Optimism Bias and World Bank Project Performance

LAVAGNON IKA* & SIMON FEENY**
*University of Ottawa, Ottawa, Canada, **Centre for International Development, RMIT University, Melbourne, Australia

(Original version submitted June 2021; final version accepted July 2022)

ABSTRACT This paper examines the correlates of optimism bias and its impact on World Bank project performance. We measure optimism bias in different ways using estimated Economic Rates of Return (ERR) of projects at approval and closure. We examine over 2,800 World Bank projects that were appraised between 1960 and 2019. We find that approximately 60% of projects in the sample were prone to optimism bias. Correlates of optimism bias include both project and country characteristics. Findings also indicate that the incidence of optimism bias reduces the chance of a satisfactory project performance rate at the time of evaluation by 17–20%. Recommendations include embracing complexity and uncertainty in considering projects for approval, providing organizational incentives for ensuring projects are successful rather than ERRs being accurate, shifting some resources from appraisal to implementation, and changing the nature of project supervision.

KEYWORDS: Optimism bias; economic rate of return; project performance; project appraisal; project evaluation

JEL CLASSIFICATION CODES: D61; D83; F35; O22

1. Introduction

Project underperformance remains an everlasting puzzle in development theory and practice (Andrews, 2018; Hirschman, 1967; Ika, 2012). Consequently, project performance has been the subject of research, with three broad areas of interest (Denizer, Kaufmann, & Kraay, 2013; Feeny & Vuong, 2017; Ika, 2018). The first stream, which dates back to the 1970s, focuses on Cost-Benefit Analyses (CBA) and typically assesses the Economic Rates of Return of projects (ERRs), prospectively at project appraisal/approval and retrospectively at project completion/closure (e.g. Del Bo & Florio, 2010). The second and relatively recent stream focuses on rigorous impact evaluation through randomized control trials (e.g. Banerjee et al., 2015). In addition to these two streams which are in the micro-economic tradition, a third look into the inner workings of the project management ‘black box’ and explores how activities and processes are actually carried out to fill the void in practical insight on how projects really get done (e.g. Ika, 2015).
However, notwithstanding the contribution of these streams of literature, the question of why there is a systematic gap between before- and after-project ERRs remains a longstanding conversation that has been stuck for 20 years. Certainly, there is mounting evidence that a gap persists between the theory and practice of CBA (Ward, 2019) and that inaccuracy, bias and uncertainty surround the assessment of ERRs (World Bank, 2010) and project outcomes (Kilby & Michaelowa, 2019). As we learn from psychologists and project scholars, bias consists of a systematic deviation between the (average) judgment of a person or a group and a true value or norm, or in most cases, a deviation from a statistical principle (Ika, Love, & Pinto, 2022). Research has shown that ‘systematic bias’ manifests in World Bank project assessments (Kilby & Michaelowa, 2019) and development policy professionals, who play a prominent role in the process, are prone to bias and most strikingly confirmation bias (Banuri, Dercon, & Gauri, 2019). Evidence suggests that a ‘systematic bias’ prevails in ERRs at approval across projects and sectors (World Bank, 2010). For example, 22% of project proposals were found to be ‘chronically’ optimistic and the time it takes to complete a project was often underestimated at appraisal by two years on average (World Bank, 1992). Thus, many “World Bank reports cited several manifestations of an optimism bias at appraisal” (World Bank, 2010, p. 26).

The notion of undue optimism or what behavioral economists now call ‘optimism bias’ refers to the belief that people, whether they are experts or laymen, are less likely to be prone to risk than statistical reality warrants when it comes to experiencing negative consequences (Kahneman, 2011; Sunstein, 2015). Flyvbjerg (2016), a key proponent of optimism bias, suggests it is THE reason for explicating cost overruns and benefit shortfalls, as it occurs in 80% of the 327 large-scale, economic infrastructure projects in his sample including roads, rails and bridges. Love, Ika, and Sing (2022) show that, at best, 57% of projects may be prone to optimism bias in their sample of 47 large-scale, social infrastructure projects including schools, hospitals and prisons and conclude that optimism bias is ‘overrated’, as competing explanations of project underperformance such as scope changes, complexity, and uncertainty play key roles. While the prevalence and magnitude of optimism bias are still a bone of contention in the project management literature, these two studies focus on project settings in industrialised countries.

Thus, we still do not know the extent to which optimism bias occurs in development projects or World Bank projects in particular, which take place in low- or middle-income countries and whose effects are more difficult to assess (Greßer & Stadelmann, 2020; Ika, 2018). A case in point, optimism bias is not discussed in Banuri et al.’s (2019) experimental study of biased policy professionals and its impact on project performance remains largely unexamined. An exception is World Bank (2010), which notes that it remains “rare to find reports that analyze and quantify bias in project analysis and appraisal” (p. 27). To date, we know very little about optimism bias and its impact on project performance, notably at the World Bank.

In this paper, we focus on World Bank projects and seek to replicate the findings of Flyvbjerg (2016) and Love et al. (2022). We however turn to the proxy of what the World Bank (2010) calls the ERR gap or the difference between the ERR at approval and closure. As World Bank (2010) observes, “The first indicator of possible bias is based on a comparison of the ERRs at project entry with those at project closing. Are the rates at closing higher or lower than the original rates?” (p. 26). Accordingly, this paper uses a dataset of over 2,800 World Bank projects, implemented in 163 developing countries covering project closings from 1960 to 2019.

The impact of optimism bias on project outcomes is assessed using World Bank project performance ratings, which measure the extent to which projects achieve their intended objectives. We find that the ERRs at approval tend to be inaccurate, biased and optimistic in our dataset, adding to the relevant empirical evidence (Del Bo & Florio, 2010; Pohl & Mihaljek, 1992; World Bank, 2010). We also build on the scant literature available and show that project-level micro characteristics such as the ERR at approval, the sector in which the project is undertaken, and the type of loan that funds it are important correlates of optimism bias. We also
examine country-level macro characteristics such as average economic growth rates, income levels, political rights, and civil liberties that prevailed during project implementation (Bulman, Kolkma, & Kraay, 2017; Denizer et al., 2013). We also control for the country or region of implementation. While many of these correlates have been identified in the literature (World Bank, 2010), this study is the first to test their significance, not only in a larger dataset but also in terms of both the incidence and intensity of optimism bias.

More precisely, we find that the ERR at approval is positively and significantly associated with optimism bias, which has tended to increase up to the mid-1980s to the 1990s and then decrease thereafter. This contrasts with Flyvbjerg (2016) who suggests there has been no sign of amelioration or learning throughout the years. Further, going beyond World Bank (2010), Flyvbjerg (2016) and Love et al. (2022), we quantify the impact of optimism bias on project performance. We show that the incidence of optimism bias or the percentage of World Bank projects that are deemed overoptimistic is at best 60% as opposed to 57% in social infrastructure (Love et al., 2022) or 80% in economic infrastructure projects (Flyvbjerg, 2016). On average, the ERRs at project closure is 22.9% compared to 25.4% at approval, demonstrating a mean ERR gap of −2.5%. Finally, optimism bias can reduce the chances of a satisfactory project performance rating by 17%–20%.

The remainder of this paper is organized as follows. Section 2 describes the practice of CBA at the World Bank. Section 3 reviews the empirical evidence not only on the ERR gap but also on optimism bias and its relationship to performance in World Bank projects. Section 4 provides details on the methodological background of the research while the findings from the estimation of the empirical models are reported in Section 5. Finally, Section 6 concludes with the practical implications of the research, presents its limitations, and offers suggestions for further research.

