of the maintenance contracts are said to be[1] outcome or performance based, meaning that the contractor is not told exactly which (or the level of) activities that are to be carried out. A fixed payment is received by the contractor who needs to meet a set of requirements with respect to the quality of maintenance. The purpose is to give the contractor an incentive to develop the maintenance production. We are therefore in a second-best situation where the client (the IM) can (and has in this case chosen to) observe the outcome rather than prescribing the input. This can, however, create a moral hazard situation as the contractor's actions may not be optimal for the client. In addition, the contractor can obtain a higher rent when information about its efficiency (technology) is not known to the client, which is the problem of adverse selection. This asymmetry in information means that the client has to make a trade-off between inducing effort and extracting rent from the contractor. The power of the incentive scheme is a central parameter in this trade-off (see Laffont and Tirole 1986). Moreover, the complexity of the project can affect the preferred power of the incentive scheme. Bajari and Tadelis (2001)

[1] This formulation is used in view of the extensive reference to regulations and provisions in the contracts.

points out that a cost-plus contract (i.e. low powered contract) is preferred when the projects are complex, while fixed price contracts can be better for simple projects.[2]

One way of providing incentives is to use a performance incentive scheme in which the contractor receives an award and/or penalty for its performance. A contractor will make a trade-off between different tasks within a project if these are rewarded differently and the tasks are substitutes; see for example the seminal paper by Holmström and Milgrom (1991). Indeed, the performance incentive schemes in the maintenance contracts in Sweden are tilted (described in section 2.1), which can affect the attention to different tasks and consequently the outcome of the project.

The theoretic work on contracts and information asymmetry in the principal-agent framework is extensive (for textbook treatments, see for example Laffont and Tirole 1993, Laffont and Martimort 2002 and Salanié 2005). Wunsch (1994) is an early example of an empirical study on contract design within the field of procurement and regulation, where menus of linear contracts are calibrated for transit firms. Gagnepain and Ivaldi (2002) study the regulatory schemes for French urban transport and compare these to the optimal policies, while Roy and Yvrande-Billon (2007) use the same study object (in a different time period) to estimate differences in technical efficiency between regulatory schemes and fixed-price and cost-plus contracts. Other examples within the transport field are the study by Dalen and Gomez-Lobo (2003) - showing that high-powered incentive schemes reduce operating costs for bus companies in Norway - and the study by Piacenza (2006) with similar results for Italian public transport.

To the author's knowledge, an econometric test of the effect of incentive intensity has not been made in field of rail infrastructure management. Nonetheless, Vickerman (2004) provides an exploration of incentives in transport infrastructure maintenance, and a case study

[2] A project that is complex such that quality is *ex ante* non-contractible might even be better to produce in-house instead of being contracted out (Hart et al. 1997).

on incentives in rail maintenance contracts is made by Stenbeck (2008). Moreover, studies on the power of incentive schemes in procurement and regulation usually compare different types of contracts (for example fixed-price contracts compared to cost-plus contracts). We can however make use of the variation in the incentive intensity in the cost-reimbursement contracts that are used for railway maintenance services in Sweden. This enables an estimation of the marginal effect of incentive intensity within the same contract type.

There is a wide literature on the effects of performance payments; see for example Lazear and Oyer (2013) for a review of theories and empirical findings on incentives and performance (among other topics) in personnel economics.[3] A recent study on procurement and performance incentives is made by Lewis and Bajari (2014), showing that penalties induced effort in high-way construction contracts (with welfare improvements and low contractor costs according to simulations). Our study adds to this literature by estimating the effects of performance incentives in rail infrastructure management, focusing on the reallocation of efforts.

The outline of the paper is as follows. Section 2 describes the main ingredients of the railway maintenance contracts that are important for this study. The research questions and modelling approach are set out in section 3, where we also specify the models we estimate. A description of the data is provided in section 4. The results are presented in section 5 followed by a discussion of our findings in section 6. Section 7 concludes.

[3] Other examples are Rosenthal et al. (2006) and van Herck et al. (2011) who provide reviews of empirical evidence in the health sector, Podgursky and Springer (2007) present evidence in the education sector and Devers et al. (2007) is a review of evidence on executive pay and firm performance.

2.0. Maintenance contract design

Most of the railway maintenance contracts in Sweden are performance-based contracts.[4] These contracts are a mix between a fixed price and a cost-plus contract, i.e. a fixed payment is received for certain activities while others have variable payments. In the competitive tendering of maintenance contracts, the firms place a bid that contain prices for activities with variable payments and a fixed payment they require for other activities. The procedure and timing is the following. (1) The client provides a description of the maintenance area and specifies the expected amount of activities with variable payments that need to be performed each year and which of these activities that entail a fixed payment. (2) The firms submit bids comprising one fixed component as well as a unit price component, i.e. a unit price on each variable activity in the contract. (3) The bid with the lowest total cost wins.

Most contracts have a fixed payment for the (expected number of) activities required when an infrastructure failure occurs. However, the cost for each activity is capped; a clause states when the cost of rectifying a failure is included in the fixed payment to the contractor. It also indicates that when the cost is higher than the cap, the contractor is paid according to the variable cost for the amount above the specified cost level. For intuition, consider the following example illustrated in Figure 1: the contractor receives a fixed payment for rectifying failures during one year. A clause states that if the cost of rectifying one failure is above 10 000 SEK, the contractor will be paid according to the direct cost of rectifying that failure (cost for labor and material resources according to prices stated in the contract) for the amount above 10 000 SEK. Hence, if the total cost of rectifying one failure is 15 000 SEK, the contractor will be paid 5 000 SEK in addition to the fixed payment. This reimbursement rule can vary between contracts, creating different levels of power in the incentives. The same

[4] A few contracts are so called design-bid-maintain contracts in which the contractor mainly executes the activities set up by the client. These contracts are used for newly built railway lines.

reimbursement rule in each contract is used for maintenance activities that prevent infrastructure failures, i.e. for fixing a defect before it becomes a failure.

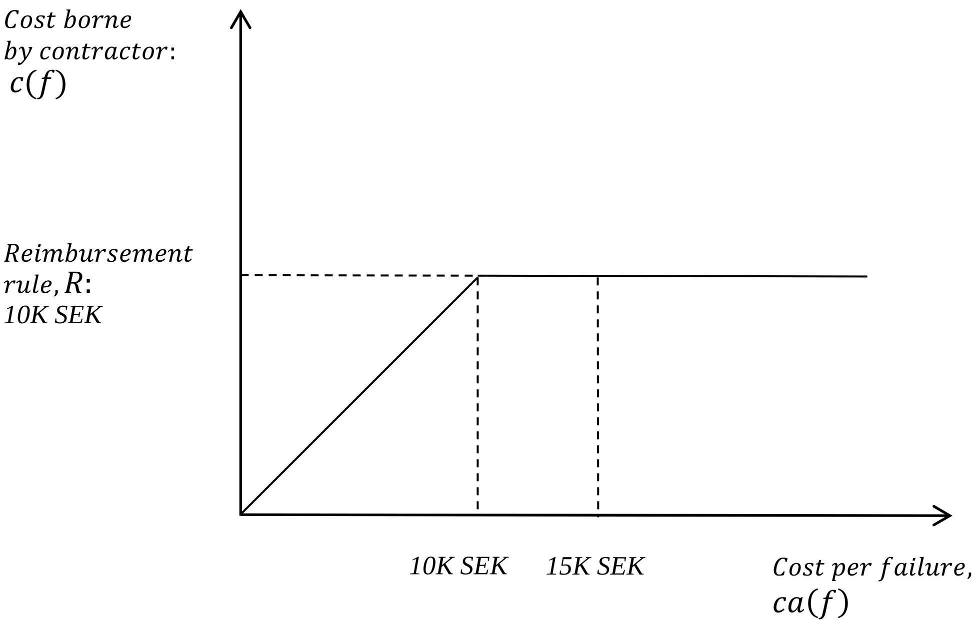

Figure 1 – Illustration of the reimbursement rule

2.1 Performance incentive schemes

Apart from capping the contractor's cost for some activities, the contracts also include a bonus and/or penalty linked to the number of failures in the maintenance area. These are tilted towards failures that cause train delays, which imply that an average train delay failure will have a larger impact on the bonus or penalty compared to an average failure not causing a train delay. For example, a contract using a performance index has the weight 1.8 for train delays while the weight is 1 for other failures and 0.2 for a measure of track geometry. The performance equation is presented in the appendix. Note that a train delay failure will affect the train delay outcome and the outcome for the number of failures - that is, it affects two parts of the equation and thus have more weight than 1.8 in the performance equation. Other contracts have bonuses and/or penalties linked to target values for train delays while failures not causing a train delay are excluded.[5]

In summary, the contracts are designed so that a contractor prefers a failure that is not causing train delays instead of a train delay failure. Note that this performance incentive scheme implies that a failure expected to cost more than the cost in the reimbursement rule (illustrated in Figure 1) has a probability of imposing an extra cost for the contractor if not rectified in time.

