Baseline Drift and Schedule Slippage in the UK Government Major Projects Portfolio, 2012–2026

Yuxin Zhang *

Alliance Manchester Business School, The University of Manchester, Booth Street West, Manchester, M15 6PB, UK

* Corresponding Author Email: yuxin.zhang-33@student.manchester.ac.uk

Abstract. Public reporting on major project performance is usually read as evidence about cost overrun. This study assembles a project-year panel from the annual Government Major Projects Portfolio (GMPP) data published by the United Kingdom government between 2012 and 2026 — 2,532 project-year records across fourteen snapshots, reconciled project-for-project against the official portfolio totals — and shows that the published fields do not support an overrun measure at all. No release, in any year, publishes a forecast whole-life cost; only an approved baseline is published. What can be measured is baseline drift, the year-on-year movement of that approved figure, which is a lower bound on overrun. Across the six snapshot pairs with reliable project linkage, median annualised baseline drift is 0.0 per cent while the mean is +53.1 per cent, and median schedule slippage is zero months against a mean of +4.6 months. The choice between median and mean, computed from the same column, reverses the conclusion. Two further results follow. The share of projects publishing a whole-life cost fell by exactly twenty percentage points, from 90.4 per cent in 2019 to 70.4 per cent in 2026; a balanced-panel decomposition attributes this to composition rather than to progressive redaction, since projects entering the portfolio disclose at 75.9 per cent while those leaving disclosed at 91.2 per cent. Delivery Confidence Assessments are strongly autoregressive but do not detectably predict subsequent drift or slippage. Five structural breaks in the published series are documented, together with the procedure required to survive them.

Keywords: Major projects, cost overrun, schedule delay, infrastructure governance, transparency data, United Kingdom.

1. Introduction

Large public infrastructure projects absorb a substantial share of national capital budgets and have a long-documented tendency to cost more and take longer than approved [1]. The United Kingdom is an informative setting for studying this because central government publishes structured annual data on its largest and riskiest projects. At the March 2026 snapshot the Government Major Projects Portfolio contained 189 projects with a combined whole-life cost of £924.2 billion, of which 68 projects and £450.0 billion sat in the infrastructure and construction category [2]. The portfolio has been reported annually since 2013 and is one of the few standardised, longitudinal, public records of major project performance maintained by any government.

That record is nonetheless close to unexploited in the scholarly literature. A full-text search of the bibliographic record for the portfolio's name returns a single peer-reviewed empirical study that analyses the data directly, and that study examines assurance routines rather than delivery outcomes, finding that delivery confidence improves during the period a project is assured [3]. The most recent institutional review of the United Kingdom scheme closes by stating that establishing its effectiveness "would require a more comprehensive follow-up study to examine the interplay of the project delivery performance with the institutional arrangement put in place to enact the project governance scheme" [4]. Adjacent quantitative work is either restricted to one asset class — 138 English road schemes completed between 2001 and 2016, with typical cost overrun of 2.4 per cent [5] — or evaluates a single appraisal reform, reporting a fall in average United Kingdom cost overrun from 38 per cent to 5 per cent following the adoption of reference class forecasting [6]. Neither covers information technology, government transformation or defence, and both end before 2017. Whether the optimistic reading generalises to the whole portfolio, and to the 2020s, is untested.

This study assembles the full published series [7] and reports three results. First, the published fields cannot support the conventional overrun measure, because no release publishes a forecast whole-life cost; the measurable quantity is baseline drift, and the distinction is consequential rather than pedantic. Second, on that measure the median project-year shows no net movement in either approved cost or approved end date, while the means are large and are produced by a small number of documented rebaselining events. Third, the share of projects whose cost is published at all has fallen by twenty percentage points since 2019, and a balanced-panel decomposition shows this to be compositional. The analysis also documents five structural breaks in the series that silently corrupt a naive panel construction, and the reason for reporting them now is that the window is closing: from 1 April 2026 the portfolio was reduced to roughly eighty projects, with assurance for the remainder devolved to departments [2].

2. Related Work

2.1. Scale and shape of the problem

The modern empirical literature begins with a large-sample study of transport projects reporting that costs were underestimated in almost nine of ten cases [8]. Subsequent work has established that the resulting distributions are not merely high on average but heavily right-tailed, to the point that mean and variance may be unstable, so that summary statistics based on central tendency mislead [9]. This has a direct methodological consequence adopted here: distributions are reported in full, and outliers are neither trimmed nor treated as errors without inspection.

