How (In)Accurate Is Cost-Benefit Analysis?

Data, Explanations, and Suggestions for Reform

By Bent Flyvbjerg and Dirk W. Bester*

Abstract

Most cost-benefit analyses assume that the estimates of costs and benefits are more or less accurate and unbiased. But what if, in reality, estimates of costs and benefits are highly inaccurate and biased? Then the assumption that cost-benefit analysis is a rational way to improve resource allocation would be a fallacy. Using the largest dataset of its kind, we test the assumption that cost and benefit estimates of infrastructure capital investments are accurate and unbiased. We find this is not the case with overwhelming statistical significance. We document the extent of cost overruns, demand shortfalls, and forecasting bias in infrastructure investments. We further assess whether such forecasting inaccuracies seriously distort effective resource allocation, which is found to be the case. We explain our findings in behavioral terms and explore their implications for infrastructure economics and policy. Finally, we conclude that cost-benefit analysis stands in need of reform, and we outline four steps to such reform.

Keywords: Cost-benefit analysis, cost-benefit fallacy, infrastructure economics, infrastructure investment, infrastructure policy, forecasting, resource allocation, welfare economics, behavioral science, behavioral economics.

Full reference: Flyvbjerg, Bent and Dirk W. Bester, 2022, "How (In)Accurate Is Cost-Benefit Analysis? Data, Explanations, and Suggestions for Reform," in José A. Gómez-Ibáñez and Zhi Liu, eds., *Infrastructure Economics and Policy: International Perspectives* (Cambridge, MA: Lincoln Institute of Land Policy), pp. 174-196, url: https://cup.columbia.edu/book/infrastructure-economics-and-policy/9781558444188.

*) Bent Flyvbjerg is first BT Professor and Inaugural Chair of Major Programme Management at Oxford University and Villum Kann Rasmussen Professor and Chair at the IT University of Copenhagen. Dirk W. Bester holds his DPhil in statistics from the Department of Statistics at the University of Oxford and works as a professional statistician in research and industry. Flyvbjerg developed the ideas for the chapter, did the research, and authored the paper. Bester developed the statistical proofs and did all statistical analyses. Flyvbjerg is corresponding author (flyvbjerg@mac.com).

Data and Methodology

For testing the thesis that cost and benefit estimates are accurate and unbiased,[1] we collected a sample of 2,062 infrastructure investments with data on cost overrun and benefit shortfall. The sample includes six investment types: bridges, buildings, bus rapid transit, dams, power plants, rail, roads, and tunnels. Geographically, the sample incorporates investments in 104 countries on six continents, covering both developed and developing nations. Historically, the data cover almost a century, from 1927 to 2013. Older investments were included to enable analyses of historical trends. For each investment in the dataset, the accuracy of cost estimates was measured by cost overrun (actual divided by estimated cost) while the accuracy of benefit estimates was measured by benefit overrun (actual divided by estimated benefits).

Data on estimated and actual costs and benefits of infrastructure investments are difficult to obtain. No statistical agency or other data service exists from which valid and reliable data may be secured. The data therefore have to be mined item by item from the source, which is time-consuming. All investments that we know of for which data on estimation accuracy were obtainable were considered for inclusion. Data were collected from a variety of sources, including annual accounts, cost and procurement accounts, revenue accounts, auditors' data, questionnaires, interviews, and other studies. Only data that could be supported by reliable documentary evidence were included, in order to avoid the well-known problems with recalled data and interviewee biases. In sum, all investments for which data were considered valid and reliable were included in the sample.

Preferably, costs and benefits would be measured over the full life-cycle of an investment. However, such data are rarely, if ever, available. The convention is therefore to measure the costs of an infrastructure investment by the proxy of construction or capital costs and benefits by the proxy of first-year benefits. This convention is followed here. Estimated costs and benefits are the estimates made at the time of decision to build (sometimes also called the "final investment decision," or FID, based on the final business case), which is the baseline in time from which cost and benefit overrun are measured. Actual costs are measured as recorded outturn costs; actual benefits, as first-year benefits or a later value as close to this as possible, if available and if first-year benefits were not available. Estimated and actual benefits are recorded in the unit of measurement that the planners of the investment decided to use for measuring benefits. Pros and cons of measuring benefits by first-year benefits, and the issue of wider benefits, are discussed in Annex 1.

Ideally, data would be available for both cost and benefit overrun for each investment in the dataset. However, data availability is far from ideal in the measurement of estimated and actual costs and benefits of infrastructure investments. For only 327 investments out of the 2,062 in the sample were data available for both cost overrun and benefit overrun. Using this ideal criterion would therefore result in scrapping large amounts of useful information for the 1,735 other investments in the sample, which would clearly be unacceptable. We therefore decided to run the statistical tests twice, first for the full sample of 2,062 investments and second for the subsample of 327 investments with data available for both cost overrun and benefit overrun for each investment. This gave us the advantage of being able to test whether results were robust across different samples, which proved to be the case.

Investments were included in the sample based on data availability, as mentioned. Using this criterion means that the results of the statistical analyses presented below are probably conservative, i.e., cost overruns in the investment population are most likely larger than in the sample, and benefit overruns smaller. These departures occur because the availability of data often indicates better-than-average investment management and because data from badly performing investments are often not released and are therefore likely to be underrepresented in the sample. This conservative bias must be kept in mind when interpreting results.

Data and methodology are described in more detail in Annex 1.

How Accurate Is Cost-Benefit Analysis?

Table 1 summarizes the results from testing the thesis that cost and benefit estimates are accurate and unbiased. Taking rail as an example, average cost overrun is listed in the table as 1.40, which means that for rail investments, actual costs turned out to be 40 percent higher than estimated costs, on average and in real terms, indicating substantial inaccuracy in cost estimates for rail. Average benefit overrun for rail is listed in the table as 0.66, which is evidence of a benefit shortfall of 34 percent, meaning that on average 34 percent of the estimated passengers never showed up on the actual trains, again indicating substantial inaccuracy.

If cost estimates were largely accurate and unbiased, cost overruns would be narrowly and more or less symmetrically distributed around one. Table 1 clearly shows this thesis to be false, which is verified by statistical tests that reject the thesis at an overwhelmingly high level of statistical significance ($p<0.001$, Wilcoxon test).[2] Cost overruns are highly inaccurate and biased for infrastructure investments, ranging from an average cost overrun for roads of 24 percent to dams at 85 percent, in real terms.[3] The Wilcoxon test was supplemented by tests of skewness (Agostino test) and kurtosis (Anscombe test). The cost overrun data has a skewness of 23.2, and the D'Agostino test confirmed this is significantly different from 0 ($p<0.0001$). The cost overrun data furthermore has a kurtosis of 724.1, and the Anscombe-Glynn kurtosis test found this is significantly different from 3 ($p<0.0001$), which is the kurtosis of a normal distribution, from which we conclude that the data have tails much fatter than the normal distribution. This finding is crucial, because whereas errors cancel out, biases compound. Biases—and especially biases with fat tails, as here—are therefore much worse news than errors in infrastructure investing.

