Quantifying Effects of Incentives in a Rail Maintenance Performance-Based Contract

Torbjörn Stenbeck, Ph.D.[1]

Abstract: Methods to quantify and evaluate quality become important when lump sum and performance contracts replace traditional unit-price or cost-plus contracts. Here, a combined graphical and mathematical method is described along with its results when applied on a Swedish rail maintenance contract with incentives. The regression analysis tools in the Excel software were used. The result of the incentives was that train delay decreased about 10% and the number of technical errors decreased about 20%. The improved quality took place without cost increase. The good relation between the owner and the contractor did not suffer from the rise of efficiency. On the contrary, it was improved. With minor modifications and clarifications the owner now intends to use it for in-house contracting, as the case studied, as well as when outsourcing to private companies. With other performance indicators, the elaborated method and lessons learned should be applicable also for other sectors, where a contractor assuring a specified service level during a period of time is paid a bonus depending on degree of fulfillment.

DOI: 10.1061/(ASCE)0733-9364(2008)134:4(265)

CE Database subject headings: Contracts; Contract management; Incentives; Quality control; Railroad engineering; Maintenance; Monitoring.

Introduction

Although this paper contains lessons applicable to performance-based contracts in general, only literature narrowed down to *rail maintenance/performance-based contracts/quality/incentives* is covered in the following, brief introduction. In 1975, Baumol published a study for Amtrak, USA, addressing the possibilities to direct contractors performances toward the final users', i.e., the train passengers', perceived utility by economic incentives. Theoretical-mathematical considerations with regard to incentives have been published (Vickerman 2004; Tarakci et al. 2006), e.g., in connection with the privatization of rail maintenance in the United Kingdom. Practical, published ex post experience of incentive effects is more scarce, however, making even case studies based on ex post data of interest for research. A critical challenge is how to define, measure, and reward quality (Stenbeck 2007). Also critical is how and by whom the true quality verification procedure is effected. In the railway sector, governmental administrations such as the Federal Rail Administration (FRA) in the United States, monitor the system performance and development, in particular to assure safety (Andersson et al. 2004). FRA's Swedish counterpart's inspection system was used as determinant of quality in the following case study, a contract between the public infrastructure owner and its production unit. Incentives in an *in-house* contracting setting, where one or both parties have the same owner, add to both the challenges and opportunities (Whitehouse 2003).

Quantifying the Quality

Defining and quantifying quality is a research domain of its own. Often human assessment scales are used, although an objective, third party measurement would be preferable (Aguilar et al. 2005; Milling et al. 1998). In the following case, quality was defined by number of errors and minutes of train delay in a rail network. The errors and train delay were monitored and registered by a third party, which can be regarded as independent of the two contracting parties.

A theoretical framework was developed as follows. The perspective of the region's transport authority was assumed representing the user's interest and was divided into two parts; U_{Now} and U_{Future}. U_{Now} represents the user's momentary interest of minimal train delay and maximum pleasure of ride. U_{Future} represents the benefits of a facility at its best at lowest cost eternally. Formula (1) adds these two interests together without weighting them

$$U = U_{\text{Now}} + U_{\text{Future}} \tag{1}$$

In the studied case, U_{Now} was represented by train delay (hours and minutes) plus vibration (the Q coefficient, which was, however, eliminated from this study, and is on hold also in the real case). U_{Future} was represented by the gross number of registered technical errors in the system requiring immediate or later attention. Any of the three U (U, U_{Now}, and U_{Future}) can be assumed to depend on A, T, C, P, I, and W according to

$$U = F(A, T, C, P, I, W) \tag{2}$$

The variables A=condition of the asset; T=traffic intensity on the line; C=cost, the resources allocated; P=end user's preferences and how well the contractor is made aware of them;

[1]CBI-Swedish Cement and Concrete Research Institute, DKV 26, SE-100 44, Stockholm, Sweden. E-mail: torste@cbi.se

Note. Discussion open until September 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on June 29, 2007; approved on September 18, 2007. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 134, No. 4, April 1, 2008. ©ASCE, ISSN 0733-9364/2008/4-265–272/$25.00.