2. The practice of cost-benefit analysis at the World Bank

Since the 1970s, CBA has featured prominently as a powerful tool for project investment justification, now mandated by Operational Policy OP 10.04 of the World Bank. Under this guidance, the World Bank should conduct CBAs for all its funded projects, unless some costs and benefits cannot be measured in monetary terms and, therefore, a cost-effectiveness analysis, which includes a comparison of the costs of alternative ways of reaching a given goal, should be conducted (World Bank, 2010). While CBA at approval is generally assessed at the project appraisal stage and finalized at the negotiations/approval stage, CBA at closure is calculated at the completion/evaluation stage of the World Bank project cycle, which consists of six stages: identification, preparation, appraisal, negotiations/approval, implementation/support, and completion/evaluation. This project cycle is provided in Figure S1 in the paper’s supplementary material.

Typically, CBA assesses whether the present value of benefits exceeds that of costs for a given project to be undertaken. CBA consists of a cash-flow analysis that quantifies over time the major benefit and cost flows and the economic rate of return (ERR) of a project or the discount rate at which its net present value (NPV) is equal to zero (World Bank, 2010). With few exceptions, the World Bank usually considers a cut-off rate of an ERR of at least 10% at approval (in constant prices), reassesses, using data from completion reports, the ERR this time after completion or start-up of normal operations, and measures the ERR gap (Del Bo & Florio, 2010; Pohl & Mihaljek, 1992). Notably, the ERRs at closure are rarely real ex-post ERRs, which may be calculated only after 10+ years when evaluation reports are prepared (Little & Mirrlees, 1990).

Yet, the ERR is not the sole criterion for project performance. While ERRs measure efficiency, (non-ERR) performance ratings, however, measure not only efficiency but effectiveness, relevance, impact and sustainability (World Bank, 2010, p. 58). Such ratings have been applied
since 1973. The World Bank initially had a binary (satisfactory and unsatisfactory) scale, which they changed to the current six-point scale in 1993 (highly satisfactory, moderately satisfactory, satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory) (Kilby & Michaelowa, 2019).

3. The empirical evidence on cost-benefit analysis, optimism bias and project performance at the World Bank

While the theory of CBA may be conceptually rigorous, it remains difficult for practitioners to get key parameter estimates right (Pohl & Mihaljek, 1992). Even Little and Mirrlees (1990) – pioneers of the CBA method – in their assessment of the use of CBA ‘twenty years on’, noted that, though it was intended to be “practical” (p. 353), it has been “hard to implement,” “unsound,” “self-defeating,” and “too complex” (p. 365), especially in the face of complexity and uncertainty of both the project and its setting, which can diminish its usefulness and “value” (p. 356). They conclude that changing organizational contexts meant that, in the 1970s, ERRs “became more and more optimistic, without justification,” a phenomenon they dubbed the “McNamara effect” as the then president of the World Bank was advocating for a push in the volume of investments (p. 364).

More concretely, a few studies have sought to measure the accuracy of CBA in World Bank projects. In their study of a sample of 1,015 infrastructure projects in the agriculture, transport, energy, industry, and urban development sectors, Pohl and Mihaljek (1992) analyzed the differences between ERRs at approval and closure when the projects were completed about 5–10 years after appraisal (between 1974 and 1977). Their regression analysis revealed a large ERR gap and thus a prevalence of a high level of uncertainty in project appraisal which tends to be biased and optimistic. Their findings suggest a high degree of optimism bias in the average project returns in the 1970s with a rise in ERRs at approval from 16% in the mid-1960s to 20–25% in the mid-and late 1970s, and a downtrend in the ERRs at closure in the 1970s.

Del Bo and Florio (2010) found that there are no statistical differences between ERRs at approval and closure for 84 projects approved in fiscal years 1988–1997, but this was not the case for 259 projects completed in years 1990–1997. They showed that the ERR gap for the more recent years was diminishing but the highest for industry projects and the lowest for road projects. The ERRs at closure, when compared against the ‘industry’ sector benchmark, revealed higher returns for roads and highways, ports and airports, energy distribution and telecommunications infrastructure. The findings also suggest ‘constant forecasting optimism’ of around 30% in the ERRs at approval for industry, water, and railways projects, whereas it is less for road projects.

World Bank (2010) looked into the ERR gap of a total of 1,938 investment projects including 1,772 projects from 11 different sectors over two periods of closing (1975–1979, 2003–2007) and 166 projects in 2008. The presence of an ERR estimate served as an indicator of whether CBA was performed. It was found that the percentage of projects with a CBA analysis dropped from 70% to 25% between 1970 and 2008. A time series between 1972 and 2008 using median values showed that the ERR gap had “virtually disappeared” around 2000 perhaps essentially due to “a rise in the upward bias in returns, improvement in overall economic conditions as measured by growth, and a rise in the degree of market orientation of the economic regime” (p. 40).

The ERR gap may be due to a ‘positive’ or ‘upward’ bias of different forms including the planning myth of “Everything Goes According to Plan”; overestimation of long-term benefits; a tendency not to report CBA for many projects, especially those with negative returns; and recalculation bias in that many projects, especially the poorly performing ones, do not have ERRs at closure. Further, both the ERRs and the performance ratings are correlated: “the returns are observed by the evaluator doing the rating” yet “the empirical overlap between them is not huge” (World Bank, 2010, p. 39). These findings that optimism bias is at work in World Bank CBA
trigger this study, which investigates the extent of optimism bias, its correlates and its impact on project performance.

4. Data and methodology

4.1. Data

This research utilizes the World Bank’s Project Performance Rating database, a publicly available dataset administered by the Independent Evaluation Group (IEG), which evaluates World Bank projects, programs and activities by examining their achievements against stated objectives. It includes some 10,500 completed projects spanning the period 1964–2019. While World Bank staff prepare Implementation Completion Reports (ICRs), which include initial project performance ratings, IEG staff carry out a review of all ICRs together with the lending/grant agreement, project appraisal documentation and the country assistance strategy. The dataset provides the estimated ERRs of projects at their approval and at their closure. While it remains difficult to measure optimism bias (e.g. Love et al., 2022), a proxy is to calculate a binary variable equal to one if the ERR of a project at appraisal is higher than at closure. If this is true for most projects, then there is evidence of optimism bias.

However, this absolute proxy, by essence, implies that projects with high ERRs at approval will be positively associated with a high likelihood of optimism bias. Following Pohl and Mihaljek (1992), we resort to a relative proxy, the percentage difference between a project’s ERR at closure relative to its ERR at approval: (ERR at closure – ERR at approval)/ERR at approval (multiplied by 100). Though this approach does not remove the measurement problem altogether, lower values of the variable imply higher expectations of the ERR at project approval than closure. Thus, the intensity rather than incidence of bias is modelled. Given its continuous nature, the variable does not distinguish between optimism and pessimism bias, the latter potentially arising when, on average, ERRs at approval are systematically lower than at closure. We therefore also include results pertaining to models with a binary variable in our examination of optimism bias, noting that a positive association with ERRs at approval is unsurprising and that robustness tests are warranted.

It is World Bank staff that calculate the ERRs at approval and closure although the IEG will use this information in their project evaluations (Kilby & Michaelowa, 2019). The IEG data only report ERRs if CBA has been applied to at least 20% of a project commitment (World Bank, 2015). Our analysis is thus restricted to the IEG ratings on those projects where CBA applies, implying that they are different in nature from others in the database. Further, as the World Bank typically doesn’t fund projects with low ERRs at approval (i.e. less than 10%), we exclude them from subsequent regression analysis. We also exclude projects that are deemed a complete failure or arbitrarily assigned an ERR at a closure of –5% (e.g. Pohl & Mihaljek, 1992). We restrict the values of the continuous optimism bias variable to between −300 and +300 to account for 16 outliers (although they are kept for the binary variable). A histogram providing the distribution of values for the continuous optimism bias variable is provided in Figure S2 of the supplementary material.

Projects in some sectors are much more likely to have ERRs at approval and closure. ‘Agriculture and Rural Development’, ‘Energy and Mining’, ‘Global Information/Communications Technology’, ‘Transport’, ‘Urban Development’ and ‘Water’ have between 25 and 67% of projects subject to CBA (see Table 1). These six can be referred to as ‘high-CBA’ sectors and the other ten as ‘low-CBA’ sectors. Consistent with World Bank (2010), about 30% of projects are subject to CBA (3,098 projects out of a total of 10,459).

