Contract design and performance of railway maintenance: empirical evidence on incentive intensity and performance payment

Kristofer Odolinski - JUNIOR SCHOLAR (PhD Student)

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

Draft version: 2015-05-25. DO NOT CIT

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, when tendered in competition, the performance payment schemes result in a decrease in infrastructure failures causing train delays, but an increase in other failures.
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/or 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. The objective with this paper is to estimate the effect of contract design on the performance of railway maintenance. More specifically, we consider two features of the design: the power of the incentive schemes and the structure of the performance payments.

The competitive tendering of railway maintenance is relatively new, as railway systems have generally been considered natural monopolies (however, the vertical separation of train operations and infrastructure management (IM) in Europe, initiated in 1991, was a first step in the deregulation of this sector in Europe). The contracting out of railway services poses some challenges, which Great Britain experienced after the privatization and subcontracting of track maintenance that led to initial cost reductions followed by concerns over quality and then increased costs; see for example Kennedy and Smith (2004). It can be argued that the optimal design of the maintenance contract is complex due to for example the dynamic aspect of maintenance (maintenance in one year can affect the maintenance required
the next year), as well as the interdependence between maintenance and renewals. This relates to the effect maintenance has on the service life of the infrastructure, which is significantly longer than the typical contract lengths in use. Moreover, there is limited access to the tracks where a balance needs to be attained between the train operator’s cost and the maintenance cost. This complexity makes it more difficult to align incentives.

With the British case in mind, learning-by-doing seems to have been a concept used in Sweden where the competitive tendering of railway maintenance was introduced gradually. The first contract was awarded in 2002, covering a small number of track sections, and the last part of the railway network was tendered in competition in 2014. This learning-process implies that different contracts designs have been used over the years, making it a fruitful case for empirical study. More precisely, the incentive intensity of contracts varies and its effect on performance can be tested.

One objective of the transfer to competitive tendering of maintenance contracts was to provide scope for innovation (Banverket 2000). To do so, contractors (firms) should be given degrees of freedom by the contracts. Most of the maintenance contracts are also 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 contractor will then have an incentive to develop the maintenance production.\(^2\) 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 generates a moral hazard situation because the client and the contractor have different main objectives, and the contractor’s actions may not be optimal for the client. In addition, the contractor can extract rent when information about its efficiency (technology) is not known to the client, which is

---

\(^1\) This formulation is used in view of the extensive reference to regulations and provisions in the contracts.

\(^2\) However, see Behn and Kant (1999) for examples on potential pitfalls with performance contracting.
the *ex ante* problem of adverse selection. This asymmetry in information means that the client has to make a trade-off between giving up rent to the efficient contractors and inefficient allocations when lowering the required effort level. The power of the incentive scheme is a central parameter in this trade-off (see Laffont and Tirole 1986), and the different levels of incentive intensity used in the maintenance contracts can result in varying performance.

Moreover, 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). Therefore, considering that the performance payments in the maintenance contracts in Sweden are weighted, which we describe in section 2.1, these can affect the attention to different tasks and hence 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). Empirical work has lagged behind, as noted by Chiappori and Salanié (2003) who provide a survey of empirical tests of contract theory. 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) uses 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), which showed 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 contract design 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 on incentives in rail maintenance contracts is made by Stenbeck (2008). Hence, our paper fills a gap in the literature by providing empirical evidence on the effect of incentive schemes on maintenance performance. Moreover, studies on the power of incentive schemes in procurement and regulation usually compare different types of contracts (for example fixed-price and cost-plus contracts). In our study, we have variation in the strength of incentives within the same contract type. Hence, we are able to estimate the marginal effect of the incentive intensity.

We propose a simple model for railway maintenance production to analyze the effect of the cost-sharing rule that creates the different levels of incentive intensity, and we provide an empirical test of its effect on performance. We also consider the performance payments in the contract that generates different incentives for different tasks. More specifically, we estimate the effect on the number of train delay failures vis-à-vis other failures due to the bonus/penalty schemes.

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. Our data is described in section 4. The results are presented in section 5, and we discuss our findings in section 6. Section 7 concludes.