3.0 Research questions and modelling approach

In this paper, performance refers to the number of infrastructure failures that needs to be fixed immediately or within two weeks. There are other types of infrastructure quality indicators such as “minor” deviations in track geometry or other defects that require a preventive maintenance, i.e. activities that prevent infrastructure failures. Thus, a lack of preventive maintenance will result in a failure that requires corrective maintenance.

Given the design of the contracts, we formulate the following research questions:

1. Do variations in the reimbursement rule affect the performance of maintenance contracts?
2. Do performance incentive schemes tilted against train delays have an effect on the relationship between the number of failures causing train delays and other failures?

Model 1, presented in section 3.1, addresses the first research question. In section 3.2 we present *Model 2* that addresses the second research question.

[5] There are also contracts that do not have any bonus connected to train delays. However, all procured maintenance contracts have a penalty for the contractor if a time limit to rectify a train delay failure is exceeded. For example, the penalty can be 10 000 SEK if it takes more than five hours to rectify a train delay failure.

3.1 Model 1

We express the cost of a maintenance project as

$$C_{ij} = M_{ij}(X,e) + CA_{ij}(F), \tag{1}$$

where i =contract and j =contractor. $M_{ij}(X,e)$ is preventive maintenance and is a function of $\boldsymbol{X}_i$ which is a vector of variables for traffic and infrastructure characteristics such as track length and rail age. We also consider preventive maintenance to be a function of the contractor's effort level $\boldsymbol{e} = (e_1, e_2, \dots, e_n)$, where $\frac{\partial M_{ij}(X,e)}{\partial e} > 0$. $CA_{ij}(F)$ is the total corrective maintenance cost which is a function of the number of failures per maintenance project.

The number of failures is represented by

$$F_{ij} = \alpha_i + \boldsymbol{\beta X}_i + \gamma_j M_{ij}(X,e) + \varepsilon_{ij}, \tag{2}$$

where α_i is a constant and $\boldsymbol{\beta}$ is a vector of parameters for the effect of the explanatory variables $\boldsymbol{X}_i$. ε_{ij} is an error term. γ_j is a parameter indicating the effect of preventive maintenance on the number of failures.

With the reservation of the caveat indicated by footnote 1, the level of preventive maintenance is not a direct decision made by the client in performance based maintenance contracts. Instead, the (winning) contractor receives a fixed payment, W_{ij}, corresponding to the bid submitted, and then decides on a level of effort which generates $M_{ij}(X,e)$ and results in a disutility for the contractor $\Psi(e)$.[6]

Preventive maintenance's marginal effect on the cost of the project is decided by γ_j. Hence, this is the contractor's efficiency parameter. It is also assumed that each contractor knows its own efficiency and that this parameter is continuous.

[6] We assume $\Psi(e)$ is convex in effort.

As described in section 2, there is a reimbursement rule that states the maximum total cost per failure the contractor will bear. The corrective maintenance cost for a failure borne by the contractor is therefore

$$c_{ijf} = \begin{cases} R_i, & if\ ca_{ijf} \geq R_i \\ ca_{ijf} & otherwise \end{cases}, \tag{3}$$

where $f = 1, .., F$ failure, R_i is the reimbursement rule, ca_{ijf} is the accounted cost of rectifying a failure (see also Figure 1), where $\sum_{f=1}^{F} ca_{ijf} = CA_{ij}(F)$. The contractor's profit is then

$$\pi_{ij} = W_{ij} - [M_{ij}(X, e) + \sum_{f=1}^{F} c_{ijf}], \tag{4}$$

This implies that the contractor bear a share $b \in [0,1]$ of the cost of rectifying a failure, which is determined by the relationship between $\sum_{f=1}^{F} c_{ijf}$ and $CA_{ij}(F)$; that is $b_{ij} = \frac{\sum_{f=1}^{F} c_{ijf}}{CA_{ij}(F)}$. We can therefore express (4) as

$$\pi_{ij} = W_{ij} - \left[M_{ij}(X, e) + b_{ij} \cdot CA_{ij}(F)\right], \tag{5}$$

which is an *incentive contract* where b_{ij} is the power of the incentive scheme. From equation (3) it is clear that an increase in the reimbursement rule R_i would increase the corrective cost borne by the contractor (c_{ijf}) for failures that cost more to rectify than the previous reimbursement rule, *ceteris paribus*. The power of the incentive scheme will therefore increase with the reimbursement rule: $\frac{\partial b}{\partial R} > 0$.

When the maintenance project is tendered, a contractor considers its objective function that it wants to maximize

$$\max_{W,e}\ U_j = \pi_{ij} - \Psi(e), \quad \text{subject to } \pi_{ij} - \Psi(e) \geq 0, \tag{6}$$

The setting outlined in equations (1)-(6) implies that an increase in the reimbursement rule increases the comparative advantage of the efficient contractor in the competitive tendering of a maintenance contract (causing a selection effect):

- The information about R_i in the quote for bids is one point of departure for the bid submitted. A higher R_i will increase the share of cost that is borne by the contractor per failure, and will therefore reduce the contractor's utility ($\frac{\partial U}{\partial R} = -\frac{\partial b}{\partial R} CA_{ij}(F) < 0$). The consequence is that the higher reimbursement rule corresponds to submitting a higher bid. However, since contractors have different efficiency levels, γ_j, the expected marginal cost savings of efficiency increases with the reimbursement rule, $\frac{\partial^2 U}{\partial R \partial \gamma} = -\frac{\partial((\partial b/\partial R)E[CA])}{\partial \gamma} > 0$, as the expected total cost of rectifying failures $E[CA] = E[E[CA|F]]$ is decreasing with the efficiency level γ_j, *ceteris paribus*.

Equations (1)-(6) also imply that the level of preventive maintenance increases with the reimbursement rule:

- When a contract has been awarded, the contractor will choose an effort level that generates preventive maintenance. This decision will depend on R_i. A contractor that bears a larger share of the cost will clearly have a stronger incentive to reduce this cost, which it can do via $M_{ij}(X, e)$. Indeed, by taking the cross partial derivative of the contractor's utility with respect to the reimbursement rule and the effort level (generating preventive maintenance) $\frac{\partial^2 U}{\partial R \partial e} = -\frac{\partial((\partial b/\partial R)E[CA])}{\partial e} > 0$, we can see that the reimbursement rule increases the marginal cost savings of preventive maintenance. Hence, a higher reimbursement rule generates effort (reduces the moral hazard problem).

3.1.1 Estimating the effect of the incentive schemes

We estimate the effect of the cost-reimbursement with the following model

$$F_{it} = \alpha_i + \boldsymbol{\beta X}_{it} + \mu R_{it} + \varepsilon_{it}, \tag{7}$$

referred to as *Model 1,* where $t = 1, .., T(i)$ years. Our hypothesis is that a higher reimbursement rule will reduce the number of failures:

$$Hypothesis\ 1: \mu < 0$$

The effect of the cost reimbursement rule can be due to a selection effect and/or because it reduces the moral hazard problem. Hence, the estimation of *Model 1* does not discriminate between these effects.