2.2. The measurement dispute

A sustained exchange has questioned whether the headline overrun figures survive scrutiny of their baselines and samples, arguing that inconsistent baselines and evolving scope, rather than deception, account for much of the measured effect [10]. The reply concedes the central methodological point while rejecting the substantive one, holding that consistent definition and measurement of overrun — as against mixing inconsistent baselines and price levels — is the precondition for any comparison [11], a position developed further in subsequent contributions on both sides [12]. A separate strand argues that the term "overrun" is used for several distinct quantities that ought to be named separately [13].

The dispute turns on a quantity that is rarely measured directly. Overrun computed against an early planning estimate is large; computed against the estimate at the decision to build it is moderate; computed against the most recently approved baseline it can be close to zero, because each re-approval resets the denominator. The three are answers to different questions and are not interchangeable, yet portfolio-level evidence on how far that denominator actually moves is scarce, in part because most datasets record a single baseline rather than a sequence of them. The present study sits inside this dispute rather than on one side of it: an annual series of approved baselines for the same projects makes the movement of the denominator observable, and that movement is what is reported here.

2.3. Front end and lock-in

Recent evidence from Swedish national investment plans finds that estimates made at the decision to build are close to final costs on average, while substantial escalation occurs earlier, during planning, with a highly skewed distribution of planning-stage changes; the mechanism proposed is lock-in of premature decisions [14]. This is the closest existing analogue to a baseline-drift study and motivates treating rebaselining as the object of measurement rather than as noise. Related work identifies front-end definition and governance as the decisions with the greatest leverage over outcomes [15], and a systematic review of megaproject performance consolidates the causes and proposed remedies [16]; a widely used taxonomy sorts explanations into technical, economic, psychological and political categories, finding political explanations dominant [17].

2.4. Selection

A formal argument shows that cost overruns and demand shortfalls can arise from the selection of projects on estimated benefit-cost ratios even when individual forecasts are unbiased [18]. This is directly applicable here, since the portfolio studied is a selected set at both ends: projects enter because they are large, complex, novel or contentious, and leave on completion, cancellation or reclassification. Entry and exit flows are therefore reported explicitly.

3. Data and Measurement

3.1. Source and assembly

The data is published in two forms, and the distinction determines whether a panel is complete. The government's data collection [7] lists 197 publications, of which 196 carry data files — 542 files in all, 196 comma-separated and the remainder spreadsheet copies — and these are departmental releases, one per department per year. Separately, a consolidated whole-portfolio file is published alongside each annual report; twelve of the fourteen snapshots have one, and ten of those twelve sit on annual-report pages outside the collection. A panel assembled from the departmental releases alone is not the portfolio: it silently omits projects, chiefly those of the Office for National Statistics and the National Crime Agency, in eight of the fourteen snapshots. The panel used here is therefore built from the consolidated file wherever one exists, and only the two snapshots without one, September 2012 and September 2015, are assembled from the departmental union. Every snapshot is then reconciled against the project count published in the corresponding annual report, and all twelve consolidated snapshots agree exactly. The result is 2,532 project-year records over fourteen snapshots from September 2012 to March 2026.

Two properties of the raw files shape the loader. Layout is not stable: of the 194 departmental comma-separated files, 120 are transposed, with field labels running down the first column and one column per project, and 74 are conventional wide tables, so orientation must be detected per file rather than assumed. The sampling frame matters as much as the parser: enumerating pages by keyword search rather than from the collection returns a different corpus, adding twelve senior-responsible-owner lists that contain no project record and omitting both annual-report files. Publication year is also not snapshot year — the 2023-24 report and its data appeared in January 2025 — so records are keyed on the snapshot date recovered from file names and publication titles, never on the publication timestamp. One reconciliation fails and is reported rather than resolved: the September 2012 files contain 181 projects, while the 2013 annual report states that the portfolio held 191 at that date.

3.2. The absence of a forecast whole-life cost

Across all fourteen snapshots the published cost fields are an in-year baseline, an in-year forecast, an in-year variance, and a whole-life cost that is a baseline with no forecast counterpart. The conventional overrun measure — final or forecast cost relative to the estimate at the decision to build [8, 11] — is therefore not computable from these data. Two quantities are computable. In-year variance is published directly but measures annual budget accuracy over a twelve-month window rather than project outcome. Baseline drift, defined for project i between consecutive snapshots t and t+1 as

$$drift = \frac{(WLC_{\mathrm{i,t+1}} - WLC_{\mathrm{i,t}})}{WLC_{\mathrm{i,t}}} \times \frac{12}{m} \quad (1)$$

where m is the elapsed months between snapshots, measures how far the approved whole-life cost moved while the project was under central assurance. Drift is a lower bound on overrun, because a project whose baseline has been raised has had part of its overrun absorbed into the new denominator. Schedule slippage is defined analogously as the movement in months of the published approved end date, annualised the same way, and is free of price-base effects.

