Benchmarking in a state-owned monopoly: estimating the
dual-level cost efficiency of railway maintenance in Sweden

Kristofer Odolinski

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

Draft version: 2015-08-08

The author is grateful to the Swedish Competition Authority for funding this research, and
also wants to acknowledge the support and advice offered by Professor Jan-Eric Nilsson.

Abstract

The cost efficiency of railways maintenance regions and contracts in Sweden is estimated. We
use a dual-level inefficiency model on a panel data set stretching from 1999-2013. The model
allows us to decompose the efficiency of railway maintenance in two parts, showing which
contract that is most cost efficient within a region and the relative cost efficiency between
regions. The results indicate considerable cost reduction possibilities if the cost inefficient
regions replicated the most efficient region, and that scale economies are not fully exploited.
Nonetheless, competitive tendering of maintenance has been beneficial according to the
estimation results, showing a further cost reduction after the first period of tendering.
1.0 Introduction

The organization of the rail sector in Europe has been subject to reforms during the last decades, initiated by the European Commission’s Directive (Dir. 91/440) in 1991 with a vertical separation of train operations and infrastructure management. Nevertheless, the infrastructure managers (IMs) have been run as state-owned monopolies, often with demands on higher efficiency being put forward, leading to changes of the organization. This was the case in Sweden where the IM was restructured in 1998 by separating the production unit from the administrative unit, with the argument that the roles within the Swedish Rail Administration (Banverket) should be refined in order to achieve a more efficient management (Prop. 1997/98:56, p.103). The reform created a client-contractor relationship between the production unit and the administrative units, which paved the way for a decision in 2001 to gradually expose the maintenance of railways to competition.

The objective of this paper is to estimate the cost efficiency of railway maintenance regions and contracts in Sweden. This analysis is made in the context of a national organization with relatively independent regional units, each administering several maintenance contract areas. This implies that there may be variability in the performance between and within these regional units, and this notion is further supported by the organizational and regulatory changes during the last two decades.

Up until 2007, Banverket was a decentralized organization with five regional units responsible for the infrastructure management within their region and they were directly accountable to the director-general. Each region was instructed to have competence to define, order and follow up the maintenance from the production unit (Banverket 2000). The responsibilities of the regional units were centralized in 2007, with the purpose of coordinating the working procedures and have a more consistent approach between the
regions (Riksrevisionen 2010).\textsuperscript{1} The regional units are still responsible for managing and following up the maintenance contracts, with a central unit in charge of the planning procedures. Indeed, there is variability in contract design (further described in section 4), indicating that the organizational structure is not completely centralized. In essence, the differences in contract design and the changes in the management of regions over time are likely to create performance differences within and between regions.

Consequently, it is important to account for the organizational structure when estimating the cost efficiency in order to find the sources of efficiency gaps and being able to make improvements. For example, the IM can benefit from knowing the variation in performance within a region, if there are differences between regions and, if that is the case, investigate if these differences have changed over time. Answers to these questions can lead to different actions for improvements: knowledge of which contract that is most cost efficient within a region is vital for implementing internal best practice, while information on the relative cost efficiency between regional units may require reforms of the management of these units.

In this paper we estimate a dual-level inefficiency model proposed by Smith and Wheat (2012a) which decomposes the inefficiency into one part for the regions and another for the contracts managed by the regions. Using an unbalanced panel over the period 1999-2013, we are able to track costs before and after competitive tendering of maintenance (introduced gradually in 2002) and costs in complete contract periods which is normally five to seven years. Odolinski and Smith (2014) showed that competitive tendering reduced maintenance costs with about 11 per cent in Sweden, using data over 1999-2011. In this paper, we revisit this question, now testing if there is a difference between the first tendering period for a contract area and subsequent periods of competitive tendering. One reason for addressing this

\textsuperscript{1} Another change occurred in 2010 when Banverket was merged with the Swedish Road administration (Vågverket) to form the Swedish Transport Administration (Trafikverket). However, this had minor implications for the administration of the contracts.
issue is the negative experience in Britain where privatization and sub-contracting led to an initial decrease in costs, but then increased due to problems with track quality (Kennedy and Smith 2004). This implies that costs in subsequent periods of tendering are expected to rise in Sweden if cost reductions in the first period were made at the expense of quality.\(^2\) Another aspect is that it is difficult to find more ways of reducing costs significantly after the cost reductions in the first period, making the cost level in subsequent periods similar to the first period. A third effect could be that costs are further reduced in subsequent periods because the contractors have more information about the infrastructure and are therefore able to place lower bids, and/or because the client has learned from mistakes such as distorted incentive structures in the contracts. An increase in the level of competition over the years can also affect costs. However, this effect would not be the main explanation for a difference between contract periods in our estimation because the exposure to competition is gradual with some areas being tendered for the first time the same year as other areas are tendered a second time.

There is a wide literature studying the efficiency and productivity of railway systems with respect to train operations. Caves and Christensen (1980) is an early example, making a comparison of public and private firms in the Canadian railroads industry, while Coelli and Perelman (1999 and 2000), Sánchez and Villarroya (2000), and Cantos and Maudos (2001) measure the efficiency of European railways. Within country comparisons are made by Farsi et al. (2006) who analyses the efficiency of railway companies in Switzerland, and Smith and Wheat (2012b) studies different rail operation contracts in Britain. Corresponding contributions on railway infrastructure are scarcer. Smith (2006) makes total factor productivity analysis of British Rail and Smith (2012) compare the cost efficiency of different infrastructure managers in European countries, while Smith and Wheat (2012a) use data on

\(^2\) We also note that the winner's curse (see for example Kagel and Levin 1986) could result in a low cost level in the first period, with an increase in the following periods, which is explained by bids being too optimistic in the first period.
North America and Europe. The maintenance performance of Swedish railways has been studied by Espling and Kumar (2008) and Åhrén and Parida (2009), using a range of performance indicators for benchmarking. However, these previous studies on Swedish railway do not include an estimation of the cost efficiency. Furthermore, cost efficiency studies on rail infrastructure maintenance have mainly used national data, except for the study by Kennedy and Smith (2004) who apply different estimation techniques (Corrected Ordinary Least Squares and Stochastic Frontier Analysis) to make comparisons of seven regional zones in Britain’s rail network. Still, that study does not compare the cost efficiency of the approximately twenty contract areas within the zones.

With access to a panel data set of regions and maintenance contracts stretching over 15 years, our paper provides a unique internal (within country) benchmarking study of rail infrastructure maintenance. Our data allows us to estimate a cost model in accordance with the organization of railway maintenance in Sweden, which is important when estimating the economies of scale. Policy implications from this paper are the possibility to make cost efficiency improvements, via for example internal benchmarking or yardstick competition (see Shleifer 1985), and with respect to the economies of scale estimate, reconsider the size of the contract areas. Moreover, the effect on costs after the first contract period of competitive tendering has a regulatory relevance, as this might call for different actions depending on the sign of the effect.