3.2 Model 2

The previous model focuses on the share of the cost per corrective maintenance activity that is borne by the contractor. However, there is also a performance incentive scheme in place (described in section 2.1). In particular, contractors may be penalised for the number of failures appearing each year and even more so for failures causing train delays. The profit for a contractor can then be formulated in the following way:

$$\pi_{ij} = W_{ij} - [M_{ij}(X, e) + b_{ij} \cdot CA_{ij}(FT + FO) + PT \cdot FT_{ij} + PO \cdot FO_{ij} + \varphi FO^2], \qquad (8)$$

The term $b_{ij} \cdot CA_{ij}(FT + FO)$ is the same as in equation (5), except that a distinction is now made between failures causing train delays

$$FT_{ij} = \alpha_{ij} + \beta_{FT} X_i + \gamma_j M_{ij}(X, e_T) + \varepsilon_{ij}, \qquad (9)$$

and other failures

$$FO_{ij} = \alpha_{ij} + \beta_{FO} X_i + \gamma_j M_{ij}(X, e_O) + \varepsilon_{ij}, \qquad (10)$$

In addition, PT and PO are the performance penalties incurred by the contractor for FT_{ij} and FO_{ij} respectively. The tilted incentive scheme implies that $PT > PO$, which will tilt the contractor's maintenance strategy towards preventing failures causing train delays. For example, consider a situation where two defects are found that have the same expected corrective maintenance cost, but one defect is more likely to cause train delays than the other (which can be due to the type or the severity of the defect). The contractor will then benefit more by first fixing the defect that is more likely to cause train delays, which increases the probability of the other defect to become a failure. However, the number of other failures must be handled in order to cap the risk of them causing train delays (for example, fixing failures require time slots on the tracks, and trains will eventually need to be rescheduled when the number of failures grows). We therefore consider other failures to have a second-order effect (φ) on profit.

We characterize the contractor's maintenance strategy by the choice of efforts on preventing train delay failures and other failures, $e = (e_T, e_O)$, where one of the efforts crowds out the other.[7] In other words, the contractor's marginal cost of exerting e_T increases with e_O.

The contractor's choice of effort will be tilted towards e_T as a result of its marginal effect on the contractor's expected cost of the project $E[Cj]$:[8]

$$\frac{\partial E[Cj]}{\partial e_T} = PT \cdot \frac{\partial FT}{\partial e_T} < PO \cdot \frac{\partial FO}{\partial e_O} + \varphi \frac{\partial FO^2}{\partial e_O} = \frac{\partial E[Cj]}{\partial e_O},$$

$$for \;\; FO \in [0, Z], \tag{11}$$

(note that $\frac{\partial F}{\partial e} < 0$, and $\frac{\partial (FO_{ij})^2}{\partial e_O} < 0$).[9] Hence, a marginal increase in e_T has a larger cost reducing effect (will be more profitable) than a corresponding increase in e_O. This relationship only holds until a certain threshold Z is breached, which means that not all effort will be allocated to the prevention of train delay failures.

7 Dewatripont et al. (2000) characterizes this two-task situation by a strictly positive cross-partial derivative of the disutility of effort $\frac{\partial^2 \Psi(e_T, e_O)}{\partial e_T \partial e_O} > 0$.

8 $E[Cj] = M_{ij}(X, e) + b_{ij} \cdot CA_{ij}(FT + FO) + PT \cdot FT_{ij} + PO \cdot FO_{ij} + \varphi(FO)^2$

9 For simplicity, we assume that $\frac{\partial M(X, e_T)}{\partial e_T} = \frac{\partial M(X, e_O)}{\partial e_O}$

3.2.1 Estimating the effect of tilted performance incentive schemes

There is no point in time where performance incentive schemes were introduced. For example, there are examples of performance clauses in contracts awarded to the in-house production units prior to the introduction of competitive tendering. It is reasonable to assume that in-house production in general had some sort of incentive structure to reduce train delays. We can however use the *sampling benefit* from competitive tendering, which imply that it is more likely that the chosen contractor is efficient (see for example Armstrong and Sappington 2007, chapter 4). This is also suggested by the results in Odolinski and Smith (2016), showing that competitive tendering reduced maintenance cost in Sweden with about 11 per cent (which of course also can be explained by other factors than just the sampling benefit).

More specifically, we note that the effect of a maintenance strategy, where the choice of e_T and e_O is tilted to the former, is increasing with the efficiency parameter γ_j (the number of failures decreases with efficiency; see equations 9 and 10). Hence, the introduction of competitive tendering can be used to test the effect of performance incentive schemes, given that tendering has resulted in more efficient contractors being awarded the maintenance projects. Our hypothesis is therefore that the use of competitive tendering will increase the effect of the tilted performance incentive schemes. We test our hypothesis by estimating two models. *Model 2a* is related to train delays

$$FT_{it} = \alpha_i \ + \ \boldsymbol{\beta X}_{it} + \vartheta_T D_{it} + \varepsilon_{it} \tag{12}$$

D_{it} is a dummy variable indicating when a maintenance area is tendered in competition and is used as a proxy variable for a change in the effect of tilted performance incentive schemes. The other model is

$$FO_{it} = \alpha_i \ + \ \boldsymbol{\beta X}_{it} + \vartheta_O D_{it} + \varepsilon_{it}, \tag{13}$$

referred to as *Model 2b*, where failures not causing train delays is the dependent variable. According to (11), the parameters ϑ_T and ϑ_O should differ between *Model 2a* and *2b*. Hence, we state the following hypothesis:

$$Hypothesis\ 2: \vartheta_T < \vartheta_O$$

3.2.2 Selection bias

We need to consider a possible selection bias when estimating *Model 2a* and *2b*. The maintenance of the Swedish railway network was gradually put out to tender, with the first contract tendered in 2002 and the last part of the network tendered in competition in 2014. The estimates from the tendering dummy variables in *Model 2a* and *2b* will be biased if there are systematic differences between areas tendered first and tendered later that are not controlled by the independent variables; omitted variable bias will be present. A selection bias can also be present if we have reverse causality; if areas tendered first were tendered because they had high (low) probability of certain failures to occur. This issue – with respect to maintenance costs - is addressed in Odolinski and Smith (2016) who did not find such bias (see also Domberger et al. 1987 and Smith and Wheat 2012). We use the same approach in this paper and construct a vector of dummy variables:

$$z_{ikt} = [D_{ik}, D_{ikt}, D_{iM}, D_Y; \boldsymbol{\vartheta}], \tag{14}$$

where $k = C$ indicate when a segment is tendered in competition, $k = F$ when tendered during 2002-2004 for the first time and $k = L$ when tendered during 2005-2013 for the first time. The time period before tendered in competition is indicated by $t = B$ and $t = O$ when tendered in competition and onwards. The dummy variable D_{iM} is used for the year when the transition from not tendered to tendered takes place, i.e. the first year an area is tendered. z_{ikt} also includes year dummies ($k = Y = 2004, \dots, 2013$). $\boldsymbol{\vartheta}$ are parameters to be estimated.

As a robustness test of *Model 2a* and *Model 2b*, we estimate ϑ_{iFB}, ϑ_{iFO} and ϑ_{iLO} and test if $\vartheta_{iFB} = 0$ which would imply that we have no systematic difference between areas tendered in 2002-2004 compared to areas tendered in 2005-2013 (before they were tendered) and areas not tendered during 2003-2013.[10]

We note that a general difference-in-differences approach would include a dummy variable indicating all areas tendered and a dummy variable for the period after tendering, as well as an interaction between these variables; see for example Greene (2012, p.155-157). We do not have a general post-tendering period as the exposure to competition was gradual. Hence, we use year dummy variables to control for general effects that occur over time which leaves the tendering variable to pick up the impact of tendering. Moreover, we also include a dummy variable indicating all areas tendered in competition sometime during 2003-2013 along with the time-specific tendering variable, in line with the difference-in-differences approach.

3.3 Regression model

As previously mentioned, the performance measure used in this paper is the number of failures on the railway infrastructure, and is also the dependent variable in the models. This variable consists of non-negative discrete values, i.e. it is a count variable. The distribution of the number of failures in Figure 2 shows that a large share of the observations has no failures.

The number of failures is clearly not normally distributed. A Poisson regression can be used for this type of count data, where the number of events (y_i) that is occurring during a certain time interval is assumed to be Poisson distributed, where $i =$individual $1, 2 \dots N$. The events can be explained by k variables x_{ik}. In a Poisson distribution, the conditional mean is

[10] The definition of areas tendered first is arbitrary because the exposure to competition was gradual, and we therefore perform sensitivity tests with respect to this definition.

equal to its conditional variance ($E[y_i \mid x_{ik}] = VAR[y_i \mid x_{ik}]$). However, the variance can be greater than the mean (overdispersion), violating the restriction in the Poisson model. Overdispersion can be present when a large fraction of the observations have a zero value, which is indicated by Figure 2 below.[11] Indeed, the sample variance is many times larger than the sample mean for the dependent variables used in *Model 1*, *2a* and *2b* (see sample means and standard deviations in Table 1).