2.0. Maintenance contract design

Most of the railway maintenance contracts in Sweden are performance-based contracts. 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

---

3 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.
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 cost level. To provide intuition, consider the following example which is 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 except for the fixed payment. The cost level (which we hereafter refer to as the cost ceiling) can vary between contracts, which creates different levels of power in the incentives. The same cost ceiling in each contract is used for maintenance activities that prevent infrastructure failures, i.e. for fixing a defect before it becomes a failure.
2.1 Performance payment

Except for capping the contractor’s cost for some activities, the contracts also provide a bonus and/or penalty linked to the number of failures in the maintenance area. These may be weighted 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. 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.

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 payment implies that a failure expected to cost more than the cost ceiling in the contract 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, which is defined as a failure that needs to be fixed immediately or within a week. There are other types of infrastructure quality indicators such as deviations in track geometry or other defects that require a preventive maintenance, i.e. activities that prevent infrastructure failures. Hence, when preventive maintenance has not been carried out, an infrastructure failure can occur which requires corrective maintenance.

The design of the incentive schemes in the contracts will affect our measure of performance. As described in section 2, different cost ceilings are used in the contracts which determine the power of the incentive schemes with respect to infrastructure failures (and defects), and performance payments may be used with higher penalty for failures causing train delays. We therefore formulate the following research questions:

1. Do variations in the cost ceiling affect the performance of maintenance contracts?
2. Do performance payments weighted 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.

3.1 Model 1

We express the cost of a maintenance project as

\[ C_{ij} = M_{ij}(X, e) + ca_{ij}(F), \]  

(1)

where \( i = \) contract and \( j = \) contractor. \( M_{ij}(X, e) \) is preventive maintenance and is a function of \( X_i \) which is a vector of variables for traffic and infrastructure characteristics such as track length and rail age. Preventive maintenance is also a function of the contractor’s effort level \( e = (e_1, e_2, ..., e_n) \), 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.\(^4\)

The number of failures is expressed as

\[ F_{ij} = \alpha_i + \beta X_i + \gamma_j M_{ij}(X, e) + \epsilon_{ij}, \]  

(2)

where \( \alpha_i \) is a constant and \( \beta \) is a vector of parameters for the effect of the explanatory variables \( X_i \). \( \epsilon_{ij} \) is an error term. \( \gamma_j \) 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, \( W_{ij}(X) \), 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 \( \Psi(e) \).\(^5\)

\(^4\) We assume that \( \frac{dca(F)}{dF} > 0, \frac{d^2ca(F)}{dF^2} < 0 \) and \( \frac{d^3ca(F)}{dF^3} \geq 0 \).

\(^5\) We assume \( \Psi(e) \) is convex in effort.
Preventive maintenance’s marginal effect on the cost of the project is decided by $\gamma_j$. 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.

In a fixed-price contract, the contractor’s profit from the project is

$$\pi_{ij} = W_{ij}(X) - [M_{ij}(X,e) + ca_{ij}(F)]$$  \hspace{1cm} (3)

However, as we described in section 2, there is a cost ceiling 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

$$c_{ijf} = \begin{cases} c_{c,i}, & \text{if } ca_{ijf} \geq c_{c,i} \\ ca_{ijf}, & \text{otherwise} \end{cases}$$  \hspace{1cm} (4)

where $f = 1, \ldots, F$ failure, $c_{c,i}$ is the cost ceiling, $ca_{ijf}$ is the accounted cost of rectifying a failure (see also Figure 1), and $\sum_{f=1}^{F} ca_{ijf} = ca_{ij}(F)$. The profit in the contract is then

$$\pi_{ij} = W_{ij}(X) - [M_{ij}(X,e) + \sum_{f=1}^{F} ca_{ijf}],$$  \hspace{1cm} (5)

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 (5) as

$$\pi_{ij} = W_{ij}(X) - [M_{ij}(X,e) + b_{ij} \cdot ca_{ij}(F)],$$  \hspace{1cm} (6)

which is an incentive contract where $b_{ij}$ is the power of the incentive scheme. From equation (4) it is clear that an increase in the cost ceiling $c_{c,i}$ would increase $c_{ijf}$ for a failure that costs more than the previous ceiling. The power of the incentive scheme will therefore increase with the cost ceiling: $\frac{db}{dc_c} > 0$.