Experiences from competitive tendering of passenger train operations has been extensively studied (see for example Alexandersson, 2009; Brenck and Peter, 2007; Kain, 2009), and the impact on costs over time is analyzed in for example Smith and Wheat (2012b). To the author’s knowledge, Odolinski and Smith (2014) is the only formal study of the cost impact of competitive tendering in rail maintenance. However, it does not consider the development of this effect over time. This effect is particularly important in sectors that
procure maintenance on structures that has a long service life. This paper adds to the literature by estimating the effect of subsequent rounds of tendering in rail maintenance, showing whether the relatively short-term maintenance contracts have been beneficial or not.

This paper is organized as follows: section 2 sets out the methodology we use. The model we estimate is presented in section 3 together with a specification of the hypothesis tests. In section 4 we describe the data and its structure. Estimation results are presented in section 5. We discuss our results and the potential sources of cost efficiency differences in section 6. Section 7 concludes.

### 2.0 Methodology

A common measure to describe the performance of a firm is productivity: output divided by input. The maximum output produced with a given set of inputs – or the minimum inputs used for a given set of output - can be represented by a production frontier. This frontier is useful for comparing the relative efficiency of firms. A firm is technically efficient if it produces outputs on the production frontier while firms producing outputs beneath the frontier are technically inefficient. Economic efficiency is a wider measure as it includes both technical efficiency and allocative efficiency; the latter being a firm’s ability to use the right share of inputs depending on the input prices and production technology (Coelli et al. 2005).

Estimating a cost function with econometric methods has the advantage of allowing economies of scale and returns to density to be accounted for (in our case, cost elasticities with respect to network size and traffic respectively), and hence also separating the effect of scale and density on performance from the inefficiency measure. For example, economies of scale and returns to density are assumed to be constant in a total factor productivity (TFP) measure. See Nash and Smith (2014) for a further discussion on this topic in the area of railway operations and infrastructure management.
An often used frontier method is the stochastic frontier introduced by Aigner et al. (1977) and Meeusen and Van den Broeck (1977):

\[ \ln C_i = \alpha + \sum_n \beta_n \ln x_{ni} + v_i + s\mu_i, \] (1)

where \( i \) = decision making unit, \( C_i \) is costs, \( x_{ni} \) is the vector of \( n \) inputs used, \( \alpha \) and \( \beta \) are parameters to be estimated. \( v_i \) and \( \mu_i \) are the error components and \( s = 1 \) (in a production function \( s \) takes the value -1). \( v_i \) is the noise component assumed to be independently identically normally distributed (\( v_i \sim \text{iid } N(0, \sigma^2_v) \)). \( \mu_i \) is a non-negative random term that is supposed to capture the inefficiency.

Estimating (1) with ordinary least squares (OLS) would result in a biased intercept coefficient \( \alpha \) if \( sE(\mu_i) \geq 0 \). Thus, we need to separate noise from inefficiency which requires distributional assumptions on \( \mu_i \), for example a half normal distribution (\( \mu_i \sim \text{iid } N^+(0, \sigma^2_\mu) \)).3 Moreover, both \( v_i \) and \( \mu_i \) are assumed to be independently distributed from each other and independent from the regressors. Armed with these distributional assumptions it is feasible to estimate the inefficiency using maximum likelihood and thus get a consistent estimate of the intercept.

With access to panel data it is possible to relax some of the distributional assumptions of the inefficiency term and still get a consistent estimate of a firm’s efficiency, using a fixed- or random effects approach. For example, we can drop the assumption that inefficiency is independently distributed over time, which is an unreasonable assumption when estimating the inefficiency of a maintenance contract; the efficiency during one year is likely to affect the efficiency level the next year within the same contract. In fact, as described in section 3.0 below, we consider the inefficiency to be constant during a contract period. Moreover, with panel data we only need to assume the inefficiencies to be non-negative. The fixed- and random effects approaches are further described in section 2.2.

---

3 Other distributions often considered are the truncated normal, exponential, and gamma distributions.
2.1 Dual-level inefficiency model


\[ \ln C_{its} = \alpha_t + f(X_{its}; \beta) + \tau_{its} + \nu_{its}, \]

where \( i = 1, \ldots, N \) firms, \( t = 1, \ldots, T(i) \) years and \( s = 1, \ldots, S(i) \) sub-companies. \( X_{its} \) is a vector of explanatory variables and \( \beta \) is a vector of parameters to be estimated. The firm specific term \( \alpha_{it} = \alpha + \mu_{it} \) comprises a constant \( \alpha \) and an inefficiency term \( \mu_{it} \) that is constant across all sub-companies that the firm consists of. \( \tau_{its} \) is the residual inefficiency which varies across the sub-companies within the same firm. \( \nu_{its} \) is white noise.

2.2 Unobserved heterogeneity

Firms operate in different environments and have different production possibilities. For example, the infrastructure characteristics and climate differ across contract areas. These are exogenous factors to each firm, and their effect on the cost frontier is important to include in an efficiency analysis. Having sought all feasible possibilities to control for heterogeneity in the production environment, one is left with unobserved heterogeneity and whether it should be treated as random or fixed effects.

First, we note that with the fixed-effects approach we can drop the assumption of \( \hat{\mu}_i \) (estimated as \( \hat{\mu}_i = \hat{\alpha}_i - \min_i \{\hat{\alpha}_i\} \) using firm-specific intercepts) being independent from the regressors and the noise term. However, time-invariant effects that are different between firms will be captured by the inefficiency term with the fixed-effects approach. This is not the case if a random-effects model is used, where inefficiency is estimated as \( \hat{\mu}_i = \bar{e}_i - \min_i \{\bar{e}_i\} \), where \( \bar{e}_i = T(i)^{-1} \sum_{t=1}^{T(i)} \bar{\varepsilon}_{it} \) (following Schmidt and Sickles 1984). However, this requires that
\( \hat{\mu}_t \) is uncorrelated with the regressors and the noise term in order to get consistent estimates. To avoid the heterogeneity bias caused by such correlation in the random effects model, we can use the formulation specified by Mundlak (1978). That is, we include group means of the explanatory variables in the model, \( \bar{X}_{its} = T^{-1}_t \sum_{t=1}^{T} X_{its} \), to capture the time-invariant effects correlated with \( X_{its} \) that should not be confounded with cost efficiency.

**3.0 Model**

A stochastic cost frontier model is estimated:

\[
\ln C_{its} = \alpha + f(Q_{its}, W_{its}, P_{its}, X_{its}, Z_{jita}, R_i; \beta, \mu_i) + \tau_{its} + v_{its},
\]

where \( i = 1, 2, \ldots, N \) regions and \( t = 1, 2, \ldots, T(i) \) years, \( a = 1, \ldots, A(i) \) contract areas, and \( s = 1, \ldots, S(i) \) contracts (see Figure 1 below for an illustration of the data structure). The distinction between \( a \) and \( s \) is that a contract area is re-tendered over time, generating several contracts within the area during the relevant period. \( Q_{its} \) is a vector of variables for traffic. The inputs price variables \( W_{its} \) and \( P_{its} \) are proxies for wages and price of materials. \( X_{its} \) is a vector of network characteristics and quality, \( Z_{jita} \) is a vector of dummy variables which includes policy variables and year dummies, where the subscript \( j \) indicates the type of contract area with respect to tendering (see equation 8 in section 3.1 below). \( R_i \) are dummy variables for regions. \( \alpha \) is a constant and \( \beta \) and \( \mu_i \) are parameters to be estimated, where the latter are parameters for the region dummies. \( v_{its} \) is white noise. \( \mu_i \) is used to obtain region-specific inefficiency, while \( \tau_{its} \) is the inefficiency of contracts, i.e. within region inefficiency. The dependent variable \( \ln C_{its} \) is the logarithm of maintenance costs for contract \( s \) in region \( i \) in year \( t \).

