Can removing human bias deliver more realistic project forecasts?

The Nichols and National Audit Office reports have once again highlighted cost overruns of transport infrastructure projects. Andrew Forster spoke to Danish planning academic Bent Flyvbjerg about why predictions of cost and demand are so often wrong and how statistical analysis can remove human bias from forecasts

Leeds Supertram; the M1 widening; Manchester Metrolink extensions; the Channel Tunnel; Edinburgh tram; the M25 widening; the West Coast Main Line upgrade; the Hastings-Bexhill link road; A3 Hindhead improvement; A303 Stonehenge improvements; South Hampshire Rapid Transit… The list of transport projects hit by significant cost escalations in recent years is long and growing. Some have been cancelled; others have been approved at higher cost; and others have been trimmed in scope to fit the available budget.

Project managers can perhaps draw comfort from the fact that poor cost estimation is not unique to the UK and nor is it unique to transport projects, as the planning of the 2012 London Olympics, or the building of the new Scottish Parliament testify.

So why do infrastructure projects so frequently end up costing more – often much more – than forecast? The recent Nichols and National Audit Office reports into road scheme planning provided some answers (*LTT* 15 Mar), reporting that around half the cost overrun on schemes in the Highways Agency's targeted programme of improvements was explained by construction price inflation running at about twice the rate of the Retail Price Index. Nichols said the remainder could be attributed roughly equally to inadequate initial cost estimates and changes in project scope and delays.

Cost overruns do not surprise Bent Flyvbjerg, a professor at Aalborg University's department of development and planning in Denmark and Delft's University of Technology in the Netherlands. Using questionnaires, interviews, project accounts and other studies, Flyvbjerg has collated cost data on hundreds of transport infrastructure projects from around the world. The data shows that cost overruns are the norm when the actual cost of projects is compared with the forecast cost when the decisions to build the projects were made. "In my research group we think statistically and what we can document is that statistically the likelihood is very high that something comes up – we don't need to predict what it is in order to take it into account as a risk," he says. "It may be inflation, it may be geology, it may be environmental demands, it may be safety demands and so forth."

Flyvbjerg's analysis of 258 transport infrastructure projects completed between 1927 and 1998 shows that costs were underestimated in nine out of ten instances. Rail projects had the largest overruns, averaging 44.7% (three-quarters of the 58 rail projects had cost escalations of at least 24%). For bridges and tunnels the average overrun was 33.8% and for roads 20.4%. As far as Flyvbjerg knows, the samples are the largest of their kind ever collected. And if the samples are biased the bias is most likely conservative, he believes.

Flyvbjerg has also studied the accuracy of opening year demand forecasts for transport projects and says this shows a strong bias towards overestimating demand for rail projects but a much more varied picture for road schemes, though road scheme forecasts were also often subject to a large margin of error (*LTT* 9 Mar 06).

What surprised him was the data suggests that forecast accuracy has not improved with time. "This was the most shocking thing for me," he says. "There had been no improvement for 70 years on the cost side and no improvement for 30 years on the demand side. How could this be? My colleagues who are in the business of demand modelling had been telling me all the time that they had much better models now than they had ten years ago or 20 years ago and I asked them, 'How come you can't see that in the results?'"

The obvious concern that Flyvbjerg's analysis raises is that the information on which decision-makers base their judgements about whether to invest in projects can often be highly inaccurate. "When cost and demand forecasts are combined, for instance in the cost-benefit analyses that are typically used to justify large transportation infrastructure investments, the consequence is inaccuracy to the second degree," he says.

Two key causes of error

Flyvbjerg says technical explanations are often used to explain away inaccuracies. Traffic forecasters, he says, are often saying: 'Well the forecast was wrong because our data are not good enough, they are old and we should have new data. Or we are using this old model because we haven't had money to develop or implement a new model yet, therefore the forecasts are off.'

"Our answer [for rail demand forecasts] is no, that's not the reason, we can document that with an incredibly high level of statistical significance. If that was the reason you'd expect a random distribution of errors and what we document is that it's not a random distribution – it's a highly biased distribution."

Bent Flyvbjerg is a professor at Aalborg University, Denmark and Delft in the Netherlands

"If you just ask experts how do you think this project is going to perform… then you get garbage in and garbage out."