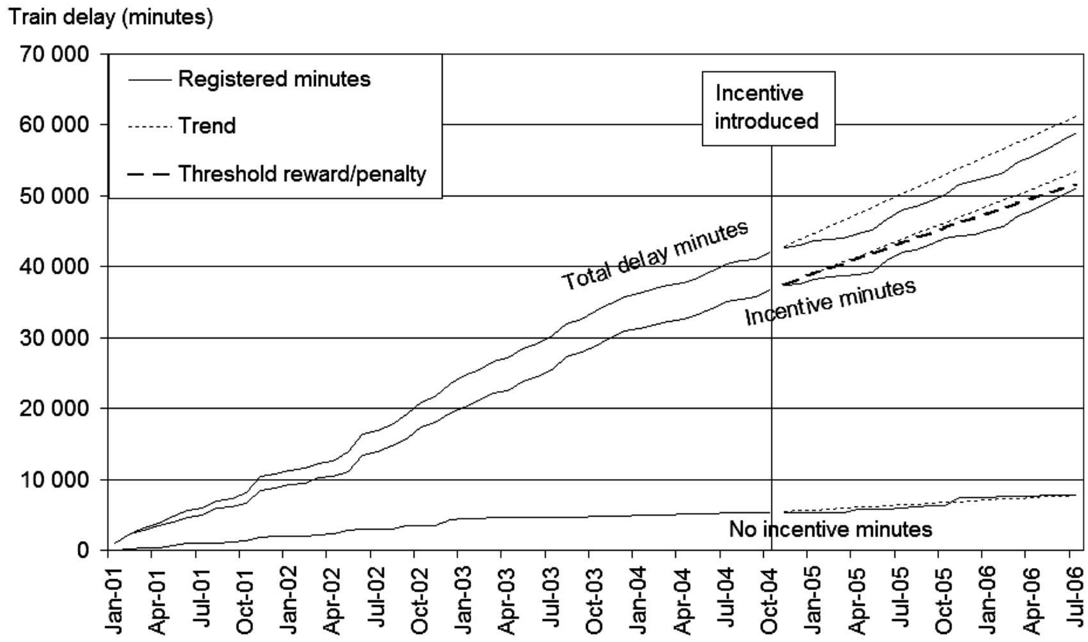

Fig. 1. Accumulated delay for the incentive segments (Segments 16–18, and 63 in area Västerås, Sweden)

I=incentives for the service provider to please the end user; and W denotes external factors like weather and sabotage.

Applying the Theoretical Framework on the Studied Case

The Swedish public rail infrastructure owner Banverket defined the output, U, by two parameters in this study; *train delay* (measured in hours and minutes, registered in *Tför*, Banverket's train delay information system) and technical *errors* (measured as discrete errors registered in *Ofelia*, Banverket's error information system). The research question was I's effect on U, why I needed to be discerned from the other variables (A, T, C, P, W) and then applied first on train delay and then on errors. I was made out by three methods used in combination:

(1) Compare output before and after incentives were introduced. The two periods will be referred to as the *preincentive period* versus the *incentive period*.

(2) Compare output of rail segments subject to incentive with output of rail segments not subject to incentives. The two categories will be referred to as *incentive segments* versus *no incentive segments*.

(3) Compare output related to causes included in the incentive with causes not included. The two categories will be referred to as *incentive causes* versus *no incentive causes*. Delay and errors connected to incentive causes will be referred to as *incentive minutes* and *incentive errors*. Delay and errors connected to no incentive causes will be referred to as *no incentive minutes* and *no incentive errors*.

Methods 1 and 3 are combined in Fig. 1 where the incentive segments only are displayed.

In Fig. 1, a trend break is visible for the middle curve but not for the bottom curve (Method 3). The incentive segments constitute 8.3% of Banverket Eastern Region, BRÖ (based on costs in July 2006). A corresponding diagram, which included all segments of Banverket's Eastern Region BRÖ, showed straight lines, i.e., no incentive effect could be noticed in November 2004 in any of the three curves (~Method 2). The three methods together support I having a noticeable effect on U. The next challenge is to quantify it and its variation over time in the incentive period.

Purpose

The purpose was to quantify the effect of incentives on quality and costs. A narrow approach with a particular rail maintenance performance-based contract was selected in order to deliver concrete and precise results. The owner and contractor had agreed on how to measure, quantify, and compensate quality. A graphical and mathematical method was tested to evaluate the result over time. In the studied case, quality was defined as total train delay (minutes) and technical error (number), why the effect in this case was subdivided into two:

(1) How much was the quality improvement, measured as delay reduction?

(2) How much was the quality improvement measured as reduction of technical errors? The effect on output should be critically compared with the change in cost for inputs:

(3) Did the increased quality have a price tag?

(4) Did it reduce or increase the number of innovations and technical development?

(5) Was the benefit at the expense of a good relation with own staff or with the other party? Short term and long term?

(6) What modifications should be considered for future contracts, based on the experience?

Although the method as well as the questions were applied on one particular case, the lessons learned should hopefully apply well beyond it, e.g., to highway and housing maintenance and construction or anywhere where continuous quality is important to the owner.

Method

Mathematical curves were fitted to the delay and error statistics and compared with diagrams and regression analysis tools pro-