The inefficiency of a contract (\( \tau_{its} \)) is considered to be time-invariant during the length of each assignment because it is the design of the contract together with the winning bid in the procurement that affect inefficiency. Any change in technical or allocative inefficiency during
a contract period would mainly affect the contractor’s rent extraction and not appear in the final cost of the contract for the IM.\footnote{For clarity reasons, we note that the contractors are firms performing the maintenance which is procured by the IM who manages the contracts via regional units.}

In our base model (\textit{Model 1}) we consider the inefficiency term for the regions to be time-invariant ($\mu_i$). Hence, the relative inefficiency between regions is constant, while the change in inefficiency over time within a region is captured by the inefficiency term $\tau_{it}$ when changing from one contract to another. We also estimate a model (\textit{Model 2}) using the specification proposed by Cornwell et al. (1990), allowing region-specific inefficiency to vary over time:

$$
\mu_{it} = \mu_{i1} + \mu_{i2}t + \mu_{i3}t^2, 
$$

where $t = 1, \ldots, T(i)$ years.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{data_structure.png}
\caption{Illustration of the data structure}
\end{figure}
We furthermore include group means of the explanatory variables in Model 3 (Mundlak 1978 formulation) in order to disentangle unobserved heterogeneity from the contract-specific inefficiency.

We use a random effects model and estimate the region-specific inefficiency using region dummies. The time-varying cost inefficiency for regions in (4) is estimated by including region dummies in (3) together with interactions between these region dummies and the time trend variables $t$ and $t^2$. We then compute $\hat{\mu}_{it}$ using the parameter estimates for $\mu_{i1}$, $\mu_{i2}$ and $\mu_{i3}$, and specify the cost frontier as $\hat{\mu}_{it}^* = \min_t \{\hat{\mu}_{it}\}$. The inefficiency of each region is then calculated as

$$\hat{\mu}_{it} = \mu_{it} - \min_t \{\hat{\mu}_{it}\},$$

With the estimated residuals $\hat{\varepsilon}_{its} = \hat{\tau}_{is} + \hat{\vartheta}_{its}$ we can compute the contract specific cost inefficiency:

$$\hat{\epsilon}_i = \hat{\varepsilon}_i - \min_{I} \{\hat{\varepsilon}_i\},$$

where $\hat{\varepsilon}_{it} = T(i)^{-1} \sum_{t=1}^{T(i)} \hat{\varepsilon}_{its}$.

Following Smith and Wheat (2012a), the overall inefficiency of a region is estimated by adding the region-specific inefficiency to a weighted average of the contract-specific inefficiency:

$$\bar{\mu}_i = \mu_i + \frac{\sum_{k=1}^{K} \tau_{ik} \epsilon_{iks}}{\sum_{k=1}^{K} \epsilon_{iks}},$$

### 3.1 Hypothesis tests

The gradual exposure to competitive tendering allows us to test its effect on costs. We test this with a dummy variable indicating when a contract area is tendered in competition. Moreover, we examine if there is a difference in costs between the first contract period in an area (when tendered in competition) and the following tender periods within the same area.
There might be a selection bias due to omitted variables or reverse causality. In our case, omitted variable bias is present if there are systematic cost differences between contract areas tendered first and areas tendered later not captured by our explanatory variables, i.e. we have an omitted variable that is correlated with the tendering variable. Reverse causality can be present if areas tendered first were tendered because they had high (low) costs.

We use the same approach as Odolinski and Smith (2014) when testing for a selection bias (see also Domberger et al. (1987) and Smith and Wheat (2012b)) and construct the following vector of variables:

$$Z_{jita} = [D_{jia}, D_{ita}, D_{it}; \beta_{jt}],$$  \hspace{1cm} (8)

The dummy variables indicate areas tendered in competition ($j = C$), areas tendered first ($j = F$), areas tendered later ($j = L$). The time period when areas are tendered in competition is indicated by $t = 0$, and $t = B$ indicates the time before tendered in competition. $t = P$ indicates when a contract area is tendered for the first time, while $t = E$ when an area is tendered the second or the third time. $D_{it}$ are a set of year dummy variables and $\beta_{jt}$ are the parameters to be estimated. Note that the variables in (9) are not mutually exclusive, and hence, not all the variables are included in the same model estimation.

A general difference-in-differences approach would include a post-competitive tendering dummy variable along with a dummy variable indicating areas tendered in competition, as well as an interaction between these variables (see for example Greene 2012, p.155-157). However, all areas are tendered in competition in our data set, except one that is tendered in 2014, and the exposure to competition was gradual which means that we do not have a post competitive tendering period for all contracts. We therefore use the interaction between $D_{cia}$ (areas tendered in competition) and $D_{t0a}$ (time period when area is tendered in competition)

---

\footnote{The definition of an area tendered first is arbitrary because the exposure to competition was gradual. We therefore perform sensitivity tests with respect to this definition.}
to capture the effect of competitive tendering and we use year dummies to control for general
effects each year not to be confounded with the effects of tendering.

Using the set of variables in (9) we can test the following hypotheses:

**Hypothesis 1** $\beta_{FB} = 0$; prior to tendering there is no difference in costs for areas
tendered first compared to the group of areas tendered later, and the
area not tendered\(^6\)

**Hypothesis 2** $\beta_{CO} = 0$; competitive tendering had no effect on costs

**Hypothesis 3** $\beta_{CP} = \beta_{CE}$; no difference in costs between the first contract period and
the following period(s) when tendered in competition

### 4.0 Data

The data set is an unbalanced panel over the period 1999-2013. Data has been obtained from
the IM (*Trafikverket*) apart from the climate variables collected from the Swedish
Metrological and Hydrological Institute (SMHI), and hourly wages and price indices of
materials used in maintenance collected from Statistics Sweden.

We use observations at the contract level, which is described in the next section, followed
by a section with descriptive statistics for the variables used in the estimations.

### 4.1 Regions and contracts

There are five regional units: East, West, South, North and Central, each responsible for
managing the maintenance contracts within their geographical area. Six contracts in our
sample are jointly managed by two different regions, which therefore generates a sixth

\(^6\) This estimation also includes $D_{Fia} \cdot D_{iOa}$ and $D_{Lia} \cdot D_{iOa}$, that is, the estimation of $\beta_{FO}$ and $\beta_{LO}$
respectively.
artificial region. The contracts consist of a varying number of track sections\(^7\) and the size of the contracts range from about 60 to 1240 track kilometers. However, since competitive tendering started, some contract areas have been redesigned, either two or three areas has merged into one, or an area has been split, creating two new areas. We have accounted for these changes when compiling the dataset.

A contract is normally tendered for 5 years with a possibility for up to two years of prolongation. Hence, no contract is observed for the whole period 1999-2013. Moreover, the contract periods usually do not start at the beginning of a calendar year, while the available data on costs, traffic, infrastructure characteristics etc. does. In order to analyze the cost efficiency of contracts, we therefore need to drop 53 observations with a mix of two different contracts. Hence, the number of observations during 1999-2013 varies, ranging from 21-33 in each year. In total, we have 420 observations.