3.3. Structural breaks

Five discontinuities were measured and are shown in Fig. 1. The snapshot date moved from 30 September to 31 March, and no September 2020 snapshot exists, so the interval between the September 2019 and March 2021 observations is eighteen months. The rating scale changed from five points to three between the March 2021 and March 2022 snapshots, the column headers of the two files stating the respective scales explicitly. File layout changed from transposed to wide one snapshot earlier, between September 2019 and March 2021. A project identifier is absent from every file before the September 2019 snapshot and present for 99 to 100 per cent of records from it onward, so earlier linkage must rely on names, which change between snapshots. Monetary values in the March 2026 file are restated in 2024/25 real prices, and the publisher states that earlier years are not directly comparable [2]. A sixth issue, a renaming of the whole-life cost field that drops the word "baseline" in the two most recent releases, is unresolved in the published guidance and is treated here as a limitation.

3.4. Disclosure

Every whole-life cost cell was classified as a published point value, a published range, withheld under a statutory exemption, not yet baselined, not provided, or blank. Only 70.4 per cent of cells in the March 2026 file carry a usable number. Fifty-one projects, 27 per cent of the portfolio, carry exemption text in place of a figure, most frequently commercial interests, cited by 39 projects and by 37 of them as the sole ground, with smaller counts citing future publication, policy formulation, national security, international relations and information supplied by security bodies. Summing the disclosed values gives £464.1 billion against a published portfolio total of £924.2 billion. All three projects the publisher identifies as exceeding ten per cent of portfolio value between them — High Speed 2 Phase 1, Dreadnought and Sizewell C — lack a published point value. Table 1 and Fig. 2 report disclosure by snapshot and by category. The equivalent field for schedule is also affected: 121 end-date cells could not be parsed, of which 35 carry exemption text, and schedule observation coverage falls from 92.9 per cent to 63.5 per cent over the analysis window.

Non-disclosure is thus concentrated in the projects that dominate portfolio value, which has a direct consequence for what can be claimed. Any statistic weighted by project value is computable only over the disclosed subsample, and that subsample excludes the largest members by construction. Equal-weighted statistics are reported throughout for this reason, and the disclosed share is stated wherever a total is given.