J. Constr. Eng. Manage., 2008, 134(4): 265-272

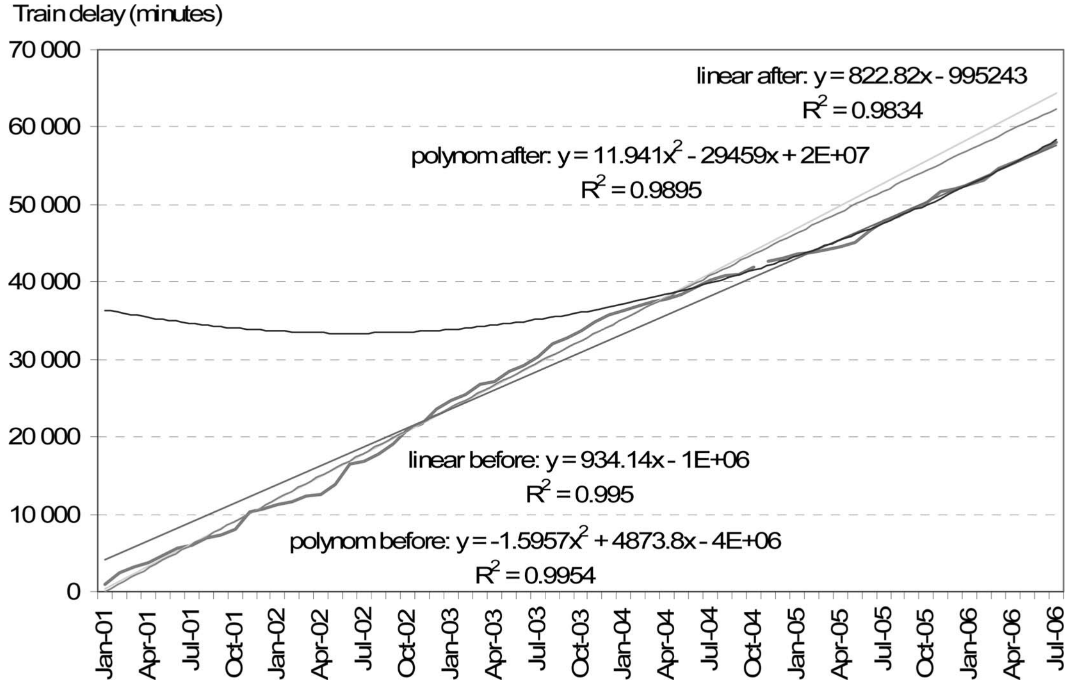

Fig. 2. Linear, logarithmic, polynomial of second, third, and fourth order, power and exponential function types were tested. The highest R^2 equations, the linear and the polynomial of the second order, are shown superimposed on the actual total delay minutes.

vided by the Excel 2006 software. The difference of the slopes between the preincentive and incentive periods would reflect the quality change. Statisticians normally recommend to select the most probable function type and calculate its goodness-of-fit to the data, rather than to test several function types on the data without preconceptions (Crown 1998), but to gain additional knowledge, push the scientific frames of reference, and to find the spread and precision of the result depending on assumption, that recommendation was violated as a first step of the research process.

To evaluate whether the improved quality had a price tag, the lump sum price of the contract plus the incentive payment and any unforeseen costs was compared to the previous costs to maintain the same rail network. To evaluate whether innovations and technical development had gained or suffered, owners and contractors were asked to mention examples of any innovations. To evaluate if quality had been gained at the expense of a good relation with the other party or with their own staff (due to pushing them harder without economic compensation for example), owner and contractor high level staff were asked to comment on these issues, which was done in a session with both parties present.

Results

The result can be expressed as the change in quality, but any such change should also be critically compared to any change of costs in the short and long term. Costs do not only include short-term financial costs, but also the long-term financial implications such as the technical development and a good relation with their own staff and the other party.

Three quality parameters were included in the contract and two were analyzed; minutes of delay the and number of errors per month. For the third parameter, train vibration, the contracting parties had not reached full agreement regarding how segments were to be weighted and added (Banverket, personal communication, January 11, 2007), why it was excluded from this study and may be excluded from the real contract as well.

Train Delay

The curve Jan-01 to Oct-04 in Fig. 1 appears to be quite linear, but not perfectly. The linear function type assumption was challenged to see what would happen, Fig. 2. The effort to include some nonlinear models was not as cumbersome as it would have been in the old days. The spreadsheet software offered the following model types: linear, logarithmic polynomial, potential, exponential, and gliding average. Out of these, only the polynomial of the second order turned out to deliver a correlation that under certain conditions was better than the linear curve's R^2 and adjusted R^2.

The equations in Fig. 2 deliver a dependent variable (y) expressed as cumulative delay minutes since January 1, 2001 with the independent variable (x) being the number of months. R^2 is just slightly higher for the polynomial curves than the linear ones.

The slopes were now compared; the $\Delta y/\Delta x$ after with the $\Delta y/\Delta x$ before the incentive was introduced on November 1, 2004. Only the slope coefficients before the independent variable need to be compared, as the intercepts depend on the starting date of the statistics, which becomes the x origin. Therefore no effort was made at this stage to change the January 1, 1900 default setting in Excel for the x origin, which has resulted in unnecessarily high intercepts in the trend line equations.

When testing only the finalists after the regression analysis, four combinations were possible. The best linear equation for the preincentive period followed by the best polynomial equation for the incentive period (LP), linear followed by linear (LL), polynomial followed by linear (PL), and polynomial followed by poly-

Table 1. The Average Incentive Effect per Month. The Values 800 and 820 Are Derived by Ocular Reading of the Polynomial Curves in Fig. 2.