The proportion of projects subject to CBA by exit year points to a downward trend since 1980, due to a gradual shift away from high-CBA sectors and a decline notably in low-CBA sectors (World Bank, 2010) (see Figure S3 in the supplementary material). Notably, the number of projects with an exiting year in the years just prior to 2020 is low, thus explaining why the...
proportion of projects subject to CBA in these years falls to zero. Table 2 provides the incidence of optimism (and pessimism, e.g. Love et al., 2022) bias and shows that, over the sample period, at best 60% of projects with an ERR at approval and closure were subject to optimism bias.

Figure 1 provides the average ERR gap in percentage terms by exit year, with lower values implying higher expectations of returns at approval than closure. Evidence of optimism bias is provided by most values falling below zero. Moreover, there was a tendency for greater optimism bias from mid-1970s to mid-1980s.

Figure 2 provides the percentage of projects subject to CBA that encountered optimism bias by the year of closure. There is a clear non-linear relationship, with the incidence of optimism bias tending to increase during the period 1960–1985 and falling thereafter. The so-called ‘McNamara effect’ along with the push for disbursement and the reorganizations the World Bank underwent in the 1980s leading to lower quality control over projects may perhaps explain this increase (Little & Mirrlees, 1990). The fall however is likely to be attributed to a tighter focus on project planning and the move toward results-based management (RBM), which gained momentum in the early 1990s (World Bank, 2017).

Out of the sample of 3,801 projects with an ERR at approval, 691 never received an estimated ERR at closure, making it impossible to calculate optimism bias and indicating the presence of sample attrition or selection bias (see Table S1 in the supplementary material). While this proportion is not great (18%), unless the omission is random, it may affect the research findings. For example, the ERRs for poorly performing projects at approval might be less likely to be re-estimated at project closure, highlighting a persistent recalculation bias in favour of

---

**Table 1.** Percentage of projects with an ERR at approval and closure by sector

| Sector                                      | %   |
|---------------------------------------------|-----|
| Agriculture and rural development           | 50.0|
| Economic policy                            | 0.3 |
| Education                                  | 2.0 |
| Energy and mining                          | 48.6|
| Environment                                | 6.8 |
| Financial sector                           | 3.1 |
| Financial and private sector development    | 2.8 |
| Global information/communications technology| 63.3|
| Health, nutrition and population            | 2.0 |
| Other                                       | 2.0 |
| Public sector governance                    | 0.9 |
| Social development                         | 5.9 |
| Social protection                          | 1.4 |
| Transport                                  | 66.6|
| Urban development                          | 25.8|
| Water                                      | 37.4|
| Total                                      | 29.6|

**Table 2.** Projects by ERR at closure relative to approval

| Economic rate or return (ERR)                      | N  | %   |
|----------------------------------------------------|----|-----|
| Project ERR higher at closure than approval (pessimism bias) | 1,007 | 35.3|
| Project ERR lower at closure than approval (optimism bias) | 1,723 | 60.4|
| Project ERR the same at closure as approval          | 123 | 4.3 |
| Total                                               | 2,853 | 100.0|

Source: The authors. Calculated from the World Bank (2020a).
high-performing projects. Common reasons for not reporting an ERR at project closure include staff belief that an ERR was not applicable or benefits could not be quantified (World Bank, 2010). Figure S3 in the supplementary material demonstrates that the proportion of projects with an ERR at both approval and closure fell during the period 1960–1990. Since 1990, no discernible pattern appears in the data. We discuss how we measure and correct for sample selection bias in Section 5.

This paper also empirically examines whether optimism bias impacts project outcomes. Again, the World Bank changed its project performance ratings from a binary (satisfactory/not satisfactory) measure to a six-point Likert scale in 1993. So, to maximize our sample size, we reclassify the projects from 1993 using the previously employed binary classification (e.g. Bulman et al., 2017). Consistent with Ika (2018) and World Bank (2017), almost 82% of the projects were deemed satisfactory at the time of their evaluation (see Table S2 in the supplementary material).

Figure 3 provides the average proportion of projects that were evaluated as satisfactory by year of closure, showing a sharp fall from the mid-1970s to the mid-1980s. Denizer et al. (2013)
reveal a downward trend between 1980 and 1985 but an upward trend since then. World Bank (2010) finds that project performance has improved since 1993. Bulman et al. (2017) show that average project success rates fluctuated between 70% and 80% between 1995 and 2015.

The Project Performance Rating database is augmented with country-level data from the World Bank’s World Development Indicators database (World Bank, 2020b) and from Freedom House (2020). Our choice of these variables is driven by the literature (e.g. Bulman et al., 2017; Isham & Kaufmann, 1999). Variable definitions and sources are provided in Table S3 and summary statistics for key variables are provided in Table S4 (see supplementary material). The projects in the sample were approved between 1960 and 2011 and completed between 1964 and 2015. On average, they took about seven years to complete after approval. The ERRs were on average 22.9% at closure versus 25.4% at approval, suggesting an ERR gap of −2.5% for the whole sample. Extant literature suggests an average ERR of 16% at closure as opposed to 22% (Pohl & Mihaljek, 1992) or 25% (Del Bo & Florio, 2010) at approval.

4.2. Methodology

In order to ascertain the correlates of optimism bias, the following model is specified and estimated:

\[ B_i = \beta_0 + \beta_1 P_i + \beta_2 C_i + \beta_3 Z_i + \mu_i \]  

Our dependent variable: \( B_i \) is a continuous variable or gap measure (of the intensity of optimism bias) or a binary variable (measuring the incidence of optimism bias).

\( P \) is a vector of project level characteristics for the project \( i \) including the: (i) ERR at approval; (ii) sector; (iii) lending between the World Bank and the borrower (e.g. funding sources include the International Bank of Reconstruction and Development, IBRD; the International Development Association, IDA); and (iv) type of investment loans (e.g. Financial Intermediary, Specific Investment, and Sector Investment and Maintenance Loans).\(^6\) It also includes the length of the project implementation period, measured in years. Given endogeneity concerns over the ERR at approval, the model is also re-estimated with this variable excluded.
$C$ is a vector of country-level variables including Gross Domestic Product (GDP) per capita, inflation (logged) (World Bank, 2020b), and extent of freedom measured by political rights and civil liberties (Freedom House, 2020).

$Z$ is a vector of control variables and includes geographic region and approval year. Note that all country-level variables are averaged during the project implementation period. The model is estimated using Ordinary Least Squares (OLS). To test for sample selection bias, we estimate a probit model using binary attrition as a dependent variable which takes the value of one for projects that received an ERR at approval but not closure and zero otherwise.

After examining the correlates of optimism bias, we assess whether it impacts the project’s evaluated outcome. The following model is estimated:

$$O_i = \alpha + \beta_1 B_i + \beta_2 P_i + \beta_3 C_i + \beta_4 Z_i + \mu_i$$ (2)

$O_i$ is a binary project outcome variable taking the value of 1 if the project is evaluated as satisfactory and zero if non-satisfactory. $B_i$ measures optimism bias for the project $i$ as discussed above. The vectors $P_i$, $C_i$ and $Z_i$ provide the same project characteristics, country and control variables as in Equation 1. Robust standard errors are used in all specifications. While results are unlikely to suffer from simultaneity bias, endogeneity might arise from omitted variables being correlated with both dependent and independent variables. Thus, Oster’s (2019) unobservable selection and coefficient stability test was applied to examine the potential role of unobserved variables.

5. Results and interpretation

5.1. ERR gap measure of intensity of optimism bias

Results examining the intensity of optimism bias from the estimation of Equation 1 using OLS are provided in Table 3. In the first column, region-fixed effects are included and standard errors are clustered at the country level. In the second column, country-fixed effects are included (without region-fixed effects) and the model is estimated with robust standard errors. In the third column, the model clusters standard errors by country. Macroeconomic variables are also included.