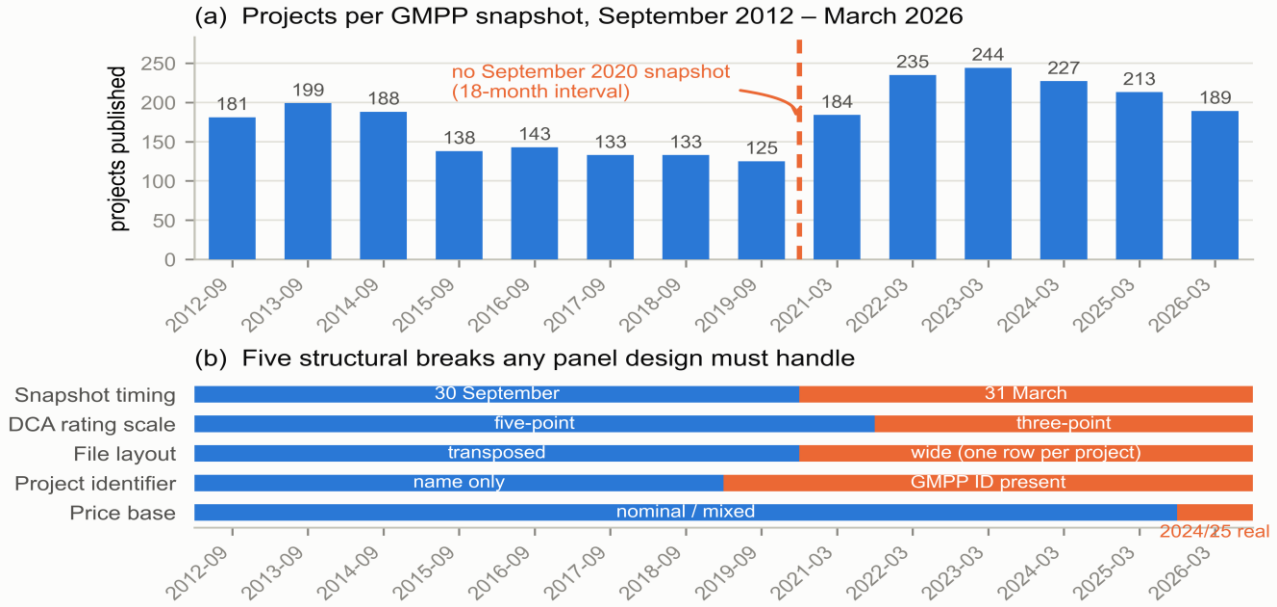


Fig. 1 Structure of the published GMPP series. (a) Projects per snapshot, September 2012 to March 2026, with the missing September 2020 snapshot marked. (b) Five structural breaks in the series.

Table 1. Disclosure of whole-life cost by snapshot, September 2012 to March 2026.

Snapshot	Projects	Disclosed	Withheld under exemption
2012-09	181	84.5%	0
2013-09	199	92.0%	15
2014-09	188	92.6%	9
2015-09	138	93.5%	9
2016-09	143	93.0%	9
2017-09	133	89.5%	14
2018-09	133	88.7%	15
2019-09	125	90.4%	11
2021-03	184	89.1%	17
2022-03	235	81.3%	37
2023-03	244	79.1%	46
2024-03	227	76.7%	46
2025-03	213	73.7%	49
2026-03	189	70.4%	51

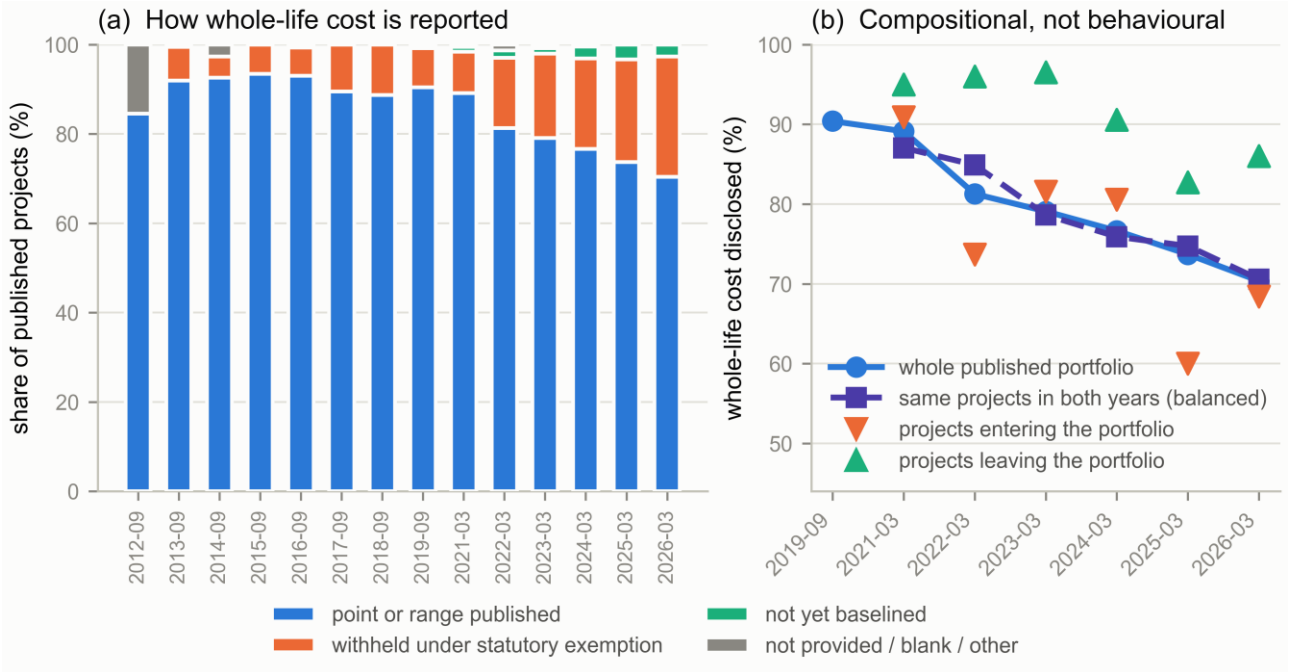


Fig. 2 Disclosure of whole-life cost. (a) Composition of the published cost field by snapshot. (b) Portfolio disclosure rate against the balanced panel, with entrant and leaver rates.