	Linear after ($y=823x$)	Polynomial after (Fig. 2, top right)
Linear before: ($y=934x$)	934−823=109 min 1−823/934=12% (LL)	934−800=134 min 1−800/934=14% (LP)
Polynomial before: (Fig. 2, bottom left)	820−823=−3 min 1−820/823=−0.4% (PL)	820−800=20 min 1−800/820=2.5% (PP)

nomial (PP). The average slope for each curve within the period was used in the comparison. The improvement of $\Delta y/\Delta x$, i.e.

$$\frac{\Delta y_{\mathrm{before}}/\Delta x_{\mathrm{before}}}{\Delta y_{\mathrm{after}}/\Delta x_{\mathrm{after}}} \tag{3}$$

for each of the four combinations are shown in Table 1. (The polynomial curves' average slopes in the studied period have been ocularly estimated from Fig. 2.)

LP has the highest value, 14% (Table 1, Row 2, Column 3), reflecting the immediate improvement in the first months of the incentive contract (Fig. 1). PL returns the lowest value, −0.4 % (Table 1, Row 3, Column 2), reflecting that some months in the incentive period had more delay than the average in the preincentive period. The average of the four values in (Table 1) LL, LP, PL, and PP is 7%. In honor of the statisticians' linear recommendations, let us assign the LL value 12% (Table 1, Row 2, Column 2) a higher rate. The LP and PP curves risk to fit worse if extrapolated into the far future. Last word to the statisticians? We are back to where we left them, but thanks to our disobedience, we are a lot wiser now than when we left. The PL, LP, and PP values are useful to approximate a precision for the selected LL. The real value may be as low as −0.4 % (PL) and as high as 14% (LP). Weighting, adding, and dividing these considerations somewhat subjectively, they can be summed up to support that the incentive effect on delay is approximately 10% ±5%, i.e., 10% with one valid digit.

A magnification of the benefits of the incentives over time is to use the derivate, the delay each month, Fig. 3. It gives a more detailed picture and the absolute value of the coefficient (the sign of the slope) is more informative. This time, resetting x origin to January 1, 2000 has lowered the trend line intercepts in Fig. 3 compared to if January 1, 1900 had been used as in Fig. 2.

The incentive effects are magnified, if errors are zeroed at the beginning of each month, i.e., not accumulated, Fig. 3. In Fig. 3 the accumulated delay is represented by the area under the curves.

Number of Errors

The reward for delay improvement was 60%, error improvement 30% (and vibration 10%). As the delay is rather an effect of the errors than vice versa, the contractor can be assumed to have better influence on the errors, leading to an expectancy of a better incentive effect for the error variable. On the other hand, the higher weight, 60%, for the train delay could favor train delay fixes at the expense of errors fixes under particular circumstances. The result was that the errors dropped more than the delay, according to the first expectancy. Fig. 4 shows how accumulated errors developed before and after the incentive was introduced.

An analogous mathematical approach as for the train delay, fitting the trend lines and their equations with the spreadsheet software in a diagram resembling Fig. 2, delivered the following values: LL=24%, LP=26%, PL=5%, and PP=7%, which was weighted and summed to 20% ±5%, i.e., 20% with one valid digit. The errors per month, the derivative, is presented in Fig. 5.

