

Cost benchmarking of Network Rail's maintenance expenditure 2024 to 2025 - Technical report on our modelling approach for cost benchmarking

2. Analytical approach for maintenance expenditure

2.1 This section explains the data used in our analysis, our model specifications and statistical results. We have adjusted historic expenditure data used in our analysis for inflation using the Consumer Prices Index (CPI) so that expenditure is in constant 2024 to 2025 prices.

Approach for our maintenance cost benchmarking analysis

2.2 Our cost benchmarking analysis estimates maintenance expenditure as a function of a range of explanatory factors. Maintenance expenditure is the dependent variable, meaning it is the outcome we are seeking to explain. The independent explanatory variables are factors that may influence this outcome.

Dependent variable

2.3 Regional maintenance expenditure has been taken from Network Rail's regulatory financial statements. Network Rail reported route level information for years 2010 to 2011 and 2018 to 2019,

which we have combined to provide a continuous time series for regional data. The change from routes to regions is explained here.

2.4 MDU maintenance expenditure uses data for financial years 2014 to 2015 and 2024 to 2025 for the MDUs into which Network Rail is divided for business planning and delivery. This excludes centrally managed expenditure (covering activities such as structures examination, major items of maintenance plant and other HQ managed activities).

Independent variables

2.5 We selected independent variables based on engineering expectations about the main drivers of maintenance expenditure. These covered network scale, usage and complexity. We then examined which were the best variables for each of these cost drivers.

2.6 Network Rail provides data for these variables through the annual cost benchmarking data collection exercise.

2.7 For regional maintenance expenditure, we selected three independent variables:

- Track length as a scale variable. This is the total track in km for each region.
- Traffic density as a usage variable. This is the total distance travelled by passenger and freight trains during the year in each region.
- Proportion of track classified as criticality 1 and 2 as a complexity variable. Network Rail categorises track into five bands of 'criticality', with criticality 1 as having the highest impact on the network (for example, urban mainline), and criticality 5 having the lowest (for example, rural branch line).

2.8 We also included time dummy variables in our regional model. These variables take the form of zero or one for individual years. They are included to capture changes in costs not accounted for by the other explanatory variables over time. Without time dummies, expenditure changes caused by common year-specific factors could be incorrectly attributed to the modelled cost drivers.

2.9 For MDU maintenance expenditure, we selected two additional complexity variables which improved model fit:

- proportion of electrified track; and
- switches and crossings density.

2.10 In our MDU model, we included dummy variables for the Covid-19 year 2020-21 and a post-Covid-19 dummy for years 2021 to 2022 and 2024 to 2025. We included a year time trend in the MDU model to capture gradual changes in maintenance expenditure over time that are not explained by the other explanatory variables.

2.11 Table 1 summarises the expected direction of the relationship between our independent variables and Covid-19 dummies, and maintenance expenditure.

Table 1: Independent and dummy variables used in the models and their expected relationship

Model(s)	Variable	Expected relationship	Comments
Region and MDU	Track-km (length of track)	Positive	A larger network requires more maintenance.
Region and MDU	Traffic density (passenger train-km + freight train-km/track-km)	Positive	Higher traffic volumes cause more wear and tear to the network.
Region and MDU	Proportion of track criticality 1 and 2 (criticality 1 + criticality 2/ track-km)	Positive	Track is split into five bands of criticality according to the impact that a loss of service would have on the network, with 1 having the greatest impact and 5 the lowest. A network with higher proportion of track

Model(s)	Variable	Expected relationship	Comments
			criticality 1 and 2 is likely to be more physically complex. For example, there will likely be more switches and crossings that require frequent maintenance.
MDU	Switches and crossings (S&C) density (number of S&C/track-km)	Positive	A network with more switches and crossings per track-km is more complex and therefore requires more costly maintenance.
MDU	Proportion of electrified track (electrified track-km/ track-km)	Positive	The presence of electricity and of power supply infrastructure is likely to increase the complexity of track maintenance work. As a pure effect, we would expect this to be positive. However, we note electrified track may be associated with newer assets so there may be some ambiguity over the sign.