Flyvbjerg says the observed biases in both cost and demand forecasting are attributable to psychological and political explanations. The psychological element is known as 'optimism bias' and the political dimension 'strategic misrepresentation'. "Optimism bias is involuntary – it's not something you do deliberately," he says. "It's just the tendency to underestimate the costs – and time – that it takes to do planned actions and at the same time to overestimate the benefits of the same actions." In some respects he thinks this self deception can be beneficial. "I think that any organisation needs to have people with rose coloured glasses who see everything in a rosy way. If you had just people who saw everything realistically I think creativity would suffer," he explains. But he believes it's essential that such people are complemented by individuals who can assess projects in a "hard-nosed way", "especially when money starts to be on the line".

"Strategic misrepresentation is deliberate: you deliberately underestimate the costs and overestimate the benefits," he explains. "It's what I call the Machiavellian formula for project approval: you underestimate the costs, you overestimate the revenues, you underestimate the negative environmental impact and you overestimate the economic regional impacts that the politicians are always touting. If you use that formula you can make the project look good on paper."

Strategic misrepresentation appears to be a euphemism for lying and Flyvbjerg confirms

that is exactly what it is. "We use this very strange term 'strategic misrepresentation' because it makes it possible to talk about it," he says. "If I call it a lie when I give talks or when I write, people are furious – I get hate mail. But if I'm in an audience of engineers, economists and planners and I say 'strategic misrepresentation' then people nod their heads and say to themselves 'Okay, we can talk about that, we know that, we had that.' So it's an Orwellian term really."

"There's this grey area between strategic misrepresentation and optimism where it's not clear whether it's one or the other," he says, adding that it is also not uncommon to find both types of behaviour within the same project organisation.

The precise mix of psychological and political explanations for forecasting bias will depend on the policy-making context, he says. "Explanations in terms of optimism bias have their relative merit in situations where political and organisational pressures are absent or low, whereas such explanations hold less power in situations where political pressures are high," he explains. "Conversely, explanations in terms of strategic misrepresentation have their relative merit where political and organisational pressures are high, while they become immaterial when such pressures are not present."

Based on qualitative interviews "with the few people who are willing to talk about these things on the record" and a lot more information obtained off the record, it is Flyvbjerg's impression that strategic misrepresentation is "a substantial problem" in the planning of major transport infrastructure projects around the world. He came across the practice early in his career. "I came across this problem very early on when I was a student intern – not in transportation but in regional planning," he says. "I thought it was extremely interesting but I also knew that I didn't want a professional life writing reports where the results were already given before you started the report because the 'higher ups' wanted you to arrive at this result. So I had the opportunity early on to take a stand on it."

Flyvbjerg stresses that he's not saying optimism bias or strategic misrepresentation are at play every time a project cost overruns or forecast demand fails to materialise. "In statistics you can never say exactly what happened in one instance," he points out. "It's really important that people understand that we are not saying that each and every project is planned this way. They may have very different reasons including the possibility of technical explanations that costs overran and traffic was low but we can say that it's near impossible that technical explanations would account for what we see in the aggregate."

How to Improve accuracy

Flyvbjerg says that the way to minimise these problems is to adopt 'reference class forecasting' techniques that reduce the opportunity for human biases to shape results. Project teams will always be at risk of taking an optimistic outlook on the costs, delivery timetables and benefits of the project, he says. Therefore, this 'inside view' of the project should be challenged by an 'outside view', whereby a class of similar or comparable projects is used to assess the likelihood of changes to the costs, benefits and time schedule affecting the project at hand *(see panel below)*.

The recent introduction of empirically-based optimism bias uplifts into costing UK transport infrastructure projects is an example of reference class forecasting. As far as Flyvbjerg is aware, the UK was the first – and remains the only – country to have systematically built optimism bias uplifts into cost estimation. This was heralded in the revised Green Book published by the Treasury in 2003. Individual Government departments subsequently commissioned their own research to develop departmental-specific optimism bias uplifts and it was Flyvbjerg and Danish consultancy COWI who prepared the advice for the DfT *(see panel, page 13)*.