Table 1: Cost-benefit estimates are inaccurate, biased, and compound each other. Accuracy is measured by cost and benefit overruns in 2,062 public investments. Cost and benefit overruns are measured as actual divided by estimated costs and benefits (A/E), respectively, in real terms.

Investment type	**Cost overrun (A/E)**			**Benefit overrun (A/E)**			**p****
	N	**Average**	**p***	**N**	**Average**	**p***	
Dams	243	1.96	< 0.0001	84	0.89	< 0.0001	<0.0001
BRT†	6	1.41	0.031	4	0.42	0.12	0.007
Rail	264	1.40	< 0.0001	74	0.66	< 0.0001	<0.0001
Tunnels	48	1.36	< 0.0001	23	0.81	0.03	0.015
Power plants	100	1.36	0.0076	23	0.94	0.11	0.0003
Buildings	24	1.36	0.00087	20	0.99	0.77	0.01
Bridges	49	1.32	0.00012	26	0.96	0.099	<0.0001
Roads	869	1.24	< 0.0001	532	0.96	< 0.0001	<0.0001
Total	**1603**	**1.39/1.43††**	**< 0.0001**	**786**	**0.94/0.83††**	**< 0.0001**	**<0.0001**

*) The p-value of Wilcoxon test with null hypothesis that the distribution is symmetrically centered around one. The Wilcoxon test was supplemented by tests of skewness (Agostino test) and kurtosis (Anscombe test), which confirmed at an overwhelmingly high level of significance (p<0.0001 for both tests) that the data are skewed and heavy tailed, both for cost and benefit overruns.

**) The p-value of the test with null hypothesis that cost overrun is balanced by benefit overrun (Mann-Whitney test). See main text for explanation.

†) Bus rapid transit.

††) Weighted and unweighted average, respectively.

Source: Flyvbjerg Database.

Table 1 further shows that benefit estimates are also inaccurate, though less so than cost estimates. Bus rapid transit and rail investments have significant benefit shortfalls, on average 58 percent and 34 percent, respectively. In contrast, for bridges, buildings, power plants, and roads, benefit estimates are fairly accurate on average. If benefit estimates were generally accurate, benefit overruns would be narrowly and more or less symmetrically distributed around one. Statistical tests reject this thesis at an overwhelmingly high level of statistical significance ($p<0.001$, Wilcoxon test). Like cost estimates, benefit estimates are inaccurate and biased. Again, the Wilcoxon test was supplemented by tests of skewness and kurtosis, and again these tests confirmed that the data on benefit overrun are skewed and heavy tailed. The benefit overrun data has a skewness of 1.1, and the D'Agostino test confirmed that this significantly differs from 0 ($p<0.0001$). The benefit overrun data furthermore has a kurtosis of 6.5, and the Anscombe-Glynn kurtosis test found this number significantly differs from that of a normal distribution ($p<0.0001$). We see that the kurtosis for cost overrun (724) is much larger than for benefit overrun (6.5), meaning that although both types of overrun are significantly fatter-tailed than the normal distribution, the data on cost overrun have a more extreme tail than the data on benefit overrun.

Considering cost and benefit overrun together, we see that the detected biases work in such a manner that cost overruns are not compensated by benefit overruns, but quite the opposite, on average. This finding is important because not only do errors not cancel out for costs and benefits viewed separately, as documented above, but inaccuracy is also accelerated by the fact that the combination of cost overrun and benefit shortfall hampers the average investment, undermining investment viability on two fronts, that is, for both costs and benefits. Table 1 includes a statistical test of the thesis that cost overrun is balanced by benefit overrun, that is, that errors of cost underestimation are compensated by similar errors of benefit underestimation (the p-values for this test are shown in the right-most column of table 1). We see that the thesis is rejected with overwhelming statistical significance ($p<0.0001$, Mann-Whitney test).[4]

Finally, we tested whether error and bias in cost-benefit estimates have improved over time and found this is not the case. In sum, the data show that the main problem with cost-benefit analysis is not error, but bias. This finding is bad news because where errors tend to cancel out each other because of their randomness, biases are systematic and therefore compound leading to results of cost-benefit analysis that are highly misleading.

It is noteworthy that for not a single of the eight investment types in table 1 did forecasters overestimate cost, and for not a single investment type did they underestimate benefits, on average. That fact shows how strong and consistent the biases are. Theories, like Albert O. Hirschman's (2014) "Hiding Hand" and similar "just-start-digging" theories, which predict that cost underestimates will be offset by benefit underestimates of similar or larger magnitude—and that therefore resource allocation will be okay despite initial errors in cost-benefit estimates—are not supported by the data; they are rejected, again at an overwhelmingly high level of statistical significance ($p<0.0001$) (Flyvbjerg 2016).

We see that the average investment is impaired by a combination of substantial cost underestimates compounded by significant benefit overestimates. Such a systematic and significant bias in cost-benefit analysis is likely to lead to resource misallocation, including initiating investments that ultimately turn out to have negative net benefits and should never have been started, as argued by Ansar et al. (2014) and the World Bank (2010).[5] If investments are large enough and the economies where they are built are fragile, then just one major misallocation of resources in this manner, for a single investment, can negatively affect the national economy for decades, as Brazil and Pakistan learned with their large-dam investments (Ansar et al. 2014) and Greece did with the 2004 Olympic Games (Flyvbjerg and Stewart 2012).

We emphasize that the finding above—that investments are on average undermined by a combination of cost underestimates leading to cost overruns and benefit overestimates resulting in benefit shortfalls—does not mean that no investments exist for which cost underestimates were fortuitously outweighed by similar or larger benefit underestimates. For instance, the German Karlsruhe-Bretten light-rail line, which is in the dataset, had a cost overrun in real terms of 78 percent but an even larger benefit overrun of 158 percent. Similarly, the Danish Great Belt toll-road bridge—the longest suspension bridge in the world at the time of completion—had a cost overrun of 45 percent combined with a benefit overrun of 90 percent, making the investment profitable. Such investments, however, are in the minority. For the vast majority of investments (80 percent), cost overrun is not compensated by benefit overrun, according to our data.

One might speculate, of course, that conceivably the one-fifth of investments where benefit overrun compensated cost overrun may have generated more benefits in the aggregate than the four-fifths that did not. We tested this thesis and found, at an overwhelmingly high level of statistical significance, that (a) not only is this not the case, but (b) the opposite is true since the net effect of the situation where cost overrun is not compensated by benefit overrun (80 percent of investments) is larger than the net effect of the situation where benefit overrun compensated cost overrun ($p=0.001$, two-sided Mann–Whitney test).

Table 1 shows that a typical ex-ante benefit-cost ratio produced by conventional methods is overestimated by between approximately 50 and 200 percent, depending on investment type. Again, this is a statistically significant bias showing that standard cost-benefit analysis consistently overestimates the net benefits of investments by a large margin and therefore may not be trusted.