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

$$\max_{W,e} U_j = \pi_{ij} - \Psi(e), \quad \text{subject to } \pi_{ij} - \Psi(e) \geq 0,$$  \hspace{1cm} (7)
The information about $c_{C,t}$ in the quote for bids is one point of departure for the bid submitted. A higher cost ceiling would increase the share of cost that is borne by the contractor per failure, and would therefore reduce its utility since $\frac{\partial U}{\partial c_C} = -\frac{\partial b}{\partial c_C} c_{aij}(F) < 0$. The consequence is that the higher cost ceiling corresponds to submitting a higher bid. However, since contractors have different efficiency levels, $\gamma_j$, the marginal profit (cost savings) of efficiency increases with the cost ceiling, $\frac{\partial^2 U}{\partial c_C \partial \gamma} = -\frac{\partial ((\partial b/\partial c_C) c_{aij}(F))}{\partial \gamma} > 0$, as the total number of failures is decreasing with the efficiency level $\gamma_j$ (which lowers the total cost of rectifying failures). We therefore make the following proposition:

**Proposition 1** – An increase in the cost ceiling increases the comparative advantage of the efficient contractor in the competitive tendering of a maintenance contract.

When a contract has been awarded, the contractor will choose an effort level that generates preventive maintenance. This decision will depend on the cost ceiling. 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, taking the cross partial derivative of the contractor’s utility with respect to the cost ceiling and the effort level (generating preventive maintenance) $\frac{\partial^2 U}{\partial c_C \partial e} = -\frac{\partial ((\partial b/\partial c_C) c_{aij}(F))}{\partial e} > 0$, we can see that the cost ceiling increases the marginal profit (cost savings) of preventive maintenance.

**Proposition 2** – The level of preventive maintenance increases with the cost ceiling.

### 3.1.1 Estimating the effect of the incentive schemes

We estimate the effect the cost ceiling has on the number of failures with the following model
\[ F_{it} = \alpha_i + \beta X_{it} + \mu C_{c, it} + \epsilon_{it}, \quad (8) \]

referred to as Model 1, where \( t = 1, \ldots, T(i) \) years. Our hypothesis is that the cost ceiling will reduce the number of failures:

Hypothesis 1: \( \mu < 0 \)

The effect of the cost ceiling can be due to a selection effect (Proposition 1) and/or because it reduces the moral hazard problem (Proposition 2). 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, as described in section 2.1, there is also a performance payment in place. 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:

\[ \pi_{ij} = W_{ij}(X) - [M_{ij}(X, e) + b_{ij} \cdot c_{aij}(F_T + F_O) + P_T \cdot F_{T,ij} + P_O \cdot F_{O,ij} + \varphi(F_O)^2]. \quad (9) \]

The term \( b_{ij} \cdot c_{aij}(F_T + F_O) \) is the same as in equation (6), except that a distinction is now made between failures causing train delays

\[ F_{T,ij} = \alpha_{F_T,ij} + \beta_{F_T} X_i + \gamma_{T,i} M_{T,ij}(X, e_T) + \epsilon_{ij}, \quad (10) \]

and other failures

\[ F_{O,ij} = \alpha_{F_O,ij} + \beta_{F_O} X_i + \gamma_{O,i} M_{O,ij}(X, e_O) + \epsilon_{ij}, \quad (11) \]

In addition, \( P_T \) and \( P_O \) are the performance penalties incurred by the contractor for \( F_{T,ij} \) and \( F_{O,ij} \) respectively. The weighted incentive scheme implies that \( P_T > P_O \), 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 ($\varphi$) on profit.\textsuperscript{6}

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.\textsuperscript{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 cost of the project

$$\frac{\partial C}{\partial e_T} = P_T \cdot \frac{\partial F_T}{\partial e_T} < P_O \cdot \frac{\partial F_O}{\partial e_O} + \varphi \frac{\partial (F_{O,ij})^2}{\partial e_O} = \frac{\partial C}{\partial e_O},$$

for $F_{O,ij} \in [0, Z]$, \hspace{1cm} (12)

(note that $\frac{\partial F}{\partial e} < 0$, and $\frac{\partial (F_{O,ij})^2}{\partial e_O} < 0$).\textsuperscript{8} Hence, a marginal increase in $e_T$ will be more profitable than a corresponding increase in $e_O$. However, 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.