Model(s)	Variable	Expected relationship	Comments
Region	Year (dummy variable for each year)	N/A	Year dummy variables capture year-specific factors affecting maintenance expenditure that are common across all regions and are not explained by observable cost drivers.
MDU	Time trend	N/A	The time trend captures gradual changes in maintenance expenditure over time that are common across MDUs and are not explained by observable cost drivers.
MDU	Covid-19 dummy variable (applies to 2020-21)	N/A	Captures the additional effect of Covid-19 beyond the underlying time trend.
MDU	Post-Covid-19 dummy variable (applies to 2021-22 to 2024-25)	N/A	Captures the additional effect of post-Covid-19 factors beyond the underlying time trend.

- 1. Where one km of double-tracked route counts as two track-km.
- 2. We disaggregated traffic density into passenger and freight density in the MDU model because it improved model fit.

Our maintenance expenditure models for 2024 to 2025

2.12 Our models are estimated using natural logarithms, which enables the results to be interpreted in percentage terms. This is in line with past academic and regulatory cost function literature.

Our regional maintenance expenditure model for 2024 to 2025:

ln (Regional Maintenance Expenditure)

$$= \beta_0 + \beta_1 \ln (\text{Track km}) + \beta_2 \ln (\text{Traffic density})$$
$$+ \beta_3 \text{Proportion of track Criticality 1 and 2} + \beta_4 (\text{Year dummies}) + e$$

2.13 The ‘β’ symbol represents the expected percentage change in the dependent variable (maintenance expenditure in this instance) for a 1% change in a specific independent variable, while holding the other variables constant. This is also referred to as a coefficient estimate.

2.14 The residual (‘e’) captures the unexplained variation in maintenance expenditure. This can include inefficiency, omitted variables and other random factors.

Table 2: Random effects coefficient estimates for regional maintenance expenditure

Variable	Regional coefficient (β)
Track-km	1.00***

Variable	Regional coefficient (β)
Traffic density	0.86***
Proportion of track criticality 1 and 2	0.47*
Year effects (included in the model to control for general time trends but are not reported in the table for brevity. Dummies for Covid-19 years are not needed as they would duplicate the year effects.)	Yes
Constant	-11.43***
Number of observations	75
R^2	0.95

*** Statistically significant at the 99% confidence level

** Statistically significant at the 95% confidence level

* Statistically significant at the 90% confidence level

2.15 As shown in Table 2, track-km and traffic density were statistically significant at the 99% confidence level (denoted by ***). This means there is less than a 1% probability that the observed relationship is due to random chance, providing strong evidence that these variables have a relationship with maintenance expenditure. The proportion of track criticality 1 and 2 is statistically significant at the 90% confidence level (denoted by *). Our findings are consistent with our expectations outlined in Table 1.

2.16 Our model included 75 observations, which corresponds to five regions over 15 years. The R² for our regional model is 0.95. This means that the model explains 95% of the observed variation in regional maintenance expenditure.

Random effects and OLS modelling approaches

2.17 We considered three different approaches to estimating econometric relationships: Ordinary Least Squares (OLS), random effects and fixed effects.

2.18 Both OLS and random effects use differences between regions (or MDUs) and changes within regions (or MDUs) over time to estimate the relationship between maintenance expenditure and its cost drivers. OLS models pool these sources of information, whereas random effects models give a different weight to each depending on the structure of the data. This is important for benchmarking as our primary purpose is to compare expenditure across regions (or MDUs) while accounting for changes over time. Fixed effect models only use changes within each region (or MDU) over time. Fixed effects does not take account of differences across regions (or MDUs). We have therefore preferred OLS and random effects models as they provide greater explanatory power.

2.19 We selected random effects for our regional maintenance expenditure model because it produced plausible coefficient estimates and benchmarking results that were broadly consistent with OLS. Random effects models are well suited to panel datasets, where observations are available for multiple regions across multiple years.