To conclude, the data show that cost and benefit estimates for infrastructure investments are highly inaccurate and biased. Cost underestimates and benefit overestimates are much more common than cost overestimates and benefit underestimates, at an overwhelmingly high level of statistical significance ($p<0.0001$). The average investment is impaired by a double whammy of substantial cost underestimates compounded by significant benefit overestimates. As a consequence, benefit-cost ratios are overestimated by between 50 and 200 percent, depending on investment type. The data show ex-ante cost-benefit analysis to be so misleading as to be worse than worthless, because decision makers might think they are being informed by such analysis when in fact they are being significantly misinformed about the return on planned investments. As a consequence, decision makers may give the green light to investments that should never have been started, leading to *mis*allocation of resources, where the whole purpose of cost-benefit analysis is the opposite: improved resource allocation. Such false green-lighting points to a central problem for welfare economics and for any type of economics or policy that relies on cost-benefit analysis: ex-ante estimates of costs and benefits are so erroneous and biased that instead of being the powerful tool for effective resource allocation and improved welfare depicted by theory, in practice cost-benefit analysis is a poor and highly misleading guide for policy. Biases creep in and derail the logic and good intentions of theory. To disregard such biases, as is common, is a fallacy. We call it the "cost-benefit fallacy" and add it to the list of other fallacies and biases in human decision making identified by behavioral science in recent years.

The Cost-Benefit Fallacy

We define the cost-benefit fallacy as the situation where individuals behave as if cost-benefit estimates are more or less accurate and unbiased, when in fact they are highly inaccurate and biased. Two questions arise. First, how does one explain the cost-benefit fallacy in more depth? Second, how does one eliminate or overcome the fallacy? In this section we focus on the first question. In the next section, we address the second.

The data presented above show with overwhelming statistical significance that it is not the fact that cost-benefit forecasts are in error that needs explaining, as is the common view for cost-benefit analysis, if the question of accuracy comes up at all (which it mostly does not). The fact that needs explaining is that the vast majority of cost-benefit forecasts are systematically biased, with underestimation for cost and overestimation for benefits. Our data go back 86 years, and for this period the bias in cost-benefit forecasts has been constant. Cost-benefit forecasters are "predictably irrational" as regards bias, in the words of Ariely (2009). To begin to understand why, consider the following example.

Recently, the CEO of one of the biggest and most successful infrastructure providers in the world explained to us why the cost forecast for one of their multi-billion-dollar investments—a high-speed rail line—had proven too low, resulting in significant cost overrun. There were three causes of the underestimate and overrun, according to the CEO. First, construction had taken longer than scheduled. Second, the investment, which involved extensive tunneling, had run into unexpected geological conditions, resulting in unplanned costs. Third, price inflation for both labor and materials had been higher than expected, the CEO explained. This is a typical and plausible way to account for cost overrun, which most would accept. Delays, geology, and price inflation are commonly stated causes of cost overrun in construction investments, as are complexity, scope changes, bad weather, archeology, and the like. This is the industry standard, in terms of explanation.

We analyzed the evidence for the specific investment and assured the CEO he was right. On the surface of things delays, geology, and price inflation caused the cost overrun. However, he would neither truly understand nor solve the problem as long as he saw it like this, or so we argued. To really understand what happened to his company's investment, he would need to think in terms of root causes. And at the level of root causes, the explanation of the overrun is much simpler, with a single source: optimism. His company was optimistic about the schedule, in assuming it could deliver the rail line several years faster than is usual for this type of investment, without having good reasons for this assumption. Second, the company was optimistic about the geological conditions, without having investigated sufficiently. Third, it was optimistic about price variations, assuming variations would be small when in fact history shows they are large. This is optimism, pure and simple, unless the company deliberately misrepresented the schedule, geology, and prices, in which case it would be strategic misrepresentation, we further argued.[6] In either case, the problem is a behavioral one, related to the company's own conduct. To understand and solve the problem, the company would need behavioral science, not a better understanding of geology or market prices or better Gantt charts, or so we advised the CEO.

Behavioral scientists would agree that schedules, geology, market prices, scope changes, and complexity are relevant to understanding what goes on in infrastructure investment projects but would not see them as root causes of inaccuracy and bias in cost-benefit forecasts. The root cause of cost underestimates and overruns, according to behavioral science, is the well-documented fact that planners and managers keep underestimating the importance of schedules, geology, market prices, scope changes, and complexity in investment after investment. From the point of view of behavioral science, the mechanisms of scope changes, complex interfaces, archaeology, geology, bad weather, business cycles, etc. are not unknown to planners of infrastructure investments, just as planners know that such mechanisms may be mitigated. However, planners often underestimate these mechanisms and overestimate the effectiveness of mitigation measures, due to well-known behavioral phenomena like overconfidence bias, the planning fallacy, and strategic misrepresentation.

In behavioral terms, scope changes and so on are manifestations of such misjudgment on the part of planners, and in this sense planners' behavior is the root cause of inaccuracy and bias in cost-benefit forecasts. But because scope changes and so on are more visible than the underlying root causes, they are often mistaken for the cause of inaccuracy and bias. In behavioral terms, with scope changes as an example, the causal chain starts with human bias (deliberate or not), which leads to underestimation of scope during planning, which leads to unaccounted-for scope changes during delivery, which lead to cost overrun. Scope changes are an intermediate stage in this causal chain through which the root causes manifest themselves. Similar reasoning applies for complexity, geology, and other so-called causes. With behavioral science, we say to infrastructure planners and managers, "Your biggest risk is you." It is not scope changes, complexity, etc. in themselves that are the main problem; it is how planners and managers misconceive and underestimate these phenomena through overconfidence bias, the planning fallacy, strategic misrepresentation, etc. Behavioral science brings this profound and proven insight to cost-benefit forecasting and infrastructure investment planning, but it is unfortunately not always readily acknowledged and integrated in cost-benefit scholarship and practice (Flyvbjerg et al. 2018).

Behavioral science entails a change of perspective: The problem with cost-benefit forecasts is not error but bias; as long as we try to understand and solve the problem as something it is not (error), we will not solve it. Forecasts, policies, and decisions need to be de-biased, which is fundamentally different from eliminating error (Kahneman et al. 2011; Flyvbjerg 2008; 2013a). The main problem is also not cost overrun, even if overrun is what hurts, is visible, and therefore gets the attention. The main problem is cost underestimation. Overrun is a consequence of underestimation, with the latter happening upstream from overrun, often years before overruns manifest. Again, if one tries to understand and solve the problem as something it is not (cost overrun), one will fail. We need to solve the problem of cost underestimation to solve the problem of cost overrun.[7] Until we understand these basic insights from behavioral science, we are unlikely to get cost-benefit forecasting and resource allocation right. As long as we allow ourselves to be blinded by optimism bias, overconfidence bias, probability neglect, and other behavioral biases, we will keep reproducing the cost-benefit fallacy. The solution to overcoming the fallacy is therefore fairly straightforward: de-biasing.[8]

Four Steps to Cost-Benefit Reform

Cost-benefit analysis may, no doubt, be a helpful tool in infrastructure economics and policy. However, cost-benefit analysis as practiced today is less than useful because it is highly biased, as documented above. Cost-benefit theory and practice were developed long before behavioral science and have yet to adapt to its findings about the nature and causes of bias. This lack of adaption needs to change. We saw above that for infrastructure investments, the biases are so substantial that the average cost-benefit analysis results in resource *mis*allocation, instead of contributing to the effective use of scarce assets, which is the whole point of cost-benefit analysis.[9] For cost-benefit analysis to become useful in infrastructure economics and policy, the following need to happen:

1. Systematic and effective de-biasing of cost-benefit forecasts;
2. Introduction of skin-in-the-game for cost-benefit forecasters;
3. Independent audits of cost-benefit forecasts; and
4. Adaption of cost-benefit forecasting to the messy, non-expert character of democratic decision-making.