\textsuperscript{6} For simplicity, we consider this relationship to be linear for train delay failures.

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

\textsuperscript{8} 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 payments

There is no point in time where performance payments was introduced, meaning that we may not perform a before-after analysis of its effects. 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. However, we can use the results from optimal auction, stating that the most efficient firm will be awarded the contract (see Laffont and Tirole 1987), which is also suggested by proposition 1 and the results in Odolinski and Smith (2014), showing that competitive tendering reduced maintenance cost in Sweden with about 11 per cent.

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 $\gamma_j$ (the number of failures decreases with efficiency; see equations 10 and 11). Hence, given that tendering has resulted in more efficient contractors being awarded the maintenance projects, the introduction of competitive tendering can be used to test the effect of performance payments. Our hypothesis is therefore that the use of competitive tendering will increase the effect of the tilted performance payments. We test our hypothesis by estimating two models. Model 2a is related to train delays

$$F_{T, it} = \alpha_{T,i} + \beta X_{it} + \vartheta_T D_{C, it} + \epsilon_{it}$$

\[ (13) \]

$D_{C, it}$ 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 payments. The other model is

$$F_{O, it} = \alpha_{O,i} + \beta X_{it} + \vartheta_O D_{C, it} + \epsilon_{it},$$

\[ (14) \]
referred to as Model 2b, where failures not causing train delays is the dependent variable. According to (12), the parameters $\vartheta_T$ and $\vartheta_O$ should differ between Model 2a and 2b. Hence, we state the following hypothesis:

$$Hypothesis 2: \vartheta_T \neq \vartheta_O$$

### 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 (2014) 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:

$$z_{ikt} = [D_{ik}, D_{ikt}, D_{iM}, D_Y; \beta_{it}, \vartheta_{IF,ikt}],$$

(15)

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 ($k = Y = 2004, ..., 2013$). $\beta_{it}$ and $\vartheta_{IF,ikt}$ are vectors of
parameters to be estimated, where \( IF = T \) for Model 2a and \( IF = O \) for Model 2b. We address the issue of selection bias by testing the following hypothesis:

\[ Hypothesis \ 3: \theta_{IF,O} = 0 \]

If we cannot reject Hypothesis 3, we conclude 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.\(^9\)

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 presented in sections 3.1 and 3.2. 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

\(^9\) This definition is arbitrary because the exposure to competition was gradual, and we therefore perform sensitivity tests with respect to this definition.
certain time interval is assumed to be Poisson distributed, where \( i = \text{individual 1, 2 \ldots N} \). The events can be explained by \( k \) variables \( x_{ik} \). In a Poisson distribution, the conditional mean is 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 value of zero, which is indicated by Figure 2 below. 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 \((\text{chi2}(1) = 3.8e + 05)\). 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 as $\text{VAR}[y_i \mid x_{ik}] = \lambda_i + \delta_i \lambda_i^2$. This model is referred to as the NB2 model in the literature (see Cameron and Trivedi 2013), where $\delta_i$ is the dispersion parameter to be estimated.\(^{10}\)

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

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

(16)

where $\alpha_i$ is the individual specific effect. If the regressors are correlated with $\alpha_i$ we could use a fixed effects model, resulting in consistent estimates of $\beta$. If $\alpha_i$ is independent of the regressors we would prefer a random effects model resulting in efficient and consistent estimates of $\beta$. In the fixed effects negative binomial model, the dispersion is constant $(1 + \delta_i)$ for each individual, while the random effects model assumes that $1/(1 + \delta_i) \sim \text{Beta}(r, s)$, which means that $\delta_i$ can vary randomly between individuals (see for example Hilbe 2011 for a specification of the negative binomial models with fixed and random effects). 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.\(^{11}\) An unconditional fixed effects negative binomial model can be used instead, which controls for the individual unobserved effects by including individual-specific dummy variables. However, with a large panel (many individuals) this can 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.