Results suggest that the ERR at approval is associated with the intensity of optimism bias. Lower levels of the dependent variable represent higher levels of optimism bias. A project with an ERR at approval 10% higher than the mean is associated with a 3%–5% higher level of optimism bias. While the coefficients attached to region dummy variables are not statistically significant in column 1, results from column 3 indicate that, relative to projects implemented in other regions, projects implemented in Latin America and the Caribbean, LAC (18% of the sample) have a higher level of optimism bias, potentially due to a higher level of regional economic volatility and political instability. ERRs may be higher for projects implemented in less distorted economic environments (Isham & Kaufmann, 1999).

The project sector appears important. The level of optimism bias is higher in the ‘Agriculture and Rural Development’ and ‘Water’ sectors relative to other sectors. Pohl and Mihaljek (1992) found that the ERR gap is smaller for transport compared to agriculture projects. This might reflect projects in the latter sector being reliant on weather and climate and therefore subject to a higher degree of volatility and uncertainty. Column 2 suggests that projects funded through IDA experience lower levels of optimism bias (relative to IBRD-funded projects). In their assessment of bias, Kilby and Michaelowa (2019) obtain mixed findings between IDA and non-IDA projects. The type of loan is also important in explaining optimism bias. Relative to ‘Financial Intermediary Loans’ (FIL), ‘Sector Investment Loans’ and ‘Sector Investment and Maintenance Loans’ show higher levels of optimism bias, suggesting that project appraisers more effectively account for downside risks associated with FIL loans.
To capture the non-linear relationship between project approval and the intensity of optimism bias (see Figure 1), the approval year and its squared terms are included in the model. An inverted U-shaped relationship prevails where the degree of optimism bias first increases with the approval year before falling. There is no significant association between the number of years a project takes to implement and the level of optimism bias. Indeed, though optimism bias and

| Table 3. Determinants of optimism bias (ERR gap measure) |
|---------------------------------------------------------|
| (1) Base | (2) Base + Country FE | (3) Base + Macroeconomic |
| ERR at approval | −0.43*** | −0.46*** | −0.43*** |
| [−7.00] | [−6.19] | [−6.20] |
| Region (omitted category = Sub-Saharan Africa) | | | |
| East Asia and the Pacific | 0.20 | [0.062] | −4.27 |
| Europe and Central Asia | −1.96 | [−0.46] | −2.86 |
| Latin America and the Caribbean | −6.33 | [−1.57] | −10.8** |
| Middle East and North Africa | 3.62 | [0.89] | 1.83 |
| South Asia | 2.87 | [0.82] | −1.78 |
| Sector (omitted category = agriculture and rural development) | | | |
| Energy and mining | 10.8*** | 10.7*** | 10.3** |
| [2.90] | [3.88] | [2.31] |
| Global information/communications technology | 24.7*** | 24.0*** | 24.1*** |
| [5.22] | [4.88] | [4.87] |
| Other | 14.2*** | 16.5** | 12.9 |
| [1.80] | [1.99] | [1.52] |
| Transport | 19.6*** | 20.1*** | 16.7*** |
| [6.42] | [8.37] | [5.13] |
| Urban development | 13.9*** | 14.3*** | 11.1* |
| [2.53] | [2.69] | [1.78] |
| Water | −1.34 | [−0.32] | −3.77 |
| [−0.26] | [−0.79] |
| Project agreement type (omitted category = IBRD) | | | |
| International development association (IDA) | 1.32 | 6.84*** | 1.83 |
| [0.50] | [2.22] | [0.61] |
| Other | −15.5 | −25.0 | −28.3* |
| [−1.14] | [−1.50] | [−1.80] |
| Lending instrument (omitted category = financial intermediary loan) | | | |
| Other | −7.32 | −7.74 | −11.2 |
| [−0.93] | [−1.08] | [−1.31] |
| Specific investment loan | −8.62* | −10.8*** | −11.1** |
| [−1.80] | [−2.67] | [−2.13] |
| Sector investment and maintenance loan | −8.98 | −11.2** | −10.7* |
| [−1.63] | [−2.42] | [−1.73] |
| Approval financial year | −161*** | −152*** | −170*** |
| [−3.96] | [−4.54] | [−3.97] |
| Approval financial year squared | 0.041*** | 0.038*** | 0.043*** |
| [3.98] | [4.56] | [3.99] |
| Implementation duration | 0.032 | −0.02 | 0.10 |
| [0.070] | [−0.45] | [0.21] |
| Inflation (logged) | 5.17** | | |
| [2.31] | | |
| GDP per capita growth | 1.56*** | | |
| [3.54] | | |
| GDP per capita (constant 2010 SUS) | | −0.59 | |
| | | [−1.54] | |
| Political rights | −2.12 | | |
| | | [−1.03] | |
| Civil liberties | 1.01 | | |
| | | [0.36] | |
| Country clustered standard errors | Yes | No | Yes |
| Country fixed effects | No | Yes | No |
| Observations | 2,816 | 2,816 | 2,436 |
| R-squared | 0.08 | 0.14 | 0.09 |

***p < 0.01, **p < 0.05, *p < 0.1.
time overruns are theoretically related (Kahneman, 2011), project appraisers generally make their entry ERR assessments at appraisal, not during execution.

Column 3 describes a model with smaller sample size and includes country-level variables, which add very little explanatory power to the model (cf. $R^2$). Projects implemented during periods of higher growth are associated with lower levels of optimism bias. This could reflect projects performing better than expected in countries experiencing high growth rates and favorable macroeconomic conditions leading to high ERRs. The average inflation rate during implementation is also associated with lower levels of optimism bias. There is no discernible relationship between political rights/civil liberties and optimism bias intensity.

Given endogeneity concerns, the models are re-estimated without the ERR at approval variable (results are available from the authors on request). Interestingly, the $R^2$-squared numbers fall by a value of just 0.02, implying that the ERR at approval explains only about 2% of optimism bias variation, which is negligible in our view. Results relating to the coefficient on other variables are nearly identical in terms of statistical significance.

5.2. The incidence of optimism bias

The models are also estimated with the binary optimism bias as the dependent variable using a linear probability model. Results are very similar (see Table 4). Considering a mechanical link between the incidence of optimism bias and the ERR at approval, the latter variable was removed as a covariate (results are available from the authors on request). All results with respect to region, sector, agreement type, lending instrument, approval year fixed effects, inflation and GDP growth are consistent. The only difference is in the magnitude of the coefficient estimates and the greater statistical significance of the variables. Again, the inclusion of the ERR at approval variable led to an increase in the $R^2$-squared numbers of around just 1%. Thus, this variable, though statistically significant, does little in explaining optimism bias variation. These results are also consistent with the estimation of a probit model. The impact of attrition and sample selection bias are examined in Insert S1 and Tables S5–S7) and Insert S2 (and Table S8) of the supplementary material to this paper. The impact of including the ERR at appraisal and the issue of mathematical coupling is also discussed in Insert S3 of the supplementary material.

5.3. Optimism bias and project outcomes

Optimism bias matters for non-ERR project performance for two reasons. Firstly, lower ERRs at closure than approval might influence evaluators’ overall assessment of the project (Kilby & Michaelowa, 2019; World Bank, 2010). Secondly, project appraisers might be influenced by organizational incentives to identify the best-performing projects, which might lead them to be over-optimistic (or deliberately misleading) with respect to ERRs (Little & Mirrlees, 1990). Tables 5 and 6 provide results examining the impact of our two measures of optimism bias on project outcomes.