First, and most importantly, behavioral science predicts that any forecast—including cost-benefit estimates—is prone to bias. If forecasts are biased, they must be de-biased before they can be reliably used in policy making and investment decisions, or policies and decisions will be biased, too. Nudges will not suffice. Deliberate and precise de-biasing will be needed. Taking our clue from Kahneman and Tversky (1979) and using Kahneman's (2011, 245) so-called "outside view," we developed methods for such de-biasing for infrastructure investments. We use the distributional information about actual estimation errors in previous investments (established via ex-post studies) to precisely assess by how much the cost-benefit estimates for a planned venture must be adjusted before they may be considered effectively de-biased (Flyvbjerg et al. 2004; Flyvbjerg 2006). In his book *Thinking, Fast and Slow*, Daniel Kahneman reviews this work and especially our advice to use empirical distributional information for de-biasing. Kahneman concludes:

> This may be considered the single most important piece of advice regarding how to increase accuracy in forecasting [of costs and benefits] through improved methods. Using such distributional information from other ventures similar to that being forecasted is called taking an "outside view" and is the cure to the planning fallacy [and thus to bias] . . . . The outside view is implemented by using a large database, which provides information on both plans and outcomes for hundreds [now thousands] of projects all over the world, and can be used to provide statistical information about the likely overruns of cost and time, and about the likely underperformance of projects of different types. (Kahneman 2011, 251-52)

The method for systematically (mathematically and statistically) taking the outside view is called "reference class forecasting" and is today used at scale in infrastructure investing around the world, including in countries where the method has been made mandatory, such as the UK and Denmark (UK Department for Transport 2006; Danish Ministry for Transport and Energy 2006; 2008). Independent evaluations of the method confirm its value and accuracy (Awojobi and Jenkins 2016; Batselier 2016; Batselier and Vanhoucke 2017; Bordley 2014; Chang et al. 2016; Kim et al. 2011; Liu and Napier 2010; Liu et al. 2010).[10] Only with such de-biasing will cost-benefit forecasts be accurate and help allocate resources effectively.

Second, better incentives for accuracy in cost-benefit forecasts should be introduced. Better methods alone will not solve the problem. Akerlof and Shiller (2009, 146) suggest "firing the forecaster" when forecasts are very wrong and the consequences severe. Flyvbjerg (2013, 771-772) goes one step further in proposing "suing the forecaster" in cases of gross neglect and gross deliberate manipulation of forecasts and investors. The first court cases in infrastructure investing were decided in 2015, for Sydney's Lane Cove toll tunnel, which went bankrupt when more than half the forecasted cars never appeared in the tunnel, and for Brisbane's Clem 7 tunnel and Airport Link, which had comparable problems (Rubin 2017; Saulwick 2014; Worthington 2012). More litigation quickly followed in the United States, treating traffic forecasters as criminals (Dezember and Glazer 2013; Evans 2010; Hals 2013; Miller 2013; Singh 2017; Wright 2014). Recently, those responsible for the Muskrat Falls dam inquiry in Canada decided to report executives to the police for possible criminal charges related to cost underestimation for the dam (LeBlanc 2020; Roberts 2020). Criminalizing planners and cost-benefit analysts this way has sent shock-waves through the global forecasting and cost-benefit community, contributing to much-needed discipline and accountability. Here we suggest, as a more general and more moderate heuristic, that often it would make sense to consider giving cost-benefit forecasters skin in the game. Lawmakers and policymakers should develop institutional setups that reward forecasters who get their estimates right and punish those who do not. We should not be surprised that cost-benefit estimates are wrong if forecasters have no incentive to get them right. True, forecasters are supposed to be neutral and unbiased. But to confuse such normativity with reality is naive, for policymakers and scholars alike.

Third, independent audits are needed to ensure that points one and two above—better methods and better incentives for more accurate forecasts—work according to intention and to adjust them when they do not. Auditing must be effectively separated from political motivations to avoid the type of political distortion of cost-benefit forecasts described in detail by the World Bank (2010), Harrington (2006), Flyvbjerg et al. (2002; 2004), Harrington et al. (2000), and Wachs (1986; 1989; 1990; 2013). If this separation does not happen, audits will not and cannot serve their purpose. Audits should allow for transparency and external scrutiny, including scrutiny by independent experts, the public, and media.

Finally, and perhaps most importantly, for cost-benefit analysis to be accepted and have impact, it must be understood and practiced *not* in the top-down, technocratic fashion that is common but instead in ways that fit with the messy, non-expert character of present-day democratic decision-making. Here cost-benefit analysis is just one of many inputs that are amalgamated in the overall decision-making process. Top-down, technocratic approaches are anachronistic and will not work in this context. Kahneman (2011, 141-42) contrasts cost-benefit technocrats with cost-benefit democrats, exemplifying technocrats with Harvard law professor Cass Sunstein and democrats with University of Oregon psychology professor Paul Slovic. Kahneman writes about Sunstein, who is a self-declared cost-benefit technocrat and whose work Kahneman otherwise holds in high regard:

> He [Sunstein] starts from the position that risk regulation and government intervention to reduce risks should be guided by rational weighting of costs and benefits , , , , he has faith in the objectivity that may be achieved by science, expertise, and careful deliberation . . . . Cass Sunstein would seek mechanisms [like the cost-benefit principle] that insulate decision makers from public pressures, letting the allocation of resources be determined by impartial experts who have a broad view of all risks and of the resources available to reduce them. (Kahneman 2011)

Kahneman here captures the key beliefs and a priori assumptions of cost-benefit technocrats, including (a) the view that cost-benefit analysis is a force for good in deciding policies; (b) faith in the objectivity of science as a basis for cost-benefit calculations; and (c) belief in impartial experts for making better decisions. Kahneman (2011, 140-45) juxtaposes this position with that of Slovic and other cost-benefit democrats, who, Kahneman (144) says, "trusts the experts much less and the public somewhat more than Sunstein does." Kahneman approvingly cites Slovic (2000) for arguing that expert-based decision-making, like that supported by cost-benefit technocrats, produces policies that the public will reject, which is an untenable situation in a democracy, according to Kahneman. "Slovic rightly stresses the resistance of the public to the idea of decisions being made by unelected and unaccountable experts," concludes Kahneman (2011, 144) with Slovic.

Kahneman (2011, 144) considers both Sunstein and Slovic "eminently sensible"; he explicitly states, "I agree with both." We agree with Kahneman (Flyvbjerg and Bester, forthcoming). Nobody in their right mind would argue that expert input, science, and cost-benefit analysis are undesirable in deciding which policies and investments to pursue in a society, including for infrastructure. But, similarly, no one in their right mind would depend on only experts in making such decisions, especially when we know that experts are subject to the cost-benefit fallacy and produce, over and over, the kind of biased cost-benefit analyses documented above. Sunstein and other cost-benefit technocrats overemphasize the importance of experts and place too much trust in them. By supplementing Sunstein with Slovic—technocracy with democracy—Kahneman (2011, 144-45) develops a more balanced position, from which he argues that democracy "is inevitably messy" and emphasizes that the public's concerns, "even if they are unreasonable, should not be ignored by policy makers." Experts, science, and cost-benefit analyses are not enough to arrive at high-quality policy decisions, argues Kahneman. The messy process of democracy is called for to merge all concerns, including unreasonable ones, because only in this manner will policy be able to muster the support necessary to be successful.