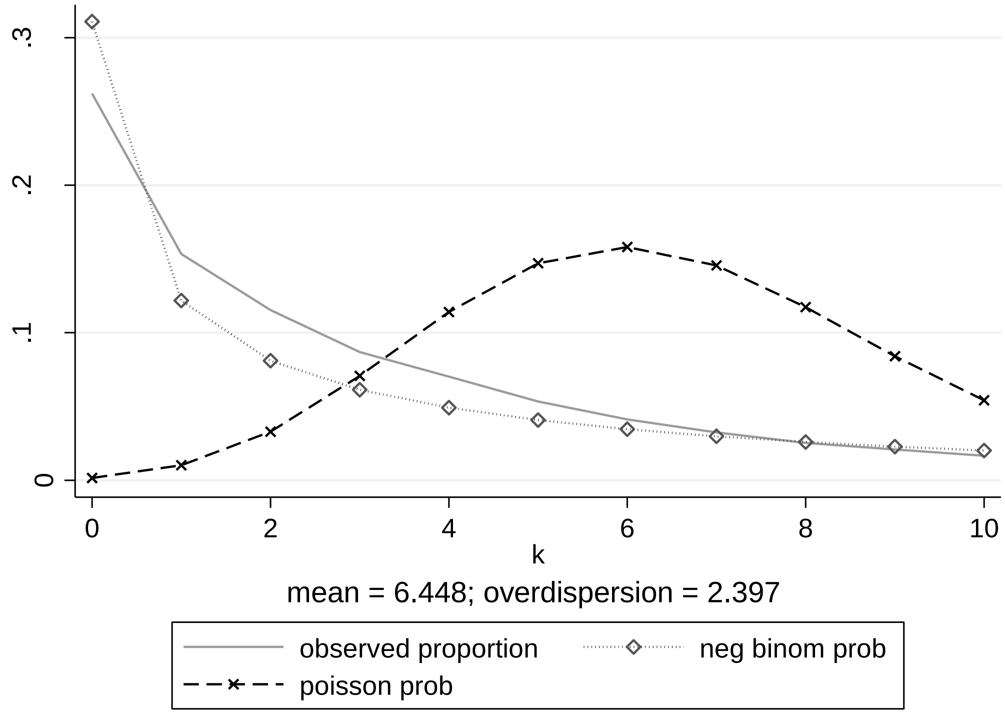

Figure 2 – Proportions of observations: observed, Poisson- and negative binomial probability

In this case, the negative binomial model is an alternative regression model in which the conditional mean is not equal to the conditional variance, which can be modelled

[11] The overdispersion in Figure 2 is estimated from the pooled negative binomial model and is significantly different from zero according to a likelihood ratio test ($chi2(1) = 3.8e + 05$).

as $VAR[y_i \mid x_{ik}] = \lambda_i + \delta_i\lambda_i^2$. This model is referred to as the NB2 model in the literature (see Cameron and Trivedi 2013), where δ_i is the dispersion parameter to be estimated.[12]

With access to panel data, i.e. a cross-section of contracts observed over time, we have a good opportunity to control for individual heterogeneity. The conditional mean is

$$E[y_{it} \mid x_{ikt}, \alpha_i] = \alpha_i\lambda_i = \alpha_i \exp(x'_{ikt}\beta_k), \tag{15}$$

where α_i is the individual specific effect. If α_i is independent of the regressors we would prefer a random effects model resulting in efficient and consistent estimates of β. If the regressors are correlated with α_i we could use a fixed effects model, resulting in consistent estimates of β.[13] However, Allison and Waterman (2002) found that the conditional fixed effects negative binomial model, proposed by Hausman et al. (1984), is not a true fixed effects model because the time-invariant variables do not drop out from the estimation. Hence, the fixed effects are not conditioned out in this model.[14] An unconditional fixed effects negative binomial model can be used instead, which controls for the individual unobserved effects by including individual-specific dummy variables. A large panel (many individuals) can however result in biased parameter estimates due to the incidental parameters problem, first presented by Neyman and Scott (1948). This leaves us with the random effects negative binomial model.

We will nevertheless have inconsistent estimates if the regressors are not independent of the individual effects α_i. A solution, first proposed by Mundlak (1978) for the linear model (see also Chamberlain (1982, 1984)), is to include averages over time of the variables. This

---

[12] The conditional variance in the NB1 model is $VAR[y_i \mid x_{ik}] = \lambda_i + \delta_i\lambda_i$

[13] In the fixed effects negative binomial model, the dispersion is constant $(1 + \delta_i)$ for each individual, while the random effects model assumes that $1/(1 + \delta_i) \sim Beta(r, s)$, which means that δ_i can vary randomly between individuals (see for example Hilbe 2011 for a specification of the negative binomial models with fixed and random effects).

[14] Guimarães (2008) showed that the fixed effects are conditioned out only if the fixed effect equals the logarithm of the dispersion parameter, i.e. $\alpha_i = \ln(\delta_i)$

model produces ‘within’ estimates using the random effects estimator and is often referred to as the correlated random effects model. For the non-linear case (see for example Papke and Wooldridge 2008, Cameron and Trivedi 2013), the individual effect can be specified as

$$\alpha_i = \exp(\bar{x}'_{ikt}\gamma_k + \varepsilon_{it}), \tag{16}$$

where $\bar{x}'_{ik} = T^{-1}\sum_{t=1}^{T} x_{ikt}$ for each $k = 1 \ldots K$. Using (16) we can then express (15) as

$$\exp(x'_{ikt}\beta_k + \bar{x}'_{ik}\gamma_k + \varepsilon_{it}), \tag{17}$$

where we are controlling for the correlation between α_i and our regressors x'_{ikt} via $\bar{x}'_{ik}$. Hence, we assume that the remaining individual effect in α_i is independent of x'_{ikt}. Furthermore, (17) assumes that the relationship between α_i and $\bar{x}'_{ik}$ is linear. We can also test interactions and quadratic terms between different $\bar{x}'_{ik}$ (Papke and Wooldrige 2008), where a failure to reject the null hypothesis of a parameter estimate equal to zero will lead us to drop the nonlinear term. Moreover, we can avoid collinearity between x'_{ikt} and $\bar{x}'_{ik}$ by using deviations from the mean

$$\exp[(x'_{ikt} - \bar{x}'_{ik})\beta_k + \bar{x}'_{ik}\gamma_k + \varepsilon_{it}], \tag{18}$$

where $T^{-1}\sum_{t=1}^{T}(x'_{ikt} - \bar{x}'_{ik}) = 0$ for each individual i (see Allison 2009 and Bell and Jones 2015).

We estimate a negative binomial regression model on a panel data set stretching from 2003-2013. The model we estimate is:

$$\Pr(F_{it} = f_{it} \,|\, x_{ikt}, z_{ikt}, \alpha_i, \delta_i), \tag{19}$$

where F is the count of failures, i = track segment $1,2, \ldots, N$ and t = year $1,2 \ldots, T(i)$. x_{ikt} is a vector of k explanatory variables, including infrastructure characteristics and traffic volume. A variable for different cost levels for compensation when rectifying a failure, i.e. the reimbursement rule, is included in *Model 1*.

Track length is an important exposure variable in the model. We expect the coefficient for track length to not be significantly different from 1, meaning that a segment with track length 2 km is twice as likely to have a failure as a segment with track length 1 km, *ceteris paribus*. If the track length coefficient was significantly above 1, the 2 km section would be more than twice as likely to have a failure compared to the 1 km section. This would reflect that track length captures other aspects than just exposure; for example if variations in infrastructure characteristics are not explained by other variables in the model estimation and are instead picked up by the track length variable.

z_{ikt} is a vector of dummy variables, containing year dummies and a dummy variable for when a track segment belongs to a contract area tendered in competition, as well as a dummy variable indicating when there is a transition from not tendered to tendered in competition (which in most cases does not happen in the beginning of a calendar year). See section 3.2.2 and equation (14) for definitions of the dummy variables. The tendering dummy is also interacted with a time trend taking the value 1 during the first year of tendering, 2 during the second year etc. The tendering dummies are included in *Models 2a* and *2b*. Finally, α_i is the individual effect as specified in (16) and δ_i is the dispersion parameter in the model, where we assume that $1/(1+\delta_i) \sim Beta(r,s)$.