Table 5 estimates the model using the intensity of optimism bias (ERR gap measure), showing that, at the mean, a 10 unit increase in the ERR gap of a project is associated with a fall in the probability of project outcomes being satisfactory by 2%–3%. Relative to Sub-Saharan Africa, evidence from two of the models suggests projects undertaken in East Asia and the Pacific and South Asia are more likely to be evaluated as satisfactory. This is also the case in Europe and Central Asia (column 1) and LAC (column 3). However, the Middle East and North Africa results are mixed. These findings are supported by the literature. About 50% of African investment projects funded by the World Bank’s private arm (IFC) are rated satisfactory as compared to 60% elsewhere in the world due to higher risk, business and political climate, below-average project environment, and social compliance (Dugger, 2007). Further,
|                    | (1) Base | (2) Base + Country FE | (3) Base + Macroeconomic |
|--------------------|----------|-----------------------|--------------------------|
| **ERR at approval**| 0.0033***| 0.0037***             | 0.0033***                |
|                    | [5.08]   | [4.87]                | [4.53]                   |
| **Region (omitted**|          |                       |                          |
| category = Sub-    |          |                       |                          |
| Saharan Africa)    |          |                       |                          |
| East Asia and the  | -0.019   |                       | 0.015                    |
| Pacific            | [-0.54]  |                       | [0.38]                   |
| Europe and Central | 0.028    |                       | 0.0056                   |
| Asia               | [0.55]   |                       | [0.11]                   |
| Latin America and  | 0.070*   |                       | 0.11**                   |
| the Caribbean      | [1.82]   |                       | [2.42]                   |
| Middle East and    | -0.045   | -0.011                |                          |
| North Africa       | [-1.10]  | [0.24]                |                          |
| South Asia         | -0.019   | 0.0096                |                          |
|                    | [-0.51]  | [0.20]                |                          |
| **Sector (omitted**|          |                       |                          |
| category =         |          |                       |                          |
| agriculture and    |          |                       |                          |
| rural development  |          |                       |                          |
| Energy and mining  | -0.084***| -0.092***             | -0.089**                |
|                    | [-2.79]  | [-3.32]               | [-2.41]                  |
| Global information/| -0.19*** | -0.18***              | -0.19***                |
| communications     |          |                       |                          |
| technology         | [-4.00]  | [-3.26]               | [-3.72]                  |
| Other              | -0.13**  | -0.17***              | -0.14**                 |
|                    | [-2.51]  | [-2.92]               | [-2.42]                  |
| Transport          | -0.18*** | -0.19***              | -0.16***                |
|                    | [-6.28]  | [-7.41]               | [-5.20]                  |
| Urban development  | -0.13*** | -0.14***              | -0.13**                 |
|                    | [-2.92]  | [-2.82]               | [-2.55]                  |
| Water              | 0.030    | 0.032                 | 0.017                   |
|                    | [0.72]   | [0.75]                | [0.35]                   |
| **Project agreement**|        |                       |                          |
| type (omitted      |          |                       |                          |
| category = IBRD)   |          |                       |                          |
| International      | -0.014   | -0.073**              | -0.010                  |
| development        |          |                       |                          |
| association (IDA)  | [-0.47]  | [-2.30]               | [-0.29]                  |
| Other              | 0.23**   | 0.28**                | 0.30**                  |
|                    | [2.21]   | [2.34]                | [2.36]                   |
| **Lending instrument**|        |                       |                          |
| (omitted           |          |                       |                          |
| category =         |          |                       |                          |
| financial          |          |                       |                          |
| intermediary loan  |          |                       |                          |
| Other              | 0.041    | 0.035                 | 0.065                   |
|                    | [0.63]   | [0.50]                | [0.85]                   |
| Specific investment| 0.077*   | 0.093**               | 0.10**                  |
| loan               | [1.73]   | [2.18]                | [2.02]                   |
| Sector investment  | 0.091**  | 0.11**                | 0.11**                  |
| and maintenance    |          |                       |                          |
| loan               | [2.05]   | [2.22]                | [2.25]                   |
| Approval financial | 1.97***  | 1.87***               | 2.04***                 |
| year               | [5.52]   | [5.60]                | [5.14]                   |
| Approval financial | -0.00050 ***| -0.00047 *** | -0.00051***             |
| year squared       | [-5.53]  | [-5.61]               | [-5.15]                  |
| Implementation     | 0.0041   | 0.0048                | 0.0014                  |
| duration           | [0.94]   | [0.94]                | [0.30]                   |
| Inflation (logged) | -0.033** |                      |                         |
|                    | [-2.08]  |                      |                          |

(continued)
projects in the Global ICT and Transport sectors are more likely to be evaluated as satisfactory relative to other sectors. Column 2 indicates that IDA-funded projects have better outcomes than IBRD ones, reflecting perhaps the use of a performance-based allocation model for IDA whereby more aid is provided to lower-income countries which deliver development benefits more effectively.

The coefficients attached to the approval year variable in columns 1 and 2 suggest that projects have become less likely to be evaluated as satisfactory over time, although the coefficients are very small. The approval year squared variable was dropped from this specification due to collinearity. Other findings indicate that projects implemented over longer periods are less likely to be satisfactory. This might reflect such projects being larger, more complex, or having experienced more difficulties during the implementation cycle. This is supported by Bulman et al. (2017) who found that the larger the project, the longer it takes time to complete, and the less likely it is to succeed. Finally, macroeconomic variables also appear to be important for project performance, as attested by Denizer et al. (2013) and Bulman et al. (2017). Column 3 indicates that projects implemented in countries with higher levels of civil liberties, income and economic growth rates are more likely to be deemed satisfactory while less likely for those implemented during periods of higher inflation. Interestingly, projects are more likely to be deemed satisfactory in countries with lower levels of political rights.

Table 6 also suggests an important role of optimism bias in project performance. Its incidence is associated with a 17%–18% lower probability of the project being assessed as satisfactory at the time of evaluation. Findings with respect to the other variables are broadly consistent with those in Table 5 examining the intensity of optimism bias. Given the binary nature of the dependent variable, probit models are also estimated (see Table S9 in the supplementary material). Again, results are consistent although the impact of optimism bias is even greater. At the mean, a 10 unit increase in the ERR gap of a project is associated with a fall in the probability of project outcomes being successful by 3%, while the incidence of optimism bias reduces the chance of a satisfactory project rating by 19%–20%. Interestingly, when the optimism bias variables are removed from the three models, the $R^2$ numbers fall to 0.03, 0.10 and 0.05, respectively. This suggests that optimism bias is important in explaining a relatively large amount of variation in project outcomes.
Table 5. Determinants of project performance (ERR gap measure)

|                          | (1)       | (2)       | (3)       |
|--------------------------|-----------|-----------|-----------|
|                          | Base      | Base + Country FE | Base + Macroeconomic |
| **ERR gap – optimism bias** | 0.0025*** | 0.0024*** | 0.0025*** |
|                          | [12.2]    | [14.0]    | [11.3]    |
| **Region (omitted category = Sub-Saharan Africa)** |           |           |           |
| East Asia and the Pacific | 0.13***   | −0.10     | 0.096***  |
|                          | [4.42]    | [−1.03]   | [3.11]    |
| Europe and Central Asia  | 0.096***  | −0.14     | 0.065     |
|                          | [2.99]    | [−1.62]   | [1.63]    |
| Latin America and the Caribbean | 0.048    | −0.051    | 0.060*    |
|                          | [1.62]    | [−0.48]   | [1.74]    |
| Middle East and North Africa | 0.073**  | −0.12***  | 0.062     |
|                          | [2.14]    | [−2.79]   | [1.60]    |
| South Asia               | 0.084***  | 0.031     | 0.090***  |
|                          | [3.26]    | [0.36]    | [3.32]    |
| **Sector (omitted category = agriculture and rural development)** |           |           |           |
| Energy and mining        | 0.0071    | 0.018     | 0.0047    |
|                          | [0.34]    | [0.80]    | [0.20]    |
| Global information/communications technology | 0.082**   | 0.082**   | 0.071     |
|                          | [2.05]    | [2.35]    | [1.59]    |
| Other                    | 0.0017    | −0.018    | −0.013    |
|                          | [0.036]   | [−0.37]   | [−0.27]   |
| Transport                | 0.047**   | 0.046**   | 0.027     |
|                          | [2.31]    | [2.35]    | [1.21]    |
| Urban development        | 0.020     | 0.00094   | 0.011     |
|                          | [0.57]    | [0.025]   | [0.32]    |
| Water                    | 0.044     | 0.032     | 0.029     |
|                          | [1.34]    | [0.87]    | [0.85]    |
| **Project agreement type (omitted category = IBRD)** |           |           |           |
| International development association (IDA) | 0.011     | 0.046*    | 0.019     |
|                          | [0.56]    | [1.91]    | [0.84]    |
| Other                    | −0.055    | −0.017    | −0.028    |
|                          | [−0.71]   | [−0.17]   | [−0.28]   |
| **Lending instrument (omitted category = financial intermediary loan)** |           |           |           |
| Other                    | −0.0051   | 0.025     | −0.0077   |
|                          | [−0.087]  | [0.45]    | [−0.12]   |
| Specific investment loan | 0.0083    | 0.022     | 0.0074    |
|                          | [0.22]    | [0.58]    | [0.18]    |
| Sector investment and maintenance loan | −0.0082  | 0.0059    | −0.00091  |
|                          | [−0.20]   | [0.14]    | [−0.021]  |
| Approval financial year  | −0.0033***| −0.0041***| −0.0038***|
|                          | [−5.18]   | [−5.09]   | [−5.57]   |
| Approval financial year squared |         |           |           |
| Implementation duration  | −0.0100** | −0.0092** | −0.0042   |
|                          | [−2.31]   | [−2.25]   | [−0.90]   |
| Inflation (logged)       |           |           | −0.031**  |
|                          |           |           | [−2.19]   |
| GDP per capita growth    |           |           | 0.010***  |
|                          |           |           | [3.38]    |
Further robustness tests