\(^{10}\) The conditional variance in the NB1 model is $\text{VAR}[y_i \mid x_{ik}] = \lambda_i + \delta_i \lambda_i^2$

\(^{11}\) Guimarães (2008) showed that the fixed effects are conditioned out only if the fixed effect equals the logarithm of the dispersion parameter, i.e. $\alpha_i = \ln(\delta_i)$
Still, we will have inconsistent estimates if the regressors are not independent of the individual effects $\alpha_i$. 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, a model that gives us the within estimates using the random effects estimator. This model 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

$$\alpha_i = \exp(\bar{x}_{ik}^t y_k + \varepsilon_{it}),$$  \hspace{1cm} (17)

where $\bar{x}_{ik} = T^{-1} \sum_{t=1}^T x_{ikt}$ for each $k = 1 ... K$. Using (17) we can then express (16) as

$$\exp(x_{ikt}^t \beta_k + \bar{x}_{ik}^t y_k + \varepsilon_{it}),$$ \hspace{1cm} (18)

where we are controlling for the correlation between $\alpha_i$ and our regressors $x_{ikt}^t$ via $\bar{x}_{ik}^t$. Hence, we assume that the remaining individual effect in $\alpha_i$ is independent of $x_{ikt}^t$.

Furthermore, (18) assumes that the relationship between $\alpha_i$ and $\bar{x}_{ik}^t$ is linear. We can also test interactions and quadratic terms between different $\bar{x}_{ik}^t$ (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. Furthermore, we can avoid collinearity between $x_{ikt}^t$ and $\bar{x}_{ik}^t$ by using deviations from the mean

$$\exp[(x_{ikt}^t - \bar{x}_{ik}^t) \beta_k + \bar{x}_{ik}^t y_k + \varepsilon_{it}],$$  \hspace{1cm} (19)

where $T^{-1} \sum_{t=1}^T (x_{ikt}^t - \bar{x}_{ik}^t) = 0$ 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:

$$\Pr(F_{it} = f_{it} \mid x_{ikt}, z_{ikt}, \alpha_i, \delta_i),$$  \hspace{1cm} (20)

where $F$ is the count of failures, $i =$ track segment $1,2,...,N$ and $t =$ year $1,2,...,T(i)$. $x_{ikt}$ 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. cost ceiling, 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.

\( z_{iikt} \) 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 (15) 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, \( \alpha_i \) is the individual effect as specified in (17) and \( \delta_i \) 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 a week. 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 include 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. Hence, 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. We were, however, 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.\textsuperscript{12}

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 ceiling 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

\textsuperscript{12} 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.
contract area varies; Figure 3 is only an illustration of the fact that \( segment\ km < section\ km < contract\ area\ km \).

![Diagram showing Track segment, Track section, and Contract area]

**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 27,501 observations on segments administered by the IM over the period 2003-2013. Only tendered contracts have a cost ceiling. In total, this information is available for a third of the observations. In the comparison of cost ceilings, 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**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEG.: Failures, total</td>
<td>27,501</td>
<td>6.45</td>
<td>20.52</td>
<td>0</td>
<td>522</td>
</tr>
<tr>
<td>SEG.: Failures, train delays</td>
<td>27,501</td>
<td>1.05</td>
<td>2.90</td>
<td>0</td>
<td>101</td>
</tr>
<tr>
<td>SEG.: Failures, not train delays</td>
<td>27,501</td>
<td>5.39</td>
<td>18.27</td>
<td>0</td>
<td>483</td>
</tr>
<tr>
<td>SEG.: Track length, metres</td>
<td>27,128</td>
<td>5,890</td>
<td>5,980</td>
<td>10</td>
<td>60,061</td>
</tr>
<tr>
<td>SEG.: Rail weight, kg</td>
<td>27,468</td>
<td>51.13</td>
<td>6.00</td>
<td>27</td>
<td>63</td>
</tr>
<tr>
<td>SEG.: Quality class, 0-5*</td>
<td>26,186</td>
<td>2.14</td>
<td>1.31</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>SEC.: Traffic density, million gross tonnes**</td>
<td>25,870</td>
<td>8.18</td>
<td>8.58</td>
<td>0.00</td>
<td>49.79</td>
</tr>
<tr>
<td>SEC.: Tendered, dummy</td>
<td>27,501</td>
<td>0.58</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SEC.: Transition to tend., dummy</td>
<td>27,501</td>
<td>0.08</td>
<td>0.28</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
### Subset of data used in Model 1