2.20 For the MDU analysis, we selected OLS after comparing its results with panel model alternatives. The random effects models produced materially different expenditure ratios for some MDUs. Some fixed effects models produced implausible coefficient estimates. The OLS model produced the most plausible and stable results overall, based on the coefficient estimates, benchmarking results and our wider understanding of relative maintenance expenditure across MDUs. We intend to investigate this further next year.

Our MDU maintenance model for 2024 to 2025

$\ln(\text{MDU Maintenance Expenditure})$

$$= \beta_0 + \beta_1 \ln(\text{Track km})$$

$$\begin{aligned}
& + \beta_2 \ln \diamond (\text{Passenger density}) + \beta_3 \ln (\text{Freight density}) + \beta_4 \ln (\text{Switches and Crossings density}) \\
& + \beta_5 \text{Proportion of track electrification} \diamond \\
& + \beta_6 \text{Proportion of track Criticality 1 and 2} + \beta_7 (\text{Year time trend}) \\
& + \beta_8 (\text{Dummy for Covid year}) + \beta_9 (\text{Dummy for post COVID years}) + e
\end{aligned}$$

Table 3: OLS coefficient estimates for MDU maintenance expenditure

Variable	MDU coefficient (β)
Track-km	0.84***
Passenger traffic density	0.41***
Freight traffic density	0.08*
Switches and crossings density	0.24*
Proportion of electrified track	0.30**
Proportion of track criticality 1 & 2	0.03
Year time trend	0.01**

Variable	MDU coefficient (β)
Dummy for 2020-21 (deviation from the annual growth rate due to Covid-19)	0.18***
Dummy for 2021-22 to 2024-25 (deviation from the annual growth rate due to post-Covid-19)	0.10***
Constant	-6.69***
Number of observations	385
R^2	0.55

*** Statistically significant at the 99% confidence level

** Statistically significant at the 95% confidence level

* Statistically significant at the 90% confidence level

2.21 The coefficient estimates for our MDU model are shown in Table 3. The coefficient on the proportion of track classified as criticality 1 and 2 is positive, as expected, but is not statistically significant at conventional confidence levels. We have retained this variable because it is our most direct measure of underlying network complexity and there are strong theoretical reasons to expect it to influence maintenance expenditure. The variable is also correlated with other measures of network complexity and usage. Removing the variable did not materially affect the coefficients, expenditure ratios or rankings. However, it would reduce the extent to which differences in underlying complexity are explicitly captured within the benchmarking framework.

2.22 The dummy variables for 2020 to 2021 and the post-Covid-19 years 2021 to 2022 and 2024 to

2025 are positive and statistically significant, indicating that maintenance expenditure was higher than expected during and following the Covid-19 pandemic. Costs are estimated to have been around 20% higher in 2020 to 2021 and around 10% higher post-2021 to 2022 relative to underlying trends. This is due to, among other factors, a significant fall in usage drivers (such as passenger traffic density) without a corresponding fall in expenditure. During the Covid-19 pandemic, the rail industry also experienced reduced productivity (this is consistent with our findings in our report on rail industry productivity for 2024 to 2025), operational disruption, and a period of catch-up work.

2.23 There are 385 observations, which corresponds to 35 MDUs over 11 years. To create a consistent panel over time following boundary changes, the estimation dataset combines some MDUs into three merged units and therefore contains 35 comparable units.

2.24 The R^2 for our MDU model is 0.55. This is lower than the regional model. This reflects the larger variation in expenditure between different MDUs compared to regions, which can reflect lumpiness in expenditure and unobserved MDU factors, such as hosting. The MDU model covers expenditure attributed directly to MDUs, which accounts for 57% of maintenance expenditure, whereas the regional model covers 95% of maintenance expenditure.

2.25 As part of our analysis, we compared the preferred model against alternative model specifications that also demonstrated good statistical performance and produced plausible benchmarking results. These specifications included different combinations of complexity variables and input prices.