4.0 Data

The data set is an unbalanced panel over the period 2003-2013 and has been provided by the Swedish Transport Administration (which we hereafter refer to as the infrastructure manager, IM). The number of failures constitutes all failures reported to the IM that needed to be fixed immediately or within two weeks. The failure reports come from the train management system and may emanate from operators, train drivers, maintenance personnel as well as the public.

There are many different causes of failures. Some are strictly exogenous with respect to maintenance such as animals or humans hit by a train, sabotage etc. These failures are out of the contractors' control and are not included, meaning that only failures occurring because of deterioration and/or poor maintenance of the infrastructure are analysed.

The data set contains failures causing train delay as well as other "regular" failures. During 2003-2009 a train had to be delayed more than 5 minutes between two stations for a failure causing the delay to be reported as such. This definition was changed to 3 minutes in 2010. To consistently analyse the number of train delay failures during 2003-2013 we only include failures causing more than a 5 minute delay. Failures causing less than 5 minutes delay are therefore defined as a regular failure in this study. Furthermore, we were not able to get consistent information about the knock-on effects of a first train being delayed, meaning that it is impossible to report the total number of delay minutes per failure from the available data.[15]

The failures reported to the IM are linked to different parts of the infrastructure and its location on the railway network, that is, between the two stations the failure was located. Each of these segments is an individual (i) in our estimations. Some of the segments have a very short track length (for example 10 metres) as they only constitute a switch or a bridge.

Not all data is reported at the segment level. The traffic volume is available at the track section level, while the cost-reimbursement rule and the tendering dummies are available at the contract area level. The length relationship between a segment, section and area is shown in Figure 3. The number of segments per track section and the number of track sections per contract area varies; Figure 3 is only an illustration of the fact that $segment\ km\ <\ section\ km\ <\ contract\ area\ km.$

[15] Moreover, in 2010-2013 we do not know if a total delay of 100 minutes (aggregated minutes of delay between several stations due to one infrastructure failure) contain any 3 minute delays between two stations or not.

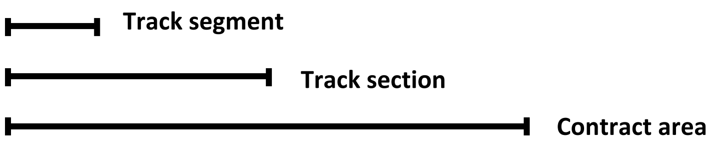

Figure 3 – Track length relationship between segments, sections and contract areas

Table 1 shows the descriptive statistics for failures and the other explanatory variables included in the estimations. In total we have 24 940 observations on segments administered by the IM over the period 2003-2013. However, only tendered contracts have a reimbursement rule. This information is available for a third of the observations. In the analysis of the reimbursement rule we exclude the design-bid-maintain contracts because these are, by definition, not performance based (which is the basis for our derived model) and are primarily used for the newly built railway lines. Moreover, we exclude yards when analysing train delay failures because these are exempted from the bonus/penalty system with respect to train delays in the maintenance contracts.

Table 1 - Data 2003-2013

Variable (24 940 obs.)	Mean	Std. Dev.	Min	Max
SEG.: Failures, total	6.32	17.75	0	482
SEG.: Failures, train delays	1.11	2.85	0	101
SEG.: Failures, not train delays	5.22	15.47	0	413
SEG.: Track length, metres	5 871	5 695	10	43 870
SEG.: Rail weight, kg	51.58	5.68	27	63
SEG.: Quality class, 0-5*	2.06	1.26	0	5
SEC.: Traffic density, million gross tonnes**	8.29	8.62	0.00	49.79
SEC.: Tendered, dummy	0.58	0.49	0	1
SEC.: Transition to tend., dummy	0.08	0.27	0	1
Subset of data used in *Model 1* (8528 obs.)				
ARE.: Cost-reimbursement rule, thousand SEK	7.66	4.04	5	20
SEG.: Failures, total	5.95	15.11	0	482
SEG.: Failures, train delays	1.12	2.85	0	101
SEG.: Failures, not train delays	4.83	12.91	0	381

SEG.: Track length, metres	6 052	6 262	10	43 077
SEG.: Rail weight, kg	51.73	5.82	32	60
SEG.: Quality class, 0-5*	2.01	1.32	0	5
SEC.: Traffic density, million gross tonnes**	7.89	9.32	0.00	49.79

SEG = segment, SEC = section, ARE = area

*A high value implies a low speed line with less strict requirement on track geometry standards compared to a high speed line (Banverket 1997)

**Traffic density = (Million gross tonnes-km/Route km)

The letters preceding each variable name denotes whether the information is available at the segment-, section-, or contract area level. As noted in Table 1, the information on traffic volume is available at the track section level. We make the assumption that each segment has the same traffic volume as the section it belongs to, which need not be the case in reality.

To provide an indication of the relationship between failures and the cost-reimbursement rule, we summarise the number of failures per million ton-km for each group of segments with the same cost-reimbursement rule. The correlation coefficient for these “cost-reimbursement groups” and failures per million ton-km is 0.50. Still, there are other factors than ton-km that can explain the number of failures. The rail weight is a proxy for the quality of the rail (and newer rails are generally heavier than the old). Moreover, the quality classification of the tracks determines the maximum speed allowed and the related track quality requirements with respect to track geometry. Other variables capturing the infrastructure’s characteristics, such as curvature and sleeper age, have been analysed but excluded from the estimations due to high collinearity between variables (for example between rail weight and sleeper age).

5.0 Results

Three models are estimated. *Model 1* considers the effect of different reimbursement rules on the number of failures, and *Models 2a* and *2b* the effect of incentive schemes with weights put on train delays.

$\hat{\lambda}(x)_{it}$ and $\hat{\lambda}(x+1)_{it}$ is estimated, which means that we estimate the expected value of failures when the explanatory variable x_{it} increases with one unit. The estimated coefficient is then $\hat{\beta} = \ln\left[\frac{\hat{\lambda}(x+1)_{it}}{\hat{\lambda}(x)_{it}}\right]$, referred to as a *semi-elasticity*. $e^{\hat{\beta}}$ is an incidence ratio (IRR), expressed as $\frac{\hat{\lambda}(x+1)_{it}}{\hat{\lambda}(x)_{it}}$. Hence, an IRR<1 indicates a negative effect. The incidence ratios are reported in Tables 2 and 3 together with standard errors for the estimated coefficients $\hat{\beta}$. All estimations are carried out with Stata 12 (StataCorp.2011).

5.1 Econometric results: Model 1

Table 2 shows the results from the estimations of the first model, which include results from both the random effects model and the preferred correlated random effects model. In the latter model, the coefficients for variables averaged over time are denoted ‘between estimates’ while the other coefficients are denoted ‘within estimates’ (referring to effects between and within segments, respectively).

The ‘within estimate’ for rail weight is not significant in our preferred model, but has the expected sign (IRR=0.9987); the number of failures is expected to decrease when heavier rails are installed (note that there is a negative correlation between rail age and rail weight as old rails are lighter). However, the average rail weight per segment (RAIL_WEIGHTbar) picks up the ‘between effects’ and has a significant IRR at 0.9625. The difference in the parameter estimates for rail weight illustrates the difference between the random effects model and the correlated random effects model. The estimate in the random effects model uses both the within and between effects, which results in a significant IRR at 0.9789, which is between the estimates for rail weight in the correlated random effects model.

The IRR for quality class – which determines the maximum speed allowed and corresponding requirements on track standards – is not significant in the estimations. Track length (TRACK_L), which is the exposure variable, has the expected IRR of 1 and is significant. The estimations includes a squared term for million gross tonne density, and the estimates reflect a non-linear relationship with the number of failures, which is shown by both the within and between estimate in the correlated random effects model.

Note that only the period 2004-2013 is included in this estimation due to missing data (reimbursement rule), which means that we include year dummy variables for 2005-2013. We tested the average values over time for the year dummy variables in the estimation because we have an unbalanced panel (Wooldridge 2013), but these were not jointly significant and dropped from the estimation.