4. Method

Projects are linked across consecutive snapshots by identifier. Because identifiers are absent before September 2019 and name matching then recovers only 11 to 56 projects per pair against portfolios of 125 to 199, all longitudinal analysis is restricted to the six pairs from September 2019 onward, where identifier matching recovers 85 to 206 projects per pair. Coverage is reported at every step and unmatched records are retained and counted rather than dropped silently.

Distributions are summarised by median, mean and selected quantiles, following the tail argument in the literature [9]. Extreme observations were inspected individually against the published narratives before any decision about their treatment, and none is trimmed from any reported distribution. The one departure is stated here: in the ordered logit alone, lagged drift is winsorized to the interval −100 to +300 per cent and lagged slippage to −36 to +60 months, because a single regressor value of

+23,486 per cent otherwise dominates the likelihood. The untrimmed specification is reported alongside it and reaches the same conclusion. Category comparisons use the Kruskal–Wallis test, with a one-sided Mann–Whitney test for the specific ordered contrast. Delivery Confidence Assessments follow the publisher's own convention — the rating supplied by the central authority where present, otherwise the rating supplied by the senior responsible owner — and the five-point scale is collapsed to three by mapping the two intermediate grades to amber. Transitions are modelled with an ordered logit on the rating at $t{+}1$, with the rating at t, portfolio category, rating source, lagged drift, lagged slippage and log whole-life cost as regressors. All estimates are reported as associations. A poor rating attracts intervention and re-planning, so the direction of influence between rating and subsequent movement is not identified by these data.

Three safeguards proved necessary rather than decorative, and each caught an error that would otherwise have been invisible. Loading fails on any unrecognised column rather than skipping it, which is how an incomplete field map was found to be discarding the rating field of the earlier governance regime — 387 cells across the September 2013 and 2014 snapshots, the outcome variable for those years. Every unparsed value is reported rather than coerced, which revealed that a unit suffix used in the earlier files was causing valid figures to be classified as non-disclosure, understating disclosure in one snapshot year as 1.1 per cent against a true 92.6 per cent. And every per-snapshot project count is asserted against the figure published in the corresponding annual report, which is what identified the departmental releases as an incomplete sampling frame.

5. Results

5.1. Distributions

Table 2 reports the outcome distributions over 744 cost and 815 schedule observations. Median annualised baseline drift is 0.0 per cent and the mean is +53.1 per cent; the tenth and ninetieth percentiles are −10.2 and +47.3 per cent, and 50.3 per cent of project-years show an increase against 27.6 per cent showing a decrease and 22.2 per cent no change at all. Median annualised slippage is zero months against a mean of +4.6 months, with a ninetieth percentile of +24 months; 61.0 per cent of project-years leave the approved end date untouched and 32.0 per cent slip. Fig. 3 shows the full distributions by category. The median is flat in both dimensions and the means are produced by a small tail, but the flat cost median is a balance of offsetting movements rather than an absence of movement.

The final snapshot pair spans a change of price base, since the March 2026 file is restated in 2024/25 real prices and its predecessor is not. It is reported separately rather than pooled silently: excluding it raises mean drift from +53.1 to +61.2 per cent over 631 observations and leaves the median at 0.0 per cent, so the pooled headline is the conservative of the two. Slippage is unaffected.

The tail was inspected rather than trimmed. The largest observations correspond to documented rebaselining events: the New Hospital Programme moves from £3,690 million to £60,008 million between the 2024 and 2025 snapshots, consistent with the reset of that programme; a prisons transfer project moves from £6.86 million to £1,618 million; a digital identity programme moves from £22 million to £305 million. These are changes of approved scope, not unit errors, and excluding them would remove precisely the events the measure exists to capture. The pattern is visible in the largest programme in the portfolio, which was reset in 2020 and again in 2025, each reset re-establishing the baseline against which subsequent performance is then judged [19]. Baseline drift measures the size of that step; it cannot measure what the outturn would have been against the original approval, because that comparison is not published.

Table 2. Annualised baseline cost drift and schedule slippage, six snapshot pairs, September 2019 to March 2026.