2.26 The results were broadly stable across variable specifications. However, random effects and fixed effects models produced materially different expenditure ratios and rankings for some MDUs. We selected OLS as our preferred approach for the MDU analysis.

2.27 We explored a range of OLS models. Across these models, the MDUs with the highest and lowest expenditure relative to model predictions remained largely unchanged, while most other MDUs experienced only modest movements in ranking and expenditure ratios. For example, Doncaster remained among the lowest-spending MDUs relative to model predictions, while Sandwell and Dudley remained among the highest.

2.28 Similar patterns were observed when comparing results over the previous five years, from 2020 to 2021 and 2024 to 2025, with the relative position of most MDUs remaining broadly unchanged. This provides confidence that the benchmarking results are not unduly driven by the

specific model chosen or the year selected for assessment.

2.29 We used some different variables in the MDU and regional models because the expenditure covered by each dataset differs. The MDU dataset covers 57% of total maintenance expenditure. A further 38% is managed by regional central teams rather than MDUs, meaning that the regional dataset covers 95% of total maintenance expenditure. Some spending decisions made at regional level are therefore not reflected in MDU expenditure. The larger MDU dataset also allows additional complexity variables, including electrification and switches and crossings density, to be estimated reliably. These variables improved model performance when tested.

Additional testing

2.30 We investigated a range of other possible independent variables in developing the final models. These included the proportion of electrified track, density of switches and crossings, number of possession days, average Network Rail salary, average rainfall, average track length, and lagged maintenance expenditure.

2.31 We excluded variables from the model that were either not statistically significant or were highly correlated with the existing variables in the model. Including highly correlated variables can make coefficient estimates less stable and increase uncertainty around the relative importance of individual cost drivers. Excluding these variables helps ensure that benchmarking results are driven by meaningful underlying relationships rather than statistical artefacts.

2.32 Testing indicated that including these variables did not materially affect the ratio of predicted to actual expenditure or the ranking of MDUs. We therefore excluded these variables from the model.

2.33 We also tested both combined and separate passenger and freight traffic density measures. For the regional model, a combined traffic density variable was adopted. However, at MDU level separate passenger and freight density variables improved model fit.

2.34 We also compared our final models with different modelling approaches, including OLS, random effects and fixed effects, and tested their stability by excluding the first and last years of data. The broad consistency of the regional results provides confidence that our findings are not unduly affected by the choice of model or time period.

2.35 For MDUs, results were broadly stable across OLS models. However, as explained in paragraph

2.20, random and fixed effects models produced materially different expenditure ratios and rankings. These results were considered alongside the plausibility of the coefficient estimates and our wider understanding of relative maintenance expenditure when selecting the preferred models. This is summarised in Table 4.

Table 4: Model specification assessment and outcomes

Model	Assessment	Outcome
Regional random effects	Uses both changes within regions over time and differences between regions to estimate the model parameters. It produced plausible coefficient estimates and a high model fit (R^2 of 0.95), with benchmarking results that were broadly consistent with OLS.	Selected because it produced plausible coefficient estimates and broadly stable benchmarking comparisons, while taking account of the panel structure of the data.
Regional fixed effects	Uses only changes within each region over time to estimate the model parameters. The Hausman tests suggested that fixed effects may be statistically preferable in some model variants. However, some fixed effects specifications produced implausible coefficient estimates and substantially reduced the differences between regions shown by the benchmarking results.	Not selected because it excluded differences between regions when estimating the model parameters and some specifications produced implausible coefficient estimates.