<table>
<thead>
<tr>
<th>ARE.: Cost ceiling, thousand SEK</th>
<th>SEG.: Failures, total</th>
<th>SEG.: Failures, train delays</th>
<th>SEG.: Failures, not train delays</th>
<th>SEG.: Rail weight, kg</th>
<th>SEG.: Quality class, 0-5*</th>
<th>SEC.: Traffic density, million gross tonnes**</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 940</td>
<td>8 940</td>
<td>8 940</td>
<td>8 940</td>
<td>8 940</td>
<td>8 940</td>
<td>8 704</td>
</tr>
<tr>
<td>7.69</td>
<td>5.82</td>
<td>1.09</td>
<td>4.73</td>
<td>51.54</td>
<td>2.07</td>
<td>7.84</td>
</tr>
<tr>
<td>4.05</td>
<td>15.16</td>
<td>2.83</td>
<td>12.96</td>
<td>5.91</td>
<td>1.35</td>
<td>9.27</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>20</td>
<td>482</td>
<td>101</td>
<td>381</td>
<td>60</td>
<td>5</td>
<td>49.79</td>
</tr>
</tbody>
</table>

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.

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

We estimate three models. *Model 1* considers the effect of different cost ceilings on the number of failures (see section 3.1), and *Models 2a* and *2b* the effect of incentive schemes with weights put on train delays (see section 3.2).

We estimate \( \hat{\lambda}(x)_{it} \) and \( \hat{\lambda}(x + 1)_{it} \), 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 $\hat{\beta} = \ln \left[ \frac{\hat{\lambda}(x+1)_{it}}{\hat{\lambda}(x)_{it}} \right]$ which is referred to as a semi-elasticity. $e^{\hat{\beta}}$ is an incidence ratio (IRR) which is 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.11).

5.1 Econometric results, Model 1

The results from the estimation of the first model are presented in Table 2, which include results from both the random effects model and the preferred correlated random effects model (using the approach by Mundlak 1978). In the latter model, the coefficients for variables averaged over time are denoted ‘between estimates’ while the other coefficients are denoted as ‘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, which is the exposure variable, has the expected IRR of 1.0000 and is significant. The estimation 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 (cost ceiling), 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, COST CEILING, 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 cost ceiling gives an incentive to increase preventive maintenance (proposition 2), which is a contract the efficient type is more likely to be awarded (proposition 1). The incidence rate ratio at 0.9610 indicates that an increase in the cost ceiling with one unit (in our case with 1000 SEK) will reduce the number of failures with 3.9 percent (100*(1-0.961)). We made sensitivity tests 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 ceiling.

Table 2 - Results Model 1

<table>
<thead>
<tr>
<th></th>
<th>Random effects</th>
<th></th>
<th>Correlated Random effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>Std. Err.</td>
<td>IRR</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>Constant</td>
<td>5.4600***</td>
<td>1.5638</td>
<td>13.6631***</td>
<td>4.6980</td>
</tr>
<tr>
<td>COST CEILING</td>
<td>0.9987</td>
<td>0.0051</td>
<td>0.9610***</td>
<td>0.0095</td>
</tr>
<tr>
<td>RAIL_WEIGHT</td>
<td>0.9789***</td>
<td>0.0049</td>
<td>0.9928</td>
<td>0.0079</td>
</tr>
<tr>
<td>QUALAVE</td>
<td>1.0151</td>
<td>0.0225</td>
<td>0.9688</td>
<td>0.0401</td>
</tr>
<tr>
<td>TRACK_M</td>
<td>1.0000***</td>
<td>0.0000</td>
<td>1.0000**</td>
<td>0.0000</td>
</tr>
<tr>
<td>MTGTDEN</td>
<td>1.1014***</td>
<td>0.0077</td>
<td>1.0385***</td>
<td>0.0123</td>
</tr>
<tr>
<td>MTGTDEN2</td>
<td>0.9983***</td>
<td>0.0002</td>
<td>0.9993**</td>
<td>0.0003</td>
</tr>
<tr>
<td>D.2005</td>
<td>1.5680***</td>
<td>0.1830</td>
<td>1.6390***</td>
<td>0.1953</td>
</tr>
<tr>
<td>D.2006</td>
<td>1.4461***</td>
<td>0.1655</td>
<td>1.5513***</td>
<td>0.1810</td>
</tr>
<tr>
<td>D.2007</td>
<td>1.7700***</td>
<td>0.2007</td>
<td>1.9625***</td>
<td>0.2265</td>
</tr>
<tr>
<td>Year</td>
<td>COST CEILING</td>
<td>RAIL_WEIGHT</td>
<td>QUALAVE</td>
<td>TRACK_M</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>2008</td>
<td>1.6242***</td>
<td>0.1840</td>
<td>1.8119***</td>
<td>0.2091</td>
</tr>
<tr>
<td>2009</td>
<td>1.5133***</td>
<td>0.1713</td>
<td>1.7102***</td>
<td>0.1971</td>
</tr>
<tr>
<td>2010</td>
<td>1.4245***</td>
<td>0.1611</td>
<td>1.6221***</td>
<td>0.1870</td>
</tr>
<tr>
<td>2011</td>
<td>1.4504***</td>
<td>0.1642</td>
<td>1.6294***</td>
<td>0.1879</td>
</tr>
<tr>
<td>2012</td>
<td>1.4578***</td>
<td>0.1653</td>
<td>1.6462***</td>
<td>0.1906</td>
</tr>
<tr>
<td>2013</td>
<td>1.5515***</td>
<td>0.1757</td>
<td>1.7608***</td>
<td>0.2045</td>
</tr>
</tbody>
</table>