Activities included in the contracts are corrective and preventive maintenance, including snow removal. Most of the contracts are performance based, with a set of requirements on track quality and train delays stated in the contract, while a few contracts are so called design-bid-maintain contracts in which the contractor mainly executes the activities set up by the client. Moreover, the design of the performance contracts varies. For example, there are differences in the payment clauses for rectifying infrastructure failures, where a fixed payment turns into a variable payment if the cost of rectifying a failure reaches a certain level. Added to this, different incentive schemes have been used, with a bonus/penalty linked to the number of failures that occur. Unfortunately, information on the specific contract designs throughout the period 1999-2013 is incomplete, which limits the possibility to include these aspects in the econometric analysis of cost efficiency. However, the differences in

---

\(^7\) Track sections are a breakdown of the railway system, and are for example used for financial follow-up.
contract design can be a major source of the efficiency differences estimated in our model, which is further discussed in section 6.

4.2 Costs

The dependent variable in our estimations is maintenance costs, and we obtained the data from the IM’s accounting system.

The wages for the employees differ depending on where in Sweden maintenance is performed. Hence, the contractors might face different labor costs, affecting their bids and most likely the final cost of the contract. We have obtained data from the Swedish Mediation Office (via Statistics Sweden) on the gross hourly wages for workers within the occupational category “building frame and related trade workers” in eight different regions. Naturally, these regions do not perfectly correspond to the maintenance contract areas, and we therefore use an average wage when a contract area is located in two wage regions. The average wage is based on the share of the contract’s track length located in each region.

Most of the materials used in maintenance are procured by the IM and sold to the firms without any price discrimination. The input prices for materials will therefore only vary over time. This variation can still affect the contracts differently depending on the size of the area. Materials often used in railway maintenance are iron, steel and copper. Price indices for these materials were obtained from Statistics Sweden. We note that the correlation coefficient for the iron and steel price index and the copper price index is 0.93. Table 1 presents the descriptive statistics of maintenance costs, wages and price indices for iron, steel and copper.

<table>
<thead>
<tr>
<th>Table 1 – Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Maintenance cost, million SEK*</td>
</tr>
<tr>
<td>Hourly wage, SEK*</td>
</tr>
</tbody>
</table>
Iron and steel, price index  
105.55  
96.08  
32.58  
52.30  
140.90  

Copper, price index  
126.70  
120.59  
53.32  
55.00  
199.10  

*2013-prices: inflation adjusted using the Swedish consumer price index

4.3 Traffic, infrastructure characteristics and weather

The output measure we use is tonne density (tonne-km/route-km), which is standard in the literature on rail infrastructure costs. We also have access to separate estimates of tonne density for freight and passenger traffic.

Different variables capturing the infrastructure characteristics have been considered, all obtained from the IM’s track information system. The descriptive characteristics of the variables we use are presented in Table 2. We also have access to a track quality class variable, which is a number assigned to the track and is mainly decided with respect to the maximum speed allowed. A temporary lowering of line speeds does not affect the quality class, and this variable is more or less time-invariant (slow-moving).

There are different requirements on track geometry standard depending on the track quality class. For example, a high linespeed will require a high track geometry standard. When a deviation from the required track geometry reaches certain limits, it will be defined as a failure. We do not have access to a consistent data set over the number of track geometry failures (see Odolinski and Smith 2014 for an off-model analysis of these failures over the period 2002-2012). In addition, including the number of failures as an explanatory variable in our cost model is problematic as this variable would be endogenous; more failures require more corrective maintenance, while more preventive maintenance decreases the number of failures that occur.

Table 2 — Traffic, infrastructure and weather

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>Mean</th>
<th>St. dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tonne density, million*</td>
<td>4.82</td>
<td>7.52</td>
<td>7.05</td>
<td>0.33</td>
<td>32.94</td>
</tr>
<tr>
<td></td>
<td>1.05</td>
<td>2.78</td>
<td>4.79</td>
<td>0.00</td>
<td>31.02</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Passenger train tonne density, million*</td>
<td>2.30</td>
<td>4.57</td>
<td>4.66</td>
<td>0.02</td>
<td>18.81</td>
</tr>
<tr>
<td>Freight train tonne density, million*</td>
<td>284.05</td>
<td>350.5</td>
<td>226.2</td>
<td>58.5</td>
<td>1237.5</td>
</tr>
<tr>
<td>Track length, km</td>
<td>235.97</td>
<td>279.47</td>
<td>171.64</td>
<td>22.95</td>
<td>1008.94</td>
</tr>
<tr>
<td>Route length, km</td>
<td>1.14</td>
<td>1.33</td>
<td>0.41</td>
<td>1</td>
<td>2.93</td>
</tr>
<tr>
<td>Ratio track length and route length</td>
<td>5.86</td>
<td>8.5</td>
<td>6.7</td>
<td>0.6</td>
<td>38.2</td>
</tr>
<tr>
<td>Switch length, km</td>
<td>3.28</td>
<td>5.4</td>
<td>7.1</td>
<td>0.5</td>
<td>40.5</td>
</tr>
<tr>
<td>Track length bridges and tunnels, km</td>
<td>51.72</td>
<td>51.5</td>
<td>3.5</td>
<td>43.7</td>
<td>59.6</td>
</tr>
<tr>
<td>Rail weight, kg (average)</td>
<td>2.17</td>
<td>2.07</td>
<td>0.83</td>
<td>0.41</td>
<td>4.00</td>
</tr>
<tr>
<td>Snow mm precip. when temp. &lt; 0 Celsius (average)</td>
<td>104.96</td>
<td>122.7</td>
<td>68.3</td>
<td>2.1</td>
<td>308.8</td>
</tr>
<tr>
<td>Max. axle load allowed (average)</td>
<td>22.5</td>
<td>23.17</td>
<td>1.86</td>
<td>19.57</td>
<td>30</td>
</tr>
<tr>
<td>Dummy var. tendered in competition</td>
<td>0</td>
<td>0.44</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy var. tendered first period</td>
<td>0</td>
<td>0.28</td>
<td>0.45</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy var. tendered subsequent periods</td>
<td>0</td>
<td>0.15</td>
<td>0.36</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy var. tendered first (before 2005), prior to tend.**</td>
<td>0</td>
<td>0.13</td>
<td>0.33</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy var. tendered first (before 2005), after tend.**</td>
<td>0</td>
<td>0.18</td>
<td>0.39</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dummy var. tendered later (after 2004), after tend.**</td>
<td>0</td>
<td>0.25</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

* Tonne-km/route-km, ** We also test corresponding dummy variables with year 2006 as the limit for when an area is considered to be tendered first.

With snow removal included in the maintenance contracts, it is important to consider that the amount of snowfall varies within and between contracts. We have therefore obtained data from SMHI which consists of temperature and precipitation data at the resolution of 4x4km. We created a variable for the amount of snowfall during a year, using mm of precipitation each day when the daily mean temperature is below zero degrees Celsius. Data from the 4x4km grids were retrieved for different locations along the tracks.