Model	Assessment	Outcome
Regional OLS	Uses both changes within regions over time and differences between regions, and produced plausible coefficient estimates and benchmarking conclusions that were broadly similar to random effects. However, OLS does not take account of the panel structure of the data.	Not selected because random effects produced similarly plausible results while also taking account of the panel structure of the data.
MDU random effects	Uses both changes within MDUs over time and differences between MDUs to estimate the model parameters, while taking account of the panel structure of the data. The Breusch-Pagan LM test supported the use of a panel model rather than pooled OLS, while the Hausman test did not support fixed effects over random effects. However, random effects produced materially different expenditure ratios and rankings for some MDUs compared with OLS, which we could not reconcile.	Not selected, despite the statistical support for using random effects with panel data, because the resulting expenditure ratios and rankings differed materially from previous benchmarking exercises and our wider understanding of relative MDU expenditure.

Model	Assessment	Outcome
MDU fixed effects	Uses only changes within each MDU over time to estimate the model parameters. The Hausman test did not support fixed effects over random effects. Some fixed-effects specifications also produced implausible coefficient estimates, including a negative traffic elasticity, and materially different expenditure ratios and rankings.	Not selected because the statistical tests did not support fixed effects over random effects and some specifications produced implausible coefficient estimates.
MDU OLS	Uses both changes within MDUs over time and differences between MDUs to estimate the model parameters. It produced economically plausible coefficient estimates and expenditure ratios that were broadly consistent across alternative OLS specifications, previous benchmarking exercises and our wider understanding of relative maintenance expenditure. However, the Breusch-Pagan LM test indicated that there were statistically significant MDU-specific effects that are not	Selected because, taking the evidence as a whole, it produced the most plausible coefficient estimates and the most stable benchmarking results. We recognise that the statistical tests provided support for a panel model and have considered this limitation when interpreting the OLS results.

Model	Assessment	Outcome
	accounted for by pooled OLS.	

2.36 To help ensure that our selected models were robust, we assessed our maintenance expenditure models against a range of statistical tests summarised in Table 5.

Table 5: Statistical tests

Potential statistical issue	Statistical test(s) used
Multicollinearity: Multicollinearity occurs when two or more explanatory variables are highly correlated with each other. This can make it difficult to distinguish their individual effects on costs and may lead to unstable coefficient estimates. Potential multicollinearity was assessed using variance inflation factors (VIFs). We did not find evidence of multicollinearity in the regional model. Where variables were found to be highly correlated in the MDU model, we considered alternative specifications and retained variables based on theoretical relevance, explanatory power and overall model performance. We also undertook sensitivity tests including other variables and found limited impact.	Variance Inflation Factor (VIF > 5)

Potential statistical issue	Statistical test(s) used
Heteroskedasticity: Heteroskedasticity occurs when the variability of the model's errors differs across observations. An error (or residual) is the difference between the value predicted by the model and the actual value observed. It represents the part of maintenance expenditure that is not explained by the variables included in the model. We found evidence of heteroskedasticity in our MDU model but not in our regional model. We managed potential issues caused by heteroskedasticity through using cluster-robust standard errors in our models.	Breusch-Pagan
Autocorrelation: Autocorrelation occurs when the model's errors are correlated over time, meaning that shocks in one period are related to those in another. Standard errors measure the uncertainty around the estimated coefficients, and autocorrelation can affect their reliability if not addressed. We found evidence of autocorrelation in both our regional and MDU models. We managed potential issues caused by autocorrelation through using cluster-robust standard errors in our models.	Wooldridge Breusch-Godfrey Durbin-Watson

Potential statistical issue	Statistical test(s) used
Normality: Normality refers to whether the model's errors follow a normal distribution. This assumption supports the reliability of some statistical tests, although it becomes less important in larger samples. It is also less important for our benchmarking results because our primary focus is on relative expenditure comparisons rather than precise statistical inference. We therefore do not consider departures from normality to materially affect the interpretation of our results.	Shapiro-Wilk
Misspecification: Misspecification occurs when the model omits an important variable or represents the relationship between expenditure and its cost drivers incorrectly. This can produce biased or unreliable coefficient estimates. We did not find evidence of misspecification in our models.	Ramsey RESET
Panel data estimator: For the regional panel model, we tested whether a random effects or fixed effects specification was more appropriate. Fixed effects models focus on changes within individual regions over time and control for all time-invariant regional characteristics. Random effects models, by contrast, use both variation within regions over time and variation	Hausman