Our core result that optimism bias reduces the likelihood that a project will be deemed successful might be biased due to the role of unobserved variables. To examine this, we adopt Oster’s (2019) coefficient stability approach, which assumes that bias arising from observed controls is more informative than that from unobserved factors. The approach allows us to calculate the bias-adjusted impact of optimism bias by specifying: (1) a value for the relative degree of selection on observed and unobserved variables ($\delta$); and (2) a value for $R_{\text{max}}$ which is the highest expected $R$-squared from a hypothetical regression of project outcomes on optimism bias and both observed and unobserved controls (see Insert S4 and Table S10 in the supplementary material). The bias-adjusted coefficients are very similar in magnitude to those from the controlled and uncontrolled regressions, suggesting that unobserved variables do not play an important role in our findings.

Finally, we undertake randomization inference which provides an alternative approach to calculating $p$-values and testing hypotheses. This exercise allows us to test whether the impact of optimism bias is likely to be observed by chance. It establishes what we would find if optimism bias was randomly assigned across the projects and therefore whether our finding from regression analysis is unusual.

The approach does not require large sample sizes nor make any distributional assumptions. It allows us to identify whether optimism bias is important by repeatedly and randomly re-sampling the assignment of project bias. For each random assignment, the test statistic is calculated as the difference in average performance ratings for projects with and without optimism bias. Repeating this exercise provides a reference distribution for the null hypothesis of optimism bias not having an effect on project performance. We perform 10,000 permutations using the methodology of Young (2019), stratifying at the country level to determine whether or not there is a statistically significant impact of optimism bias on project performance. The null hypothesis is that there is no effect and the $p$-value is calculated as the proportion of times the placebo optimism bias effect is larger than the estimated effect. Results (available upon request) from using 10,000 permutations revealed that there were no instances where the re-sampled test statistic was more extreme than the test statistic for the data for the two measures of optimism bias. This implies that the impact of optimism bias on project performance ratings is highly unlikely due to chance.

| (1) Base | (2) Base + Country FE | (3) Base + Macroeconomic |
|----------|-----------------------|-------------------------|
| GDP per capita (constant 2010 $US) | 0.0073*** | 0.0073*** |
| | [2.82] | [2.82] |
| Political rights | 0.033** | 0.033** |
| | [2.48] | [2.48] |
| Civil liberties | -0.038** | -0.038** |
| | [-1.99] | [-1.99] |
| Country clustered standard errors | Yes | No | Yes |
| Country fixed effects | No | Yes | No |
| Observations | 2,816 | 2,816 | 2,436 |
| R-squared | 0.13 | 0.18 | 0.14 |

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. 

Table 5. (Continued)
### Table 6. Determinants of project performance (Incidence measure)

|                          | (1) Base       | (2) Base + Country FE | (3) Base + Macroeconomic |
|--------------------------|----------------|-----------------------|--------------------------|
| Binary – optimism bias   | -0.18***       | -0.17***              | -0.17***                 |
|                          | [-11.8]        | [-12.3]               | [-11.0]                  |
| **Region (omitted category = Sub-Saharan Africa)** | | | |
| East Asia and the Pacific| 0.12***        | -0.12                 | 0.088***                 |
|                          | [4.31]         | [-1.43]               | [2.81]                   |
| Europe and Central Asia  | 0.10***        | -0.16**               | 0.060                    |
|                          | [3.05]         | [-2.13]               | [1.60]                   |
| Latin America and the Caribbean | 0.046         | -0.022                | 0.052                    |
|                          | [1.44]         | [-0.29]               | [1.44]                   |
| Middle East and North Africa | 0.071**      | -0.18***              | 0.064*                   |
|                          | [2.06]         | [-4.28]               | [1.67]                   |
| South Asia               | 0.084***       | 0.00089               | 0.083***                 |
|                          | [3.17]         | [0.014]               | [3.07]                   |
| **Sector (omitted category = agriculture and rural development)** | | | |
| Energy and mining        | 0.019          | 0.028                 | 0.016                    |
|                          | [0.90]         | [1.21]                | [0.65]                   |
| Global information/communications technology | 0.11***       | 0.11***               | 0.10**                   |
|                          | [2.68]         | [3.18]                | [2.14]                   |
| Transport                | 0.062***       | 0.061***              | 0.039*                   |
|                          | [3.01]         | [3.00]                | [1.68]                   |
| Urban development        | 0.032          | 0.013                 | 0.019                    |
|                          | [0.92]         | [0.34]                | [0.53]                   |
| Water                    | 0.050          | 0.038                 | 0.025                    |
|                          | [1.53]         | [1.02]                | [0.74]                   |
| **Project agreement type (omitted category = IBRD)** | | | |
| International development association (IDA) | 0.013         | 0.051**               | 0.023                    |
|                          | [0.68]         | [2.07]                | [1.00]                   |
| Other                    | -0.047         | -0.028                | -0.043                   |
|                          | [-0.61]        | [-0.27]               | [-0.40]                  |
| **Lending instrument (omitted category = financial intermediary loan)** | | | |
| Other                    | -0.0097        | 0.017                 | -0.019                   |
|                          | [-0.16]        | [0.29]                | [-0.28]                  |
| Specific investment loan | 0.0034         | 0.015                 | 0.00040                  |
|                          | [0.089]        | [0.39]                | [0.0097]                 |
| Sector investment and maintenance loan | -0.014       | -0.0026               | -0.0082                  |
|                          | [-0.33]        | [-0.061]              | [-0.17]                  |
| Approval financial year  | -0.0025***     | -0.0032***            | -0.0029***               |
|                          | [-3.88]        | [-3.89]               | [-3.92]                  |
| Approval financial year squared | 0.0085*     | -0.0081*              | -0.0031                  |
| Implementation duration  | [-1.93]        | [-1.96]               | [-0.67]                  |
| Inflation (logged)       |                |                      | -0.025                   |
|                          |                |                      | [-1.49]                  |
| GDP per capita growth    | 0.013***       |                      | [4.20]                   |
| GDP per capita (constant 2010 SUS) | 0.0067**   |                      | [2.46]                   |
| Political rights         | 0.029**        |                      | [2.17]                   |
| Civil liberties          | -0.037*        |                      | [-1.92]                  |
| Country clustered standard errors | Yes      | No                    | Yes                      |
| Country fixed effects    | No             | Yes                   | No                       |
| Observations             | 2,832          | 2,832                 | 2,449                    |
| R-squared                | 0.08           | 0.14                  | 0.09                     |

***p < 0.01, **p < 0.05, *p < 0.1.
6. Conclusion

This paper examines the correlates of optimism bias and its impact on World Bank project performance. The paper analyzes both the intensity (magnitude) and incidence (prevalence) of optimism bias for over 2,800 World Bank projects that were appraised between 1960 and 2019. Our main findings are worth reiterating for their contributions.