Optimism bias uplifts should make project costings more realistic but Flyvbjerg warns that they could reduce project management discipline so that the usual cost overruns still occur. "Then you're in a much worse situation because you're getting a percentage increase on top of a higher budget which means you have a much worse cost overrun than you would otherwise have had. So you'd have been better off doing nothing." Flyvbjerg and COWI also warned that optimism bias uplifts would create budgets for some projects that would be more than adequate for delivery. "This may in itself have an incentive effect that works against tight cost control if the total budget reservation is perceived as being available to the project," the report noted.

Meanwhile, when strategic misrepresentation is a problem, Flyvbjerg says optimism bias adjustments could motivate project promoters to underestimate project costs even more "so that even with the optimism bias you will have a low budget and you will look good on paper".

"[Reference class forecasting] may be considered the single most important piece of advice regarding how to increase the accuracy of forecasting."

"I don't expect optimism bias adjustments to work if you don't change the incentive structure for project approval," says Flyvbjerg. "If you keep an incentive structure in place that encourages people to misrepresent cost and benefit why should we expect them not to do it?" Getting the incentives right means "exactly what you're doing in the UK now", he says. "Punish people for not doing realistic estimates and reward people who do realistic estimates."

The Flyvbjerg/COWI report argued that the major scheme approval procedures for local authorities in England encouraged project promoters to make their project look as good as possible on paper because they were in competition with one another for the DfT's limited resources (*LTT* 15 Jul 04). "Local authorities that cost their schemes realistically high... run the risk of not receiving funding," the report concluded.

Since 2004 the DfT has made significant changes to the approval process that are intended to reduce cost overruns (*LTT* 6 Apr 06). Full project approval used to be granted before procurement of a construction contract began but the DfT now only grants full approval when a firm tender price has been established. Furthermore, local authorities must now find at least 10% of the capital costs for schemes locally (25% for light rail) which, says the Department, acts as a "clear incentive to ensure that the right structures

Continued on page 13

Reference class forecasting: removing human bias

The practice of applying optimism bias adjustments based on empirically derived figures from other projects is an example of 'reference class forecasting'. The foundations of this approach were laid by psychologist Daniel Kahneman, who won the Nobel prize in economics in 2002 for his work on decision-making under uncertainty.

Flyvbjerg explains that reference class forecasting avoids human biases such as optimism and strategic misrepresentation by focusing not on the specific context of the planned action (the 'inside view' of a project) but on the experiences of similar projects that have already been completed (the 'outside view' of a project). "This may be considered the single most important piece of advice regarding how to increase the accuracy of forecasting," he says.

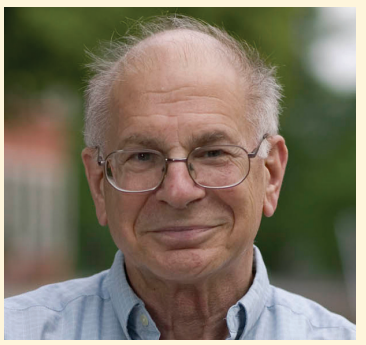
An outside view doesn't try to forecast the specific uncertain events that will affect a particular project. Instead, it places the project in a statistical distribution of outcomes from a group of reference projects. The technique requires three steps: identification of a relevant reference class of projects; the establishment of a probability distribution for the reference class cases that allows statistically meaningful conclusions to be drawn (normally at least ten projects); and the placement of the planned project at an appropriate point in the reference class distribution.

Flyvbjerg says a lot of risk assessments used in project management fall far short of this best practice. "A lot of risk assessment methods are not empirically based, they are subjective," he says. "You just ask experts, 'What do you think the risk is here?' and then you pool the experts' evaluations and that's your risk." But Flyvbjerg says Kahneman found that 'experts' were as liable to suffer from optimism bias as lay people.

"You can really only trust the numbers if the risk assessment is empirically based – looking at other comparable projects and how they performed – that's what I mean by empirically based," says Flyvbjerg. "If you just ask experts how do you think this project is going to perform... then you get garbage in and garbage out."

The DfT's guidance note, *The estimation and treatment of scheme costs*, published last October, endorses the use of both empirically-based risk assessment and probability distributions.