Potential statistical issue	Statistical test(s) used
between regions when estimating relationships between expenditure and its cost drivers. The Hausman test indicated that a fixed effects specification may be statistically preferable. However, the fixed effects models estimate cost relationships using only variation within regions over time and produced implausible elasticity estimates. This limits their suitability for benchmarking analysis. Our primary objective is to compare expenditure across regions after controlling for observable differences in scale, usage and complexity. We therefore adopted a random-effects specification, which uses both changes within regions over time and differences between regions, and produced plausible coefficient estimates and benchmarking results.	

Testing additional variables and comparisons with other work

2.37 Our modelling takes account of scale, usage and complexity drivers of costs. We have also investigated using other possible cost drivers. These included asset condition, weather disruption and performance. These variables were only available at a regional level.

2.38 Data for asset condition, weather disruption and performance were taken from Network Rail's annual return. We were unable to include these variables in our regional model because data was only available from 2019 to 2020 onwards, which would reduce our sample size from 75 to 30. This would make our dataset too small to draw robust benchmarking conclusions. Nevertheless, we have examined the potential impact of each of the three variables on regional maintenance expenditure below.

Asset condition

2.39 Asset condition could be a potential explanatory variable for differences in regional maintenance expenditure. Network Rail uses a Composite Reliability Index (CRI) measure of the short-term condition and performance of its assets including track, signalling, points, electrification, telecoms, buildings, structures and earthworks. CRI measures asset reliability in each year compared to the end of the previous Control Period. A higher CRI score means assets experienced fewer service affecting failures.

2.40 We expect a negative relationship between CRI and maintenance expenditure as a more reliable network could mean less need for maintenance interventions. Conversely, higher maintenance expenditure may itself contribute to improved reliability, suggesting a positive relationship between CRI and maintenance expenditure.

2.41 We found no clear or consistent relationship between CRI and variances in our modelled regional maintenance expenditure.

Weather disruption

2.42 We examined extreme weather as a potential explanatory factor for regional maintenance expenditure. Network Rail records extreme weather days as periods of weather that are outside of a normal range, taking account of prolonged and intense rainfall, prolonged dry and hot periods, and periods of repeated freezing and thawing. These events increase the likelihood that a network asset may suffer performance loss, experience accelerated degradation, or to fail completely.

2.43 A higher number of extreme weather days is expected to increase maintenance expenditure due to the need to undertake more proactive and reactive maintenance activities.

2.44 However, we found no clear or consistent relationship between extreme weather days and variances in our modelled regional maintenance expenditure.

Performance

2.45 We considered whether differences in train performance could help explain variations in regional maintenance expenditure. Train cancellations were used as an indicator of performance, as poor asset performance may contribute to service disruption and could be associated with higher maintenance activity. However, train cancellations can be influenced by a range of factors,

many of which are outside the control of Network Rail's need for maintenance.

2.46 Our analysis found no clear or consistent relationship between train cancellations and the unexplained variation in regional maintenance expenditure.

Input prices

2.47 We investigated the inclusion of input price effects in our analysis. This did not result in any improvement to model fit largely because input prices were not found to vary across business units. An important component of this is that most of Network Rail's procurement and wage setting is undertaken at a national level, so we would not particularly expect significant regional variances relating to input prices. We will continue to explore the use of input prices in our analysis.

Comparison with CEPA's study on using econometrics to calculate variable charges

2.48 We compared our results with the existing literature in this area, including a recent CEPA report commissioned by the ORR, which uses econometrics to calculate variable charges. Although this is not an exact comparison to our work, it provides a useful point of comparison.

2.49 We found that the traffic density coefficients were broadly consistent with our estimates, supporting the approach adopted in this report.

- ← Previous Cost benchmarking of Network Rail's maintenance expenditure 2024 to 2025 - Technical report on our modelling approach for cost benchmarking
- → Next 3. Data issues