The four measures for better cost-benefit analysis outlined above are not ivory tower theory. They have recently been implemented in innovative, full-scale experiments around the world, with encouraging results, as documented by Kahneman (2011, 249-53), Flyvbjerg et al. (2016; 2018), HM Treasury (2003), the UK Department for Transport (2006), the Danish Ministry for Transport and Energy (2006; 2008), and the Swiss Association of Road and Transportation Experts (2006). More such work is in the pipeline, finally bringing the findings of behavioral science to welfare economics and cost-benefit forecasting, resulting in more realistic and more useful forecasts.

Summary and Conclusions

In this chapter, we set out to test the accuracy of cost-benefit forecasts in capital infrastructure investments. We found a fallacy at the heart of conventional cost-benefit analysis: Forecasters, policymakers, and scholars tend to assume that cost-benefit forecasts are more or less accurate, when in fact they are highly inaccurate and biased, at an overwhelmingly high level of statistical significance. In infrastructure investing, the fallacy results in average overestimates in ex-ante benefit-cost ratios of 50 to 200 percent, depending on investment type. Individual estimates of benefit-cost ratios are routinely off by much more than these averages.

Cost-benefit analysis can undoubtedly be a useful tool in infrastructure economics and policy, but only after the cost-benefit fallacy has been acknowledged and corrected for. For this to happen, the following must become part and parcel of cost-benefit practice: (a) new methods for effective de-biasing of ex-ante cost-benefit estimates, based on behavioral science; (b) incentives for using these methods, including skin-in-the-game; (c) independent audits to check that the methods and incentives work according to intention; and, finally, (d) integration of results in messy, real-life democratic decision-making processes, which consider other concerns than the results of cost-benefit analyses. The good news is that the theoretical and methodological rationale for this type of behaviorally based cost-benefit analysis has already been developed and is finding good use in policy and practice. The bad news is that the cost-benefit fallacy seems to be as ingrained in human behavior as the many other biases and fallacies identified by behavioral science and will therefore be as difficult to root out as these. That obstacle should not keep us from trying, needless to say, which is what we have done with the research reported in the present chapter.

Acknowledgments

We wish to thank Tony Gómez-Ibáñez, Zhi Liu, and Edward Glaeser for their penetrating and highly useful comments and suggestions for improvements to an earlier version of the chapter. Further, we would like to thank our colleagues Atif Ansar, Alexander Budzier, and Chantal Cantarelli for kindly helping to collect parts of the data used in the statistical analyses.

Annex 1: Sample and Statistical Tests

The sample for testing accuracy and bias in cost-benefit analysis includes 2,062 infrastructure investment projects across eight investment types: bridges, buildings, bus rapid transit, dams, power plants, rail, roads, and tunnels. Geographically, the investments are located in 104 countries on six continents, including both developed and developing nations. Historically, the data cover the period from 1927 to 2013. Older investments were included to enable analyses of historical trends. For each investment in the dataset, the accuracy of cost estimates was measured by cost overrun (actual divided by estimated cost, in real terms) while the accuracy of benefit estimates was measured by benefit overrun (actual divided by estimated benefits).

Data on estimated and actual costs and benefits of infrastructure investments are difficult to obtain. No statistical agency or other data service exists from which valid and reliable data may be secured. The data therefore have to be mined item by item from the source, which is extremely time-consuming. Data collection for the present study systematically followed international standards. Investments were included in the sample on the basis of data availability. We considered for inclusion all investments that we knew for which data on estimation accuracy were obtainable. Data were collected from a variety of sources, including annual accounts, cost and procurement accounts, revenue accounts, auditors' data, questionnaires, interviews, and other studies. Only data that could be supported by reliable documentary evidence were included to avoid the well-known problems with recalled data and interviewee biases. Even so, substantial amounts of data had to be rejected due to insufficient data quality, approximately 25 percent of cost data and 50 percent of benefit data.

Recently, we and colleagues discovered that some ex-post studies of cost forecasting have been contaminated by significant statistical errors, for instance in pooling data with incompatible baselines and excluding outliers that should have been included (Flyvbjerg et al., 2018, 2019). Such studies cannot be trusted. The same holds for meta-studies that included these studies without being aware of their shortcomings. This unhappy situation—which signifies the arrival of junk science in infrastructure scholarship—is accounted for in Flyvbjerg et al. (2018, 2019) together with recommendations on how to avoid the problem going forward. The importance of rooting out studies with faulty data and statistics from the body of work in infrastructure scholarship cannot be overemphasized if the field wants to enjoy continued credence in the academy, policy making, and with the general public. Data from studies that proved contaminated are not included in the present research, needless to say.

In sum, we included in our sample all investments for which data were considered valid and reliable. Data collection and the dataset are described in more detail in Flyvbjerg et al. (2002; 2005) and Flyvbjerg (2016).

Preferably, costs and benefits would be measured over the full life-cycle of an investment. However, such data are rarely, if ever, available. Convention is therefore to measure the costs of an infrastructure investment by the proxy of construction or capital costs, and benefits by the proxy of first-year benefits. This convention is followed here. Estimated costs and benefits are the estimates made at the time of decision to build (sometimes also called the "final investment decision," or FID, based on the final business case), which is the baseline in time from which cost and benefit overrun are measured. Actual costs are measured as recorded outturn costs; actual benefits are measured as first-year benefits, or a later value as close to this as possible, if available and if first-year benefits were not available. Estimated and actual benefits are recorded in the unit of measurement that the planners of the investment decided to use for measuring benefits.

First-year benefits may seem a narrow proxy to use for benefits, and it has been criticized as such. In fact, however, first-year benefits have proven a reliable proxy for overall benefits, which is fortunate because the existence of benefits data for later years is so rare that benefit measurement would be rendered impossible for large samples if one had to rely on later-year data. For investments for which data are available on estimated and actual benefits covering more than one year after operations began, it turns out that investments with lower-than-estimated benefits during the first year of operations also tend to have lower-than-estimated benefits in later years (Flyvbjerg 2013, 766–767). Using the first year as the basis for measuring benefits therefore appears rarely to result in the error of identifying investments as underperforming in terms of benefits that would not be identified as such if a different time period were used as the basis for comparison. Actual benefits do not seem to quickly catch up with estimated benefits for investments with overestimated first-year benefits, and mostly they never do. Ramp-up of benefits is commonly assumed but often does not happen or occurs only partly. To take a typical example, for the Channel tunnel between the UK and France, more than five years after opening to the public, train passengers numbered only 45 percent of those forecasted for the opening year; rail freight traffic was 40 percent of the forecast; and actual numbers had not caught up with the forecasts after 20 years, and they never have. In conclusion, we would prefer to measure benefits for all years, but data for this are generally unavailable. Given lack of such data, the good news is that first-year benefits appear to be a fair proxy, with data more readily available than for later years.