Measure and category	n	Median	Mean	p10	p90	Share above zero
Baseline cost drift (%)						
All observations	744	+0.0%	+53.1%	-10.2%	+47.3%	50.3%
Excluding the 2025–26 price-base break	631	+0.0%	+61.2%	-8.0%	+48.7%	50.1%
Government transformation	290	+0.0%	+107.6%	-18.8%	+63.9%	47.9%
ICT	99	+0.1%	+28.1%	-5.4%	+59.9%	52.5%
Infrastructure & construction	235	+0.0%	+16.3%	-3.9%	+24.7%	43.4%
Military capability	120	+3.1%	+14.0%	-8.4%	+35.3%	67.5%
Schedule slippage (months)						
All observations	815	+0.0	+4.6	+0.0	+24.0	32.0%
Excluding the 2025–26 price-base break	707	+0.0	+4.5	+0.0	+24.0	30.0%
Government transformation	286	+0.0	+2.9	+0.0	+13.0	31.5%
ICT	113	+0.0	+6.1	+0.0	+24.0	44.2%
Infrastructure & construction	244	+0.0	+9.7	+0.0	+28.0	29.5%
Military capability	172	+0.0	-0.6	-11.0	+45.0	28.5%

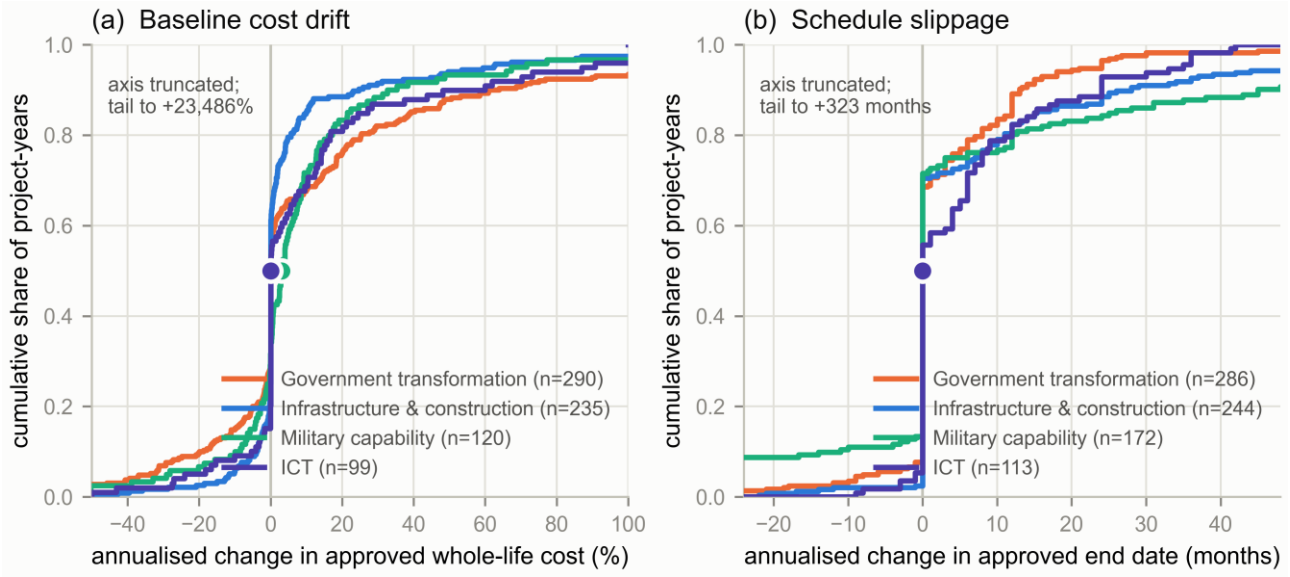


Fig. 3 Empirical distributions of (a) annualised baseline cost drift and (b) annualised schedule slippage, by portfolio category. Markers show category medians at the 50th percentile.

5.2. Categories

Government transformation shows the highest mean drift at +107.6 per cent against a median of 0.0 per cent. Military capability shows the lowest mean, +14.0 per cent, but the highest median at +3.1 per cent and the highest share of project-years with any increase, 67.5 per cent. Infrastructure and construction sits between them on cost, at a mean of +16.3 per cent, but has the worst schedule performance, at a mean of +9.7 months per year against +6.1 for information and communications technology, +2.9 for government transformation and −0.6 for military capability. Defence baselines therefore move steadily and by modest amounts, while transformation baselines move rarely and then by large amounts.