### 5.0 Results

Three models are estimated: Model 1 assumes the region-specific inefficiencies to be time-invariant and Model 2 allows these inefficiencies to vary over time as specified in (5). Model 3 includes the Mundlak (1978) formulation to control of heterogeneity bias. We focus on the
results from the deterministic part of the model in section 5.1, while the inefficiency estimates are presented in section 5.2. All estimations are carried out with Stata 12 (StataCorp.11).

5.1 Deterministic frontier

The models are estimated with random effects so that time-invariant variables can be included in the estimation (tendering variables and region dummies). However, the formulation in Model 3 implies that the first order coefficients are identical to the fixed effects estimates (see proof in Mundlak 1978). Moreover, the models are estimated with robust standard errors, since a plot of the residuals and total tonnage density indicate presence of heteroskedasticity, however, not substantial.

We start with a translog model, and test down. In the end we keep a squared track length variable, a squared variable for maximum axle load and an interaction variable between wage and tonne density. Note that the iron and steel variable only vary over time, and is therefore collinear with one of the dummy variables. Consequently, only interactions with this input price variable could be included in the estimations, but these turned out to be non-significant.

First, we note that the coefficients for the interactions between region dummies and time trends in Model 2 are jointly significant (chi2(10)=26.45, prob>chi2=0.003), which indicate that the region-specific inefficiency is time-varying. The group means in Model 3 are jointly significant (chi2(13)=48.58, prob>chi2=0.000). However, the Variance Inflation Factor (VIF) of the explanatory variables for which group means have been included, increases substantially with a mean VIF at 66.8 compared to 5.6 in Model 2. Hence, we have problems with multicollinearity in Model 3. We therefore focus on the deterministic frontier results from Model 2, and return to Model 3 in the following section on efficiency results.
Table 3 - Estimation results