Turning to the 'within estimate' for the contract design variable, REIMBR, we see that it has a negative effect on the number of failures (IRR=0.9610, p-value=0.000); we cannot reject $Hypothesis$ 1. Hence, the estimation results suggest that an increase in the reimbursement rule gives an incentive to increase preventive maintenance (reduces moral hazard), which is a contract the efficient type is more likely to be awarded (selection effect). The incidence rate ratio at 0.9610 indicates that an increase in the reimbursement rule with one unit (in our case with 1000 SEK) will reduce the number of failures with 3.9 percent (100*(1-0.961)). The average number of failures per contract and year is 340 in the sample (5.95 per segment).[16] Hence, the estimated effect of an increase in the reimbursement rule implies around 13 fewer failures per year for the average contract. We made sensitivity tests

[16] The average track length in a contract area in the sample is about 340 km and the average segment length is 6 km

using either train delay failures or other types of failures as the dependent variable, which did not have an effect on the estimate for the cost-reimbursement rule.

Table 2 - Results Model 1

	Random effects		*Correlated Random effects*	
	IRR	Std. Err.	IRR	Std. Err.
Constant	5.4600***	1.5638	13.6631***	4.6980
REIMBR	0.9987	0.0051	0.9610***	0.0095
RAIL_WEIGHT	0.9789***	0.0049	0.9928	0.0079
QUALAVE	1.0151	0.0225	0.9688	0.0401
TRACK_L	1.0000***	0.0000	1.0000**	0.0000
MTGTDEN	1.1014***	0.0077	1.0385***	0.0123
MTGTDEN2	0.9983***	0.0002	0.9993**	0.0003
D.2005	1.5680***	0.1830	1.6390***	0.1953
D.2006	1.4461***	0.1655	1.5513***	0.1810
D.2007	1.7700***	0.2007	1.9625***	0.2265
D.2008	1.6242***	0.1840	1.8119***	0.2091
D.2009	1.5133***	0.1713	1.7102***	0.1971
D.2010	1.4245***	0.1611	1.6221***	0.1870
D.2011	1.4504***	0.1642	1.6294***	0.1879
D.2012	1.4578***	0.1653	1.6462***	0.1906
D.2013	1.5515***	0.1757	1.7608***	0.2045
REIMBRbar	-	-	1.0084	0.0059
RAIL_WEIGHTbar	-	-	0.9625***	0.0061
QUALAVEbar	-	-	1.0289	0.0285
TRACK_Lbar	-	-	1.0001***	0.0000
MTGTDENbar	-	-	1.1335***	0.0100
MTGTDEN2bar	-	-	0.9978***	0.0002

***, **, *: Significance at 1%, 5%, 10% level, respectively

Log likelihood: Random effects model= -18 882.341, Correlated Random effects model= -18 841.726

Number of observations = 8 528

Number of segments=1 836

Definition of variables in table 2:

REIMBR[a] = reimbursement rule stating the cost of rectifying a failure that is included in the fixed payment (see section 2.0)

RAIL_WEIGHT[a] = Rail weight, kg

QUALAVE[a] = Average quality class; a high value of average quality class implies a low linespeed

TRACK_L[a] = Track length, metres

MTGTDEN[a] = Million gross tonnage density (gross tonnes-km/track km)

MTGTDEN2[a] = MTGTDEN^2

D.2005-D.2013[a] = Year dummy variables, 2005-2013

REIMBRbar = $T^{-1}\sum_{t=1}^{T}$ REIMBR

RAIL_WEIGHTbar = $T^{-1}\sum_{t=1}^{T}$ RAIL_WEIGHT

QUALAVEbar = $T^{-1}\sum_{t=1}^{T}$ QUALAVE

TRACK_Lbar = $T^{-1}\sum_{t=1}^{T}$ TRACK_L

MTGTDENbar = $T^{-1}\sum_{t=1}^{T}$ MTGTDEN

MTGTDEN2bar = $T^{-1}\sum_{t=1}^{T}$ MTGTDEN^2

[a] Deviations from the mean are used in the correlated random effects model $x'_{ikt} - T^{-1}\sum_{t=1}^{T} x_{ikt}$ (see section 3.3)

5.2 Econometric results: Model 2

The estimation results from *Models 2a* and *2b* using correlated random effects are presented in Table 3 (results from the random effects model are presented in the appendix). In *Model 2a*, the number of failures causing train delays is used as the dependent variable. The dependent variable in *Model 2b* is failures not causing a train delay.

Rail weight has the expected effect on the number of failures; fewer failures with higher rail weight. The effect of track length and traffic is similar to the results in *Model 1* (however the ‘within estimate’ for track length is not significant). The ‘within estimate’ for quality class is significant in *Model 2a*, showing an increase in train delay failures for lower linespeeds (lower requirements on track standards), while the ‘between estimate’ show a significant decrease in train delay failures for lower linespeeds. These results suggest that the required maintenance is not sufficient with respect to the number of failures when there is a change in average quality class within a segment. However, when comparing different segments, lower linespeeds (low requirements on track standard) is associated with fewer train delay failures.

There are differences in the effect of the competitive tendering between *Model 2a* and *Model 2b* according to the estimation results. In *Model 2a*, the IRR for competitive tendering, CTEND, is 0.9574 (p-value=0.172) with a 95 per cent confidence interval at [0.8996, 1.0190], which indicates a negative effect on the number of failures causing train delays, yet not statistically significant. The IRR for competitive tendering in *Model 2b* is 1.0729 (p-value=0.000) with a 95 per cent confidence interval at [1.0351, 1.1121], which implies that the number of failures - that has not caused a train delay - is increasing when tendered in competition. The lower parameter estimate in *Model 2a* compared to *2b,* and more importantly the non-overlapping 95 per cent confidence intervals, is in line with our hypothesis in section 3.2 (we cannot reject *Hypothesis* 2).[17] An incentive scheme weighted against train delays will affect the contractor's maintenance strategy.

Note that we include a dummy variable indicating all segments tendered sometime during 2003-2013 (DTEND) to control for any general feature among these segments that also apply before tendering. The corresponding IRR is below one in both models, but not statistically significant (p-values at 0.746 and 0.669 in the respective models). Moreover, no selection bias was found using the dummy variable approach described in section 3.2.2. That is, the parameter estimate of interest $\vartheta_{IF,iFB}$ is not significantly different from zero (p-value=0.883 and 0.183 in *Model 2a* and *2b* respectively).

The effect of competitive tendering over time was estimated using a trend variable. The parameter estimate was not significant in either of the model estimations and this variable is excluded in our preferred models.

[17] In fact, estimating the 99 per cent confidence interval for the difference between the estimates we can conclude that these are significantly different even at the 1 per cent level (see Cohen et al. 2003, p.46-47).

Table 3 - Results Model 2a and 2b: Correlated Random Effects

	Model 2a - Train delay failures		*Model 2b - Other failures*	
	IRR	Std. Err.	IRR	Std. Err.
Constant	1.7860	0.9482	3.2110***	1.1629
MIXTEND	1.0117	0.0337	1.0467**	0.0197
CTEND	0.9574	0.0305	1.0729***	0.0196
DTEND	0.8767	0.3559	0.9030	0.2152
RAIL_WEIGHT	0.9644**	0.0054	0.9776***	0.0032
QUALAVE	1.0846**	0.0402	1.0110	0.0205
TRACK_L	1.0000	0.0000	1.0000	0.0000
MTGTDEN	1.0581***	0.0120	1.0368***	0.0071
MTGTDEN2	0.9993***	0.0002	0.9992***	0.0001
D.2004	0.9002***	0.0354	0.9165***	0.0195
D.2005	0.9359*	0.0367	0.8499***	0.0186
D.2006	0.9406	0.0375	0.8292***	0.0186
D.2007	1.1342***	0.0445	1.0402*	0.0227
D.2008	1.0492	0.0441	0.9772	0.0227
D.2009	0.9904	0.0436	0.9763	0.0235
D.2010	1.2482***	0.0552	0.9169***	0.0232
D.2011	1.3242***	0.0604	0.8726***	0.0230
D.2012	1.2271***	0.0577	0.7626***	0.0210
D.2013	1.3915***	0.0651	0.7984***	0.0220
MIXTENDbar	2.1022	1.2261	2.2887*	0.9754
CTENDbar	0.9799	0.1003	0.9980	0.0813
DTENDbar	0.9702	0.1225	1.0440	0.1065
RAIL_WEIGHTbar	0.9656***	0.0058	0.9726***	0.0046
QUALAVEbar	0.9230***	0.0213	1.0145	0.0192
TRACK_Lbar	1.0000***	0.0000	1.0000***	0.0000
MTGTDENbar	1.2122***	0.0093	1.1214***	0.0070
MTGTDEN2bar	0.9962***	0.0002	0.9976***	0.0002