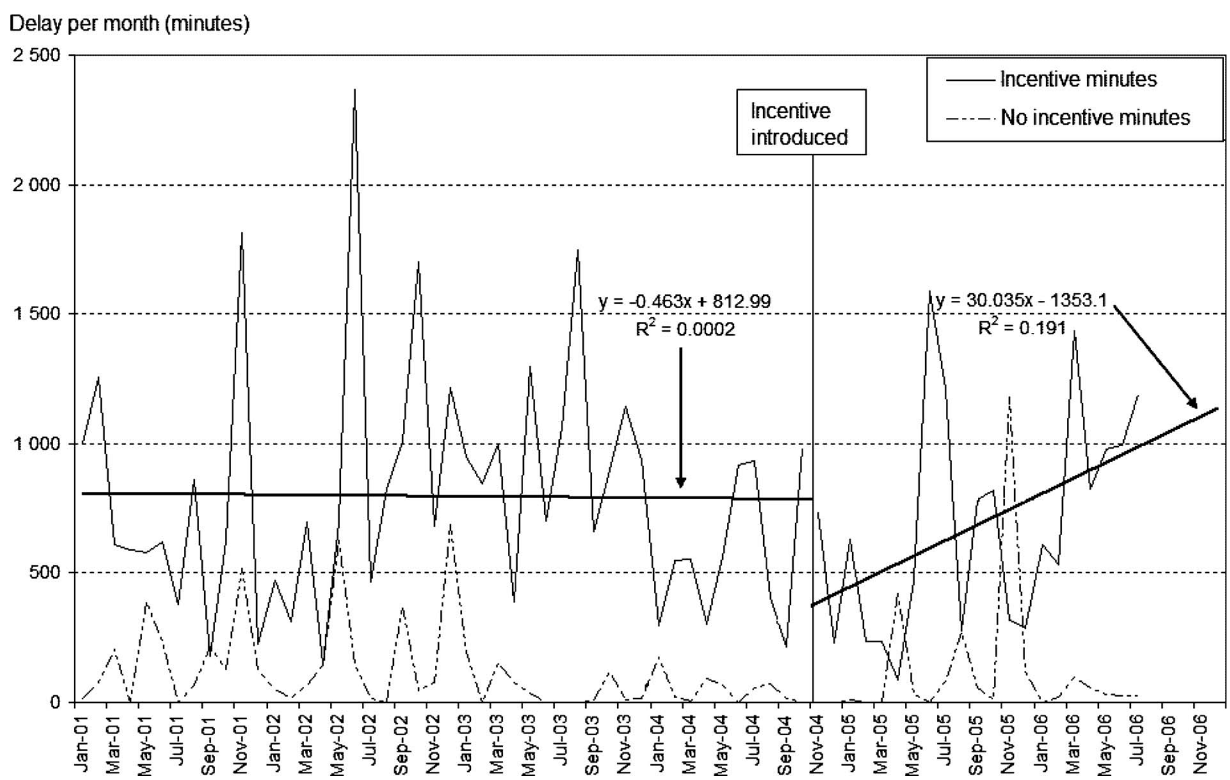

Fig. 3. Delay per month for the segments subject to the incentive contract. The bold curves are the linear trends for the incentive minutes.

J. Constr. Eng. Manage., 2008, 134(4): 265-272

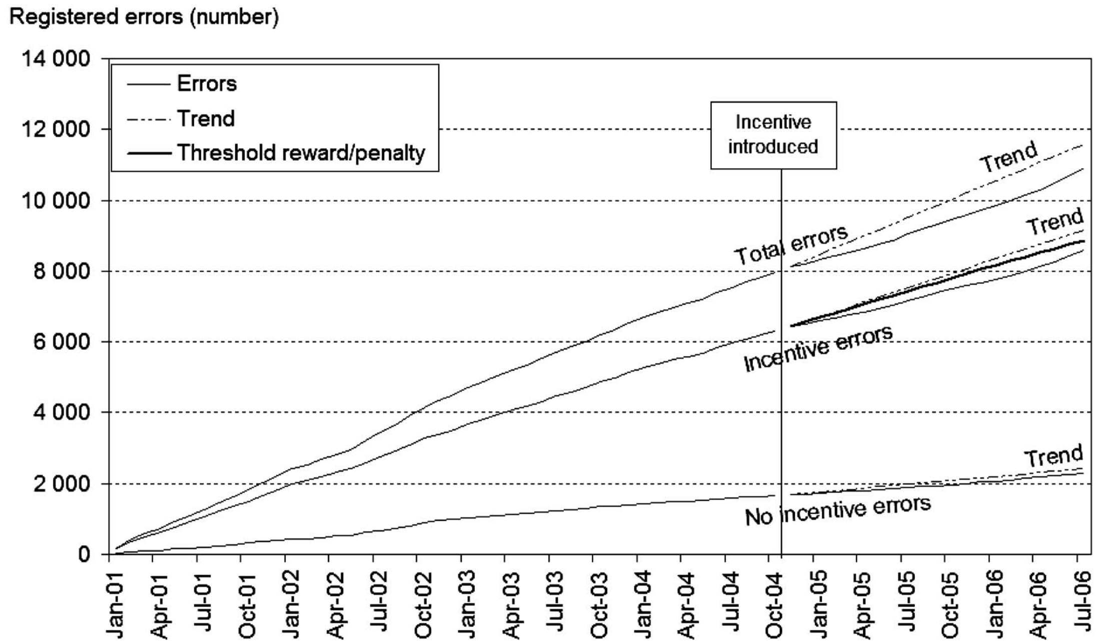

Fig. 4. Accumulated number of errors for the segments subject to the incentive contract

Quality Correction Factors

It seems that the effect of the incentive decreases with time. The almost zero slope for the trend line to the left in Fig. 3 shows how constant the number of delay minutes was. The drop of delay immediately after introduction of the incentive followed by a relative increase has resulted in an upwards trend line to the right. If comparing July 2006 with July 2004 (Fig. 5), the error rate is higher in 2006. Are all external factors under control, or should the result be corrected with some factor to compensate for external variation of the conditions after all?

A more detailed analysis of the frequency of each error code was compared with weather and particular circumstances. Two explanations were supported by the data. In March 2006, bad winter storms caused a lot of delay at Stockholm Central, the bottleneck of the Swedish railway system. This weather may not have been local to Stockholm, but its effect was accentuated there with its turnouts and traffic. Even if weather was included and treated as a "no-excuse" in the contract, its variation matters for the scientific analysis. If removing the delay and errors in March 2006, the statistics for the contract type effects look a lot better.