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th>Model 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>17.5523***</td>
<td>0.1442</td>
<td></td>
<td>17.4082***</td>
<td>0.1354</td>
<td></td>
<td>17.3762***</td>
<td>0.1658</td>
</tr>
<tr>
<td>WAGE</td>
<td>0.5608</td>
<td>0.7420</td>
<td></td>
<td>0.0700</td>
<td>0.7138</td>
<td></td>
<td>-0.0145</td>
<td>0.7057</td>
</tr>
<tr>
<td>TGT_DEN</td>
<td>0.2126***</td>
<td>0.0517</td>
<td></td>
<td>0.1722***</td>
<td>0.0532</td>
<td></td>
<td>0.0567</td>
<td>0.1478</td>
</tr>
<tr>
<td>TRACK_L</td>
<td>0.4261***</td>
<td>0.0756</td>
<td></td>
<td>0.4328***</td>
<td>0.0758</td>
<td></td>
<td>0.4587**</td>
<td>0.2092</td>
</tr>
<tr>
<td>RATIO_LRO</td>
<td>-0.2110</td>
<td>0.1885</td>
<td></td>
<td>-0.0834</td>
<td>0.1752</td>
<td></td>
<td>-0.1802</td>
<td>0.2665</td>
</tr>
<tr>
<td>RAILW</td>
<td>-0.5699</td>
<td>0.8032</td>
<td></td>
<td>0.1795</td>
<td>0.7062</td>
<td></td>
<td>0.3742</td>
<td>1.1739</td>
</tr>
<tr>
<td>QUALAVE</td>
<td>0.0028</td>
<td>0.1231</td>
<td></td>
<td>0.0605</td>
<td>0.1155</td>
<td></td>
<td>-0.0670</td>
<td>0.2140</td>
</tr>
<tr>
<td>SW_L</td>
<td>0.2767***</td>
<td>0.0745</td>
<td></td>
<td>0.2622***</td>
<td>0.0741</td>
<td></td>
<td>0.3742***</td>
<td>0.0891</td>
</tr>
<tr>
<td>STRUCT_L</td>
<td>0.0793***</td>
<td>0.0317</td>
<td></td>
<td>0.0700**</td>
<td>0.0347</td>
<td></td>
<td>0.6622**</td>
<td>0.0247</td>
</tr>
<tr>
<td>MAXAXL</td>
<td>-0.7727</td>
<td>0.5428</td>
<td></td>
<td>-0.5601</td>
<td>0.5311</td>
<td></td>
<td>-0.5091</td>
<td>0.6338</td>
</tr>
<tr>
<td>MMPRECIP</td>
<td>0.0619***</td>
<td>0.0195</td>
<td></td>
<td>0.0523**</td>
<td>0.0239</td>
<td></td>
<td>0.0622**</td>
<td>0.0247</td>
</tr>
<tr>
<td>TRACK_L^2</td>
<td>0.1300**</td>
<td>0.0519</td>
<td></td>
<td>0.1458***</td>
<td>0.0531</td>
<td></td>
<td>-0.0849</td>
<td>0.1303</td>
</tr>
<tr>
<td>MAXAXL^2</td>
<td>2.2983</td>
<td>1.7346</td>
<td></td>
<td>2.5652</td>
<td>1.5663</td>
<td></td>
<td>0.3917</td>
<td>1.6309</td>
</tr>
<tr>
<td>WAGE_TGT</td>
<td>0.4375*</td>
<td>0.2251</td>
<td></td>
<td>0.4277*</td>
<td>0.2480</td>
<td></td>
<td>0.4694*</td>
<td>0.2704</td>
</tr>
<tr>
<td>D_FIRST_P</td>
<td>-0.1247*</td>
<td>0.0741</td>
<td></td>
<td>-0.1130</td>
<td>0.0705</td>
<td></td>
<td>-0.1627</td>
<td>0.1000</td>
</tr>
<tr>
<td>D_SUBS_P</td>
<td>-0.2354**</td>
<td>0.1032</td>
<td></td>
<td>-0.2654**</td>
<td>0.1086</td>
<td></td>
<td>-0.3062**</td>
<td>0.1437</td>
</tr>
<tr>
<td>D.2000</td>
<td>-0.0674</td>
<td>0.0616</td>
<td></td>
<td>-0.0123</td>
<td>0.0694</td>
<td></td>
<td>-0.0003</td>
<td>0.0701</td>
</tr>
<tr>
<td>D.2001</td>
<td>-0.0730</td>
<td>0.0761</td>
<td></td>
<td>0.0387</td>
<td>0.1051</td>
<td></td>
<td>0.0598</td>
<td>0.1022</td>
</tr>
<tr>
<td>D.2002</td>
<td>0.1082</td>
<td>0.0879</td>
<td></td>
<td>0.2494*</td>
<td>0.1311</td>
<td></td>
<td>0.2909**</td>
<td>0.1305</td>
</tr>
<tr>
<td>D.2003</td>
<td>0.1270</td>
<td>0.0938</td>
<td></td>
<td>0.2836**</td>
<td>0.1445</td>
<td></td>
<td>0.3411**</td>
<td>0.1436</td>
</tr>
<tr>
<td>D.2004</td>
<td>0.0775</td>
<td>0.1075</td>
<td></td>
<td>0.2787</td>
<td>0.1705</td>
<td></td>
<td>0.3548**</td>
<td>0.1738</td>
</tr>
<tr>
<td>D.2005</td>
<td>0.1053</td>
<td>0.0980</td>
<td></td>
<td>0.3097*</td>
<td>0.1787</td>
<td></td>
<td>0.3970**</td>
<td>0.1786</td>
</tr>
<tr>
<td>D.2006</td>
<td>0.0675</td>
<td>0.1186</td>
<td></td>
<td>0.2748</td>
<td>0.1974</td>
<td></td>
<td>0.3720*</td>
<td>0.1939</td>
</tr>
<tr>
<td>D.2007</td>
<td>0.1417</td>
<td>0.1419</td>
<td></td>
<td>0.3694*</td>
<td>0.2140</td>
<td></td>
<td>0.4819**</td>
<td>0.2127</td>
</tr>
<tr>
<td>D.2008</td>
<td>0.2098</td>
<td>0.1643</td>
<td></td>
<td>0.4309*</td>
<td>0.2293</td>
<td></td>
<td>0.5516**</td>
<td>0.2270</td>
</tr>
<tr>
<td>D.2009</td>
<td>0.2602</td>
<td>0.1953</td>
<td></td>
<td>0.4866*</td>
<td>0.2526</td>
<td></td>
<td>0.6064**</td>
<td>0.2493</td>
</tr>
<tr>
<td>D.2010</td>
<td>0.2673</td>
<td>0.1932</td>
<td></td>
<td>0.4839*</td>
<td>0.2475</td>
<td></td>
<td>0.5829**</td>
<td>0.2379</td>
</tr>
<tr>
<td>D.2011</td>
<td>0.3439*</td>
<td>0.1792</td>
<td></td>
<td>0.5335**</td>
<td>0.2410</td>
<td></td>
<td>0.6495**</td>
<td>0.2305</td>
</tr>
<tr>
<td>D.2012</td>
<td>0.4947**</td>
<td>0.2015</td>
<td></td>
<td>0.6740**</td>
<td>0.2673</td>
<td></td>
<td>0.7616***</td>
<td>0.2585</td>
</tr>
<tr>
<td>D.2013</td>
<td>0.5741***</td>
<td>0.1988</td>
<td></td>
<td>0.7143***</td>
<td>0.2661</td>
<td></td>
<td>0.7999***</td>
<td>0.2567</td>
</tr>
<tr>
<td>D.EAS</td>
<td>-0.1494</td>
<td>0.1198</td>
<td></td>
<td>-0.2247*</td>
<td>0.1250</td>
<td></td>
<td>-0.1975</td>
<td>0.1370</td>
</tr>
<tr>
<td>D.WES</td>
<td>0.0357</td>
<td>0.1051</td>
<td></td>
<td>0.2463**</td>
<td>0.1221</td>
<td></td>
<td>0.1828</td>
<td>0.1160</td>
</tr>
<tr>
<td>D.STH</td>
<td>-0.1468</td>
<td>0.1032</td>
<td></td>
<td>-0.0254</td>
<td>0.1456</td>
<td></td>
<td>-0.0020</td>
<td>0.1696</td>
</tr>
<tr>
<td>D.NTH</td>
<td>-0.0073</td>
<td>0.1264</td>
<td></td>
<td>0.2773*</td>
<td>0.1637</td>
<td></td>
<td>0.2265</td>
<td>0.2290</td>
</tr>
<tr>
<td>D.CTR</td>
<td>-0.0080</td>
<td>0.1083</td>
<td></td>
<td>0.1845</td>
<td>0.1265</td>
<td></td>
<td>0.2282</td>
<td>0.1667</td>
</tr>
<tr>
<td>TR.EAS</td>
<td>-</td>
<td>-</td>
<td></td>
<td>0.0256</td>
<td>0.0497</td>
<td></td>
<td>0.0165</td>
<td>0.0481</td>
</tr>
</tbody>
</table>
We have transformed the data by dividing with the sample median prior to taking logs. The first order coefficients can therefore be interpreted as elasticities at the sample median.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAGETGTbar</td>
<td>0.1309</td>
<td>0.5190</td>
</tr>
<tr>
<td>TRACK_Lbar</td>
<td>0.0003</td>
<td>0.2337</td>
</tr>
<tr>
<td>RAILWbar</td>
<td>-</td>
<td>0.2469</td>
</tr>
<tr>
<td>SW_Lbar</td>
<td>-</td>
<td>0.1261</td>
</tr>
<tr>
<td>STRUCT_Lbar</td>
<td>-</td>
<td>0.0754</td>
</tr>
<tr>
<td>MAXAXLbar</td>
<td>-</td>
<td>1.0659</td>
</tr>
<tr>
<td>SNOWMMbar</td>
<td>-</td>
<td>0.1213</td>
</tr>
<tr>
<td>TRACK_L^2bar</td>
<td>0.2716**</td>
<td>0.1342</td>
</tr>
<tr>
<td>MAXAXL^2bar</td>
<td>8.1072**</td>
<td>3.2195</td>
</tr>
<tr>
<td>WAGEbar</td>
<td>0.1804</td>
<td>1.0752</td>
</tr>
<tr>
<td>TGTDENbar</td>
<td>0.1858</td>
<td>0.1719</td>
</tr>
<tr>
<td>TRACK_Lbar</td>
<td>0.0003</td>
<td>0.2337</td>
</tr>
<tr>
<td>RAILWbar</td>
<td>-</td>
<td>0.2469</td>
</tr>
<tr>
<td>SW_Lbar</td>
<td>-</td>
<td>0.1261</td>
</tr>
<tr>
<td>STRUCT_Lbar</td>
<td>-</td>
<td>0.0754</td>
</tr>
<tr>
<td>MAXAXLbar</td>
<td>-</td>
<td>1.0659</td>
</tr>
<tr>
<td>SNOWMMbar</td>
<td>-</td>
<td>0.1213</td>
</tr>
<tr>
<td>TRACK_L^2bar</td>
<td>0.2716**</td>
<td>0.1342</td>
</tr>
<tr>
<td>MAXAXL^2bar</td>
<td>8.1072**</td>
<td>3.2195</td>
</tr>
<tr>
<td>WAGEbar</td>
<td>0.1804</td>
<td>1.0752</td>
</tr>
<tr>
<td>TGTDENbar</td>
<td>0.1858</td>
<td>0.1719</td>
</tr>
<tr>
<td>TRACK_Lbar</td>
<td>0.0003</td>
<td>0.2337</td>
</tr>
<tr>
<td>RAILWbar</td>
<td>-</td>
<td>0.2469</td>
</tr>
<tr>
<td>SW_Lbar</td>
<td>-</td>
<td>0.1261</td>
</tr>
<tr>
<td>STRUCT_Lbar</td>
<td>-</td>
<td>0.0754</td>
</tr>
<tr>
<td>MAXAXLbar</td>
<td>-</td>
<td>1.0659</td>
</tr>
<tr>
<td>SNOWMMbar</td>
<td>-</td>
<td>0.1213</td>
</tr>
<tr>
<td>TRACK_L^2bar</td>
<td>0.2716**</td>
<td>0.1342</td>
</tr>
<tr>
<td>MAXAXL^2bar</td>
<td>8.1072**</td>
<td>3.2195</td>
</tr>
<tr>
<td>WAGEbar</td>
<td>0.1804</td>
<td>1.0752</td>
</tr>
</tbody>
</table>

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

We have transformed the data by dividing with the sample median prior to taking logs. The first order coefficients can therefore be interpreted as elasticities at the sample median.