Reference class forecasting is more straightforward for cost estimation than demand forecasting. "Demand forecasting is more difficult because with cost forecasting you can operate with unit costs – what does one mile of motorway or one mile of Metro cost," says Flyvbjerg. "There's not the equivalent in demand forecasting. It's more unique from project to project."



Daniel Kahneman: advocate of reference class forecasting

Nevertheless, by comparing similar projects in similar circumstances, informative reference class demand forecasts are possible, he says.

"If somebody did a forecast for a ten-mile stretch of metro, we could talk to the project's forecasters and present them with our data and say, 'OK, here's 15 other places who have done projects like the one you are forecasting. The [average] actual traffic was 40% short of the forecast, why wouldn't we get the same in your situation?'"

How optimism bias adjustments were derived and how they work

Project cost forecasts should not be trusted unless they have been adjusted for optimism bias, says Bent Flyvbjerg. "This is very clear. This is what our research shows. In nine out of ten cases you cannot trust them because there was cost overrun."

The 2004 report for the DfT by Flyvbjerg and consultant COWI, *Procedures for dealing with optimism bias in transport planning*, recommended optimism bias uplifts for road, rail and fixed link (tunnel/bridge) projects at full business case submission stage, based on a probability distribution of cost overrun in Flyvbjerg's database of European and North American projects (252 projects in all comprising 172 road projects, 46 rail projects and 34 fixed links). The report recommended different uplifts for different percentile risks of cost overrun. For instance, if the Government was prepared to accept a 50% risk of cost overrun (the '50% percentile') on a road scheme then an uplift of 15% should be applied. But if the Government was only prepared to accept a 20% risk of cost overrun (the '80% percentile') then a higher uplift of 32% was recommended.

Last October, the DfT published new guidance, *The estimation and treatment of scheme costs*, that recommends uplifts for projects at three stages of development *(see table)*. The uplifts should be applied to the 'risk-adjusted cost' of a scheme (that is, the base cost adjusted for risk using Quantified Risk Assessment).

The Stage 2 uplifts are the 50% percentile values recommended by Flyvbjerg/COWI but the Stage 1 and 3 values were recommended by consultant Mott MacDonald in research for the Treasury in 2002 (based on a study of 50 UK infrastructure projects built between 1982 and 2002).

Flyvbjerg/COWI and Mott MacDonald used different techniques to arrive at their values. For instance, the Mott MacDonald values were based on a comparison of outturn costs with estimated costs at the strategic outline case or outline business case, earlier stages in the project cycle than the full business case used by Flyvbjerg/COWI. Furthermore, Mott MacDonald's upper values (Stage 1) represent the average underestimation of costs from the study, without a published probability distribution.

The DfT plans to publish new uplifts for rail schemes shortly (*LTT* 29 Mar).

The DfT's recommended optimism bias uplifts

Type of project	Stage 1	Stage 2	Stage 3
Motorways, trunk, local roads, bicycle facilities, park-and-ride, bus lane schemes, guided buses	44%	15%	3%
Metro, light rail, guided buses on tracks, conventional rail, high speed rail	66%	40%	6%
Bridges and tunnels	66%	23%	6%
Stations and terminal buildings	51%	–	4%
IT projects	200%	–	10%

Notes:
Stage 1 = 'Programme entry' for local authority major schemes; TPI entry/Preferred route decision for Highways Agency schemes; and GRIP Stage 3 – pre-feasibility for rail schemes.
Stage 2 = Conditional approval for local authority schemes; Order Publication/Works commitment for HA schemes; and GRIP Stage 4 – option selection for rail schemes.
Stage 3 = Full approval for local authority schemes; Works Commitment for HA schemes; and GRIP Stage 5 – Design development for rail schemes.

Continued from page 11

and resources are in place to bring it to fruition to time and budget". Meanwhile, a scheme's approved budget includes an additional risk allowance that is 50% of the optimism bias adjustment and local authorities must cover 50% of any expenditure within this allowance. Any additional cost overruns above this risk layer must be met by local authorities in full.

"Hopefully these changes will create improvements in the way projects are planned and delivered and hopefully it will stop bad projects and have the good projects built," says Flyvbjerg. "I don't see that anywhere else in the world. There's an interest in the problem but the UK is way ahead."