Two consequences follow. A simple ranking of categories by "worse performance" is not supported: which category appears worst depends on whether cost or schedule, and mean or median, is the statistic. And the ordering does not match the standing appraisal prior. United Kingdom guidance on optimism bias assigns upper-bound capital uplifts of 200 per cent to equipment and development including information technology, against 44 to 66 per cent for civil engineering [20],

which anticipates that technology-led work will be the worst performer on cost. In these data the technology and transformation categories have the heaviest tails but the flattest medians, while defence — which the guidance does not single out — has the highest median drift and the highest incidence of any increase. Central tendency and tail behaviour separate by category, and a single uplift factor cannot represent both.

5.3. Disclosure

The share of projects publishing a whole-life cost is stable between 84.5 and 93.5 per cent from 2012 to 2019, then falls for six consecutive snapshots from 90.4 per cent in September 2019 to 70.4 per cent in March 2026, while the count of projects withheld under a statutory exemption rises from 11 to 51 (Table 1, Fig. 2a). Disclosure is uneven across categories: 83.6 per cent for government transformation, 72.7 per cent for information and communications technology, 66.2 per cent for infrastructure and construction and 55.3 per cent for military capability.

A balanced-panel decomposition (Fig. 2b) locates the mechanism. Among projects present in both years of a pair, disclosure moves from 79.6 to 78.6 per cent on average, approximately one point per year against a portfolio-level fall of about three. Across the whole series there are 49 within-project switches from disclosed to withheld and 40 in the opposite direction. Over the six post-2019 pairs, projects entering the portfolio disclose at 75.9 per cent while projects leaving disclosed at 91.2 per cent. The fall is therefore compositional: well-disclosed projects complete and leave, and less-disclosed projects enter, without existing projects being progressively redacted. This is a portfolio-level analogue of the selection argument [18], observed directly rather than inferred.

5.4. Ratings

Ratings are persistent: 74.0 per cent of project-years remain in the same band, and a red-rated project returns to green in the following year in 1.8 per cent of cases. Projects rated red are followed by higher median drift, +3.8 against 0.0 per cent, and higher mean slippage, +15.5 against +0.2 months for green, with 39.5 against 30.2 per cent slipping. The differences are not statistically detectable: Kruskal–Wallis across the three bands returns $p = 0.183$ for drift and $p = 0.324$ for slippage, and the one-sided red-versus-green contrast returns $p = 0.080$ and $p = 0.102$. The tests rest on 51 green, 614 amber and 74 red observations for drift and 53, 671 and 81 for slippage, so this is a failure to detect rather than evidence of no association. In the ordered logit on 676 complete cases the rating at t dominates — green −2.434 and red +1.642, both $p < 0.001$ — while lagged drift ($p = 0.250$), lagged slippage ($p = 0.131$), log whole-life cost ($p = 0.116$) and rating source ($p = 0.848$) are not significant. Estimating the same model on untrimmed regressors changes nothing that matters: lagged drift and slippage remain insignificant at $p = 0.707$ and $p = 0.246$, and the lagged rating coefficients move by less than 0.03. Military capability is the only category with a significant association with a worse subsequent rating, +0.862 at $p = 0.003$. The assessment is, on published data, largely a function of its own previous value.

6. Discussion

The median and the mean of the same column support opposite headlines. The median project-year in the portfolio shows no net change to its approved cost or completion date, which reads as a controlled portfolio; the mean shows approved costs growing by more than half each year, which does not. Both figures are correct. This is the measurement dispute [10, 11, 12, 13] instantiated on national data rather than argued in the abstract, and it suggests that the productive question is not which side is right but which statistic a given claim depends on.

The result also qualifies the optimistic recent readings. Reported improvements in United Kingdom performance rest on a single asset class ending in 2016 [5] and on the evaluation of one appraisal reform [6]. Across the whole portfolio and into the 2020s the median remains flat, which is

consistent with those readings, while the tail remains heavy, which is not usually how they are summarised.

Two findings bear on policy. Since the rating is essentially autoregressive and adds nothing detectable beyond its own lagged value, an early-warning capability built on these published fields is unlikely to outperform the previous rating; that does not speak to what is achievable on the fuller data held internally, but it does bound what external scrutiny can accomplish. And because the disclosure decline is compositional, remedies aimed at reporting behaviour are misdirected: the portfolio is being repopulated with categories whose costs are exempt from publication, which is a consequence of what government is now choosing to build.