***, **, *: Significance at 1%, 5%, 10% level, respectively

Note: D.2004bar-D.2013bar are jointly significant and included in the estimations, but dropped from table 3 for expositional convenience

Deviations from the mean are used in both models: $x'_{ikt} - T^{-1}\sum_{t=1}^{T} x_{ikt}$ (see section 3.3)

Log likelihood: *Model 2a* = -27 635, *Model 2b* = -51 402

Number of observations in *Model 2a* and *2b*: 24 940

Number of segments *Model 2a* and *2b*: 2 806

Definition of variables in table 3:

MIXTEND = Dummy for years when mix between tendered and not tendered in competition, which is the year when tendering starts for a segment
CTEND = Dummy indicating when a segment is tendered in competition
DTEND = Dummy indicating all segments tendered in competition sometime during 2003-2013
MIXTENDbar = $T^{-1}\sum_{t=1}^{T}\text{MIXTEND}$
CTENDbar = $T^{-1}\sum_{t=1}^{T}\text{CTEND}$
DTEND=$T^{-1}\sum_{t=1}^{T}\text{DCTEND}$

6.0 Discussion

The design of contracts is vital for the outcome of the maintenance projects, which places high demands on the IM as a client. The presence of hidden information and hidden action can result in inefficient outcomes if not judiciously handled. Incentive contracts linear in costs can be used to alleviate the problems incurred by these information asymmetries. Indeed, this type of contract is used by the IM in the tendering of maintenance contracts in Sweden, where different reimbursement rules have been used over the years of competitive tendering creating different incentive intensities. The estimation results show that an increase in the reimbursement rules reduces the number of infrastructure failures.

Does this result imply that we should have a high reimbursement rule in all maintenance contracts? Not necessarily. A high reimbursement rule indicates that we move closer to a fixed price contract which induces effort, but will make it easier for the efficient contractor to extract rent. Moreover, we will have a low level of competition if inefficient types do not take part in the bidding when reimbursement rules are too high, with the efficient type(s) being able to extract higher rents.

An important task of the contractors is to prevent infrastructure failures that are causing train delays. A robust and reliable railway infrastructure is an objective often stated by the IM. See for example the IM's report on strategic challenges in 2012-2021 (Trafikverket 2011). Undeniably, this objective is reflected in the design of the maintenance contracts, with performance incentive schemes tilted against train delays. This makes it beneficial for the contractors to focus on this class of infrastructure failures. The estimation results confirm our hypothesis, suggesting that effort is tilted towards preventing failures causing train delays at the expense of preventing other failures. This is in line with the results from the multitask principal-agent model by Holmström and Milgrom (1991); increased incentives for one task can result in a reallocation of attention from other tasks.

Are the performance incentive schemes beneficial with respect to the performance of the railway infrastructure? Unfortunately, we are not able to answer this question. For example, we do not have consistent information on total train delay minutes that each failure caused, which is an important overall measure of railway performance. A reduction in the number of train delay failures does not *per se* imply that the number of train delay minutes has decreased. Nevertheless, it is fair to say that a reduction in the number of train delay failures is a good sign of improved performance (note, however, that the estimate was not statistically significant). Still, the number of failures not causing a train delay has increased and possible consequences of this observation need to be further studied. For example, will this have an effect on the life cycle cost of the infrastructure? This is especially relevant considering the negative experience in Britain where misaligned incentive structures led to a deteriorating asset condition.

7.0 Conclusion

This paper offers evidence on the effect of different contract designs in rail maintenance services. It contributes to the existing literature by providing empirical evidence on the marginal effect of incentive intensity in the rail maintenance contracts, as well as the effect of tilted performance incentive schemes. More precisely, we have shown that a higher reimbursement rule increases the comparative advantage of the efficient contractor in the bidding for contracts and that it induces effort, generating a higher level of preventive maintenance. The econometric results show that the reimbursement rule reduces the number of failures. However, we do not know if the effect is due to differences in efficiency among the contractors and/or differences in effort level. Still, the results show that the marginal effect of an increase in the incentive intensity in the contracts - corresponding to a 1000 SEK increase per failure in the cost-reimbursement rule - is a 3.9 per cent reduction of the number of failures.

The econometric test of the tilted performance incentive schemes confirms our hypothesis that it has an effect on the relationship between the number of failures causing train delays and other failures. We can conclude that this contract design seems to have been beneficial with respect to the number of train delay failures, yet at the expense of other types of failures.

Our findings are informative in considerations on the design of railway maintenance contracts, especially for other IMs across Europe that plan to use competitive tendering. Setting a reimbursement rule too low can be costly for the IM with respect to the number of failures that occur, while a high reimbursement rule can induce rent extraction. Moreover, when using tilted performance incentive schemes, the IM needs to contemplate the reallocation of attention from other tasks. For example, its effect on future maintenance costs needs to be considered.

The results from this paper gives some support to the finding that the 11 per cent cost reduction in Sweden due to competitive tendering was not associated with a lower quality (see Odolinski and Smith 2016), as measured by the number of train delay failures. An important caveat is, however, that we do not know its effect on train delay minutes, or the long-term effects of an increasing number of other types of failures. Whether the reallocation between different efforts is cost efficient or not needs to be further investigated. In general, the effect of different designs on cost efficiency in railway maintenance - considering both user and producer costs - is an area for future research. Such considerations are critical in the study of optimal contract design within this field.

Acknowledgements

The author is grateful to the Swedish Transport Administration (Vivianne Karlsson and Anders Nilsson) for providing data. Special thanks to Jan-Eric Nilsson, Svante Mandell, Lars Hultkrantz and Johanna Jussila Hammes who have provided helpful comments on earlier version of this paper. The paper has also benefited from comments by participants at the ITEA conference in Oslo, June 17th 2015, at the First Nordic Meeting in Transport Economics in Oslo, November 26th 2014, and during a seminar at VTI, Stockholm, March 4th 2015. All remaining errors are the responsibility of the author.

References

Allison, P. D. and R. P. Waterman (2002): ‘Fixed-Effects Negative Binomial Regression Models’, *Sociological Methodology*, 32, 247-265

Allison, P. D. (2009): ‘Fixed effects regression models’, Series: Quantitative Applications in the Social Sciences; 160, SAGE Publications

Armstrong, M. and D. E. M. Sappington (2007): ‘Recent developments in the theory of regulation’, In: Armstrong, M. and R. H. Porter (Eds.), ‘Handbook of Industrial Organization, Volume 3, North-Holland, Elsevier, 1560-1678.

Bajari, P., and S. Tadelis (2001): ’Incentives versus Transaction Costs: A Theory of Procurement Contracts’, *The RAND Journal of Economics*, 32(3), 387-407.

Banverket (1997): ‘Spårlägeskontroll och kvalitetsnormer – Central mätvagn Strix’, Föreskrift BVF 587.02, (in Swedish).

Banverket (2000): ‘Utvecklingsstrategi för Banverkets produktionsverksamhet: Förutsättningar och strategier för konkurrensutsättning av Banverkets produktionsverksamhet’, GD-stab, GD00-2827/01, 2000-09-25 (In Swedish).

Bell, A. and K. Jones (2015): ’Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data’, *Political Science Research and Methods*, 3(1), 133-153

Cameron, A. C. and P. K. Trivedi (2013): ‘Regression Analysis of Count Data’, Second Edition, Cambridge University Press,

Chamberlain, G. (1982): ‘Multivariate Regression Models for Panel Data’, *Journal of Econometrics*, 18(1), 5-46

Chamberlain, G. (1984): ‘Panel Data’, in *Handbook of Econometrics*, Volume 2, eds. Z. Griliches and M.D. Intriligator. Amsterdam, North Holland, 1248-13-18

Cohen, J., P. Cohen, S. G. West and L. S. Aiken (2003): ‘Multiple Regression/Correlation Analysis for the Behavioral Sciences’, Third Edition, Lawrence Erlbaum Associates, Inc.

Dalen, D. M., and A. Gómez-Lobo (2003): ‘Yardsticks on the road: Regulatory contracts and cost efficiency in the Norwegian bus industry’, *Transportation*, 30, 371-386

Devers, C. E., A. A. Canella Jr., G. P. Reilly, and M. E. Yoder (2007): ‘Executive Compensation: A Multidisciplinary Review of Recent Developments’, *Journal of Management*, 33(6), 1016-1072.