Using first-year benefits has also been criticized for not considering wider development effects, for instance increased real estate values following from improved transport services. Such effects undoubtedly exist. However, if transport infrastructure has wider benefits, these must be expected to be roughly proportional to traffic, so that if traffic has been overestimated, the development benefits must also have been, which means that the case for proceeding with the investment was exaggerated. The difference between estimated and actual traffic, including first-year patronage, would therefore be a good proxy for assessing the extent of the problem. The same applies to other types of infrastructure. For the full argument and for further documentation regarding wider benefits, see Vickerman (2017) and Flyvbjerg (2005; 2013).

Ideally, data would be available for both cost and benefit overruns for each investment in the dataset. However, data availability is far from ideal in the measurement of estimated and actual costs and benefits of infrastructure investments. For only 327 investments out of the 2,062 in the sample were data available for both cost and benefit overrun. Using this ideal criterion would therefore result in scrapping large amounts of useful information for the 1,735 other investments in the sample, which would clearly be unacceptable. We therefore decided to run the statistical tests twice, first for the full sample of 2,062 investments and second for the subsample of 327 investments with data available for both the cost overrun and the benefit overrun for each investment. This approach gave us the advantage of being able to test whether results were robust across different samples, which proved to be the case.

Investments were included in the sample based on data availability, as mentioned. This means that the results of statistical analyses presented in the main text are probably conservative, i.e., cost overruns in the investment population are most likely larger than in the sample, and benefit overruns smaller. One explanation of conservatism bias is that availability of data often indicates better-than-average investment management. Another is that data from badly performing investments are often not released and are therefore likely to be underrepresented in the sample. The existence of such conservatism must be kept in mind when interpreting results. See further Flyvbjerg et al. (2002, 2005) and Flyvbjerg (2005, 2016).

Standard statistical tests are described in the main text. In addition to these, we tested results for the influence of investment type, geography, and time using Bayesian modeling.[11] We found only few significant differences across investment type, geography, and time; none of them ran counter to the main conclusion that cost and benefit estimates are inaccurate and biased and compound each other. This finding is unsurprising, given the overwhelmingly high level of statistical support for the main conclusion. In sum, results from the statistical tests proved robust across both conventional and Bayesian testing and across different samples.

References

Akerlof, George A., and Robert J. Shiller. 2009. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism.* Princeton: Princeton University Press.

Albalate, Daniel, and Germa Bel. 2014. *The Economics and Politics of High-Speed Rail.* New York: Lexington Books.

Altshuler, Alan and David Luberoff. 2003. *Mega-Projects: The Changing Politics of Urban Public Investment.* Washington, DC: Brookings Institution Press.

Ansar, Atif, Bent Flyvbjerg, Alexander Budzier, and Daniel Lunn. 2014. "Should We Build More Large Dams? The Actual Costs of Hydropower Megaproject Development." *Energy Policy* 69 (March): 43--56.

Ariely, Dan. 2010. *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* New York, NY: HarperCollins.

Awojobi, Omotola, and Jenkins, Glenn P. 2016. "Managing the Cost Overrun Risks of Hydroelectric Dams: An Application of Reference Class Forecasting Techniques." *Renewable and Sustainable Energy Reviews* (63): 19--32.

Bain, Robert. 2009. *Toll Road Traffic and Revenue Forecasts: An Interpreter's Guide.* (Published by author).

Bain, Robert, and Michael Wilkins. 2002. *Traffic Risk in Start-up Toll Facilities.* London: Standard and Poor's.

Batselier, Jordy. 2016. *Empirical Evaluation of Existing and Novel Approaches for Project Forecasting and Control*. PhD diss. University of Ghent.

Batselier, Jordy, and Vanhoucke, Mario. 2017. "Improving Project Forecast Accuracy by Integrating Earned Value Management with Exponential Smoothing and Reference Class Forecasting." *International Journal of Project Management* 35(1): 8–43.

Bordley, Robert F. 2014. "Reference Class Forecasting: Resolving Its Challenge to Statistical Modeling." *The American Statistician* 68(4): 221--229.

Chang, Welton, Eva Chen, Barbara Mellera, and Philip Tetlock. 2016. "Developing Expert Political Judgment: The Impact of Training and Practice on Judgmental Accuracy in Geopolitical Forecasting Tournaments." *Judgment and Decision Making* 11(5): 509--526.

Cantarelli, Chantal C., Bent Flyvbjerg, and Søren L. Buhl. 2012. "Geographical Variation in Project Cost Performance: The Netherlands versus Worldwide." *Journal of Transport Geography* 24: 324–331.

Danish Ministry for Transport and Energy, Transport- og Energiministeriet. 2006. *Aktstykke om nye budgetteringsprincipper* (Act on New Principles for Budgeting). Aktstykke nr. 16, Finansudvalget, Folketinget, Copenhagen, October 24.

Danish Ministry for Transport and Energy, Transport- og Energiministeriet. 2008. "Ny anlægsbudgettering på Transportministeriets område, herunder om økonomistyrings–model og risikohåndtering for anlægsprojekter," Copenhagen, November 18.

Dantata, Nasiru A., Ali Touran, and Donald C. Schneck. 2006. "Trends in US Rail Transit Project Cost Overrun." *Transportation Research Board Annual Meeting*, Washington, DC: National Academies.

Dezember, Ryan, and Emily Glazer. 2013. "Drop in Traffic Takes Toll on Investors in Private Roads." *Wall Street Journal*, 20 November, [ ], https://www.wsj.com/articles/SB10001424052702303482504579177890461812588, retrieved December 17, 2020..

Evans, Anthony G. 2010. "Declaration of Anthony G. Evans, Chief Financial Officer of South Bay Expressway, L.P., in Support of the Debtors' Chapter 11 Petitions and First Day Motions," *case 10-04516-LA11*, filed 22 March 2010, United States Bankruptcy Court Southern District of California.

Federal Transit Administration. 2003. *Predicted and Actual Impacts of New Starts Projects: Capital Cost, Operating Cost and Ridership Data*, Office of Planning and Environment with support from SG Associates, Inc. Washington, DC: Federal Transit Administration.

Federal Transit Administration. 2008. *Before and After Studies of New Starts Projects*. Report to Congress. Washington, DC: Federal Transit Administration.

Federal Transit Administration. 2013. *Before and After Studies of New Starts Projects*. Report to Congress. Washington, DC: Federal Transit Administration.

Flyvbjerg, Bent. 2005. "Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp Up and Sampling." *Transportation Research A* 39(6): 522--530.

Flyvbjerg, Bent. 2006. "From Nobel Prize to Project Management: Getting Risks Right." *Project Management Journal* 37(3): 5--15.

Flyvbjerg, Bent. 2007. "How Optimism Bias and Strategic Misrepresentation in Early Project Development Undermine Implementation." In *Beslutninger på svakt informasjonsgrunnlag: Tilnærminger og utfordringer i projekters tidlige fase* (Decisions based on weak information: Approaches and challenges in the early phase of projects), ed., Kjell J. Sunnevåg, 41—55. Trondheim, Norway: Concept Program, The Norwegian University of Science and Technology.