A second bias found was that the error registration principles

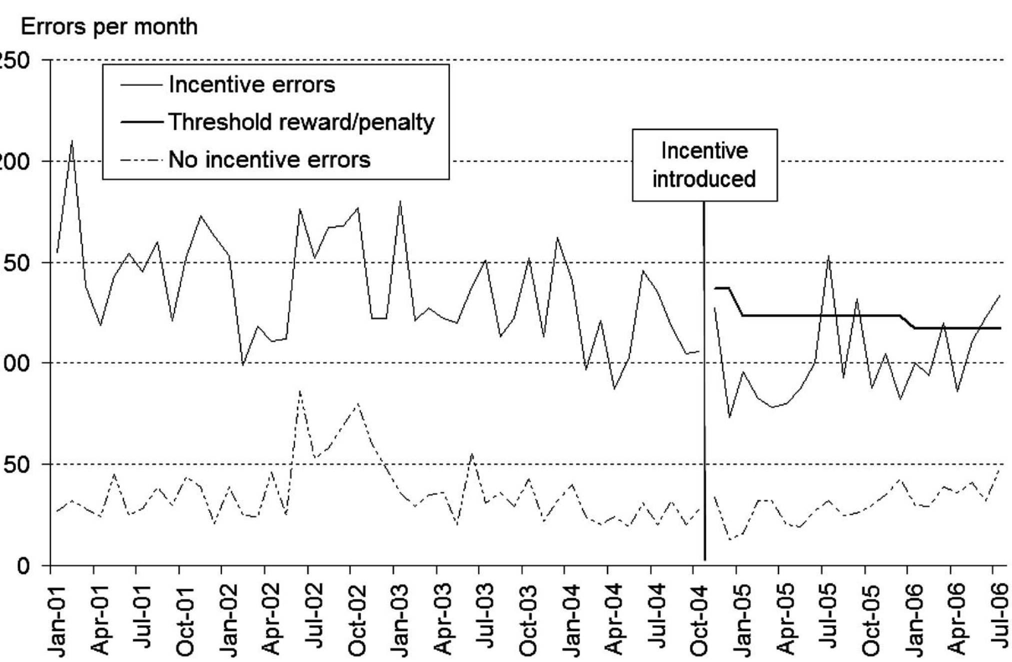

Fig. 5. Errors per month for the segments subject to the incentive contract. The bold curve indicates the threshold for reward/penalty to the contractor.

J. Constr. Eng. Manage., 2008, 134(4): 265-272

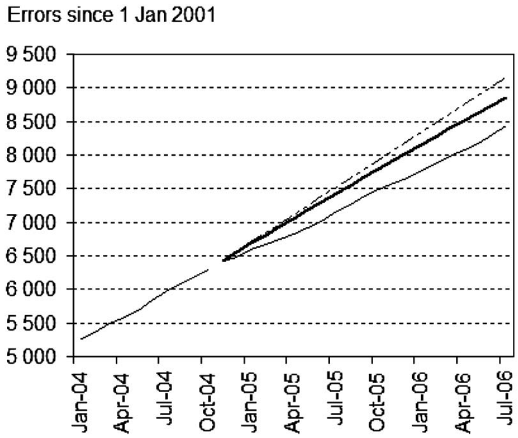

Fig. 6. Cumulative errors corrected for the increased error-registration intensity during 2006 with the erase method (bottom curve). The bottom curve now retains its lower slope compared to the preincentive level (top curve). The distance between the bottom curve and the reward/penalty threshold (middle curve) will determine the bonus to the contractor.

changed between the compared periods. Errors that resolved without any service ("error-disappeared" registrations), were registered in Tför, the delay registration information system, but not consistently transferred to Ofelia, the error registration information system. Increased efforts to couple such delays to a cause, resulted in an increased number of errors in Ofelia in 2006 compared to 2001–2005. This increased coupling creates an unfair change of the conditions for the contractor in this contract. As the contractual delay minutes are based on Tför, unaffected by the new practices, the increased coupling does not matter for the *delay* part but only for the *error* part. The owner should compensate the contractor for the effect and, in any case, science should compensate for it. This study must also compensate for it also regarding the delay, as Ofelia was used also for the *delay* calculations. That method simplification turned out to create an unexpected bias.

Correction Considerations

Correction for the change of circumstances can be made in two main ways. One is to remove all error-disappeared from all the data. We will call this the *erase-forwards* method. The other is to add fictive errors-disappeared to the period of the past when they were underrepresented. We will call this the *extrapolate-backwards* method. *Erase-forward* is easier and mathematically better, as circumstances that prevailed during the relatively longer period receive a higher weight (60 months of data to correct the following 6 months should be better than 6 months of data to correct 60 months). Fig. 6 shows the curve when all error-disappeared for the whole period 2001–2006 are erased, which is the easiest way to carry out the erase-forward principle in this case to create comparability. The slope increase June–July 2006 is now gone, or at least less marked.

There are good reasons to use the second method, too. If we extrapolate backwards now, no erase-forward corrective operation will be needed next time (at least as far as this problem is concerned, but leaving room for new problems to deal with is an additional good reason). With the second method, however, not even six months but only three months of data are available to be extrapolated backwards, as the increased error-registration in Ofelia was *gradually* introduced in the first three months of 2006, most easily handled by removing that period of transition from the operations. The extrapolate-backwards method then consists of calculating the fraction of errors-disappeared during the last three months only, April–June 2006 when the new practices were fully in place, and add the same fraction to the pre-2006 data. The number of fictive errors to be added to the data before 2006 then becomes

$$\text{(Errors per month 2001 – 2005)} \times \left(\frac{\text{Error-disappeared April – June 2006}}{\text{Error total April – June 2006}}\right) \qquad (4)$$

When we have extrapolated backwards once, future data can be used as they are without correction. In a few years time, say 2009, 2001–2005 will be history and need not be considered anymore.