6.1. Research contributions, limitations and outlook

First, we demonstrate that the incidence of optimism bias is at best 60%, which is closer to comparable figures for social infrastructure projects undertaken in the developed world (57% in Love et al., 2022), which share, though at varying degrees, similar features like the intangibility of their goals and the socio-political complexity of their settings, with development projects (Ika, 2018). However, in terms of the intensity of optimism bias, on average, the ERR at approval appears to be about 2.5% higher than at closure, as opposed to 6% in a Pohl and Mihaljek (1992) study of World Bank projects probably reflecting a fall in the level optimism bias over time. These findings offer not only new empirical evidence on the incidence but more importantly on the intensity of optimism bias in projects (Flyvbjerg, 2016; Love et al., 2022).

Second, consistent with existing literature, we find that the ERRs at approval tend to be significantly biased, partly due to optimism bias. Further, we find that the incidence and the intensity of optimism bias have decreased over the last three decades, perhaps due to a tighter focus on project appraisal, planning and RBM, suggesting, as noted by World Bank (2010), that learning takes place over time in contrast to Flyvbjerg’s (2016) results. Similar to the work of Kilby and Michaelowa (2019) on evaluation bias, we also find that project and country-level characteristics are significant correlates of optimism bias.

Third and finally, we reveal that optimism bias can reduce the chances of a satisfactory project performance rating by 17%–20%. This finding indicates that, in contrast to the theory that optimism bias is the main reason why projects experience cost overruns and benefit shortfalls (Flyvbjerg, 2016), additional influences beyond optimism bias might explain project underperformance. While the World Bank dataset does not include measures of competing explanations such as scope changes, complexity and uncertainty (Bulman et al., 2017; Denizer et al., 2013; Ika, 2018; Love et al., 2022), the literature suggests the quality of project appraisal and supervision processes also matters (Ika, 2015; Kilby, 2000). Indeed, experience seems to suggest that there are organizational incentives and practices that induce cost underestimations and/or benefit overestimations, which may not be due to optimism bias but the way organizations operate. Those producing the benefit-cost “acceptable” ratios might play according to the rules of the game. And this does not necessarily mean ‘cheating’, because there are several degrees of freedom when making ex ante estimates in conditions of high uncertainty. This observation indicates the need for more research on what explains project (under) performance in general and the role of optimism bias in particular. “Indeed, a perennial concern for World Bank management has been the frequency of projects exhibiting “disconnect” – projects that were rated as satisfactory throughout the implementation process but were then ultimately rated as unsatisfactory upon completion” (Denizer et al., 2013, p. 295).

6.2. Policy recommendations

Our policy recommendations are fourfold and go beyond World Bank project settings. First, we do not call for a CBA ‘revolution’ in the sense of a sweeping and broad application (Sunstein, 2018) but evolution to embrace complexity and uncertainty in project appraisal. The World Bank should thus distinguish between risk and uncertainty settings (see Feinstein, 2020; Hirschman, 1967). In risk settings or the realm of known unknowns, where events and their probabilities can be known, the World Bank should apply the CBA policy, particularly in high-CBA sectors. But the World Bank should expect some inaccuracies, whether they are due to
bias (e.g. optimism) or error (e.g. mistakes) (Ika et al., 2022). Using Kahneman’s (2011) ‘outside view’ (instead of ‘inside view’ of project appraisers), the World Bank might de-bias CBA over-estimates by acquiring objective past ERR data. To this end, the World Bank may implement the risk management technique of Reference Class Forecasting (RCF) advocated by Flyvbjerg (2013), which looks at the cost and benefit outcomes of a reference class of past projects similar to the one under consideration for a CBA. The World Bank may also use scenario building, a framework that weighs different scenarios with different probabilities (Feinstein, 2020).

However, under (Knightian) uncertainty, the realm of unknown unknowns, where events and their probabilities remain unknown or in instances where “we simply do not know” as Keynes would say, the reductionist CBA and RCF would fall short as the calculation cannot substitute informed and considered judgment, and cater to socio-political complexities or the human element. So, the World Bank might apply CBA, whenever relevant, along with other rules of thumb for project appraisal such as collective deliberation. But, the World Bank should expect to adapt to changing circumstances, whether they are welcome or unwelcome surprises, that is revisit assumptions, adjust plans, and make changes, particularly when projects veer from assigned paths and take complex out-turns (Hirschman, 1967; Ika et al., 2022).

Second, organizational incentives matter and project appraisers and supervisors understand their importance for their own career paths (Little & Mirrlees, 1990). While others have proposed firing or even “suing the forecaster” (Flyvbjerg, 2013, pp. 771–772), we believe error happens and forecasters may simply get their forecast wrong, especially when they face complexity and uncertainty, which are common features of development projects (Feinstein, 2020; Ika, 2018). The World Bank should instead put in place incentives not for the accuracy of CBA, which remains a means to an end, but for delivering ultimate success in projects. In other words, the World Bank should reward project appraisers and supervisors – the Task Team Leaders – who, in the face of unforeseen circumstances, reappraise and adjust their projects, and eventually ‘stumble into success’ (Hirschman, 1967).

Finally, getting the right incentives alone in place is not sufficient to make projects successful. Resources matter for CBA practice (Little & Mirrlees, 1990). We know project appraisal contributes to development impact and project supervision matters for implementation success or the delivery of specific objectives under time and cost constraints (Ika, 2015; Kilby, 2000). But resources are limited. So, to fully embrace complexity and uncertainty, the World Bank needs to shift some resources from project appraisal to implementation. As Feinstein (2020) notes, this shift means more opportunities for team leaders to reappraise projects and facilitate adjustments in the implementation phase and thus overcome a detrimental ‘project arrow’ or linear appraisal-implementation process. This move suggests a further complexification of supervision so that it allows for monitoring and evaluation during implementation as a real-time learning instrument that complements and refines appraisal in the face of changing circumstances. This entails giving project appraisers and supervisors the flexibility or the latitude (Hirschman, 1967) to ‘learn by implementing’ and try different best-fit modalities of implementation in different contexts, for example, in high versus low CBA sectors, high versus lower-income countries, and in countries with varying degrees of civil liberties or political rights (Feinstein, 2020).

Notes
1. While there are micro-economic papers that look into project characteristics such as project preparation and supervision (e.g., Kilby, 2000), they only measure how much effort or time is devoted to those activities (e.g., World Bank staff weeks). In so doing, these papers fail to unravel the ‘black box’ of management processes and thus do not explain how outputs are really derived from inputs in projects (Ika, 2015). As such, as Coase (2012) argues, they leave World Bank staff with their own management acumen, gut-feelings, and rules of thumb for making projects successful.
2. Thanks to two anonymous referees for identifying this purely mechanical effect.
3. As noted earlier, the ‘optimism bias debate’, notably in infrastructure project settings, has focused a good deal on its incidence.
4. There were 138 projects in the database that had an ERR at approval (133 also had an ERR at closure) of less than 10%; this is not surprising; for example, in their sample, Pohl and Mihaljek (1992) found 46 such projects out of 1,015 projects).

5. There were 154 projects assigned with this value. Results remain consistent to their inclusion.

6. Financial Intermediary Loans provide long-term resources to local financial institutions to finance real sector investment needs. Specific Investment Loans support the creation, rehabilitation, and maintenance of economic, social, and institutional infrastructure. Sector Investment and Maintenance Loans focus on public expenditure programs in particular sectors. They aim to bring sector expenditures, policies, and performance in line with a country’s development priorities (World Bank, 2015). The dataset does not include any measure of project complexity. While project size (e.g., financial cost) could have been an ‘imperfect proxy’ for complexity (Denizer et al., 2013), this information was not available for a large number of projects. The same goes for the only proxy for uncertainty in the dataset: risk to development outcomes.