COST CEILINGbar = COST CEILING - \sum_{t=1}^{T} COST CEILING
RAIL_WEIGHTbar = RAIL_WEIGHT - \sum_{t=1}^{T} RAIL_WEIGHT
QUALAVEbar = QUALAVE - \sum_{t=1}^{T} QUALAVE
TRACK_Mbar = TRACK_M - \sum_{t=1}^{T} TRACK_M
MTGTDENbar = MTGTDEN - \sum_{t=1}^{T} MTGTDEN
MTGTDEN2bar = MTGTDEN^2 - \sum_{t=1}^{T} MTGTDEN^2

***, **, *: Significance at 1%, 5%, 10% level
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:
COST CEILING\(^a\) = cost ceiling 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 speed line
TRACK_M\(^a\) = Track length, metres
MTGTDEN\(^a\) = Million gross tonnage density (gross tonnes-km/Track km)
MTGTDEN2\(^a\) = MTGTDEN\(^a\)^2
D.2005-D.2013\(^a\) = Year dummy variables, 2005-2013

\(x_{ikt}^\prime - \sum_{t=1}^{T} x_{ikt}\) (see section 3.3)
5.2 Econometric results, Model 2

The estimation results from Models 2a and 2b are presented in Table 3. 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 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. This may be due to the fact that these segments have a lower traffic density (train-km per route-km), with a correlation coefficient at -0.4874, which is not captured by the tonne density variable.

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 0.9428 (p-value=0.067) which implies a negative effect on the number of failures causing train delays. The IRR for competitive tendering in Model 2b is 1.0608 (p-value=0.001) which implies that the number of failures that has not caused a train delay is increasing when tendered in competition. The opposite effects in Model 2a and 2b are in line with our hypothesis in section 3.2 (we cannot reject Hypothesis 2). An incentive scheme weighted against train delays will affect the contractor’s maintenance strategy and performance.
Note that we include a dummy variable indicating all segments tendered sometime during 2003-2013 to control for any general feature among these segments (that also apply before tendering). The corresponding IRR (DTEND) is above one, but not statistically significant. Moreover, no selection bias was found using the dummy variable approach described in section 3.2.2, that is, we cannot reject Hypothesis 3: \( \theta_{IF,IFO} = 0 \) (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 models, so this variable is excluded from the estimations shown in Table 3.