The choice of method also has an incentive dimension. The contractor could have an influence on the remedy of an error, and there should not be a financial gain in pretending that an error disappeared without any remedial action. Such an incentive emerges if the contractor knows beforehand that the erase-forward method will be used when assessing the performance.

Since for the time being, the first method is both easier and better, it was used this time. If later continuing this study with more data for the time after 2006, the calculations may have to be redone with the second method.

Price and Other Costs

The positive effect on quality should be critically compared with the change of costs. Did the quality change have a price tag? Unforeseen costs and the incentive costs were added to the offered lump sum price of the new contract. This total was compared with the previous costs to maintain the same rail network. The result was that the old contract was more expensive even after these additions had been made (Banverket, private communication, January 2007). Hence, the quality improvement appears to have been achieved at no or negative cost.

Did the contract affect innovativeness or technical development? Both owner and contractor thought that long-term development had gained or would gain from the contract form, but they could not mention any particular innovations or technical development. A cautious interpretation of this input is that neither a positive nor a negative effect has been observed.

Did the quality increase happen at the expense of a good relation with one's own staff or the other party? The owner's staff was worried and reluctant to release responsibility, but accepted it. Instead, more of their time would now be released to determine what the contractor should achieve rather than how to achieve it. The contractor superiors expressed that they were pleased with the increased responsibility, speaking on behalf of their staff, but that motivation had suffered during months when the delay and error statistics looked bad, in particular when due to bad weather, sabotage, and other external factors. The formal and final performance November 1, 2004 to December 31, 2005, according to the owner's notes, was that 321 delay hours was the target and 347 was delivered. For errors, 1,555 was the target and 1,434 were delivered. (The third parameter, vibration, was neutralized and set to 1.)

J. Constr. Eng. Manage., 2008, 134(4): 265-272

Analysis

The remote continuous monitoring by a third party relative to the two contracting parties probably helped the result and avoided sampling errors, which can become an issue of controversy between owner and contractor (Schexnayder and Ohrn 1997). Even when kept as simple as possible, incentive contracts can become complicated. Financial implications and the related interests of certain interpretations add to the complexity problem. In this case, the experimental, partnering situation with two bodies at the top level united into the same legal unit most probably helped to resolve problems along the way, which might have ended up in court if the contractor had been an independent, private organization. The propensity for misunderstandings will be higher when the contract form is used in a true private, outsourcing setting. Feelings of unfairness aroused among contractor staff as their bonus was affected by weather and circumstances out of their control. This may have lowered their performance and motivation for certain months of the contract. Still, both contractor and owner express that the net effect was positive and want to continue with the contract form. The increasing curve in the summer of 2005 and March to June 2006 proved to be fully explained by extraordinary thunderstorms, snowstorms, and the unveiled "errors-disappeared."

What had the owner expected? The owner maximizes profit if the contract type maintains or improves the condition of the asset at the lowest price possible in the long run. The incentives and their effects are not quite as high but of similar magnitude as the effect on delivery time obtained by incentives in highway Design-Build projects (Shr et al. 2004; Shr and Chen 2004).

Calculating the Incentive to Be Paid

With the incentive formula used, the official owner figures regarding the outcome are to be weighted together as follows; $0.6\ (321/347)+0.3\ (1{,}555/1{,}434)+0.1\ (1/1)=0.983$. 98.3% is 1.7% short of the target (100%) which should normally lead to an equivalent negative correction of the amount paid out to the contractor. Some hard feelings when this is effected are likely. If looking at the matter from a scientific viewpoint, disregarding how the owner and contractor distribute the money between them, a net gain of quality was achieved by the incentive contract, at least as it seems so far. Since the targets had already anticipated a 10% reduction, the quality improvement is $0.6\ (321\times 1.1/347)+0.3\ (1{,}555\times 1.1/1{,}434)+0.1=1.07$, a 7.1% quality improvement. The targeted improvement was $0.6\times 1.1+0.3\times 1.1+0.1=1.09$, i.e., 9%.

The financial outcome looks like a good deal for the owner in the short run, as quality improved at a negative cost (not only the incentive result, but also the lump sum was less than previous costs for the area). In the long term, however, leaving the contractor with a good feeling of achievement might have been better. An evil spiral regarding the development of the lump sum price and morals could arise, if contractors feel the requirements are arbitrary, unreasonable, or beyond their control. Correcting this afterwards with a gratuity to the contractor for the past contract will, however, probably make things worse. Grounded reason for a correction, such as compensating the error registration bias, is better. The winter storm compensation is also possible, but creates borderline problems for the future; what weather to consider normal and what weather to exclude, and how to exclude it, from the basis of the incentive. A no-excuse philosophy is probably better in the long run for such biases that are probable to recur.