7. We also included an implementation duration squared variable to account for a non-linear relationship, but it remained statistically insignificant.

8. While inflation can impact significantly on project costs and benefits, it is unanticipated inflation, a sort of unforeseen uncertainty (Pohl & Mihaljek, 1992), which is important to the analysis of optimism bias.

Acknowledgments

Thanks to Professors Peter E.D. Love, Alberto Posso, Ashton de Silva, Osvaldo Feinstein, Christophe Bredillet, Michele Alacevich, Anna Zhu, editor Oliver Morrissey, and two anonymous reviewers for comments and suggestions. All errors are our own. The dataset is available upon request by email to authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research has been funded by the Major Projects Observatory of the Telfer School of Management (uOttawa).

ORCID

Lavagnon Ika http://orcid.org/0000-0002-5543-9489
Simon Feeny http://orcid.org/0000-0002-5955-0045

References

Andrews, M. (2018). Public policy failure: ‘How often’ and ‘what is failure, anyway’? A study of World Bank project performance. Harvard University Center for International Development Paper No. 344. Harvard MA; Harvard University Center for International Development.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., … Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six Countries. Science, 348(6236), 1260799–1260799. doi:10.1126/science.1260799

Banuri, S., Dercon, S., & Gauri, V. (2019). Biased policy professionals. The World Bank Economic Review, 33(2), 310–327. doi:10.1093/wber/lhy033

Bulman, D., Kolkma, W., & Kraay, A. (2017). Good countries or good projects? Comparing macro and micro correlates of World Bank and Asian Development Bank project performance. The Review of International Organizations, 12(3), 335–363. doi:10.1007/s11558-016-9256-x

Coase, R. (2012). Saving economics from the economists. Harvard Business Review. The Magazine. Massachusetts: Harvard.

Del Bo, C., & Florio, M. (2010). Cost-benefit analysis and the rates of return of infrastructure projects: Evidence from international organizations. Transition Studies Review, 17(3), 587–610. doi:10.1007/s11300-010-0162-2
Denizer, C., Kaufmann, D., & A. Kraay, A. (2013). Good countries or good projects? Macro and micro correlates of world bank project performance. *Journal of Development Economics*, 105, 288–302. doi:10.1016/j.jdeveco.2013.06.003

Dugger, C. W. (2007). *World Bank finds its Africa projects are lagging*. New York Times Retrieved August 2, 2007, from https://www.nytimes.com/2007/08/02/world/africa/02worldbank.html.

Feeny, S., & Vuong, V. (2017). Explaining aid project and program success: Findings from Asian development bank interventions. *World Development*, 90, 329–343. doi:10.1016/j.worlddev.2016.10.009

Feinstein, O. (2020). Development and radical uncertainty. *Development in Practice*, 30(8), 1105–1113. doi:10.1080/09614524.2020.1763258

Flyvbjerg, B. (2003). Quality control and due diligence in project management: Getting decisions right by taking the outside view. *International Journal of Project Management*, 21(5), 760–774. doi:10.1016/j.ijproman.2012.10.007

Flyvbjerg, B. (2016). The fallacy of beneficial ignorance: A test of Hirschman’s Hiding Hand. *World Development*, 84, 176–189. doi:10.1016/j.worlddev.2016.03.012

Freedom House. (2020). *Freedom in the world 2020: A leaderless struggle for democracy*. Washington, DC: Freedom House. Freedom in the World 2020: A Leaderless Struggle for Democracy (freedomhouse.org).

Greßer, C., & Stadelmann, D. (2021). Evaluating water-and health-related development projects: A cross-project and micro-based approach. *The Journal of Development Studies*, 57(7), 1221–1239. doi:10.1080/00220388.2020.1849621

Hirschman, A. O. (1967). *Development projects observed*. Washington DC: The Brookings Institution.

Ika, L. (2018). Beneficial or detrimental ignorance: The straw man fallacy of Flyvbjerg’s test of Hirschman’s Hiding Hand. *World Development*, 103, 369–382. doi:10.1016/j.worlddev.2017.10.016

Ika, L. A. (2015). Opening the black box of project management: Does World Bank project supervision influence project impact? *International Journal of Project Management*, 33(5), 1111–1123. doi:10.1016/j.ijproman.2015.01.005

Ika, L. A. (2012). Project management for development in Africa. Why projects are failing and what can be done about it. *Project Management Journal*, 43(4), 27–41. doi:10.1002/pmj.21281

Ika, L. A., Love, P. E. D., & Pinto, J. K. (2022). Moving beyond the planning fallacy: The emergence of a new principle of project behavior. *IEEE Transactions on Engineering Management*, 1–16. doi:10.1109/TEM.2019.3040526

Isham, J., & Kaufmann, D. (1999). The forgotten rationale for policy reform: The productivity of investment projects. *The Quarterly Journal of Economics*, 114(1), 149–184. doi:10.1162/003355399555972

Kaheman, D. (2011). *Thinking fast and slow*. Toronto: Doubleday Canada.

Kilby, C. (2000). Supervision and performance: The case of World Bank projects. *Journal of Development Economics*, 62(1), 233–259. doi:10.1016/S0047-2468(00)00082-1

Kilby, C., & Michaelowa, K. (2019). What influences World Bank project evaluations?. In N. Dutta & C. R. Williamson (eds): *Lessons on foreign aid and economic development: Micro and macro perspectives* (pp. 109–150). New York: Palgrave Macmillan.

Little, I. M. D., & Mirrlees, L. A. (1990). Project appraisal and planning twenty years on. In: *Proceedings of the World Bank annual conference on development economics* (pp. 351–382). Washington, DC: World Bank. doi:10.1093/wber/4.suppl_1.1351

Love, P. E. D., Ika, L. A., & Sing, M. C. P. (2022). Does the planning fallacy prevail in social infrastructure projects? Empirical evidence and competing explanations. *IEEE Transactions on Engineering Management*, 1–15. doi:10.1109/TEM.2019.2944161

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204. doi:10.1080/07350015.2016.1227711

Pohl, G., & Mihaljek, D. (1992). Project evaluation and uncertainty in practice: A statistical analysis of rate-of-return divergences in 1,015 World Bank projects. *The World Bank Economic Review*, 6(2), 255–277. doi:10.1093/wber/6.2.255

Sunstein, C. R. (2015). Albert Hirschman’s hiding hand. Foreword. In A.O., Hirschman. *Development projects observed*, 3rd edition (pp. vii–xiii). Washington, DC: Brookings Institution.

Sunstein, C. R. (2018). *The cost-benefit revolution*. Cambridge, MA: MIT Press.

Ward, W. A. (2019). Cost-benefit analysis theory versus practice at the World Bank 1960 to 2015. *Journal of Benefit-Cost Analysis*, 10(1), 124–144. doi:10.1017/bca.2019.3

World Bank. (1992). *Effective implementation: Key to development impact* (Report No. 11536). Washington, DC: World Bank.

World Bank. (2010). *Cost-benefit analysis in World Bank projects*. Washington, DC: World Bank Independent Evaluation Group.

World Bank. (2015). *World Bank project performance ratings – Codebook*. Washington, DC: World Bank https://ieg.worldbankgroup.org/Data/reports/ieg-wb-project-performanceratings-codebook_092015.pdf

World Bank. (2017). Results and performance of the World Bank Group 2016. *An independent evaluation*. Washington, DC: World Bank, Independent Evaluation Group.

World Bank. (2020a). Projects database. Retrieved December 2020, from http://go.worldbank.org/KPMU/DAVVT.

World Bank. (2020b). *World development indicators*. Washington, DC: World Bank.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2), 557–598. doi:10.1093/qje/qjy029