Number of observations = 420

Definition of variables in table 3:

- **WAGE** = ln(Average gross hourly wage)
- **TGTDEN** = ln(train tonnage density)
- **TRACKL** = ln(Track length)
- **RATIOTLRO** = ln(Track length/route length)
- **RAILW** = ln(Average rail weight)
- **QUALAVE** = ln(Average quality class); note a high value of average quality class implies a low speed line
- **SW_TL** = ln(Track length of switches)
- **STRUCTL** = ln(Track length of structures (tunnels and bridges))
- **MAXAXL** = ln(Average maximum axle load allowed)
- **MMPRECIP** = ln(Average mm of precipitation (liquid water) when temperature < 0°Celsius)
- **TRACKL2** = ln(Track length)^2
MAXAXL^2 = ln(Average maximum axle load allowed)^2
WAGETGT = ln(Average gross hourly wage)*ln(train tonnage density)
D.FIRST_P = Dummy for years when area is tendered in competition for the first time
D.SUBS_P = Dummy for years when area is tendered in competition second and third time
D.2000,...,D.2013 = Year dummy variables, 2000-2013
D.EAS,...,D.CTR = Dummy variables region East, West, South, North and Central
TR.EAS,...,TR.CTR = (Time trend*D.EAS),...,(Time trend*D.CTR)
TR2.EAS,...,TR2.CTR = ((Time trend^2)*D.EAS),...,((Time trend^2)*D.CTR)
WAGEbar,..., WAGETGTbar = T_{t-1} \sum_{t=1}^{T} ln(wage)_{its},..., T_{t-1} \sum_{t=1}^{T} ln(wage)_{its} + ln(tgtden)_{its}

First, we note that most the coefficients for the year dummies are positive and significantly different from the baseline year 1999, which also holds when changing the baseline to year 2000 or 2001. This indicates that we have a general increase in maintenance costs during the studied period compared to these baselines. Moreover, coefficients for years 2012 and 2013 are positive and significantly different from any other year.

The parameter estimate for wages has the expected sign, but is quite low and not statistically significant. However, the coefficient for the interaction between wage and tonne density (WAGETGT) is significant and positive. This implies that the effect of wages is higher when tonne density increases, which can be due to a more labor intense maintenance when the traffic volume is high (shorter time slots for maintenance) and/or more work during nighttime.

Total tonne density is used instead of tonnage measures for freight and passenger trains. The cost elasticities for the separate measures of outputs were similar, and because we lose 13 observations using passenger tonne density (zero values that we cannot use after the log transformation), we prefer the total tonne density measure. The estimated cost elasticity of 0.17 is in line with previous estimates on long panels of Swedish data (see Odolinski and...
Smith 2014, Odolinski and Nilsson 2014), but is slightly below the interval of the estimates for a number of European countries, which is between 0.20 and 0.35 (Wheat et al. 2009).

The parameter estimates for track length (TRACK KM), switch length (SW KM) and track length of tunnels and bridges (STRUCT KM) have the expected signs and are statistically significant. We can use these estimates to examine if we have economies of scale by taking their sum (which is 0.76 at the sample median) and test the null hypothesis of no economies of scale (testing if the sum is equal to 1), using a Wald test. We reject the null hypothesis (Chi²(1)=18.79, P-value=0.000) and conclude that we have economies of scale at the sample median. Note that the estimate for the second order term for track length is 0.1458 (p-value 0.006). Hence, contract areas that are sufficiently larger than the sample median have constant economies of scale (=1).

The coefficient for the ratio between track length and route length (RATIOTL) is negative but not significantly different from zero (p-value 0.634). Moreover, the results show that the coefficient for rail weight (RAIL W) is not significant. Rail weight and rail age have a negative correlation coefficient (-0.5466), implying that newer rails are heavier, and the wear and tear on heavier rail is lower compared to lighter rails, ceteris paribus. Hence, in this respect costs should decrease with heavier rail, but we note that the correlation coefficient between rail weight and tonne density is 0.7937 (dropping the rail weight variable does not change our results). This suggests that the accumulated tonne density on light (and old) rails does not increase at the same rate as on heavier rails, which can counter the effect of lower wear and tear on heavier rail. Moreover, we note that the coefficient for maximum axle load allowed (MAXAXL) is not significant (p-value=0.292), yet the second order term is nearly significant at the 10 per cent level (p-value=0.101).

The expected sign of the quality class coefficient (QUALAVE) is not a priori evident. A higher value on this variable implies lower linespeed (resulting in lower forces on the
track), and also lower requirements on track geometry standard, which by itself increases the wear and tear. Hence, we have two conflicting effects. A coefficient not significantly different from zero (which is the case in our model estimation) can be due to a good balance between requirements on track standard and linespeed.

Differences in weather can have an effect on the maintenance cost, especially since snow removal is included in the maintenance contracts. The variable capturing the amount of snowfall (MMPRECIP) has the expected sign and is significant at the 5 per cent level.

Turning to the policy variables included in the estimation results presented in Table 3 (D.FIRST_P and D.SUBS_P), we can see that competitive tendering reduced costs, yet the coefficient for the first period is not statistically significant (p-value=0.109). Interestingly, the parameter estimates show a larger effect on costs in the subsequent periods compared to the first period, and this difference is significant (Wald test in Model 2: chi2(1)=2.91, prob>chi2=0.088). We can therefore reject Hypothesis 3. The estimated effect for the first period of competitive tendering is -0.1068, with p-value=0.109, and the effect for the following periods is -0.2331 with p-value=0.015.\(^8\)\(^9\) Moreover, we test the presence of selection bias, which could be due to systematic cost differences between areas tendered first and areas tendered later (omitted variables or reverse causality). However, the parameter estimate of interest is not significant (P-value=0.285) and we can reject Hypothesis 1.\(^10\) Using a general dummy variable for competitive tendering (not discriminating between first and subsequent periods) shows that competitive tendering decreased costs with 13.4 per cent (the coefficient is -0.1439, with p-value=0.040), and we can therefore reject Hypothesis 2.\(^11\)

\(^8\) Exp(-0.1130)-1=-0.1068  
\(^9\) Exp(-0.2654)-1=-0.2331  
\(^10\) Changing the definition of areas tendered first does not change this conclusion (the p-value is 0.887 for the coefficient in Hypothesis 1)  
\(^11\) Exp(-0.1439)-1=-0.1340
With respect to the estimated effects of tendering, it should be noted that administrative costs and other types of overhead costs are not included in the presented estimation results. However, data at the national level for these costs during 1999-2012 has been retrieved from the IM. Allocating these costs to the contracts using each contract’s share of total maintenance costs as weights and re-estimating our model for years 1999-2012 did not change our results (which was not surprising given that there is no strong trend in these costs).

5.2 Efficiency results

Taking account of the cost drivers as well as factors lowering maintenance costs, presented in Table 3, we can estimate the cost efficiency in regions and contracts within each region. In Model 1, the efficiency gap between regions is assumed to be constant over time, while in Models 2 and 3 the region-specific inefficiency is allowed to vary over time for each region. The region-specific inefficiencies are estimated according to equation (5) and the expression we use to calculate the efficiency scores is $\exp(-\hat{u}_i)$. These scores are presented in Table 4, ranging from 0 to 1 where a region on the frontier has an efficiency score of 1.