"It will be very interesting to go back in some years and study the projects that had optimism bias uplifts applied and see whether they actually came out with more accurate outturn costs as compared to budget," he adds. "If, after a number of projects, it turns out that the majority still have overrun then we are in the bad situation that using the optimism bias uplift is actually an incentive for a further cost overrun."

This is the fear of many project managers who, he says, believe that if funding is available to a project, that funding will be spent. Flyvbjerg says some project managers believe that it is therefore better to underestimate costs than add an uplift. "They say, 'the more we underestimate the cost the more pressure we put on the project organisation and on the contractors to keep costs down. So even if we have a low balled budget (as they call it) we get a cheaper project in the end, even if we have a large cost overrun.'"

Flyvbjerg is not persuaded by this argument and says there are ways to avoid the pitfalls associated with optimism bias uplifts. Governments, for instance, need only release contingency funds for a scheme if the project promoter can provide good documented reasons for needing the extra funds.

Flyvbjerg's latest research is examining whether forecasts produced for Public Private Partnership projects are any more realistic than traditional public sector schemes. "Hopefully and logically one would think that if the private sector starts to scrutinise projects, do due diligence and only invest in the projects that have reliable forecasts, it should produce better results," he says, though he points out that this isn't necessarily borne out by experience. "From the case study evidence we have, there's no evidence that the private sector is better at this," he says, citing the Channel Tunnel by way of example.

He nonetheless supports the use of private capital in infrastructure projects even if it doesn't deliver more accurate forecasts. "One reason I don't hesitate to recommend private capital is that even if the private sector's not better [at cost estimation], at least it's the people who decided to put their money in the project that are losing the money and not the unprotected taxpayer."

Honesty with numbers

Flyvbjerg concedes that the nature of his research work might lead some people to wrongly conclude that he must be opposed to large-scale infrastructure projects – megaprojects as he calls them. "Because we document such a poor track record, people make the wrong assumption that we are against megaprojects," he says, describing that view as "guilt by association".

"I'm all for ambitious projects as long as they can be documented to make sense – as a professional that's what I would say," he says. "As a citizen, as a voter, as a minimum I want politicians who speak clearly and don't misinform me about the projects."

Sometimes his expectations are not met. "Right now we have a big debate about a bridge between Denmark and Germany – a hugely expensive project across the Baltic Sea – and the minister and other project promoters are again and again saying that this will be a good investment, it will be paid back by tolls and so on, even though there are very strong indications that this will not be possible because there is not a lot of traffic in that corridor. So I'm saying [to politicians], why don't you just say you want the bridge, you want to use it for European integration and you don't care whether it's going to lose money for a while. But they don't want to do that, they want to hide behind cost:benefit analyses documenting that they should build it."

We try and teach our students about the interplay between power and rationality. You need to know how to navigate when you are exposed to that.

Teaching the next generation

Flyvbjerg teaches on the Masters course in 'Urban planning and management' at Aalborg and some of his students are likely to find themselves working on major transport projects during their careers. So, to sensitise them to the political pressures they may face, one semester is devoted to the subject of 'Power and Planning'. "Rationality is the concept that is usually emphasised at universities and it should be," says Flyvbjerg. "At the same time we know that the reality that our students work in is not run by rationality alone. We therefore try and teach our students about the interplay between power and rationality. You need to know how to navigate when you are exposed to that."

He says he offers all sorts of advice to his students. "First of all I say if you don't like this you shouldn't get into it. Don't make a professional career out of doing misleading cost-benefit analyses. There are other types of planning where this issue is not as pressing as it is in the field of megaprojects. If you are a municipal planner or a land-use planner you don't have these issues to the same degree. There are similar issues but here [the big transport projects] they are again and again and again – there is much more political and organisational pressure on the big projects.

"If you do get into it, what do you do?" he asks rhetorically. "If a minister asks you to do this, what do you do? Well, you can refuse to do it. You can contact your professional organisation – they actually have rules for this. You can discuss it with colleagues. You can write memos – 'The minister asked me to do it so I felt I had to do it but here's what happened'."

"I don't want to put down every job as a policy-maker or planner or civil servant or anything like that," he stresses. "I think that they're doing a very important job that we need done but there is the problem that we've been talking about here."