Flyvbjerg, Bent. 2008. "Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice." *European Planning Studies* 16(1): 3--21.

Flyvbjerg, Bent. 2013. "Quality Control and Due Diligence in Project Management: Getting Decisions Right by Taking the Outside View." *International Journal of Project Management* 31(5): 760–774.

Flyvbjerg, Bent. 2014. "What You Should Know about Megaprojects and Why: An Overview," *Project Management Journal* 45(2): 6--19.

Flyvbjerg, Bent. 2016. "The Fallacy of Beneficial Ignorance: A Test of Hirschman's Hiding Hand," *World Development* 84(May): 176–189.

Flyvbjerg, Bent, and Alexander Budzier. 2011. "Why Your IT Project May Be Riskier than You Think." *Harvard Business Review* September, 601--603.

Flyvbjerg, Bent, Atif Ansar, Alexander Budzier, Søren Buhl, Chantal Cantarelli, Massimo Garbuio, Carsten Glenting, Mette Skamris Holm, Dan Lovallo, Daniel Lunn, Eric Molin, Arne Rønnest, Allison Stewart, Bert van Wee, 2018, "Five Things You Should Know about Cost Overrun," *Transportation Research Part A: Policy and Practice* 118 (December): 174--190.

Flyvbjerg, Bent, Atif Ansar, Alexander Budzier, Søren Buhl, Chantal Cantarelli, Massimo Garbuio, Carsten Glenting, Mette Skamris Holm, Dan Lovallo, Eric Molin, Arne Rønnest, Allison Stewart, and Bert van Wee. 2019. "On De-bunking 'Fake News' in the Post-Truth Era: How to Reduce Statistical Error in Research." *Transportation Research Part A: Policy and Practice* 126 (August): 409--411.

Flyvbjerg, Bent, and Dirk W. Bester. in progress. "The Cost-Benefit Revolution: Fact or Fallacy?"

Flyvbjerg, Bent, Carsten Glenting, and Arne Rønnest. 2004. *Procedures for Dealing with Optimism Bias in Transport Planning: Guidance Document.* London: UK Department for Transport.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Søren L. Buhl. 2002. "Underestimating Costs in Public Works Projects: Error or Lie?" *Journal of the American Planning Association* 68 (3 Summer): 279--295.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Søren L. Buhl. 2005. "How (In)accurate Are Demand Forecasts in Public Works Projects? The Case of Transportation." *Journal of the American Planning Association* 71(2 Spring): 131--146.

Flyvbjerg, Bent, Chi-keung Hon, and Wing Huen Fok. 2016. "Reference Class Forecasting for Hong Kong's Major Roadworks Projects." *Proceedings of the Institution of Civil Engineers* 169 (November CE6): 17—24.

Flyvbjerg, Bent, and Allison Stewart. 2012. "Olympic Proportions: Cost and Cost Overrun at the Olympics 1960--2012." Working Paper, Saïd Business School, University of Oxford.

Fouracre, P. R., R. J. Allport, and J. M. Thomson. 1990. *The Performance and Impact of Rail Mass Transit in Developing Countries*, Research Report no. 278. Crowthorne, U.K.: Transport and Road Research Laboratory.

Gao, Nan and Ali Touran. 2020. "Cost Overruns and Formal Risk Assessment Program in US Rail Transit Projects," *Journal of Construction Engineering and Management* 146(5): pp. 1-11.

HM Treasury. 2003. *The Green Book: Appraisal and Evaluation in Central Government*. Treasury Guidance. London: TSO.

Hals, Tom. 2013. "Detroit Windsor Tunnel Operator Files for Bankruptcy." *Reuters*, 25 July, .

Harrington, Winston. 2006. "Grading Estimates of the Benefits and Costs of Federal Regulation." *Discussion Paper* RFF DP 06-39. Washington, DC: Resources for the Future.

Harrington, Winston, Richard D. Morgenstern, and Peter Nelson. 2000. "On the Accuracy of Regulatory Cost Estimates." *Journal of Policy Analysis and Management* 19(2): 297--322.

Hirschman, Albert O. 2014. "The Principle of the Hiding Hand," originally published in *The Public Interest*, Winter 1967, pp. 10-23. In *Megaproject Planning and Management: Essential Readings*, vol. 1, Bent Flyvbjerg, ed., 149--162. Cheltenham, UK and Northampton, MA: Edward Elgar.

Huo, Tengfei, Hong Ren, Weiguang Cai, Geoffrey Qiping Shen, Bingsheng Liu, Minglei Zhu, and Hengqin Wu. 2018. "Measurement and Dependence Analysis of Cost Overruns in Megatransport Infrastructure Projects: Case Study in Hong Kong." *Journal of Construction Engineering and Management* 144(3): 1--10.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York, NY: Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. "Intuitive Prediction: Biases and Corrective Procedures." In *Studies in the Management Sciences: Forecasting*, vol. 12, ed. S. Makridakis and S. C. Wheelwright, 313--327. Amsterdam: North Holland.

Kim, Byung-Cheol, and Kenneth F. Reinschmidt. 2011. "Combination of Project Cost Forecasts in Earned Value Management." *Journal of Construction Engineering and Management* 137 (11): 958--966.

Leavitt, Dan, Sean Ennis, and Pat McGovern. 1993. "The Cost Escalation of Rail Projects: Using Previous Experience to Re-evaluate the CalSpeed Estimates," Working paper 567. Berkeley: Institute of Urban and Regional Development, University of California.

LeBlanc, Richard D. 2020. *Muskrat Falls: A Misguided Project*, vols. 1--6. Province of Newfoundland and Labrador, Canada: Commission of Inquiry Respecting the Muskrat Falls Project,.

Lee, Jin-Kyung. 2008. "Cost Overrun and Cause in Korean Social Overhead Capital Projects: Roads, Rails, Airports, and Ports." *Journal of Urban Planning and Development* 134 (2 June) 59--62.

Liu, Li, and Zigrid Napier. 2010. "The Accuracy of Risk-Based Cost Estimation for Water Infrastructure Projects: Preliminary Evidence from Australian Projects." *Construction Management and Economics* 28(1): 89--100.

Liu, Li, George Wehbe, and Jonathan Sisovic. 2010. "The Accuracy of Hybrid Estimating Approaches: A Case Study of an Australian State Road and Traffic Authority." *The Engineering Economist* (55): 225--245.

Miller, Eric. 2013. "American Roads LLC Files for Bankruptcy." *Transport Topics*, 5 August, https://www.ttnews.com/articles/american-roads-llc-files-bankruptcy-0, retrieved December 17, 2020.

National Audit Office. 1992. *Department of Transport: Contracting for Roads.* London: National Audit Office.

Nijkamp, Peter, and Barry Ubbels. 1999. "How Reliable are Estimates of Infrastructure Costs? A Comparative Analysis." *International Journal of Transport Economics* 26(1): 23--53.

Pickrell, Don. 1990. *Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost.* Washington, DC: US Department of Transportation.

Pickrell, Don. 1992. "A Desire Named Streetcar: Fantasy and Fact in Rail Transit Planning." *Journal of the American Planning Association* 58(2): 158--176.