Four limitations are material. Linkage before September 2019 is unreliable and the longitudinal results are therefore confined to six pairs. Baseline drift understates overrun by construction, and the extent is unknown because the counterfactual is unpublished. The definition of the whole-life cost field may have changed with its renaming in the two most recent releases, and the published guidance does not settle it. And every disclosure statistic here is a share of projects, not of value: the value-weighted equivalent is not computable, because the projects that dominate portfolio value are the ones withheld.

7. Conclusion

The United Kingdom's major projects transparency data does not measure what it is usually read as measuring. No release in fourteen years publishes a forecast whole-life cost, so cost overrun in the conventional sense cannot be computed from it; what can be computed is the movement of the approved baseline itself, which is a lower bound on overrun and a direct measurement of rebaselining.

On that measure the median project-year shows no net movement in approved cost or approved completion date, while the mean shows large growth generated by a small number of documented resets. Schedule performance is worst in infrastructure and construction, and cost behaviour differs qualitatively between defence, which drifts steadily, and government transformation, which drifts rarely and then by large amounts. Delivery Confidence Assessments are highly persistent and do not detectably predict subsequent movement in either dimension. The share of projects publishing a whole-life cost has fallen from 90.4 to 70.4 per cent since 2019, and the mechanism is the composition of the portfolio rather than the redaction of individual projects.

Two implications follow for the reporting regime. Publishing a forecast whole-life cost alongside the approved baseline would make outcome measurement possible at negligible cost, since the forecast already exists internally. And with the portfolio reduced to roughly eighty projects from April 2026 and assurance for the remainder devolved to departments, the series analysed here is unlikely to be extended on the same basis, which makes the case for preserving and documenting it now.

References

[1] B. Flyvbjerg: Project Management Journal Vol. 45 (2014), p. 6.

[2] National Infrastructure and Service Transformation Authority: *Major Projects Annual Report 2025-26* (HM Treasury and Cabinet Office, UK 2026).

[3] H. Vo, R.J. Kirkham, T.M. Williams, A. Howells, R. Forster and T. Cooke-Davies: International Journal of Managing Projects in Business Vol. 14 (2021), p. 865.

[4] R. Kirkham and N. Young, in: *Governance Frameworks for Major Public Projects*, edited by V. Zerjav, G.H. Volden and M. Welde, Wiley (2026), p. 41.

[5] P. Chapman: Project Leadership and Society Vol. 5 (2024), p. 100128.

[6] J.E. Park: European Journal of Transport and Infrastructure Research Vol. 21 (2021), No. 2.

[7] GOV.UK: *Major Projects Data* (Cabinet Office and HM Treasury, UK 2013-2026).

[8] B. Flyvbjerg, M.S. Holm and S. Buhl: Journal of the American Planning Association Vol. 68 (2002), p. 279.

[9] B. Flyvbjerg, A. Budzier and D. Lunn: Environment and Planning A: Economy and Space Vol. 53 (2021), p. 233.

[10] P.E.D. Love and D.D. Ahiaga-Dagbui: Transportation Research Part A Vol. 113 (2018), p. 357.

[11] B. Flyvbjerg, A. Ansar, A. Budzier, S. Buhl, C. Cantarelli, M. Garbuio et al.: Transportation Research Part A Vol. 118 (2018), p. 174.

[12] P.E.D. Love, M.C.P. Sing, L.A. Ika and S. Newton: Transportation Research Part A Vol. 122 (2019), p. 1.

[13] D.C. Invernizzi, G. Locatelli and N.J. Brookes: Proceedings of the Institution of Civil Engineers – Civil Engineering Vol. 171 (2018), p. 85.

[14] J. Eliasson: Transportation Research Part A Vol. 198 (2025), p. 104532.

[15] K. Samset and G.H. Volden: International Journal of Project Management Vol. 34 (2016), p. 297.

[16] J. Denicol, A. Davies and I. Krystallis: Project Management Journal Vol. 51 (2020), p. 328.

[17] C.C. Cantarelli, B. Flyvbjerg, E.J.E. Molin and B. van Wee: European Journal of Transport and Infrastructure Research Vol. 10 (2010), p. 5.

[18] J. Eliasson and M. Fosgerau: Transportation Research Part B Vol. 57 (2013), p. 105.

[19] National Audit Office: *High Speed Two Reset*, HC 52, Session 2026-27 (National Audit Office, UK 2026).

[20] HM Treasury: *Supplementary Green Book Guidance: Optimism Bias* (HM Treasury, UK 2003).