Dewatripont, M., I. Jewitt, and J. Tirole (2000): ‘Multitask agency problems: Focus and task clustering’, *European Economic Review*, 44, 869-877

Domberger, S., S. Meadowcroft, and D. Thompson (1987): ‘The Impact of Competitive Tendering on the Costs of Hospital Domestic Services’, *Fiscal Studies,* 8(4), 39-54

European Commission (2011): ‘Public Procurement indicators 2011’, Brussels, 5 December 2012:

http://ec.europa.eu/internal_market/publicprocurement/docs/modernising_rules/public-procurement-indicators-2011_en.pdf

Greene, W.H. (2012): 'Econometric Analysis, 7th edition', Prentice-Hall.

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

Guimarães, P. (2008): 'The fixed effects negative binomial model revisited', *Economic Letters*, 99, 63-66

Hart, O., A. Shleifer, and R. W. Vishny (1997): 'The proper scope of government: theory and an application to prisons', *The Quarterly Journal of Economics*, 112(4), 1127-1161.

Hausman, J., B.H. Hall and Z. Griliches (1984): 'Econometric Models for Count Data with an Application to the Patents-R & D Relationship', *Econometrica*, 52(4), 909-938

Hilbe, J. M. (2011): 'Negative Binomial Regression', 2nd edition, Cambridge University Press

Holmström, B. and P. Milgrom (1991): 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design', *Journal of Law, Economics, & Organization*, 7, 24-52

Kennedy, J. and A.S.J. Smith (2004): 'Assessing the Efficient Cost of Sustaining Britain's Rail Network. Perspectives based on zonal Comparisons', *Journal of Transport Economics and Policy*, 38(2), 157-190.

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

Laffont, J.-J., and J. Tirole (1986): 'Using Cost Observation to Regulate Firms', *Journal of Political Economy*, 94(3), 614-640

Laffont, J.-J., and J. Tirole (1987): 'Auctioning Incentive Contracts', *Journal of Political Economy*, 95(5), 921-937

Laffont, J.-J., and J. Tirole (1993): ‘A Theory of Incentives in Procurement and Regulation’, The MIT Press, Cambridge, Massachusetts

Lazear, E. P. and P. Oyer (2013): ‘Personnel Economics’, In: Gibbons, R. and J. Roberts (Eds), ‘Handbook of Organizational Economics’, Princeton University Press, Princeton and Oxford, 479-519.

Lewis, G. and P. Bajari (2014): ‘Moral Hazard, Incentive Contracts, and Risk: Evidence from Procurement’, *Review of Economic Studies*, 81(3), 1201-1228.

Mundlak, Y. (1978): ‘On the Pooling of Time Series and Cross Section Data’, *Econometrica*, 46(1), 69-85

Odolinski, K. and A. S. J. Smith (2016): ‘Assessing the cost impact of competitive tendering in rail infrastructure maintenance services: evidence from the Swedish reforms (1999-2011)’, *Journal of Transport Economics and Policy*, 50(1), 93-112.

Neyman, J., and E.L. Scott (1948): ‘Consistent estimation from partially consistent observations’, *Econometrica*, 16, 1-32

Papke, L.E. and J.M. Wooldridge (2008): ‘Panel data methods for fractional response variables with an application to test pass rates’, *Journal of Econometrics*, 145, 121-133

Piacenza, M. (2006): ‘Regulatory contracts and cost efficiency: Stochastic frontier evidence from the Italian local public transport’, *Journal of Productivity Analysis*, 25, 257-277

Podgursky, M. J. and M. G. Springer (2007): ‘Theacher performance pay: a review’, *Journal of Policy Analysis and Management*, 26(4), 909-949.

Rosenthal, M. B. and R. G. Frank (2006): ‘What Is the Empirical Basis for Paying for Quality in Health Care?’, Medical Care Research and Review, 63(2), 135-157.

Roy, W. and A. Yvrande-Billon (2007): ‘Ownership, Contractual Practices and Technical Efficiency: The Case of Urban Public Transport in France’, *Journal of Transport Economics and Policy*, 41(2) 257-282

Salanié, B. (2005): ‘The Economics of Contracts, A Primer’, Second edition, The MIT Press, Cambridge, Massachusetts

Smith, A.S.J., and P. Wheat (2012): ‘Evaluating Alternative Policy Responses to Franchise Failure. Evidence from the passenger rail sector in Britain’, *Journal of Transport Economics and Policy*, 46(1), 25-49.

StataCorp.2011. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP

Stenbeck, T. (2008): ’Quantifying Effects of Incentives in a Rail Maintenance Performance-Based Contract’, Journal of Construction Engineering and Management, 134(4), 209-516

Trafikverket (2010): ’Entreprenadkontrakt. Drift och underhåll av järnvägsanläggning Holmsund-Boden inklusive tvärbanor’, TRV2010/30428, F09-12859/IN60, Trafikverket 2010-11-16 (In Swedish)

Trafikverket 2011: ‘Strategiska utmaningar 2012-2021’, TDOK 2011:478, Trafikverket (In Swedish)

Van Herck, P., D. De Smedt, L. Annemans, R. Remmen, M. B. Rosenthal, and W. Sermeus (2010): ’Systematic review: Effects, design choices, and context of pay-for-performance in health care’, BMC Health Services Research, 10:247

Vickerman, R. (2004): ’Maintenance incentives under different infrastructure regimes’, *Utilities Policy*, 12, 315-322

Wooldridge, J.M. (2013): ’Introductory Econometrics: A Modern Approach, fifth edition’, South-Western, CENGAGE Learning

Wunsch, P. (1994): ‘Estimating Menus of Linear Contracts for Mass Transit Firms’, Mimeo, CORE

Appendix

Performance equation in maintenance contracts:

$$I = \beta_D \left(\frac{1}{2 + \frac{D_{outcome}}{D_{target\ value}}} \right) + \beta_F \left(\frac{1}{2 + \frac{F_{outcome}}{F_{target\ value}}} \right) + \beta_Q \left(\frac{1}{2 + \frac{Q_{target\ value}}{Q_{outcome}}} \right), \tag{21}$$

where D =train delay hours, F =number of failures, Q =quality number related to track geometry, where a higher number imply a better track geometry quality (which is why the target value is in the numerator in relation to the outcome value). Their respective coefficients are $\beta_D = 1.8$, $\beta_F = 1.0$ and $\beta_Q = 0.2$. If $I = 1$ the contractor has reached the target values, and will receive a bonus when $I > 1$.

Table 3 - Results Model 2a and 2b: Random Effects

	Model 2a - Train delay failures		*Model 2b - Other failures*	
	IRR	Std. Err.	IRR	Std. Err.
Constant	4.7074***	1.0991	7.9314***	1.2655
MIXTEND	1.0199	0.0339	1.0400**	0.0197
CTEND	0.9517	0.0290	1.0548***	0.0188
DTEND	1.0462	0.0866	1.1668**	0.0778
RAIL_WEIGHT	0.9696***	0.0039	0.9799***	0.0026
QUALAVE	0.9529***	0.0179	1.0002	0.0133
TRACK_M	1.0000***	0.0000	1.0000***	0.0000
MTGTDEN	1.1611***	0.0075	1.0861***	0.0048
MTGTDEN2	0.9974***	0.0002	0.9983***	0.0001
D.2004	0.8991***	0.0357	0.9169***	0.0198
D.2005	0.9321*	0.0368	0.8502***	0.0189
D.2006	0.9274*	0.0371	0.8274***	0.0188
D.2007	1.1170***	0.0437	1.0347	0.0227
D.2008	1.0243	0.0426	0.9690	0.0225
D.2009	0.9706	0.0423	0.9708	0.0234
D.2010	1.2162***	0.0531	0.9111***	0.0230
D.2011	1.3121***	0.0590	0.8781***	0.0231
D.2012	1.2371***	0.0570	0.7724***	0.0211
D.2013	1.3830***	0.0631	0.8024***	0.0219

***, **, *: Significance at 1%, 5%, 10% level, respectively

Log likelihood: *Model 2a* = -27 728, *Model 2b* = -51 475

Number of observations in *Model 2a* and *2b*: 24 940

Number of segments *Model 2a* and *2b*: 2 806