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

<table>
<thead>
<tr>
<th></th>
<th>Model 2a - Train delay failures</th>
<th>Model 2b - Other failures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>Constant</td>
<td>1.8884</td>
<td>1.0044</td>
</tr>
<tr>
<td>MIXTEND</td>
<td>1.0012</td>
<td>0.0333</td>
</tr>
<tr>
<td>CTEND</td>
<td>0.9428*</td>
<td>0.0303</td>
</tr>
<tr>
<td>DTEND</td>
<td>1.0986</td>
<td>0.1313</td>
</tr>
<tr>
<td>RAIL_WEIGHT</td>
<td>0.9643***</td>
<td>0.0054</td>
</tr>
<tr>
<td>QUALAVE</td>
<td>1.0850**</td>
<td>0.0402</td>
</tr>
<tr>
<td>TRACK_M</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>MTGTDEN</td>
<td>1.0584***</td>
<td>0.0120</td>
</tr>
<tr>
<td>MTGTDEN2</td>
<td>0.9993***</td>
<td>0.0002</td>
</tr>
<tr>
<td>D.2004</td>
<td>0.9018***</td>
<td>0.0355</td>
</tr>
<tr>
<td>D.2005</td>
<td>0.9394</td>
<td>0.0371</td>
</tr>
<tr>
<td>D.2006</td>
<td>0.9487</td>
<td>0.0384</td>
</tr>
<tr>
<td>D.2007</td>
<td>1.1461***</td>
<td>0.0458</td>
</tr>
<tr>
<td>D.2008</td>
<td>1.0625</td>
<td>0.0449</td>
</tr>
<tr>
<td>D.2009</td>
<td>1.0004</td>
<td>0.0440</td>
</tr>
<tr>
<td>D.2010</td>
<td>1.2623***</td>
<td>0.0559</td>
</tr>
<tr>
<td>D.2011</td>
<td>1.3419***</td>
<td>0.0613</td>
</tr>
<tr>
<td>D.2012</td>
<td>1.2426***</td>
<td>0.0586</td>
</tr>
<tr>
<td>D.2013</td>
<td>1.4100***</td>
<td>0.0661</td>
</tr>
<tr>
<td>MIXTENDbar</td>
<td>1.6542</td>
<td>0.9225</td>
</tr>
<tr>
<td>CTENDbar</td>
<td>0.8590</td>
<td>0.0867</td>
</tr>
</tbody>
</table>
RAIL_WEIGHTbar 0.9650*** 0.0058 0.9723*** 0.0047
QUALAVEbar 0.9211*** 0.0213 1.0133 0.0191
TRACK_Mbar 1.0000*** 0.0000 1.0000*** 0.0000
MTGTDENbar 1.2108*** 0.0093 1.1210*** 0.0070
MTGTDEN2bar 0.9962*** 0.0002 0.9976*** 0.0002

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

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 = -27 633.284, Model 2b = -51 402.596

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 = $T^{-1} \sum_{t=1}^{T} MIXTEND$

CTENDbar = $T^{-1} \sum_{t=1}^{T} CTEND$

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 cost ceilings have been used over the years of competitive tendering creating different incentive intensities. The estimation results show that an increase in the cost ceiling reduces the number of infrastructure failures.

Does this result imply that we should have a high cost ceiling in all maintenance contracts? Not necessarily. A high cost ceiling 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 due to information asymmetry. Moreover, we will have a low level of competition if inefficient types do not take part in the bidding when cost ceilings are too high, with the efficient type(s) being able 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 payments weighted against train delays, which makes it beneficial for the contractors to focus on this class of infrastructure failures. The estimation results confirm our hypothesis, showing that the performance payments affect the outcome, 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), showing that increased incentives for one task can result in a reallocation of attention from other tasks.

Has the performance payments been 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 decrease in the number of train delay failures does not _per se_ imply that the number of train delay minutes has reduced. Nevertheless, it is fair to say that a reduction in the number of train delay failures is a good sign of improved performance. 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?
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 payments. More precisely, we have shown that a higher cost ceiling increases the comparative advantage of the efficient contractor in the bidding for contracts and that it induces an increase in effort, generating a higher level of preventive maintenance. The econometric results show that the cost ceiling 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 ceiling - is a 3.9 per cent reduction of the number of failures.

The econometric test of the tilted performance payments 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 has been beneficial with respect to the number of train delay failures, while other failures have increased.

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 the cost ceiling too low can be costly for the IM with respect to the number of failures that occur, while a high cost ceiling can induce rent extraction. Moreover, when using tilted performance payments, 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 strengthen 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 2014), as measured by the number of train delay failures. An important caveat is, however, that it has not been possible to measure the effect on train delay minutes.

The effect of different designs on cost efficiency and performance with respect to train delay minutes is an area for future research. Such considerations are critical in the study of optimal contract design within this field.

References


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


**Appendix**

Performance equation in maintenance contracts:

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

(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 $I > 1$. 