Recommended Modifications

Two scenarios for the future are possible, one being positive and one negative. The positive one is that the performance criteria were well selected and the results so far relevant for true quality also long term. The increased performance may make it more difficult to improve the asset further, advocating that the threshold for bonus is lowered accordingly to maintain the incentives arising from having reasonable chances to reach and exceed the bonus threshold by the contractor. Another scenario is that the condition of the asset deteriorates by factors not captured by the monitoring, because the selected performance criteria did not measure true performance in a sustainable way, or that the contractor can postpone necessary actions to the next contract period in an unanticipated way. If so, more profound adaptations to the contract than adjusting the reward/penalty threshold will be needed. In addition the formula may need to be developed to counterbalance game theory arising when the private sector is invited to play the game. A hypothesis is that these and other dynamic factors explain some of the problems with the British rail maintenance outsourcing initiative, which was stopped in 2003/2004 after a series of accidents.

The 10% improvement target should rather have been 5% or maybe even 0% (as was the case with the third vibration parameter that was later excluded). A 0% threshold is easier to handle and less vulnerable to game theory as it does not require adjustment with changed conditions and previous achievements. The contractor has the same economic incentive with a low as a high threshold, making the positive feelings arising from having beaten the target favor a low threshold. The threshold level only matters for the owner short term, or not at all, if assuming that contractors will offer a higher lump sum price to compensate a too high threshold. Another advantage with a 0% threshold is that it can be used also for nonquantifiable parameters.

Conclusion and Future Research

A keep-it-simple method was tested to quantify the incentive effect on quality in a performance-based rail maintenance contract. Delay was reduced 10±5% and the number of errors by 20±5% at no additional total cost for the owner. Costs were even slightly reduced compared to the period 2001–2004. No examples of machinery or method development, or other tangible effect on technical development, could be recalled by the owner and contractor representatives when asked why the productivity and quality improvements seem to be attributable to organization and human factors rather than technical development. The contractor staff was generally pleased but the atmosphere had suffered during times when the measured performance was low due to factors out of their control. The owner's target was a 9% improvement when the three parameters were weighted together. As a 7.1% improvement was achieved, the lump sum price will be reduced by the difference, unless the parties agree otherwise. To maintain long-term contractor motivation, the owner is recommended to look through if the causes are appropriate and consider a lower reward threshold in future contracts.

When quality is measured, subjective statements from those directly involved have often been treated like objective data, a

J. Constr. Eng. Manage., 2008, 134(4): 265-272

debatable approximation of verity. More studies where quality is objectively measured and quantified in physical or economic terms are welcome to accelerate the science around performance-based contracts, not the least to better separate it from politics. Further research to fine-tune profit-sharing thresholds, bases for incentives, and incentive levers is also looked forward to.

With minor modifications and clarifications the owner intends to use similar contract provisions for its in-house contracting, as in the case studied, but also when outsourcing to private companies. With other performance indicators, the elaborated method and lessons learned should be applicable also for other sectors, where a contractor assuring a specified service level during a period of time is paid a bonus depending on degree of fulfillment.

References

Aguilar, M., Estrada, G. C., and Myers, J. J. (2005). "Decision-support quantitative models for valuing incentives in performance based contracts." *MBA—Rep.*, Naval Postgraduate School, Monterey, Calif.

Andersson, M., Murray, M., Ferreira, L., and Lake, N. (2004). "Collection and use of railway track performance and maintenance data." *Conf. on Railway Engineering*, Sec. 2.3, second paragraph, Darwin, Australia.

Baumol, W. J. (1975). "Payment by performance in rail passenger transportation: An innovation in Amtrak's operations." *Bell J. Econom.*, 6(1), 281–298.

Crown, W. H. (1998). *Statistical models for the social and behavioral sciences, multiple regression and limited-dependent variable models*, Praeger, Westport, Conn.

Milling, P. M., Maier, F. H., and Hasenpusch, J. (1998). "Total productive maintenance: An international analysis of implementation and performance." *Proc., Industry Seminar*, Mannheim Univ., Germany, ⟨http://is.bwl.uni-mannheim.de/Forschung/Publikationen/tpm.pdf⟩ (Aug. 15, 2007).

Schexnayder, C., and Ohrn, L. G. (1997). "Highway specifications—Quality versus pay." *J. Constr. Eng. Manage.*, 123(4), 437–443.

Shr, J.-F., and Chen, W. T. (2004). "Setting maximum incentive for incentive/disincentive contracts for highway projects." *J. Constr. Eng. Manage.*, 130(1), 84–93.

Shr, J.-F., Ran, B., and Sung, C. W. (2004). "Method to determine minimum contract bid for A+B+I/D highway projects." *J. Constr. Eng. Manage.*, 130(4), 509–516.

Stenbeck, T. (2007). *Promoting innovation in transportation infrastructure maintenance—Incentives, contracting and performance-based specifications*, KTH Royal Institute of Technology, Stockholm, Sweden.

Tarakci, H., Tang, K., Moskowitz, H., and Plante R. (2006). "Incentive maintenance outsourcing contracts for channel coordination and improvement." *IIE Trans.*, 38(8), 619–632.

Vickerman, R. W. (2004). "Maintenance incentives under different infrastructure regimes." *Util. Policy*, 12(4), 315–322.

Whitehouse, L. (2003). "Railtrack is dead—Long live network rail? Nationalization under the third way." *J. Law Soc.*, 30(2), 217–235.

J. Constr. Eng. Manage., 2008, 134(4): 265-272