The results in table 3 include coefficients for all regions except the mixed region which is the baseline, and Figure 2 shows the development of these region-specific estimates in relation to the mixed region (value 0) from the Model 2 results. The overall trend is that the region-specific estimates has decreased for region East and Central over the studied period, while the estimates for other regions has increased since 2006. However, based on the Model 1 estimates, only region West has a significantly lower time-invariant estimate than region East ($\chi^2(1)=3.53$, prob$>\chi^2=0.060$) and region South ($\chi^2(1)=4.82$, prob$>\chi^2=0.028$). There are no major differences in the region-specific efficiency scores between our three models, with region East and South appearing to be more efficient than the other regions.
We now turn to the contract-specific cost efficiency, which means that we consider the within region differences instead of the relative difference between regions. These inefficiencies are time-invariant in both models: $\exp(-\hat{\tau}_{t_0})$. The minimum efficiency scores are 0.275, 0.266 and 0.318 in Models 1, 2 and 3 respectively. The maximum scores (the cost frontier) are 1 by construction. Hence, we have considerable cost efficiency differences between contracts. Table 4 contains the weighted average of the contract-specific inefficiency for each region. The efficiency scores are a bit higher in Model 3 compared to the other models, which can be due to the Mundlak formulation controlling for unobserved heterogeneity that end up as inefficiency in Models 1 and 2.
Table 4 – Efficiency scores

<table>
<thead>
<tr>
<th>Region</th>
<th>Region-specific efficiency</th>
<th>Contract-specific efficiency</th>
<th>Overall efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
</tr>
<tr>
<td>East</td>
<td>1.000</td>
<td>0.935</td>
<td>0.941</td>
</tr>
<tr>
<td>South</td>
<td>0.997</td>
<td>0.958</td>
<td>0.951</td>
</tr>
<tr>
<td>North</td>
<td>0.868</td>
<td>0.853</td>
<td>0.895</td>
</tr>
<tr>
<td>Central</td>
<td>0.868</td>
<td>0.802</td>
<td>0.796</td>
</tr>
<tr>
<td>West</td>
<td>0.831</td>
<td>0.789</td>
<td>0.838</td>
</tr>
<tr>
<td>Mix</td>
<td>0.861</td>
<td>0.808</td>
<td>0.783</td>
</tr>
</tbody>
</table>

Lastly, we calculate the overall efficiency score for the regions using both the regional- and contract-specific inefficiency estimates according to equation (7). These are also presented in Table 4. The difference between the overall efficiency scores is an indication of the possible improvement that can be made if best practice was replicated within and between regions. Being aware of the uncertainty in these efficiency scores, where most estimates are not significantly different from each other, we calculate the cost reduction that is possible. We use each region’s share of total maintenance cost as weights and calculate the change in cost efficiency:

\[
\sum_{i} \frac{C_{ij}(\bar{\mu}_i - \bar{\mu}_j)}{\sum_{i} C_{ij} \bar{\mu}_j},
\]

where \( i = \text{Region South} \) and \( j = \text{Region East, West, Central, North and the mixed region} \). This translates into a 13 per cent possible cost reduction according to Model 2, while it is 9 per cent in Model 3.

6.0 Discussion

Our results show that the cost efficiency between regions varies over time and that there are differences within regions and considerable improvements within regions can be made according to the large efficiency gaps between contracts. A striking result is the efficiency improvement for regions East and Central over the studied period, while the other regions’ cost efficiencies have declined since 2006. As we described earlier, the organization became
more centralized after a change in 2007 that created a central unit above the regional units. The effect of this change has not been directly estimated in the model. Nonetheless, this change is not associated with lower cost efficiency according to the results.

A potential source of performance variation between contracts (and regions) is the differences in contract design, which we were not able to include in the model estimation due to limited data availability. Hence, some of these differences will be captured by the inefficiency terms in the estimation, and we cannot estimate the effect a certain contract design has on either the location of the cost frontier, or the distribution of the cost inefficiency terms (which relates to the question whether a contract design should be considered a variable mainly affecting the production possibility set or the distance from the frontier). Nevertheless, the estimation results are informative in showing which contract(s) within a region is efficient and can also be used to further investigate if there is a general contract design and/or management practice that is specific for the most (in)efficient region.

One needs to ask why best practice does not disseminate within the organization; what are the reasons for the efficiency gaps? Gibbons and Henderson (2012) lists explanations to why differences in performance persist. The explanations are problems with perception, inspiration, motivation and implementation. This means that managers do not know they are inefficient (perception); they do not know what needs to be done to narrow the inefficiency gap (inspiration) and there are not enough incentives to do so (motivation); and/or managers have problems with implementing changes (implementation). The problem of perception is perhaps one of the major reasons for efficiency differences in railway maintenance in Sweden. To our knowledge, an estimation of the cost efficiency of railway maintenance contracts and regions in Sweden that accounts for differences in the production environment

12 The authors also suggest that there are barriers to creating so called relational contracts, which is roughly a “...shared understanding of each party's role in and rewards from achieving cooperation…” (Gibbons and Henderson 2012: p. 24).
has not been made. Hence, the result from this paper is a move towards a better perception of differences in cost efficiency in railway maintenance. Unfortunately, we cannot state the specific causes to these differences. However, given that the IM knows which regions and contracts that are cost efficient, an answer to the inspiration problem is to implement internal benchmarking/yardstick competition. Moreover, we found that the economies of scale are not fully exploited, which suggests that some contract areas should be redesigned.

Finding out whether a regulatory change has been beneficial or not is important for future policy regarding infrastructure management. According to the results in this paper, the exposure to competition has resulted in a further reduction in costs after the first period of tendering in a contract area. Learning-by-doing is a possible explanation for this result. Hence, the fear of initial decreases in costs at the expense of quality resulting in a need for more maintenance activities in the future (which was the case in Britain after the deregulation) has not been realized in Sweden at this stage with respect to competitive tendering.

**7.0 Conclusion**

We have estimated the relative cost efficiency between and within railway maintenance regions in Sweden, using a dual-level inefficiency model in which the organizational structure of the infrastructure management is taken into account. More specifically, we estimate the time-invariant cost efficiency during a contract period, while allowing the region-specific cost efficiency to vary over time. The results are informative for the management of the regions, indicating that three regions are on the wrong path while two other regions have improved their cost efficiency during the studied period. Moreover, the estimated differences in cost efficiency confirm that there are barriers to spreading best practice within an organization. Results from our study are a way forward to overcome these barriers and improve cost
efficiency in infrastructure management, for example via internal benchmarking. Furthermore, our results show that there is reason to reconsider the size of some contract areas in order to fully exploit the economies of scale.

Exposing a state-owned monopoly to competition by contracting out some of its services is another measure that can be taken with the aim of improving cost efficiency. Our results show that this has been a beneficial regulatory change for railway maintenance in Sweden, with the striking result that competitive tendering lowered costs significantly more in contract periods after the first period of tendering.

**References**


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