Plummer, M. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing,* https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf, retrieved December 17, 2020.

Plummer, M. 2012. "rjags: Bayesian Graphical Models Using MCMC." R package version 3--7.

R Core Team. 2012. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Riksrevisionen. 2011. "Kostnadskontroll i stora järnvägsinvesteringar?" Swedish National Audit Bureau Audit, Report, RiR 6. Stockholm: Riksrevisionen.

Riksrevisionsverket. 1994. "Infrastrukturinvesteringar: En kostnadsjämförelse mellan plan och utfall i 15 större projekt inom Vägverket och Banverket", RRV23. Stockholm: Riksrevisionsverket.

Roberts, Terry. 2020. "Scathing Muskrat Falls Inquiry Report Lays Blame on Fired Executives." *CBC News*, March 10, https://www.cbc.ca/news/canada/newfoundland-labrador/muskrat-falls-inquiry-misguided-project-1.5492169?__vfz=medium%3Dsharebar,.

Rubin, Debra K. 2017. "Suit over Failed Brisbane P3 Focuses on Traffic Projections." *Engineering News-Record*, 18 October.

Saulwick, Jacob. 2014. "Trial to Start on $144 Million Lane Cove Tunnel Debacle." *Sydney Morning Herald*, 10 August.

Singh, Anil. C. 2017. "Syncora Guar. Inc. v. Alinda Capital Partners LLC," *2017 NY Slip Op 30288(U)*, Supreme Court, New York Country: Docket Number 651258/2012.

Slovic, Paul. 2000. *The Perception of Risk.* Sterling, VA: EarthScan.

Swiss Association of Road and Transportation Experts. 2006. *Kosten-Nutzen-Analysen im Strassenverkehr*, Grundnorm 641820, valid from August 1. Zürich: Swiss Association of Road and Transportation Experts.

UK Department for Transport. 2006. The Estimation and Treatment of Scheme Costs: Transport *Analysis Guidance*, TAG Unit 3.5.9. London: Department for Transport.

Vickerman, Roger. 2017. "Wider Impacts of Megaprojects: Curse or Cure?" In *The Oxford Handbook of Megaproject Management,* ed. Bent Flyvbjerg, 389--405, Oxford: Oxford University Press.

Walmsley, D. A., and M. W. Pickett. 1992. *The Cost and Patronage of Rapid Transit Systems Compared with Forecasts*, Research Report 352. Crowthorne, UK: Transport Research Laboratory.

Wachs, Martin. 1986. "Technique vs. Advocacy in Forecasting: A Study of Rail Rapid Transit." *Urban Resources* 4(1): 23--30.

Wachs, Martin. 1989. "When Planners Lie with Numbers." *Journal of the American Planning Association* 55(4): 476--479. *

Wachs, Martin. 1990. "Ethics and Advocacy in Forecasting for Public Policy." *Business and Professional Ethics Journal* 9(1--2): 141--157.

Wachs, Martin. 2013. "The Past, Present, and Future of Professional Ethics in Planning." In *Policy, Planning, and People: Promoting Justice in Urban Development,* ed. Naomi Carmon and Susan S. Fainstein, 101--119. Philadelphia, PA: University of Pennsylvania Press.

Ward, William A. 2019. "Cost-Benefit Analysis: Theory versus Practice at the World Bank 1960 to 2015." *Journal of Benefit-Cost Analysis* 10(1): 124–144.

World Bank. 1994. *World Development Report 1994: Infrastructure for Development.* Oxford: Oxford University Press.

World Bank. 2010. *Cost-Benefit Analysis in World Bank Projects.* Washington, DC: The World Bank.

Worthington, Elise. 2012. "Clem7 Tunnel Investors Launch $150m Class Action," *ABC News*, 1 June.

Wright, Robert. 2014. "US Infrastructure Crisis Drives Toll Road Bankruptcy." *Financial Times*, 21 September, https://www.ft.com/content/7991d488-40da-11e4-9ce5-00144feabdc0, retrieved December 17, 2020.

Endnotes

---

[1] Here and throughout the paper, we use the term "bias" in the sense it is used in behavioral science, including behavioral economics. This should not be confused with the way the term is used in statistics, in the frequentist-estimator sense, especially for the statistical tests reported in the main text and the annex.

[2] Significance is here defined in the conventional manner, with $p \leq 0.05$ being significant, $p \leq 0.01$ very significant, and $p \leq 0.001$ overwhelmingly significant. In addition to standard statistical tests using p-values, we also tested results for the influence of investment type, geography, and time period using Bayesian modeling. Finally, we ran the tests using different samples. We found that the results reported in the main text proved highly robust across investment type, geography, and time period. Results were also robust across both conventional and Bayesian testing, and across different samples. See further Annex 1.

[3] These findings correspond to other findings based on smaller samples, which we take to attest to the robustness of results (see, e.g., Albalate and Bel 2014; Altshuler and Luberoff, 2003; Ansar et al. 2014; Bain 2009; Bain and Wilkins 2002; Cantarelli et al. 2012; Dantata et al. 2006; Federal Transit Administration 2003, 2008, 2013; Flyvbjerg et al. 2002, 2004, 2005; Fouracre et al. 1990; Gao and Touran 2020; Huo et al. 2018, Leavitt et al. 1993; Lee 2008; National Audit Office 1992; Nijkamp and Ubbels 1999; Pickrell 1990, 1992; Riksrevisionen 2011; Riksrevisionsverket 1994; Walmsley and Pickett 1992; World Bank 1994).

[4] We ran the same tests with similar results for a subsample of 327 investments for which data were available for both cost overrun and benefit overrun for each investment. See further Annex 1.

[5] Negative impact on GDP is not limited to public infrastructure investments. Misallocation of resources similarly happens in the private sector with similar consequences. For private-sector examples, see Flyvbjerg and Budzier (2011).

[6] Strategic misrepresentation is deliberate, while optimism is non-deliberate. See more in Flyvbjerg (2007).

[7] Cost overrun may also originate with subpar delivery, needless to say. However, if the original cost estimate for an investment is truly optimistic, as is common, then no matter how good the delivery team is they will not be able to deliver to the estimate (Flyvbjerg 2016).

[8] For further explanations of the cost-benefit fallacy, see Flyvbjerg and Bester (in progress).

[9] For an in-depth case study of how cost-benefit analysis may lead to resource misallocation, even in an organization with deep experience with and extensive use of the method, see the World Bank (2010). See also Flyvbjerg and Bester (in progress).

[10] A typical assessment of reference class forecasting (RCF) concludes: "The conducted evaluation is entirely based on real-life project data and shows that RCF indeed performs best, for both cost and time forecasting, and therefore supports the practical relevance of the technique" (Batselier and Vanhoucke 2016, 36).

[11] Parameters for the Bayesian models were estimated using MCMC. The language JAGS was used for this, through the rjags interface to R (Plummer 2003,; 2012; R Core Team 2012). Statistical significance for the Bayesian tests was measured by the Bayes Factor (BF) instead of by p-values, where $12 < BF \leq 150$ indicates a statistically significant result and $BF > 150$ indicates a highly significant result.